MÉTODOS QUANTITATIVOS APLICADOS AO COMPORTAMENTO ORGANIZACIONAL

LUIS FELIPE DIAS LOPES

Santa Maria - RS
2018

EDITORA Voix

MÉTODOS QUANTITATIVOS APLICADOS AO COMPORTAMENTO ORGANIZACIONAL

LUIS FELIPE DIAS LOPES
Professor titular do Departamento de Ciências Administrativas da Universidade Federal de Santa Maria (UFSM), Doutor em Engenharia de Produção pela Universidade Federal de Santa Catarina (UFSC), bolsista de produtividade em pesquisa do CNPQ nível 2.

EDITORA
Voix

Santa Maria – RS
2018

CONSELHO EDITORIAL

Gerência Operacional, Layout, Editoração: Leandro Dorneles dos Santos
Revisão: Luis Felipe Dias Lopes
Arte final capas: Editora VOIX

L864m Lopes, Luis Felipe Dias

Métodos quantitativos aplicados ao comportamento organizacional [recurso eletrônico] / Luis Felipe Dias Lopes. – Santa Maria : Voix , 2018.

266 p.

ISBN 978-85-94414-07-6

1. Método quantitativo. 2. Estatística. 3. Comportamento organizacional. I. Título

CDU: 519.22:658.3

CATALOGAÇÃO: Bibliotecária Fernanda Ribeiro Paz - CRB 10/1720
Ilustração de capa: Pixabay

EDITORA Voix

editoravoix@editoravoix.com.br

Santa Maria – RS
2018

Este livro tem a finalidade de auxiliar pesquisadores
e apreciadores de métodos quantitativos,
contendo, conceitos básicos,
gráficos, equações e
exemplos
reais.

SUMÁRIO

1 CONCEITOS

1.1 MÉTODOS QUANTITATIVOS

É um ramo do conhecimento com a finalidade de resolver problemas de decisão em várias áreas do conhecimento, por exemplo nas ciências sociais e humanas, e em particular no comportamento das organizações.

Método de pesquisa indispensável, uma vez que apresenta ferramentas matemáticas e estatísticas para a tomada racional de decisões gerenciais, substituindo as decisões empíricas utilizadas na maioria dos casos.

Pesquisar quantitativamente, significa quantificar dados, fatos e opiniões, na forma de coleta de informações, utilizando recursos simples como frequencia (contagem), porcentagem, medidas descritivas: moda, mediana, média, desvio padrão, ou mais sofisticados como regressão e correlação, ou quem sabe mais complexos como técnicas multivariada de análise de dados como análise fatorial, componentes principais, equações estruturais, etc.

Métodos quantitativos são frequentemente utilizados no desenvolvimento das pesquisas nos campos social, de opinião, de mercado e administrativo, garantindo a precisão dos resultados, evitando distorções na interpretação dos mesmos (OLIVEIRA, 2002).

1.2 O QUE É ESTATÍSTICA?

“A Estatística é um conjunto de técnicas destinado à coleta, organização, resumo, apresentação e análise de dados de observação, bem como tomada de decisões razoáveis baseadas em tais análises” (FONSECA e MARTINS, 1993).

“A Estatística é um conjunto de processos ou técnicas empregadas na investigação e análise de fenômenos coletivos ou de massa” (MONTGOMERY e RUNGEL, 2012).

Segundo o estatístico Ronald Aylmer **Fisher** "A Estatística é a matemática aplicada aos dados de observação”.

Já para o estatístico David John **Finney** "Estatística é a ciência que nos auxilia na pesquisa de problemas quantitativos nos quais a variabilidade do material obscurece relações simples".

1.3 DIVISÃO DA ESTATÍSTICA

A estatística divide-se em duas partes geral e aplicada.

$$\text{Geral ou metodológica}\begin{cases}-\ \text{Descritiva}\\-\ \text{Indutiva ou Inferencial}\\\text{(planejamento de experimentos)}\end{cases}$$

$$\text{Aplicada}\begin{cases}-\textit{Bimétrica}\\-\textit{Econométrica}\\-\textit{Mecânica estatística}\\-\textit{Demografia}\\-\textit{Psicometria}\\-\textit{Sociometria}\end{cases}$$

Embora a estatística se divida nessas duas grandes áreas de estudo, que de uma maneira ou de outra sempre estão ligadas entre si, não se pode deixar de citar, aqui, a **Teoria da Probabilidade**, que está intimamente ligada à estatística, pois sem ela seria impossível fazer inferências. A probabilidade procura quantificar a incerteza existente em determinada situação, ora usando números, ora usando uma função matemática.

1.3.1 Estatística Geral ou Metodológica

Visa elaborar métodos gerais aplicáveis a todas as fases do estudo dos fenômenos de massa, tendo por finalidade o estudo das propriedades matemáticas desses fenômenos e a dedução e demonstração rigorosa dos procedimentos e equações usadas frequentemente.

- **Estatística Descritiva**

Suponha que se tenha informações de um conjunto de notas de estudantes matriculados em uma disciplina de estatística. Na terminologia estatística, o "conjunto de notas" desses estudantes é chamado de **conjunto de dados**, e a "nota individual" de cada estudante é chamada de **observação**.

O conjunto de dados, se torna mais maleável, ao se construir tabelas, gráficos ou sumarizando os seus valores através de medidas descritivas, como a média. A parte da estatística que nos ajuda neste tipo de análise é chamada de **estatística descritiva**.

- **Estatística Inferencial**

O conjunto de todos os elementos de interesse é chamado de **população**. A retirada de uma parte dessa população é chamada de **amostra**.

A maior parte dos objetivos estatísticos, como decisões, inferências e previsões sobre populações são baseadas em resultados obtidos de amostras.

A área da estatística que tem por objetivo tomar decisões, com base em amostras, é chamada **de estatística inferencial** ou **estatística indutiva**.

1.3.2 Estatística Aplicada

A diversidade de atuação é um dos grandes atrativos da estatística, que pode promover a melhoria da eficiência e também, a solução de vários problemas práticos em quase todas as áreas do saber: das ciências exatas às sociais e humanas.

O que modernamente se conhece como Estatística, é um conjunto de técnicas e métodos de pesquisa que, entre outros tópicos, envolve o planejamento do experimento a ser realizado, a coleta qualificada dos dados, a inferência, o processamento e análise das informações e a disseminação dessas informações.

- **Biométrica**

É a parte da estatística que investiga atributos biológicos quantitativos, pertinentes a uma população de seres vivos.

- **Econometria**

É um conjuntos de ferramentas estatísticas com o objetivo de entender a relação entre variáveis econômicas através da aplicação de um modelo matemático ou estatístico.

- **Mecânica estatística**

É o ramo da física que utiliza a teoria das probabilidades, estudando o comportamento de sistemas mecânicos macroscópicos compostos por um elevado número de entidades constituintes microscópicas a partir do comportamento destas entidades, quando seus estados são incertos ou indefinidos.

- **Demografia**

É uma área da ciência geográfica que estuda a dinâmica populacional humana, englobando as dimensões estatísticas, estrutura e distribuição das populações humanas. Estas não são estáticas, variando devido a natalidade, mortalidade, migrações e envelhecimento, entre outros, ou de um grupos específico segundo a educação, nacionalidade, religião ou grupo étnico.

- **Psicometria**

É o conjunto de técnicas utilizadas para mensurar, de forma adequada e comprovada experimentalmente, um conjunto de comportamentos que se deseja conhecer melhor.

- **Sociometria**

É uma ferramenta analítica que explora, mapeia e mensura relações ou vínculos estabelecidos entre forças sociais individuais, atuando em redes de interação no seio de um grupo de uma determinada organização (empresa, sala de aula ou grupamento de militares, por exemplo). A sociometria pode ser entendida também como o estudo dos vínculos existentes entre indivíduos, enquanto formadores sociais.

O desenvolvimento e o aperfeiçoamento de técnicas estatísticas, de obtenção e de análise de informações, permitem o controle e o estudo adequado de fenômenos, fatos, eventos e ocorrências, em diversas áreas do conhecimento. A estatística tem por objetivo fornecer métodos e técnicas para lidar, racionalmente, com situações sujeitas à incertezas.

1.4 POPULAÇÃO X AMOSTRA

- **População (N):** Conjunto de todos os elementos de um determinado fenômeno que possuem pelo menos uma característica em comum. A população é o conjunto, ou seja, o **universo**, podendo ser finita ou infinita.

- **Finita** - apresenta um número limitado de observações, que é passível de contagem.
- **Infinita -** apresenta um número ilimitado de observações que é impossível de contar e, geralmente, esta associada a processos.

- **Amostra (n):** É um subconjunto finito da população. A amostra deve ser selecionada seguindo certas regras e deve ser representativa, de modo que represente todas as características da população como se fosse uma fotografia desta.

Uma população pode, mediante processos operacionais, ser definida como finita ou infinita, pois a definição irá depender da relação tamanho da amostra x tamanho da população. Se a frequência relativa (fr) entre amostra e população for **menor ou igual** que 5% ela será considerada infinita, portanto o processo de retirada dos elementos da população será **COM** reposição, se a frequência relativa for **maior** do 5% será considerada finita, com processo de retirada **SEM** reposição, mais detalhes sobre esse processo será apresentado no capítulo 6.

$$\mathrm{fr} = \frac{\text{tamanho da amostra}}{\text{tamanho da população}} \mathrm{x}100 \; .$$

1.5 CENSO X AMOSTRAGEM

- **Pesquisa Estatística:** É qualquer informação retirada de uma população ou amostra, podendo ser através de **censo** ou **amostragem**.

- **Censo:** É a coleta detalhada de informações das "N" unidades populacionais.

- **Amostragem:** É o processo de retirada de informações dos "n" elementos amostrais, no qual deve seguir um método criterioso e adequado (tipos de amostragem (subcapítulo 6.3)).

1.6 DADO X VARIÁVEL

- **Dado estatístico**: É qualquer característica (quantitativa ou qualitativa) que possa ser observada ou medida de alguma forma. As matérias-primas da estatística são os dados observáveis.
- **Variável:** É o que se deseja observar para se tirar algum tipo de conclusão. Geralmente, a variável para estudo é selecionada por processos de amostragem. Os símbolos utilizados para representar as variáveis são as letras maiúsculas do alfabeto, tais como: X, Y, Z, ... que pode assumir qualquer valor de um conjunto de dados. As variáveis podem ser assim classificadas:

$$Tipos\ de\ variáveis \begin{cases} Qualitativas \begin{cases} Nominal \\ Ordinal \end{cases} \\ \\ Quantitativas \begin{cases} Discretas \\ Contínuas \end{cases} \end{cases}$$

- Qualitativas (ou Atributos): São características de uma população que não pode ser medida.

> **Nominal:** são utilizados símbolos, ou números, para representar determinado tipo de dados, mostrando, assim, a qual grupo ou categoria eles pertencem. Ex.: sim = 1; não = 2.

Ordinal (ou por Postos): quando uma classificação for dividida em categorias ordenadas, em graus convencionados, havendo uma relação entre as categorias do tipo “maior do que”, “menor do que”, “igual a”. Os dados por postos consistem de valores relativos atribuídos para denotar a ordem de primeiro, segundo, terceiro e, assim, sucessivamente. Ex.: 1 = ruim; 2 = regular; 3 = bom.

- Quantitativas: São características populacionais que podem ser quantificadas, sendo classificadas em discretas e contínuas.

Discretas: São aquelas variáveis que podem assumir somente valores inteiros num conjunto de valores. É gerada pelo processo de **contagem**, como por exemplo, o número de veículos que passa em um posto de pedágio ou o número de estudantes numa instituição de ensino. Ex.: número de filhos.

Contínuas: São aquelas variáveis que podem assumir, teoricamente, qualquer valor dentro de um intervalo de valores. É gerada pelo processo de **medição**. Neste caso serve como exemplo o volume de água em um reservatório ou o peso de um pacote de cereal. Ex.: peso e altura.

- Nível de mensuração de uma variável

Nível de mensuração representa a escala em que foi medida a variável objeto da investigação. São quatro os níveis de mensuração: nominal, ordinal, intervalar e de razão.

O registro das ocorrências de estudos científicos necessita de um meio de representação adequada dos acontecimentos e fenômenos, o que é realizado através de chamadas “escalas de mensuração ou escalas numéricas”. Basicamente, trata-se de modo de expressar qualidade e/ou quantidade.

Para que possam cumprir adequadamente a sua tarefa, as escalas precisam apresentar duas propriedades:

- *Exaustividade:* Abrangência que permita representar todos os fatos e ocorrências possíveis.

- *Exclusividade:* Coerência para que qualquer fato ou acontecimento só possa ser representado de uma única maneira.

Em suma, as escalas numéricas precisam ser capazes de exaurirem todas as possibilidades e, ao mesmo tempo, serem mutuamente excludentes.

1.7 MENSURAÇÃO DE ESCALAS

Os conceitos de mensuração de escalas estão baseados em propriedades numéricas, porém, como será visto, nem todas as propriedades podem ser estendidas para todos os tipos de escalas.
O sistema de números obedece a três propriedades básicas:

Propriedade dos números	**Exemplo**
os números obedecem a uma ordem.	4 < 9 < 12
o intervalo que separa um par de números adjacentes é o mesmo que separa outros dois pares.	o intervalo entre 4 e 8 é o mesmo que o intervalo entre 11 e 15; o intervalo entre 32 e 20 é duas vezes o intervalo entre 12 e 6.
a interpretação dos resultados da razão, quando um número é dividido em outro, indica a magnitude relativa dos dois números	8 é duas vezes maior que 4 e 9 é um terço menor que 27.

Há quatro tipos básicos de escalas: escalas nominais, escalas ordinais, escalas intervalares e escalas razão. A seguir, são apresentados os conceitos gerais de cada uma.

- **Escalas nominais ou Escalas classificatórias**

Uma escala nominal é aquela em que os números servem apenas para nomear, identificar e (ou) categorizar dados sobre pessoas, objetos ou fatos.

Os números que formam as escalas nominais não passam de rótulos usados para identificar diferentes categorias de respostas. Ilustra-se esse tipo de escala com o seguinte exemplo:

Qual dos seguintes tipos de mídia mais influenciam sua decisão de compra?
1. Televisão
2. Rádio
3. Jornal
4. Revista
5. Internet

A única operação matemática permitida com as respostas das escalas nominais é a contagem do número de respostas em cada categoria, por isso, a **moda** é a única medida de tendência central que pode ser calculada.

- **Escalas ordinais ou Escalas de postos**

As escalas ordinais, como o próprio nome diz, ordenam os objetos que estão sendo estudados de acordo com certas características, segundo um processo de comparação, além de servirem também para nomear, identificar e (ou) categorizar pessoas, objetos ou fatos.

A escala ordinal é mais poderosa que a escala nominal, pois os números também possuem a propriedade de ordenar. A escala ordinal indica a ordem de um grupo de itens associados a determinadas características, mas não indica a medida das diferenças relativas entre os itens nas ordens.

Essa escala permite concluir que o produto da marca A **é melhor** que o produto da marca B, segundo a preferência dos consumidores, mas não permite saber **quanto** A é melhor que B.

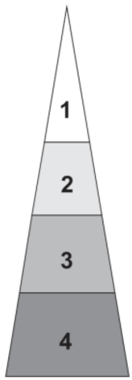
Quanto tempo você gasta lendo jornal em um fim de semana típico?	
1. menos que 5 minutos	< 5
2. de 5 minutos a 15 minutos	≥ 5 e < 15
3. de 15 minutos a 30 minutos	≥ 15 e < 30
4. mais de 30 minutos.	≥ 30

Os valores da escala 1, 2, 3 e 4 além de determinar as categorias de resposta, também fornecem indicações da extensão do tempo de leitura de jornais. Portanto, pode-se dizer que as pessoas que estão enquadradas na categoria 4 passam mais tempo lendo jornais em um fim de semana típico do que aquelas que estão na categoria 3. Entretanto, a resposta não indica quanto tempo a mais os primeiros gastam lendo jornais que os últimos. As pessoas que lêem durante 15 minutos por dia e as que lêem durante 25 minutos estão enquadradas na categoria 3, porém elas lêem quantidades de tempo diferentes. Uma vez que não há como inferir o tempo de leitura exato dos respondentes, não há nenhum significado de interpretação para os intervalos de tempo.

- *Numerais:* Números naturais, tais como 1, 2, 3..., onde "1" é o menor nível, "2" é o nível imediatamente acima, sendo "3" o nível seguinte, e assim sucessivamente até atingir um nível máximo, arbitrariamente, definido.

- *Ranking*: Ordenações do tipo 1º, 2º, 3º ..., onde o maior valor é o primeiro, o valor imediatamente abaixo é o segundo, e assim sucessivamente.

- *Rótulos:* Nomes que designam uma ordenação de intensidade, tais como “ausente, leve, moderado e grave” ou então ☆☆☆☆ e ☆☆☆☆☆".



As medidas de tendência central que podem ser calculadas em escalas ordinais são: moda e as seperatrizes (mediana, quartis e percentis).

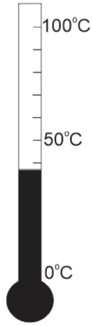
- **Escalas intervalares**

Uma escala de intervalo é aquela em que os intervalos entre os números dizem a posição e quanto as pessoas, objetos ou fatos estão distantes entre si em relação a determinada característica. Ela permite comparar diferenças entre as medições, mas não permite concluir quanto à magnitude absoluta das medições

As medidas são calculadas a partir de um ponto zero, fixado arbitrariamente. Comparativamente, as escalas de intervalo são mais poderosas que as escalas ordinais porque além de apresentarem todas as propriedades de uma escala ordinal, as diferenças entre valores das escalas de intervalo podem ser interpretadas significativamente.

Um exemplo típico de escala de intervalo é a temperatura.

Na escala Celsius, existe um ponto zero, arbitrariamente definido como sendo a temperatura exata na qual a água começa a congelar, e uma unidade de medida (grau) definida como sendo um centésimo da diferença entre a temperatura a partir da qual a água congela (0°C) e a temperatura a partir da qual a água ferve (100 °C). A temperatura de 30°C não é 50% maior do que a de 20°C, contudo, a diferença entre 20 e 30°C é a mesma que existe entre 50 e 60°C, permanecendo essa equivalência para quaisquer dois intervalos iguais na escala. Isso implica em dizer que a diferença de temperatura entre 20 e 30°C é duas vezes maior do que, digamos, a diferença existente entre 70°C e 75°C, ou entre 10 e 15°C.



Nas escalas de intervalo, a média, a moda e a mediana são todas medidas legítimas de tendência central. Os números formados por uma escala de intervalos, além de possuirem os atributos relativos à escala ordinal, também permitem o cálculo da média, do desvio padrão, do coeficiente de correlação e de testes de significância.

- **Escalas de razão**

Quando uma escala tem todas as características de uma escala de intervalos e, além disso, tem um verdadeiro ponto zero como origem é chamada escala de razões.

Em função disso, as medidas tomadas nestas escalas permitem concluir quanto a sua magnitude absoluta, além de informar a posição e quanto as pessoas, quanto objetos ou fatos estão distantes entre si em relação a determinada característica.

As respostas quantificadas que formam uma escala razão são as mais versáteis analiticamente, elas possuem todas as propriedades das escalas apresentadas anteriormente, mas com a diferença da razão entre os números possuir significado de interpretação.

Pode-se utilizar as escalas razão para renda, idade, altura, número de consumidores, número de lojas, quantidade de produtos consumidos, número de vezes que o produto é comprado ao mês, tamanho da empresa, preço, volume de vendas, lucros e participação no mercado, etc.

Nas escalas de razão, a média, a média geométrica, a moda e a mediana são todas medidas legítimas de tendência central.

- **Resumo**

Quadro 1.1 - Resumo das escalas com suas respectivas medidas

Escala	Característica	Exemplos típicos	Medidas
Nominal	Identificação.	Feminino/ masculino, usuário/ não usuário e ocupação.	Moda.
Ordinal	Ordenação.	Preferência de marcas e classe social.	Moda e Mediana.
Intervalar	Comparação de intervalos.	Escala de temperatura, atitudes em relação a marcas e conhecimento de propaganda.	Moda, Mediana e Média aritmética
Razão	Comparação de magnitudes absolutas.	Unidades vendidas, número de compradores e probabilidade de compra.	Moda, Mediana e Médias.

A Figura 1.1 apresenta a relação entre as escalas:

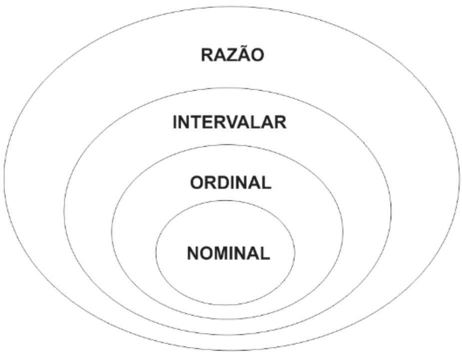


Figura 1.1 - Relação entre as escalas

1.8 ESCALAS DE ATITUDE DE AUTO-RELATO

Todas as variáveis encontradas em pesquisas de levantamento de dados podem ser classificadas em quatro categorias: atributos, variáveis comportamentais, crenças e atitudes.

Atributos: são características pessoais ou demográficas dos respondentes do estudo tais como nível educacional, idade, tamanho da família e número de filhos.

Variáveis comportamentais: relacionam-se com atividades envolvidas no estudo, por exemplo, frequência de visitas em uma loja ou a quantidade de leitores que um tipo de revista alcança.

Crenças: são os estados de conhecimento que os respondentes do estudo consideram verdadeiros, independente de serem, na realidade, corretos.

Atitude: são os estados mentais capazes de influenciar a escolha de ações de uma pessoa e mantê-la de forma consistente com essas ações.

Atitudes e crenças fazem parte do estado mental das pessoas, sendo que cada atitude origina-se de uma crença ou de uma série de crenças. Pelo fato de serem conceitos muito relacionados, a crença é um componente integral da definição formal de atitude.
Os princípios gerais das técnicas de escalas de atitudes podem também ser usados para mensurar outras variáveis internas tais como crenças, opiniões, preferências, motivos e intenções de compra.

Apresenta-se aqui uma revisão de alguns dos principais modelos de escala encontrados na literatura usados para medição de percepção e comportamento. A revisão desses modelos contribui para a elaboração dos conceitos de avaliação de percepção (MACRAE, 1988; ANDERSON, 1990; ANDRICH, 1990; ROBERTS, 1997).

As principais escalas de atitudes de auto-relato são: escalas de Thurstone, escalas diferencial semântico, escalas de Stapel, escalas de Likert, escalas de Guttman e escalas de Alpert.

1.8.1 Escalas de Thurstone

Thurstone é um dos criadores da teoria de medição da atitude moderna, o qual define atitude como sendo a quantidade de afeição ou sentimento a favor ou contra certo estímulo. A escala sugerida por Thurstone em 1928 é considerada uma das primeiras a serem desenvolvida para medir atitudes. Nesse sentido, criou a Escala de Intervalos Aparentemente Iguais, que consistem num conjunto de declarações onde cada uma possuiu um valor predefinido na escala e são apresentadas aos respondentes para que delas concordem ou discordem.

- **Exemplo de escala de Thurstone:**

Assinale se você concorda ou discorda da afirmação:

Afirmativa	Concordo	Discordo
É um café puro		
É um café forte		
Seu aroma é delicioso		
É torrado no ponto certo		
Sua embalagem é bonita		
É um produto moderno		

1.8.2 Escalas de diferencial semântico

As escalas de diferencial semântico ou escalas de Osgood foram elaboradas por Osgood; Suci e Tannenbaun (1957). Nesta técnica, os entrevistados mostram a posição de sua atitude em relação ao objeto da pesquisa em uma escala de sete pontos, o que revela a força e a direção da atitude. As extremidades do contínuo são ancoradas por um par de adjetivos polarizados ou declarações adjetivas, com a alternativa 'neutro' no centro. O escore do entrevistado é a soma dos escores em todas as escalas para esse conceito.

- **Exemplo de escala de diferencial semântico:**

Comparação dos usuários do *Google Maps* e *Yahoo Maps* em 2016

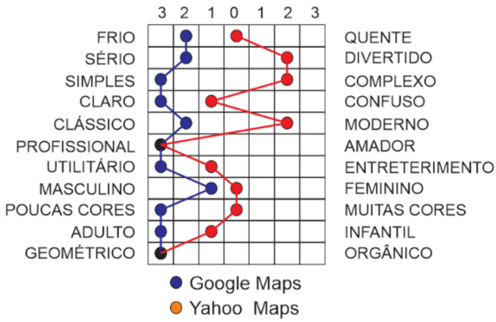


Figura 1.2 - Exemplo de escala diferencial semântico

A aplicação das escalas de diferencial semântico para produtos concorrentes permite a construção de gráficos de análises comparativos entre marcas ou produtos, o que é uma vantagem. A principal desvantagem dessas escalas está na sua construção. Para se obter resultados válidos, as escalas devem ser compostas de pares de adjetivos/frases verdadeiramente bipolares; pode acontecer de alguns dos pares escolhidos não serem verdadeiramente opostos nas mentes dos entrevistados.

1.8.3 Escalas de Stapel

A escala de Stapel foi publicada em 1950, é uma variação da escala de diferencial semântico, consiste na utilização de uma escala de pontuação verbal unipolar de 10 pontos com valores de + 5 a -5 que medem simultaneamente a força e a direção da atitude.

- **Exemplo de escala de Stapel:**

Avalie os adjetivos que descrevem a sala de aula:

+5	+5	+5
+4	+4	+4
+3	+3	+3
+2	+2	+2
+1	+1	+1
Sinalização	**Iluminação**	**Ventilação**
-1	-1	-1
-2	-2	-2
-3	-3	-3
-4	-4	-4
-5	-5	-5

1.8.4 Escalas de Likert

A escala Likert (*checklist*) foi desenvolvida pelo educador e psicólogo americano Likert (1932).

A escala Likert é uma escala psicométrica das mais conhecidas e utilizada em pesquisa quali-quantitativa, já que pretende registrar o nível de concordância ou discordância com uma declaração dada. Cada um dos itens é um ITEM LIKERT. A pontuação final da escala será a soma de todas as pontuações de cada item.

Em sua forma normal, tenta gravar um nível de concordância com as declarações do entrevistado. Por exemplo, uma afirmativa pode ser respondida das seguintes formas:

1 – Discordo totalmente
2 – Discordo
3 – Não concordo e nem discordo
4 – Concordo
5 – Concordo totalmente

1 – Nunca
2 – Poucas vezes
3 – Algumas vezes
4 – Quase sempre
5 – Sempre

1 – Nunca
2 – Ocasionalmente
3 – Frequentemente
4 – Sempre

1 – Sem importância
2 – Importante
3 – Muito importante

- **Exemplo de escala Likert sem utilização de números:**

Alienação dos consumidores em relação ao mercado.					
CT = Concordo totalmente; C = concordo; I = indiferente; D = discordo; DT = discordo totalmente					
1. As empresa recebem poucas cartas de reclamação porque não fazem nada para satisfazer os consumidores individualmente.	CT	C	I	D	DT
2. As empresas não se preocupam tanto comigo ao ponto de melhorar os produtos que vendem.	CT	C	I	D	DT
3. A satisfação que tenho experimentando novos produtos acaba em um curto período de tempo após a compra.	CT	C	I	D	DT

A escala Likert tradicional não é a melhor maneira de conhecer uma opinião já que esta pode estar situada nos intervalos que existem entre duas opções, por exemplo, caso seja usado a média dos respondentes.

Além disso, pode ser que uma mesma pessoa que marque “concordo” em dois itens diferentes não tenha o mesmo nível de opinião sobre ambos, ou questiona-se como avaliar instrumentos com escalas Likerts diferentes, ou seja, um instrumento com “x” questionamentos que utiliza uma escala com 5 pontos se relacionar com um instrumento com “y” questionamentos com escala de 7 pontos.

Uma forma de solucionar esse problema, no que se refere em avaliar ou relacionar esse tipo de instrumentos, com esse tipo de escala(s), pode ser utilizada a técnica de padronização de escores, proposta no subcapítulo 1.9.

1.8.5 Escalas de Guttman

A escala de Guttman foi publicada em 1943, é um método de escala acumulativa que procura definir mais precisamente a área neutra de uma escala de atitude, levando em consideração as atitudes dos respondentes em relação aos atributos.

A escala de Guttman é similar à escala de Likert, a diferença entre elas é que na técnica de Guttman as afirmações selecionadas incorporam a idéia da afirmação anterior.

Esta escala também pode representar a intensidade das respostas graficamente ou por meio de desenhos.

SIM	SIM	SIM	NÃO	NÃO	**NÃO**

- **Exemplo de escala de Guttman:**

Você estaria disposto a aceitar um estrangeiro(a), como visitante em seu país, como seu vizinho, como amigo pessoal, como esposo(a)?					
Entrevistado	Visitante	Vizinho(a)	Amigo(a)	Esposo(a)	**Erros**
1	1	1	0	0	**2**
2	1	1	1	1	**0**
3	0	0	1	1	**2**
4	1	1	1	0	**1**
Erros	**1**	**1**	**1**	**2**	**5**

1 = sim; 0 = não

É conveniente avaliar a reprodutibilidade (R) da escala de Guttmann, isto é, verificar se a escala tem ampla capacidade preditiva. O índice de reprodutibilidade deve ser superior a 0.9 e é de cálculo por:

$$R = 1 - \frac{E}{Q \,.\, N} = 1 - \frac{5}{4 \,.\, 4} = 1 - 0{,}3125 \cong 0{,}7$$

onde:

E = número de erros, ou respostas não afirmativas;
Q = número de perguntas ou questões;
N = número de entrevistados.

1.8.6 Escala de Alpert

A escala de Alpert foi publicada em 1971, é uma escala baseada em um método que compara marcas ou produtos dentro de uma espécie de "continuum". Há uma lista de atributos referentes ao objeto que está sendo avaliado associada a três dimensões diferentes: importância do atributo, satisfação em relação ao atributo e diferença do atributos entre os objetos que estão sendo avaliados.

A cada dimensão é atribuída uma pontuação de 1 a 5, sendo que 1 significa nenhuma importância/ nenhuma satisfação/ nenhuma diferença e 5 significa totalmente importante/ total satisfação/ totalmente diferente.

As respostas atribuídas a cada dimensão são multiplicadas, sendo que a pontuação máxima que cada atributo pode atingir é 125 (5 x 5 x 5) e a mínima é 1 (1 x 1 x 1).

A partir desses dados pode-se construir uma matriz relacionando as atitudes dos respondentes dentro de três dimensões para todos os atributos (matriz similar a da matriz da escala de diferencial semântico).

A vantagem dessa escala é a possibilidade de avaliar o atributo segundo três dimensões, possibilitando uma análise mais abrangente. Porém, se não for devidamente esclarecida, ela pode confundir o entrevistador e o respondente no momento de sua aplicação.

O estudo do comportamento humano tem grande aplicação e importância no mundo organizacional. A compreensão das atitudes de consumidores torna-se uma questão cada vez mais estratégica para as empresas. Grande interesse tem sido dado à influência que as atitudes têm sobre os hábitos de compra de consumidores e por isso os pesquisadores têm se esforçado para medir mais precisamente as motivações, as atitudes e as preferências dos consumidores.

- **Exemplo de escala de Alpert:**

Avalie, segundo uma escala de 1 a 5			
1 significa nenhum e 5 significa totalmente: 1) a importância do atributo apresentado em relação a uma sala de cinema; 2) a satisfação que o cinema A proporciona quanto ao atributo apresentado; 3) a diferença do atributo apresentado do cinema A para o cinema B em valor modular			
Atributos	**Importância**	**Satisfação**	**Diferença**
Conforto da sala de cinema	1 2 3 4 5	1 2 3 4 5	2
Luminosidade da sala de cinema	1 2 3 4 5	1 2 3 4 5	0
Som da sala de cinema	1 2 3 4 5	1 2 3 4 5	3
Geral	2 . 3 . 5 = 30	4 . 3 . 2 = 24	6

1.9 PADRONIZAÇÃO DE ESCORES

Quando se estiver utilizando ou avaliando instrumentos de pesquisas com ou sem subdivisores chamados (fatores, dimensões, aspectos ou constructos), independente do número de pontos da escala utilizada, pode-se optar por transformar os dados categóricos (1, 2, 3, ...), em dados contínuos, de 0 a 10 ou de 0 a 100, utilizando-se da técnica chamada de padronização dos escores.

A padronização dos escores é obtida por meio da seguinte operação matemática: soma-se os valores válidos subtraídos da menor soma possível, o resultado é dividido pela maior soma possível subtraído da menor soma possível, multiplicado por 10 ou 100 (valor em percentual). O cálculo Escore Padronizado (Ep_i) é representado através das seguintes equações:

$$Ep_i = 10.\left(\frac{\text{SOMA - MÍNIMO}}{\text{MÁXIMO - MÍNIMO}}\right), i = 1, 2, ..., n.$$

ou

$$Ep_i = 100.\left(\frac{\text{SOMA - MÍNIMO}}{\text{MÁXIMO - MÍNIMO}}\right), i = 1, 2, ..., n.$$

onde:

i = numero do fator;
SOMA = Somatório das respostas válidas;
MÍNIMO = menor soma possível das respostas válidas;
MÁXIMO = maior soma possível das respostas válidas.

Uma regra que poderá ser utilizada para caracterizar como "respostas válidas" ao instrumento ou ao fator, é que o entrevistado tenha respondido pelo menos a "metade mais UM" do número de questões do instrumento ou fator.

O escore padronizado auxilia a esclarecer o nível atingido por cada conjunto de variáveis (fator). A padronização dos escores permite a qualificação da soma dos resultados em medidas classificatórias. De posse dos Ep_i's, pode-se transformar os escores em quantas categorias (classificações) o pesquisador desejar.

- **Exemplo de categorização:**

- 2 categorias:

Multiplicador	Baixo(a)	Alto(a)
x 10	0,00 a 5,00	5,01 a 10,00
x 100	0,00 a 50,00	50,01 a 100,00

- 3 categorias:

Baixo(a)	Médio(a) ou Intermediário(a)	Alto(a)
0,00 a 3,33	3,34 a 6,67	6,68 a 10,00
0,00 a 33,33	33,34 a 66,67	66,68 a 100,00

- 5 categorias: escala de comportamento empreendedor

Sem comportamento	Com comportamento			
	Muito Baixo	Baixo	Alto	Muito Alto
0,00 a 5,00	5,01 a 6,25	6,26 a 7,50	7,51 a 8,75	8,76 a 10,0
0,0 a 50,0	50,1 a 62,5	62,6 a 75,0	75,1 a 87,5	87,6 a 100,0
PRETO	**VERMELHO**	**LARANJA**	**AMARELO**	**VERDE**

1.10 PARÂMETROS X ESTATÍSTICAS

- **Parâmetros:** São medidas populacionais utilizadas quando se investiga a população em sua totalidade, neste caso não há como fazer inferências, pois toda a população foi investigada. Simbolizados por letras gregas.

Ex.: Média = μ; desvio padrão = σ; coeficiente de correlação = ρ

- **Estatísticas ou Estimadores:** São medidas obtidas da amostra, torna-se possível neste caso utilizar-se das teorias inferenciais para que se possam fazer conclusões sobre a população. Simbolizados por letras arábicas

Ex.: Média = $\overline{x}$; desvio padrão = s; coeficiente de correlação = r

1.11 ARREDONDAMENTO DE DADOS

As regras de arredondamento de dados, foram editadas por uma portaria n. 36 do Instituto Nacional de Pesos e Medidas (INPM) de 06/07/1965.

1ª) Se o primeiro algarismo após aquele que será arredondado for de 0 a 4, conservamos o algarismo a ser arredondado e desprezamos os seguintes.

Ex.: 7,**3**4856 (para décimos) $\rightarrow$ 7,**3**

2ª) Se o primeiro algarismo após aquele que será arredondado for de 6 a 9, acrescenta-se uma unidade no algarismo a ser arredondado e desprezamos os seguintes.

Ex.: 1,**2**734 (para décimos) $\rightarrow$ 1,**3**

3ª) Se o primeiro algarismo após aquele que será arredondado for 5, seguido apenas de zeros, conserva-se o algarismo a ser arredondado se ele for **par** ou acrescenta-se uma unidade se ele for **ímpar**, desprezando os seguintes.

Ex.: 6,**2**500 (para décimos) $\rightarrow$ 6,**2**

12,**3**50 (para décimos) $\rightarrow$ 12,**4**

Obs.: Se o 5 (5+) for seguido de outros algarismos dos quais, pelo menos um deles for diferente de zero, aumenta-se uma unidade ao algarismo a ser arredondado e despreza-se os demais.

Ex.: 8,**2**502 (para décimos) → 8,**3**

8,**4**503 (para décimos) → 8,**5**

4ª) Quando, for arredondado uma série de parcelas, e a soma ficar alterada, deve-se fazer um novo arredondamento (por falta ou por excesso), na maior parcela do conjunto, de modo que a soma fique inalterada.

Ex.: 17,4% + 18,4% + 12,3% + 29,7% + 22,2% = 100,0%

arredondando para inteiro:

17% + 18% + 12% + **30%** + 22% = 99%

17% + 18% + 12% + **31%** + 22% = 100%

- **Resumo:**

Par ←	**5**	→ **Ímpar**
↓		↓
Mantém a unidade 0, 1, 2, 3 ou 4		Acrescenta uma unidade 6, 7, 8, 9 ou 5+*

* 5+ significa acompanhado de qualquer número diferente de zero.

1.12 NOTAÇÃO SIGMA

Muitos dos processos estatísticos exigem o cálculo da soma ou do produto de um conjunto de dados. A letra grega Σ (sigma) é usada para denotar uma soma (somatório) e a letra Π (pi) para denotar um produto (produtório).

Quando se quer somar apenas uma parte dos valores, usa-se índices para indicá-los:

Exemplo:

$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$

$$\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5$$

$$\sum_{i=3}^{4} x_i = x_3 + x_4$$

Para indicar a soma de todas (n) as observações, usam-se a seguinte representação:

$$\sum_{i=1}^{n} x_i = \sum x_i \text{ ou } \sum x$$

Propriedades:

1) Quando cada valor de uma variável for multiplicado ou dividido por uma constante, essa constante poderá ser aplicada após a soma dos valores.

$$\sum_{i=1}^{n} c \, . \, x_i = c \, . \sum_{i=1}^{n} x_i$$

Assim:

$$\sum_{i=1}^{n} 2 \, . \, x_i = 2x_1 + 2x_2 + 2x_3 + \ldots + 2x_n$$

$$= 2(x_1 + x_2 + x_3 + \ldots + x_n) = 2.\sum_{i=1}^{n} x_i$$

2) A soma de uma constante (somada n vezes) é igual ao produto da constante pelo número n.

$$\sum_{i=1}^{n} c_i = n \, . \, c$$

Assim:

$$\sum_{i=1}^{4} 5_i = 5 + 5 + 5 + 5 = 4 \, . \, 5 = 20$$

3) A soma de uma soma (ou diferença) de duas variáveis é igual a soma (ou diferença) dos somatórios individuais das duas variáveis.

$$\sum_{i=1}^{n}(x_i \pm y_i) = \sum_{i=1}^{n} x_i \pm \sum_{i=1}^{n} y_i$$

- **Exemplo**

i	x	y	x – y
1	8	5	3
2	3	2	1
3	4	0	4
4	5	4	1
Σ	20	11	9

$$\sum_{i=1}^{4}(x_i - y_i) = 9$$

$$\sum_{i=1}^{4}(x_i) - \sum_{i=1}^{4}(y_i) = 20 - 11 = 9$$

Quando se quer multiplicar apenas uma parte dos valores, usa-se índices para indicá-los:

- **Exemplo**

$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$

$$\prod_{i=1}^{5} x_i = x_1 \,.\, x_2 \,.\, x_3 \,.\, x_4 \,.\, x_5$$

$$\prod_{i=1}^{4} x_i = x_1 \,.\, x_2 \,.\, x_3 \,.\, x_4$$

Se quiser indicar o produto de todas (n) observações, usa-se a seguinte abreviatura:

$$\prod_{i=1}^{n} x_i = \prod x_i \text{ ou } \prod x$$

i	x	y	∏ = x . y
1	8	5	40
2	3	2	6
3	4	2	8
4	5	4	20
∏	480	80	38.400

$$\prod_{i=1}^{4}(x_i) . \prod_{i=1}^{4}(y_i) = \prod_{i=1}^{4} x_i . y_i$$

480 . 80 = 38.400

1.13 FASES DO MÉTODO ESTATÍSTICO

A estatística é um instrumento lógico, fundamentado no método indutivo, que tem por objetivo a descoberta em forma de valores das leis, dos fenômenos coletivos de multidão, qualquer que seja o campo experimental a que eles pertençam.

O método estatístico abrange as seguintes fases:

a) Definição do Problema

Consiste na:

- formulação correta do problema;
- examinar outros trabalhos realizados no mesmo campo (revisão da literatura);
- saber exatamente o que se pretende pesquisar, definindo o problema corretamente (variáveis, população, hipóteses, etc.)

b) Planejamento

Determinar o procedimento necessário para resolver o problema:

- Como levantar informações;
- Tipos de levantamentos: por Censo (completo);
por Amostragem (parcial).
- Cronograma, Custos etc.

c) Coleta ou levantamento dos dados

Consiste na obtenção dos dados referentes ao trabalho que desejamos fazer.
A coleta pode ser: direta: na fonte;
indireta: feita através de outras fontes.

Os dados podem ser obtidos pela própria pessoa (primários) ou se baseia no registro de terceiros (secundários).

d) Apuração dos dados ou sumarização

Consiste em resumir os dados, através de uma contagem e agrupamento. É um trabalho de ordenação e de tabulação, que pode ser feita de forma manual, mecânica, eletrônica ou eletromecânica.

e) Apresentação dos dados

É a fase em que vamos mostrar os resultados obtidos na coleta e na organização.
Podendo ser: tabular (apresentação numérica)
gráfica (apresentação geométrica)

f) Análise e interpretação dos dados

É a fase mais importante e também a mais delicada. Leva a conclusões que auxiliam o pesquisador a resolver seu problema.

2 APRESENTAÇÃO DE DADOS

Quando se realiza uma pesquisa e se quer apresentar os resultados, pode-se optar por três maneiras: tabelas ou quadros e/ou gráficos. Cada um destes tipos de apresentação possui suas características próprias, as quais serão mostrados no decorrer do capítulo.

2.1 APRESENTAÇÃO TABULAR

Consiste em dispor os dados em linhas e colunas, distribuídas de modo ordenado. A elaboração de tabelas obedece à Resolução n. 886, de 26 de outubro de 1966, do Conselho Nacional de Estatística. As normas de apresentação são editadas pelo Instituto Brasileira de Geografia e Estatística (IBGE).

2.1.1 Representação

Exemplo:

Evasão dos cotistas da UFSM em 2013 em relação ao seu ano de Ingresso

Ano de Ingresso	Evadidos
2008	61
2009	42
2010	80
2011	128
2012	185
2013	67

Fonte: Centro e Processamento de Dados - 2014

2.1.2 Elementos de uma tabela

- **Título**

Evasão dos cotistas da UFSM em 2013 em relação
ao seu ano de Ingresso

O título é a parte superior da tabela e deve conter o suficiente para responder a três perguntas:

- **O que?** Assunto a ser representado;

Evasão

- **Onde?** O lugar onde ocorreu o fenômeno;

UFSM

- **Quando?** A época que ocorreu o fenômeno.

2008 a 2013

- **Cabeçalho**

É a parte da tabela na qual se indica a natureza do conteúdo de cada coluna.

Ano de Ingresso	Evadidos

- **Corpo**

É a parte da tabela composta por linhas e colunas.

Linhas

É a parte do corpo que contém uma seqüência horizontal de informações.

Colunas

É a parte do corpo que contém uma seqüência vertical de informações.

- **Coluna indicadora**

Contém as discriminações correspondentes aos valores distribuídos pelas colunas numéricas.

- **Casa ou célula**

É a parte da tabela formada pelo cruzamento de uma linha com uma coluna.

- **Rodapé**

É o espaço aproveitado abaixo na tabela, onde são colocados detalhes do conteúdo da tabela de natureza informativa (fonte, notas de observações).

Fonte: Centro e Processamento de Dados - 2014

Fonte

Refere-se à entidade que organizou ou forneceu os dados.

Fonte: Centro e Processamento de Dados - 2014

Notas ou chamadas de observações

São detalhes ou observações inseridas na tabela

nota – esclarecimento amplo, ou seja, para toda a tabela;
chamada – esclarecimento em relação a uma célula.

2.2 SÉRIES ESTATÍSTICAS

É a classificação das tabelas, ou seja, é um conjunto de números, associado a um fenômeno expressando quantidades absolutas ou grandezas, dispostos em correspondência com um critério de modalidade. Os elementos que podem variar são: o *tempo*, o *local* e o *fato,* os elementos identificam o tipo de série. Se variar apenas um desses três elementos a série será simples, se variarem dois elementos ou mais na mesma tabela a série será mista ou composta.

- **Séries simples**

2.2.1 Série temporal ou cronológica

É a série na qual os dados estão dispostos em correspondência com o tempo, ou seja, varia o tempo e permanecem constantes o fato e o local.

- **Exemplo**

Porcentagem de alunos cotistas evadidos da UFSM de 2008 a 2013

Ano de Ingresso	Ingressantes	Evadidos	Percentual
2008	490	61	12,45
2009	917	42	4,58
2010	926	80	8,64
2011	1.161	128	11,02
2012	1.172	185	15,78
2013	1.428	67	4,69
Total	6.094	563	9,24

Fonte: Centro e Processamento de Dados - 2014

2.2.2 Série geográfica ou territorial

É a série na qual os dados estão dispostos em correspondência com o local, ou seja, varia o local e permanecem constantes a época e o fato.

- **Exemplo**

Número de universitários no Brasil por região em 2010

Região	Quantidade	Percentual
Norte	123.034	4,29
Centro-Oeste	253.413	8,85
Nordeste	464.989	16,23
Sul	580.937	20,28
Sudeste	1.441.474	50,35
Total	2.863.847	100,00

Fonte: Ministério da Educação (2011)

2.2.3 Série específica ou qualitativa

É a série na qual os dados estão dispostos em correspondência com a espécie ou qualidade, ou seja, varia o fato e permanecem constantes a época e o local.

- **Exemplo**

Vagas ofertadas e o número de ingressantes por Cota na UFSM em 2013

Cota	Ofertada	Ingressantes
Cota B	148	23
Cota D	14	7
Cota EP1	437	505
Cota EP2	405	668
Cota EP1A	355	173
Cota EP2A	331	117
Universal	2.699	2.971
Total	4.375	4.464

Fonte: Comissão Permanente do Vestibular (2013)

B - Candidato com deficiência que apresente necessidade educacional especial;
D - Indígena residente em território nacional;
EP1 - Candidato egresso do Sistema Público de Ensino Médio com renda familiar bruta mensal igual ou inferior a 1,5 salário-mínimo nacional per capita;
EP2 - Candidato egresso do Sistema Público de Ensino Médio com renda familiar bruta mensal superior a 1,5 salário-mínimo nacional per capita;
EP1A - Candidato egresso do Sistema Público de Ensino Médio, autodeclarado preto, pardo ou indígena (PPI), com renda familiar bruta mensal igual ou inferior a 1,5 salário-mínimo nacional per capita;
EP2A - Candidato egresso do Sistema Público de Ensino Médio, autodeclarado preto, pardo ou indígena (PPI), com renda familiar bruta mensal superior a 1,5 salário-mínimo nacional per capita;
UNIVERSAL - Demais candidatos que não se encaixam nas cotas anteriores.

2.2.4 Série mista ou composta

A combinação entre duas ou mais séries formam novas séries denominadas mistas ou compostas e são apresentadas em tabelas de dupla entrada. O nome dessas séries é definida conforme a combinação de, pelo menos, dois elementos.

- **Exemplo**

Série Mista Específica

Formação dos Docentes em atividade no Brasil Por tipo de instituição em 2010

Titulação	Universidades		Centro Universitários		Faculdades	
	n	%	n	%	n	%
Especialização	49.197	25	13.198	37	60.982	48
Mestrado	65.583	33	16.645	47	52.140	41
Doutorado	80.984	42	5.597	16	13.600	11
Total	195.764	100	35.440	100	126.722	100

Fonte: Instituto Nacional de Estudos e Pesquisa Educacionais, 2011

2.3 DISTRIBUIÇÃO DE FREQUÊNCIA

É o tipo de série estatística na qual permanecem constantes todos seus elementos: o fato, o local e a época. Os dados são colocados em classes pré calculadas, registrando a frequência de ocorrência. Uma distribuição de frequência pode ser classificada em discreta e intervalar.

2.3.1 Distribuição de Frequência Discreta ou Pontual

É uma série de dados agrupados na qual o número de observações está relacionado com um número real.

Qualidade dos equipamentos utilizados pelos alunos do Curso de Administração da UFSM em 2016

x_i^*	f_i
1	0
2	8
3	32
4	78
5	27
Σ	145

Fonte: Curso de Administração (2016)
* 1 – muito ruim; 2 – ruim; 3 – regular; 4 – bom; 5 – muito bom

2.3.2 Distribuição de Frequências Intervalar

Na distribuição de frequência, os intervalos parciais deverão ser apresentados de maneira a evitar dúvidas quanto à classe a que permanece determinado elemento.

O tipo de intervalo mais usado é do tipo fechado a esquerda e aberto a direita, representado pelo símbolo: |—.

Altura em centímetros de 160 alunos do Curso de Administração da UFSM – 2015

Altura (cm)	x_i	f_i
150 |— 158	154	18
158 |— 166	162	25
166 |— 174	170	20
174 |— 182	178	52
182 |— 190	186	30
190 |— 198	194	15
Σ	—	160

Fonte: Departamento de Estatística (2015)

2.3.3 Elementos de uma Distribuição de Frequências

- **Classe ou Classe de Frequência (K):** É cada subintervalo (linha) na qual divide-se o conjunto de dados. Podendo variar de 4 a 20.

- **Limite de Classe (l_i ou L_i):** São os valores extremos de cada classe.

l_i = limite inferior da i-ésima classe;
L_i = limite superior da i-ésima classe;

- **Amplitude do intervalo de classe (h):** É a diferença entre dois limites inferiores ou superiores consecutivos.

$h = l_k - l_{k-1}$ ou $h = L_k - L_{k-1}$.

Obs.: A amplitude do intervalo de classe deve ser constante em toda a distribuição de frequência intervalar.

- **Amplitude total (H):** É a diferença entre o limite superior da última classe e o limite inferior da 1ª classe, ou a diferença entre o último e o primeiro elemento de um conjunto de dados postos em ordem crescente.

$$H = L_k - l_1.$$

- **Ponto médio de classe (x_i):** É a média aritmética simples do limite inferior com o limite superior de uma mesma classe.

$$X_i = \frac{l_i + L_i}{2},$$

ou a partir do x_1 os demais pontos médios pode ser determinado por:

$$x_k = x_{k-1} + h.$$

Obs.: Quando substituirmos os intervalos de classes pelos pontos médios (x_i), ter-se-á uma **distribuição de frequência pontual**.

- **Frequência absoluta (f_i):** É a quantidade de elementos distribuídos em cada classe.

$$n = \sum_{i=1}^{k} f_i = f_1 + f_2 + \ldots + f_k.$$

- **Frequência acumulada (F_i):** É o somatório da frequência absoluta da i-ésima classe com a frequência absoluta das classes anteriores, ou a frequência acumulada da classe anterior.

$$F_k = \sum_{i=1}^{k} f_i = n.$$

- **Frequência relativa (fr_i):** É o quociente entre a frequência absoluta da i-ésima classe pelo somatório das frequências.

$$fr_i = \frac{f_i}{\sum_{i=1}^{k} f_i}$$

Obs.: $\sum_{i=1}^{k} fr_i = 1$

- **Frequência relativa acumulada (Fr_i):** É o somatório da frequência relativa da i-ésima classe com as frequências relativas das classes anteriores.

$$Fr_k = \sum_{i=1}^{k} fr_i = 1 \cdot$$

Etapas para a construção de uma distribuição de frequências

1ª) Coleta dos dados

Consiste em obter os **dados brutos**, que são os dados coletados na ordem na qual aparecem e que ainda não estão prontos para que se realize uma análise mais detalhada.

2ª) Formação do rol

É a organização dos dados brutos, em uma determinada ordem, que poderá ser crescente ou decrescente.

3ª) Determinar o número de classes ($n(k)$)

É aconselhável usar de 4 a 20 classes.

Para se determinar o número de classes ($\mathbf{n(k)}$), a partir do rol, usa-se a *Fórmula de Sturges* ou o *Método da Raiz.*

Fórmula de Sturges
$n \geq 30$

$$n(k) = 1 + \log_2 n = 1 + \frac{\log n}{\log 2} = 1 + 3{,}3 \log n$$

Método da Raiz
$n < 30$

$$n(k) = \sqrt{n}$$

onde **n** é o nº de observações coletadas.

Obs.: Além da Regra de Sturges, existem outras fórmulas empíricas para resolver o problema para determinação do número de classes. Entretanto, a verdade é que essas fórmulas não nos levam a uma decisão final; esta vai depender na realidade de um julgamento pessoal, que deverá estar ligado a natureza dos dados, procurando, sempre que possível, evitar classes com frequências nulas ou frequências relativas grandes.

4ª) Amplitude do intervalo de classe

$$h = \frac{H}{n(k)},$$

sendo: H = (maior – menor) valor coletado (amplitude total)

Obs.: A amplitude do intervalo de classe poderá sofrer um arredondamento adequado em função do tipo de dado coletado. Esse valor geralmente será arredondado para cima, de preferência na casa decimal dos dados. O intervalo de classe deverá ser preferencialmente constante em toda a distribuição de frequência.

5ª) Intervalo de classe

Consiste em definir a simbologia de representação do intervalo de classe, bem como os limites de classe, em função do número de classes estabelecidas. (|— ,—| , |—| e —)

| = fechado; — = aberto .

6ª) Frequência absoluta

É a número de indivíduos por classe. Deve-se cuidar a contagem dos indivíduos nas classes, em função do tipo de intervalo utilizado.

2.4 GRÁFICOS DE INFORMAÇÃO

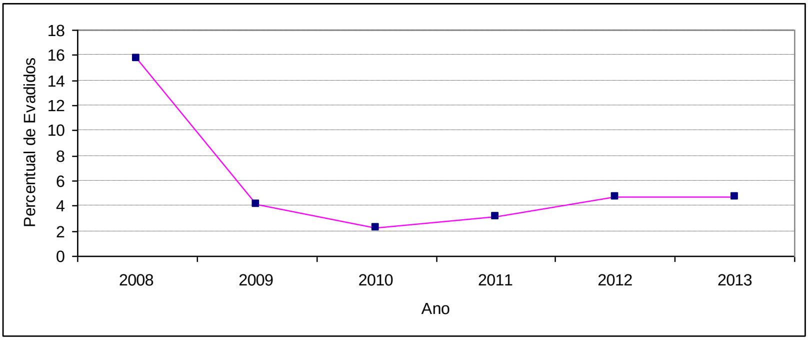
A representação gráfica é uma forma de apresentação visual dos dados. Normalmente, contém menos informações que as tabelas, mas são de fácil leitura. O tipo de gráfico depende da variável em estudo.

a) Gráficos de Linhas
Serve para representar séries simples ou compostas, geralmente utilizado para ilustrar uma série temporal. Quando se utiliza um gráfico

de linhas compostas, ele servirá tanto para informação quanto para se fazer comparações.

Exemplo 1:

Porcentagem de alunos evadidos na UFSM de 2008 a 2013

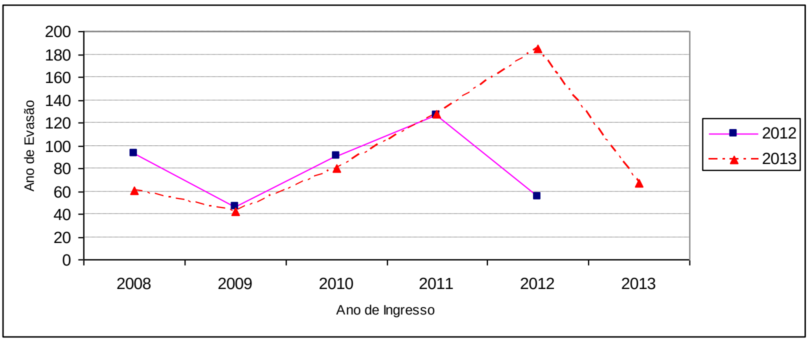


Fonte: Centro e Processamento de Dados - 2014

Figura 2.1 – Exemplo de gráfico de linhas

Exemplo 2:

Número de alunos evadidos na UFSM em 2012 e 2013
pelo ano de ingresso de 2008 a 2013



Fonte: Centro e Processamento de Dados - 2014

Figura 2.2 – Exemplo de gráfico de linhas comparativas

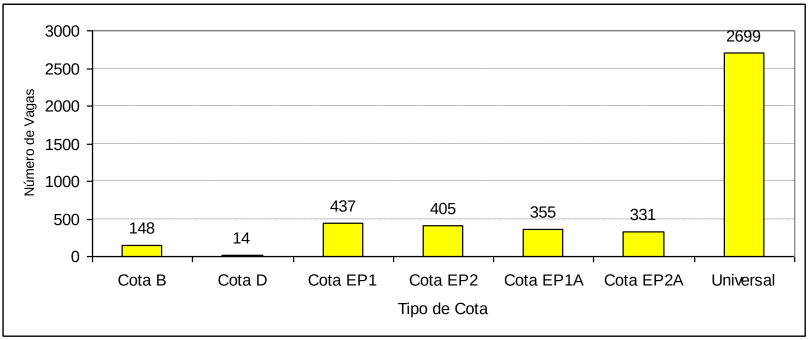
b) Gráficos de Colunas

Os gráficos de colunas são formados por retângulos no eixo horizontal. Pode-se construir gráficos de colunas simples, que serve para a representação de uma série simples e o gráfico de colunas compostas, que é indicado para séries compostas, podendo ser de colunas justapostas ou sobrepostas. Esses tipos de gráficos compostos são utilizados para ilustrar qualquer tipo de série e também servem para comparação.

b.1) Colunas simples

Exemplo:

Vagas ofertadas por Cota na UFSM em 2013



Fonte: Comissão Permanente do Vestibular (2013)

Figura 2.3 – Exemplo de gráfico de colunas

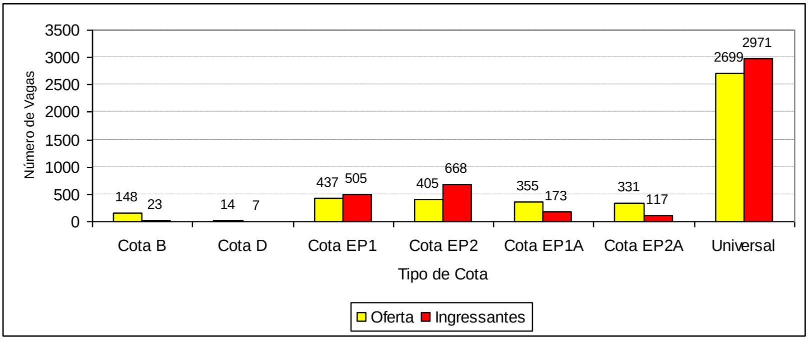
As larguras das colunas devem ser todas iguais e não têm nenhum significado neste caso, podendo ser adotada qualquer dimensão conveniente, desde que não se superponham.

O número no topo de cada coluna pode, ser omitido. Se for mostrada, a escala vertical pode ser omitida.

b.2) Colunas justapostas

Exemplo:

Vagas ofertadas e o número de ingressantes por Cota na UFSM em 2013



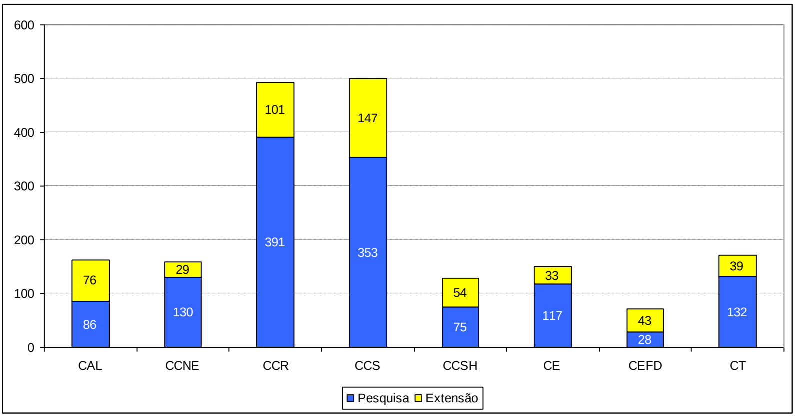
Fonte: Comissão Permanente do Vestibular (2013)

Figura 2.4 – Exemplo de gráfico de colunas justapostas

b.3) Colunas sobrepostas

Exemplo:

Número de projetos de pesquisa e extensão em andamento na UFSM em 2014



Fonte: Pró-reitoria de Planejamento (2015)

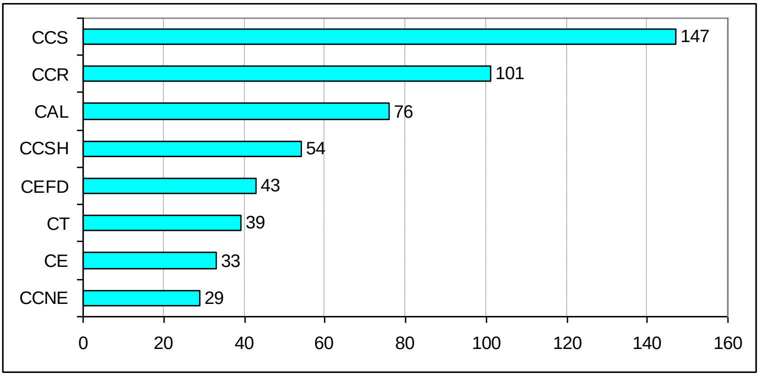
Figura 2.5 – Exemplo de gráfico de colunas sobrepostas

c) Gráfico de barras

As regras usadas para o gráfico de barras são equivalentes a usadas no gráfico de colunas, porém com a inversão dos eixos.

Exemplo:

Número de projetos de extensão em andamento na UFSM em 2014



Fonte: Pró-reitoria de Planejamento (2015)

Figura 2.6 – Exemplo de gráfico de barras

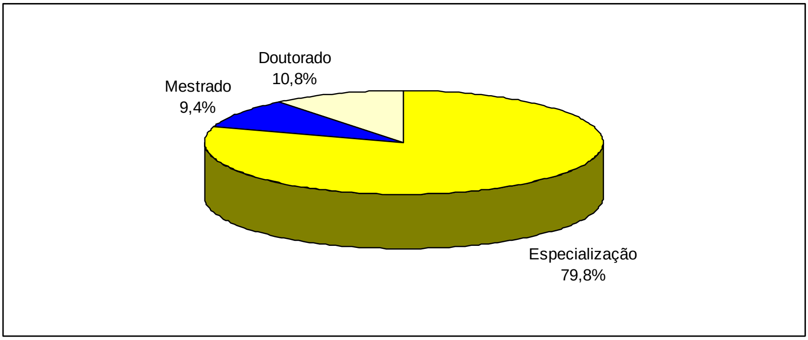
d) Gráfico de setores (*pizza*)

É a representação gráfica da frequência relativa (percentagem) de cada categoria dos dados. Este gráfico é utilizado para variáveis nominais e ordinais. Poderá ser uma opção ao gráfico de barras, quando se pretende dar ênfase à comparação das percentagens de cada categoria.

A construção do gráfico de setores segue uma regra de três simples, onde as frequências de cada classe correspondem ao ângulo que se deseja representar em relação à frequência total que representa o total de 360º.

Exemplo:

Percentual de alunos diplomados em pós-graduação na UFSM em 2013



Fonte: Pró-reitoria de Planejamento (2014)

Figura 2.7 – Exemplo de gráfico de setores

Características:

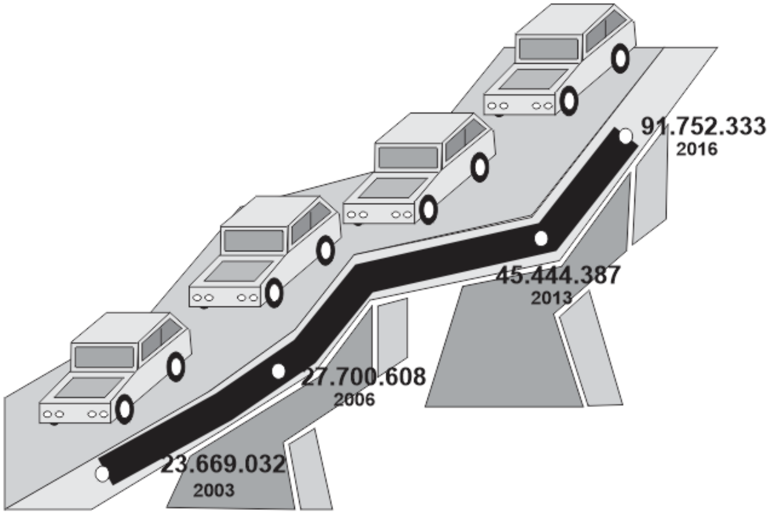
- A área do gráfico equivale à totalidade de casos (100% = 360º);

- Cada 'fatia' representa a percentagem de cada categoria representada.

e) Gráficos pictoriais

Tipo de gráfico cuja característica principal é a analogia entre o dado representado e o tipo de figura utilizado na sua representação. É bastante utilizado na propaganda, fazendo o apelo visual e percepção imediata do que se está falando. Tem por objetivo despertar a atenção do público em geral. Muitos desses gráficos apresentam grande dose de originalidade e de habilidade na arte de apresentação dos dados.

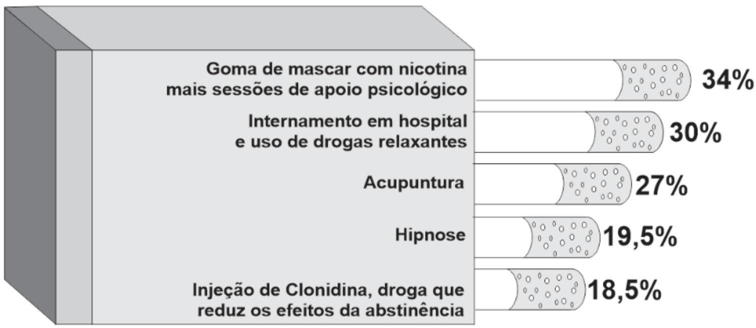
Evolução da frota nacional de carros no Brasil de 2003 à maio de 2016



Fonte: Detran (2017)

Figura 2.8 – Exemplo de gráfico pictorial de linhas

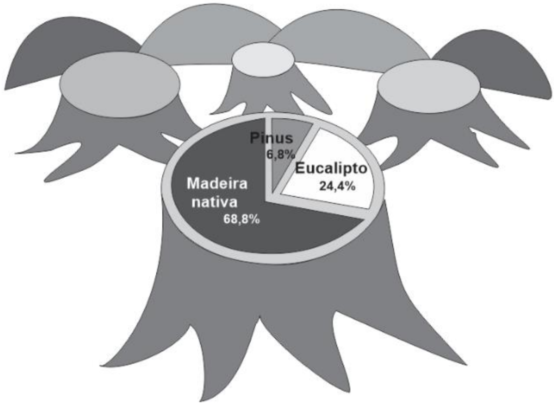
Os métodos mais eficientes para deixar de fumar segundo 30.000 fumantes entrevistados no Canadá em 2010



Fonte: Sem origem da informação

Figura 2.9 – Exemplo de gráfico pictorial de barras

Devastação Selvagem: extração de madeiras no Brasil em 2010



Fonte: Sociedade Brasileira de Silvicultura

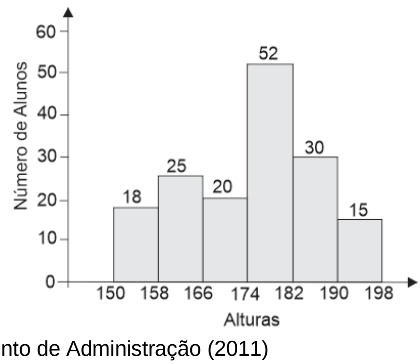
Figura 2.10 – Exemplo de gráfico pictorial de setores

2.5 GRÁFICOS DE ANÁLISE

2.5.1 Histograma

Destina-se a representar uma distribuição de frequência intervalar. Os dados são representados por colunas justapostas. Onde a base representa os intervalos e altura apresenta as frequências absolutas ou frequências relativas dentro de cada intervalo.

Altura em centímetros de 160 alunos do Curso de Administração da UFSM – 2010

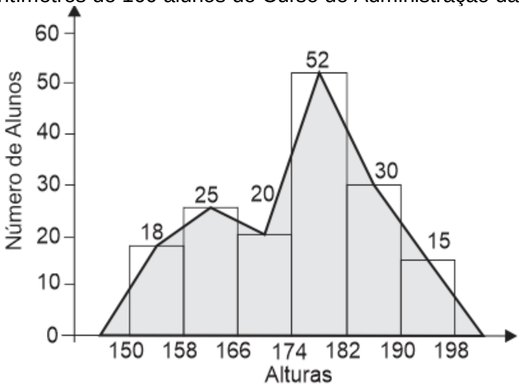


Fonte: Departamento de Administração (2011)

Figura 2.11 – Exemplo de histograma

2.5.2 Polígono de Frequências

Altura em centímetros de 160 alunos do Curso de Administração da UFSM – 2010



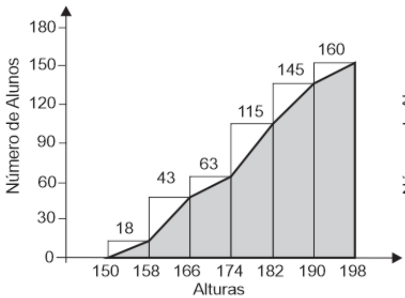
Fonte: Departamento de Administração (2011)

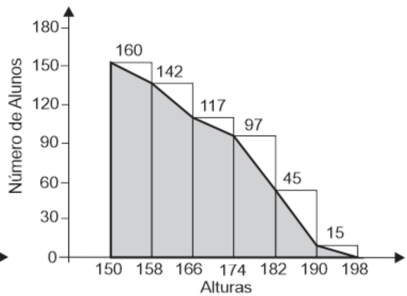
Figura 2.12 – Exemplo de polígono de frequencias

2.5.3 Ogivas

Ogiva Crescente Ogiva Decrescente

Altura em centímetros de 160 alunos do Curso de Administração da UFSM – 2010





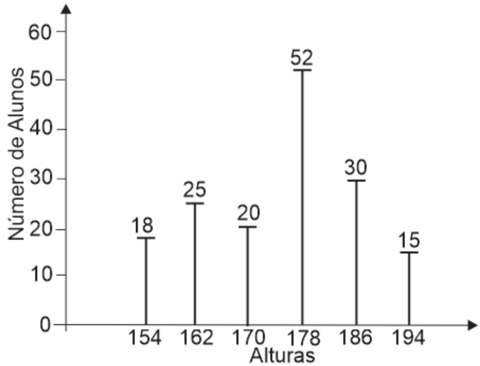
Fonte: Departamento de Administração (2011)

Figura 2.13 – Exemplo de ogivas

2.5.4 Gráfico em segmentos de reta vertical

É utilizado para representar uma distribuição de frequência pontual, onde os segmentos de reta são proporcionais às respectivas frequências absolutas.

Altura em centímetros de 160 alunos do Curso de Administração da UFSM – 2010



Fonte: Departamento de Administração (2011)

Figura 2.14 – Exemplo de gráfico de segmentos de reta vertical

2.5.5 Gráfico box e whiskers

É um tipo de gráfico que apresenta os valores centrais dos dados e alguma informação a respeito da amplitude deles. Em sua forma mais simples, um *box-plot* oferece uma representação gráfica de dados através do resumo de cinco medidas estatísticas (mínimo, 1º quartil, mediana, 3º quartil e máximo) ou (mínimo, -1 desvio padrão, média, +1 desvio padrão, máximo).

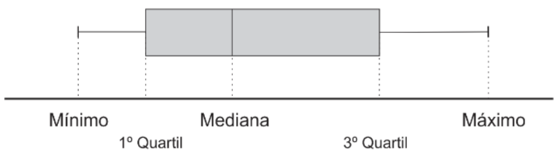


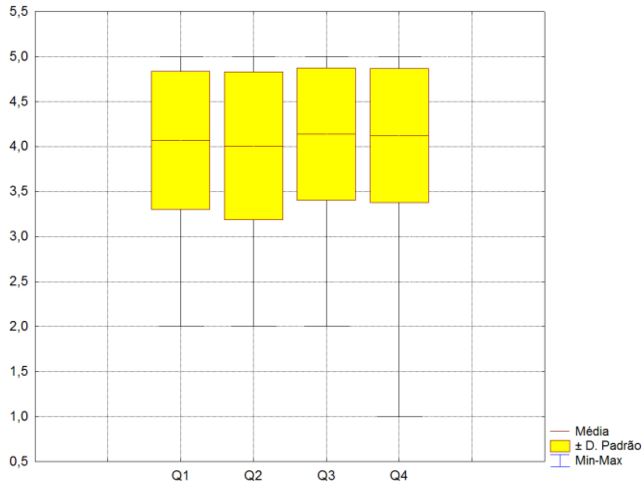
Figura 2.15 - Exemplo de box e whiskers

Algumas vezes, as retas externas (chamados de "*whiskers*") têm diferentes métodos de construção, mas os 25%, 50% e 75% percentis são sempre calculados.

Interpretação:

- a caixa central “box” inclui os 50% dos dados centrais;
- os "whiskers" mostram a amplitude dos dados, isto é, a distância entre o maior e o menor valor;
- a simetria é indicada pela caixa, o traço no centro da caixa é a localização da mediana ou da média;
- é relativamente fácil comparar grupos ou variáveis, construindo-se diagramas de caixa lado a lado.

Avaliação da Qualidade do Curso de Administração da UFSM - 2016
no quesito Coordenação



Q1 - Nível de relacionamento com os alunos
Q2 - Habilidade em dar informações
Q3 - Divulgação das atividades relacionadas ao curso (site, eventos, palestras, ...)
Q4 - Presteza no atendimento ao aluno

Fonte: Pesquisa de opinião (2016)

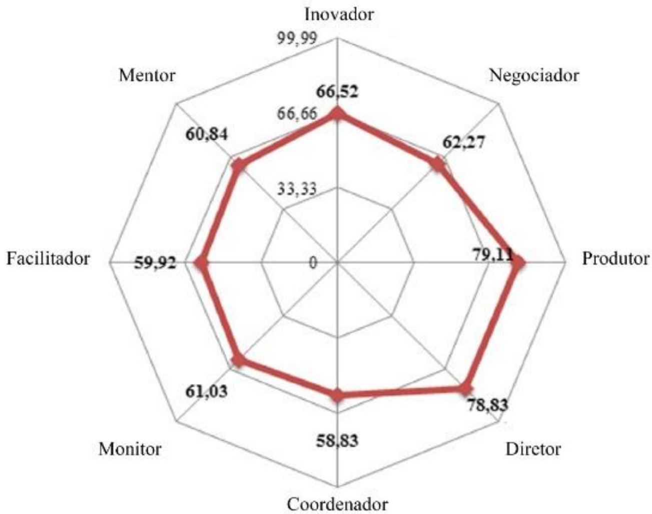
Figura 2.16 - Exemplo de box e whiskers comparativo

Obs.: No caso do box e whiskers para a média e o desvio padrão sempre haverá simetria na caixa, podendo variar a amplitude dos dados.

2.5.6 Gráfico Polar

Um gráfico polar exibe uma série de dados como um conjunto de pontos agrupados por categoria em um círculo de 360 graus. Os valores são representados pela posição do ponto, a partir do centro do círculo. Quanto mais distante o ponto está do centro, maior é o seu valor. São exibidos rótulos de categoria no perímetro do gráfico.

Média geral dos papéis dos gestores das *Startups*



Fonte: FIGUERA (2016)

Figura 2.17 - Exemplo de gráfico polar

2.6 QUADRO

É uma forma diferenciada de expressar elementos de uma pesquisa, geralmente ele expressa palavras ou frases, evitando-se números. Um quadro é semelhante a uma tabela só que fechado nas laterais, deve apresentar elementos como título, corpo e rodapé.

- **Exemplo 1**

Avaliação da Qualidade do Curso de Administração da UFSM - 2016

Fatores	Questionamentos
Coordenação	Nível de relacionamento com os alunos
	Habilidade em dar informações
	Divulgação das atividades relacionadas ao curso (site, eventos, palestras, ...)
	Presteza no atendimento ao aluno
Professores	Conhecimento teórico do assunto
	Conhecimento prático do assunto
	Clareza de explicação
	Facilidade de comunicação e de relacionamento com a turma
	Pontualidade
	Capacidade de incentivar a troca de experiências e conhecimentos
	Atendimento e esclarecimento de dúvidas individuais
	Coerência entre o programa de curso e a discussão feita em sala de aula

Fonte: Elaborado pelo autor

Quadro 1.2 - Exemplo de apresentação de quadro

- **Exemplo 2**

Modelos teóricos utilizados na pesquisa sobre Síndrome de *Burnout*

MODELOS TEÓRICOS	DESCRIÇÃO
***Maslach Burnout Inventory* (MBI-GS) -** elaborado por Maslach e Jackson (1981) e adaptado por Tamayo (2002)	A escala possui 16 variáveis relativas ao modo como os profissionais sentem-se em relação ao trabalho, o qual permite avaliar de maneira independente as dimensões de: • Exaustão Emocional (EE) (6 variáveis); • Despersonalização (DP) (4 variáveis); • Envolvimento Pessoal no Trabalho (EPT) (6 variáveis). As variáveis são mensuradas de maneira quantitativa, dispostas em escala *Likert* variando de 0 (nunca) a 6 (todo dia) conforme a frequência do sentimento.

Fonte: OBREGON (2016)

3 MEDIDAS DESCRITIVAS

Tem por objetivo descrever um conjunto de dados de forma organizada e compacta que possibilita a visualização do conjunto estudado por meio de suas medidas estatísticas.

Pode-se classificar as medidas descritivas em: medidas de posição, medidas de variabilidade, medidas de assimetria e medidas de curtose. As medidas descritivas podem ser classificadas conforme o esquema abaixo:

3.1 MEDIDAS DE POSIÇÃO

$$\text{Medidas representativas (médias)} \left\{ \begin{array}{l} \text{- Média Aritmética} \\ \text{- Média Geométrica} \\ \text{- Média Harmônica} \end{array} \right.$$

$$\text{Separatrizes} \left\{ \begin{array}{l} \text{- Mediana} \\ \text{- Quartis} \\ \text{- Decis} \\ \text{- Centis ou Percentis} \end{array} \right.$$

$$\text{Dominantes (moda)} \left\{ \begin{array}{l} \text{- Moda bruta} \\ \text{- Moda de Czuber} \end{array} \right.$$

3.1.1 Representativas - Médias

São medidas descritivas que têm por finalidade representar um conjunto de dados por um único valor. São chamadas de medidas de tendência central pelo fato de sempre tenderem a se localizar no centro do conjunto de dados considerado.

a) *Média Aritmética:* Amostral ($\bar{x}$); Populacional (μ)

- **Dados Não Tabelados**

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \text{ ou } \mu = \frac{\sum_{i=1}^{N} x_i}{N}.$$

- **Valores Ponderados**

Média Aritmética Ponderada (x_W), (onde w_i é o peso)

$$x_W = \frac{\sum_{i=1}^{n} x_i . w_i}{\sum_{i=1}^{n} w_i}.$$

- **Dados Tabelados**

Altura em centímetros de 160 alunos do Curso de Administração da UFSM – 2016

Altura (cm)			x_i	f_i	$x_i \, . \, f_i$
150	|---	158	154	18	2.772
158	|---	166	162	25	4.050
166	|---	174	170	20	3.400
174	|---	182	178	52	9.256
182	|---	190	186	30	5.580
190	|---	198	194	15	2.910
Σ			----	160	27.968

Fonte: Departamento de Administração (2016)

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i . f_i}{\sum_{i=1}^{k} f_i}$$

$$\bar{x} = 174{,}8 \cong 175$$

Características da Média Aritmética Simples

1ª) A média aritmética simples está entre o menor e o maior valor observado;

$$x_{min.} \leq \bar{x} \leq x_{max.}$$

2ª) A soma algébrica dos desvios calculados entre os valores observados e a média aritmética é igual a zero; $\text{desvio} = d = (x_i - \bar{x})$

$$\sum_{i=1}^{n} d = \sum_{i=1}^{n} (x_i - \overline{x}) = \text{zero}$$

3ª) Somando-se ou subtraindo-se todos os valores (x_i) da série com uma constante "k" ($k \neq 0$), a nova média aritmética será igual à média original somada ou subtraída por esta constante "k".

$$\begin{array}{ccc} x_i & & y_i = x_i \pm k \\ \Downarrow & & \Downarrow \\ \overline{x} & & \overline{y} = \overline{x} \pm k \end{array}$$

4ª) Multiplicando-se ou dividindo-se todos os valores (x_i) da série com uma constante "k" ($k \neq 0$), a nova média aritmética será igual à média original multiplicada ou dividida por esta constante "k".

$$\begin{array}{ccc} x_i & y_i = x_i \times k & y_i = x_i \div k \\ \Downarrow & \Downarrow & \Downarrow \\ \overline{x} & \overline{y} = \overline{x} \times k & \overline{y} = \overline{x} \div k \end{array}$$

5ª) A soma dos quadrados dos desvios, em relação à média, é um mínimo.

$$\sum_{i=1}^{n} \left(x_i - \overline{x}\right)^2 < \sum_{i=1}^{n} \left(x_i - a\right)^2, \forall\, a \neq \overline{x}$$

b) *Média Geométrica:* (x_g):

A aplicação da média geométrica deve ser feita, quando os valores do conjunto de dados considerado se comportam segundo uma progressão geométrica (P.G.) ou dela se aproximam.

- **Dados Não Tabelados**

$$x_g = \sqrt[n]{\prod_{i=1}^{k} x_i} = \sqrt[n]{x_1 . x_2 . \ldots . x_k} \; .$$

- **Dados Tabelados**

$$x_g = \sqrt[n]{\prod_{i=1}^{k} x_i^{\ f_i}} = \sqrt[n]{x_1^{\ f_1}.x_2^{\ f_2}. \ ... \ .x_k^{\ f_k}}$$

Usando um artifício matemático, pode-se usar para calcular a média geométrica a seguinte fórmula:

$$x_g = 10^{\frac{1}{n}\left(f_1.\log x_1 + f_2.\log x_2 + ... + f_k.\log x_k\right)} = 10^{\frac{1}{n}\sum_{i=1}^{k} f_i.\log x_i}$$

c) Média Harmônica **(**x_h**)**

É usada para dados inversamente proporcionais.

Ex.: Velocidade média ou preço de custo médio

- **Dados Não Tabelados**

$$x_h = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_n}}.$$

- **Dados Tabelados**

$$x_h = \frac{\sum_{i=1}^{k} f_i}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = \frac{f_1 + f_2 + ... + f_k}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + ... + \frac{f_k}{x_k}}.$$

Obs.: Deve-se observar esta propriedade entre as médias

$$\overline{x} \geq x_g \geq x_h$$

3.1.2 Dominantes - Modas (Mo)

É definida como sendo a observação com o maior número de repetições ou de maior frequência.

a) Dados não tabelados

Ex.:

3 4 4 4 5 5 6 6 7 8	$\Rightarrow Mo = 4$ (unimodal)
5 6 7 8 9 10 11 12 13	$\Rightarrow Mo = \nexists$ (amodal)
1 1 2 2 3 3 3 4 5 5 5	$\Rightarrow Mo_1 = 3 \quad Mo_2 = 5$ (bimodal)
5 5 6 6 7 7 8 8	$\Rightarrow Mo = \nexists$ (amodal)
5 5 6 6 7 7 8 9	$\Rightarrow Mo_1 = 5 \quad Mo_2 = 6 \quad Mo_3 = 7$ (trimodal)

Obs.: Acima de três modas usa-se o termo multimodal ou polimodal.

b) Dados Tabelados

- **Distribuição de frequências pontual**

- Moda Bruta (Mo_b): é o ponto médio da classe de maior frequência

- **Distribuição de frequências intervalar**

- Moda de Czuber (Mo_c)

O processo para determinar a moda usada por Czuber leva em consideração as frequências anteriores e posteriores à classe modal.

$$Mo_c = l_{Mo} + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right).h \Rightarrow \begin{cases} \Delta_1 = f_{Mo} - f_{ant} \\ \Delta_2 = f_{Mo} - f_{pos} \end{cases}$$

onde:

l_{Mo} $\Rightarrow$ limite inferior da classe modal;
f_{Mo} $\Rightarrow$ frequência absoluta da classe modal;
h $\Rightarrow$ amplitude do intervalo de classe;
f_{ant} $\Rightarrow$ frequência absoluta da classe anterior a classe modal;
f_{pos} $\Rightarrow$ frequência absoluta da classe posterior a classe modal;

Altura em centímetros de 160 alunos do Curso de Administração da UFSM - 2016

Altura (cm)			x_i	f_i
150	|---	158	154	18
158	|---	166	162	25
166	|---	174	170	20
174	|---	182	178	**52**
182	|---	190	186	30
190	|---	198	194	15
Σ			----	160

Fonte: Departamento de Administração (2016)

$$\Delta_1 = 52 - 20 = 32$$

$$\Delta_2 = 52 - 30 = 22$$

$$Mo_c = 174 + \left(\frac{32}{32+22}\right).8 = 174 + 4{,}74 = 178{,}74 \cong 179$$

3.1.3 Separatrizes

São medidas de posição que dividem o conjunto de dados em partes proporcionais, quando os mesmos são ordenados.

a) Dados não tabelados

Antes de determinarmos as separatrizes devemos em primeiro lugar encontrar a posição da mesma.

- Se o número de elementos for par ou ímpar, as separatrizes seguem a seguinte ordem:

$$\text{Posição} = \frac{i(n+1)}{S},$$

se for **mediana** $\begin{cases} i = 1 \\ S = 2 \end{cases}$

se for **quartis** $\begin{cases} 1 \leq i \leq 3 \\ S = 4 \end{cases}$

se for **decis** $\begin{cases} 1 \leq i \leq 9 \\ S = 10 \end{cases}$

se for **centis** $\begin{cases} 1 \leq i \leq 99 \\ S = 100 \end{cases}$

b) Dados Tabelados

- **Distribuição de frequências pontual**

Segue a mesma regra usada para dados não tabelados.

- **Distribuição de frequências intervalar**

$$\text{Posição} = \frac{i(n)}{S} \,,$$

$$S_i = l_{S_i} + \frac{\left(\frac{i.n}{S} - F_{ant}\right).h}{f_{S_i}} \,,$$

para:

$S_i = Md \Rightarrow i = 1;$ $\qquad S_i = Q_i \Rightarrow 1 \leq i \leq 3;$

$S_i = D_i \Rightarrow 1 \leq i \leq 9;$ $\qquad S_i = C_i \text{ ou } P_i \Rightarrow 1 \leq i \leq 99$

l_{S_i} $\Rightarrow$ limite inferior da classe que contém a separatriz;

F_{ant} $\Rightarrow$ frequência acumulada da classe anterior a que contém a separatriz;

h $\Rightarrow$ amplitude do intervalo de classe;

f_{S_i} $\Rightarrow$ frequência absoluta da classe que contém a separatriz;

Ex.:

Altura em centímetros de 160 alunos do Curso de Administração da UFSM - 2016

Altura (cm)			x_i	f_i	F_i
150	|---	158	154	18	18
158	|---	166	162	25	43
166	|---	174	170	20	63
174	|---	182	178	52	115
182	|---	190	186	30	145
190	|---	198	194	15	160
Σ			----	160	----

Fonte: Departamento de Administração (2016)

Mediana (S_i = Md) (i = 1) (S = 2)

$$\text{Posição} = \frac{1(160)}{2} = 80 \quad (4^a \text{ classe})$$

$$\text{Md} = 174 + \frac{(80-63).8}{52} = 174 + 2,62 = 176,62 \cong 177$$

3.1.4 Posição relativa da média, da mediana e da moda

Quando uma distribuição é simétrica, as três medidas coincidem. Porém, a assimetria torna-as diferentes e essa diferença é tanto maior quanto maior for a assimetria. Assim, em uma distribuição de frequência tem-se:

$\overline{X} \cong \text{Md} \cong \text{Mo}$ $\rightarrow$ curva simétrica

$\overline{X} < \text{Md} < \text{Mo}$ $\rightarrow$ curva assimétrica à esquerda ou negativa

$\text{Mo} < \text{Md} < \overline{X}$ $\rightarrow$ curva assimétrica à direita ou positiva

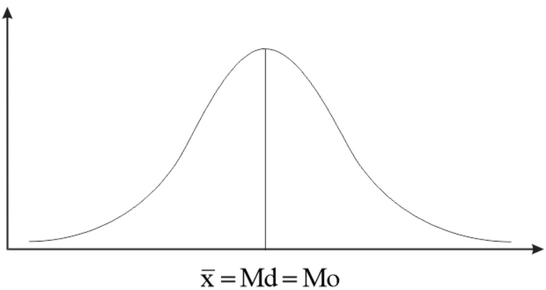


Figura 3.1 – Curva de simetria

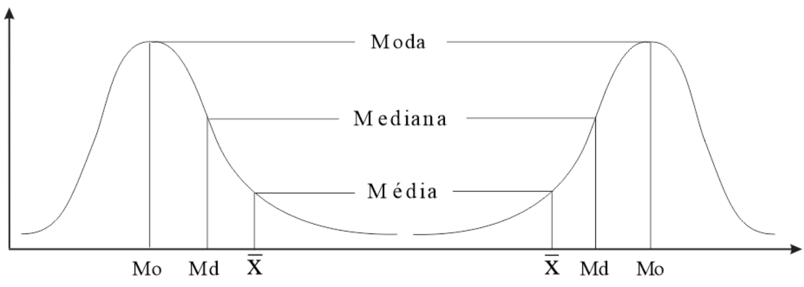


Figura 3.2 – Curvas de assimetria

No exemplo (tabela de frequências) tem-se $\overline{x}$ = 175, Md = 177 e Mo_c = 179, que vem a ser um caso de assimetria à esquerda ou negativa.

3.2 MEDIDAS DE VARIABILIDADE OU DISPERSÃO

Visam descrever os dados no sentido de informar o grau de dispersão ou afastamento dos valores observados em torno de um valor central representativo chamado média. Informa se um conjunto de dados é homogêneo (pouca variabilidade) ou heterogêneo (muita variabilidade).

As medidas de dispersão podem ser:

- Absoluta;
- Relativa;

3.2.1 Medidas de dispersão absoluta

$$\text{Absoluta} \begin{cases} \text{- Desvio extremo} = \text{Amplitude} \\ \text{- Desvio médio} \\ \text{- Desvio Quadrático ou Variância} \\ \text{- Desvio Padrão} \end{cases}$$

Para estudar as medidas de variabilidade para dados não tabelados usar-se-á um exemplo prático. Supõem-se que uma empresa esteja querendo contratar um funcionário, e no final da concorrência sobraram dois candidatos para uma única vaga. Então foram dadas 4 tarefas para cada um, onde as mesmas tiveram como registro o tempo (em minutos) de execução.

TAREFAS	**1**	**2**	**3**	**4**
OPERÁRIO 1 (TEMPO)	45	55	48	52
OPERÁRIO 2 (TEMPO)	30	70	40	60

- **Análise Gráfica**

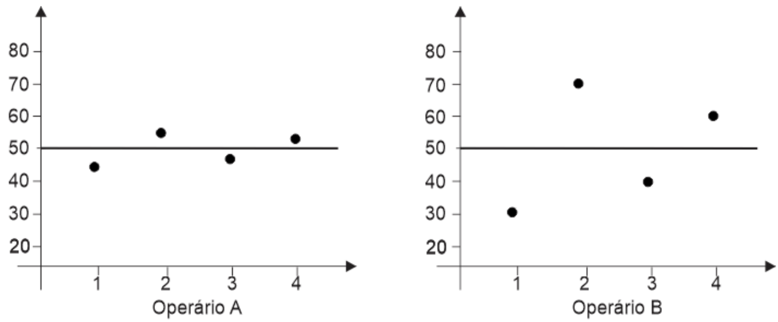


Figura 3.3 – Gráfico de dispersão por operário

- **Medidas de Dispersão Absoluta:**

a) Desvio Extremo ou Amplitude de Variação (H): É a diferença entre o maior e o menor valor de um conjunto de dados.

$$H = x_{max} - x_{min}\,.$$

b) Desvio Médio ($\bar{d}$): Em virtude do $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$, usa-se para calcular o desvio médio o valor modular, ou seja:

- **Para dados não tabelados**

$$\bar{d} = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \ldots + |x_n - \bar{x}|}{n}\,.$$

- **Para dados tabelados**

$$\bar{d} = \frac{\sum_{i=1}^{k}\left(f_i\left|x_i - \bar{x}\right|\right)}{\sum_{i=1}^{k} f_i} = \frac{f_1\left|x_1 - \bar{x}\right| + f_2\left|x_2 - \bar{x}\right| + \ldots + f_k\left|x_k - \bar{x}\right|}{\sum_{i=1}^{k} f_i}.$$

c) Desvio Quadrático ou Variância (Var): s^2 (amostra)
σ^2 (população)

- **Para dados não tabelados**

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(x_i - \mu\right)^2}{N} = \frac{\left(x_1 - \mu\right)^2 + \left(x_2 - \mu\right)^2 + \ldots + \left(x_n - \mu\right)^2}{N};$$

$$s^2 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1} = \frac{\left(x_1 - \bar{x}\right)^2 + \left(x_2 - \bar{x}\right)^2 + \ldots + \left(x_n - \bar{x}\right)^2}{n-1}.$$

- **Para dados tabelados**

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i\left(x_i - \mu\right)^2}{\sum_{i=1}^{k} f_i} = \frac{f_1\left(x_1 - \mu\right)^2 + f_2\left(x_2 - \mu\right)^2 + \ldots + f_k\left(x_k - \mu\right)^2}{\sum_{i=1}^{k} f_i};$$

$$s^2 = \frac{\sum_{i=1}^{k} f_i\left(x_i - \bar{x}\right)^2}{\sum_{i=1}^{k} f_i - 1} = \frac{f_1\left(x_1 - \bar{x}\right)^2 + f_2\left(x_2 - \bar{x}\right)^2 + \ldots + f_k\left(x_k - \bar{x}\right)^2}{\sum_{i=1}^{k} f_i - 1}.$$

- Desvio Padrão: s (amostra) ou σ (população)

- **Para dados não tabelados**

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \mu\right)^2}{N}} = \sqrt{\frac{\left(x_1 - \mu\right)^2 + \left(x_2 - \mu\right)^2 + \ldots + \left(x_n - \mu\right)^2}{N}};$$

$$s=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n-1}}=\sqrt{\frac{\left(x_1-\bar{x}\right)^2+\left(x_2-\bar{x}\right)^2+\ldots+\left(x_n-\bar{x}\right)^2}{n-1}}.$$

- **Para dados tabelados**

$$\sigma=\sqrt{\frac{\sum_{i=1}^{k}f_i\left(x_i-\mu\right)^2}{\sum_{i=1}^{k}f_i}}=\sqrt{\frac{f_1\left(x_1-\mu\right)^2+f_2\left(x_2-\mu\right)^2+\ldots+f_k\left(x_k-\mu\right)^2}{\sum_{i=1}^{k}f_i}};$$

$$s=\sqrt{\frac{\sum_{i=1}^{k}f_i\left(x_i-\bar{x}\right)^2}{\sum_{i=1}^{k}f_i-1}}=\sqrt{\frac{f_1\left(x_1-\bar{x}\right)^2+f_2\left(x_2-\bar{x}\right)^2+\ldots+f_k\left(x_k-\bar{x}\right)^2}{\sum_{i=1}^{k}f_i-1}}.$$

Obs.: (n - 1) ou $\sum_{i=1}^{k}f_i-1$ é usado como um fator de correção, onde deve-se considerar a variância amostral como uma estimativa da variância populacional.

- **Propriedades da Variância**

1ª) Somando-se ou subtraindo-se uma constante k a cada valor observado a variância não será alterada;

2ª) Multiplicando-se ou dividindo-se por uma constante k cada valor observado a variância ficará multiplicada ou dividida pelo **quadrado** dessa constante.

Ex.:

TAREFAS	1	2	3	4
OPERÁRIO 1 (TEMPO)	45	55	48	52
OPERÁRIO 1 (TEMPO + 2)	47	57	50	54
OPERÁRIO 1 (TEMPO x 2)	90	110	96	104

Média = 50 e Var = 14,5 (TEMPO + 2) Média = 52 e Var = 14,5

(TEMPO x 2) Média = 100 e Var = 14,5 x 2^2 = 58

3.3 MEDIDAS DE DISPERSÃO RELATIVA (MDR)

$$\text{MDR} \begin{cases} \text{- Variância Relativa (VR)} \\ \text{- Coeficiente de Variação de Pearson (CVP)} \end{cases}$$

É a medida de variabilidade que em geral é expressa em porcentagem, e tem por função determinar o grau de concentração dos dados em torno da média, geralmente utilizada para se fazer a comparação entre dois conjuntos de dados em termos percentuais, esta comparação revelará o quanto os dados estão próximos ou distantes da média do conjunto de dados.

a) Variância Relativa

$$VR = \frac{\sigma^2}{\mu^2} \text{ ou } VR = \frac{s^2}{\bar{x}^2} \cdot$$

b) Coeficiente de Variação de Pearson

$$CVP = \frac{\sigma}{\mu}.100 \quad \text{ou} \quad CVP = \frac{s}{\bar{x}}.100 \cdot$$

Obs.: CVP $\leq$ 50% $\Rightarrow$ a média é representativa;
CVP $\cong$ 0 $\Rightarrow$ é a maior representatividade da média (s $\cong$ 0).

3.4 MOMENTOS

São medidas descritivas de caráter mais geral e dão origem às demais medidas descritivas, como as de tendência central, dispersão, assimetria e curtose. Conforme a potência considerada, tem-se a ordem ou o grau do momento calculado.

3.4.1 Momento simples ou centrado na origem (mr)

O momento simples de ordem “r” é definido como:

- **Para dados não tabelados**

$$m_r = \frac{\sum x_i^r}{n} \cdot$$

• **Para dados tabelados**

$$m_r = \frac{\sum x_i^r f_i}{\sum f_i} .$$

onde r é um número inteiro positivo;

$m_0 = 1$;
m_1 = média aritmética.

3.4.2 Momento centrado na média (Mr)

O momento de ordem "r", centrado na média, é definido como:

• **Para dados não tabelados**

$$M_r = \frac{\sum (x_i - \bar{x})^r}{n} = \frac{\sum d_i^r}{n} .$$

• **Para dados tabelados**

$$M_r = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^r}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} d_i^r f_i}{n} ,$$

onde: $M_0 = 1$;

$M_1 = 0$;

M_2 = variância (s^2).

3.4.3 Momento abstrato (αr)

São definidos da seguinte forma:

$$\alpha_r = \frac{M_r}{s^r} ,$$

onde: s = desvio padrão.

Ex.:

Altura em centímetros de 160 alunos do Curso de Administração da UFSM - 2016

Altura (cm)	x_i	f_i	$x_i \cdot f_i$	$f_i (x_i - \overline{X})^2$
150 |--- 158	154	18	2.772	7.938
158 |--- 166	162	25	4.050	4.225
166 |--- 174	170	20	3.400	500
174 |--- 182	178	52	9.256	468
182 |--- 190	186	30	5.580	3.630
190 |--- 198	194	15	2.910	5.415
Σ	----	160	2.796	22.176

Fonte: Departamento de Administração (2016)

$$m_1 = \frac{\sum x_i^1 f_i}{\sum f_i} = \frac{2796}{160} = 174{,}8 \cong 175 = \text{média}$$

$$M_2 = \frac{\sum_{i=1}^{k} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{k} f_i} = \frac{\sum_{i=1}^{k} d_i^2 f_i}{n} = \frac{22176}{160} = 138{,}6 = \text{Variância}$$

$$\alpha_2 = \frac{M_2}{s^2} = \frac{s^2}{s^2} = 1$$

3.5 MEDIDAS DE ASSIMETRIA

Como visto no item 3.1.4 uma distribuição de valores sempre poderá ser representada por uma curva (gráfico). Essa curva, conforme a distribuição, pode apresentar várias formas. Considerando o valor da moda da distribuição, como ponto de referência, vê-se que esse ponto sempre corresponde ao valor de ordenada máxima. Logo, a curva será analisada quanto a sua assimetria.

3.5.1 Distribuição simétrica

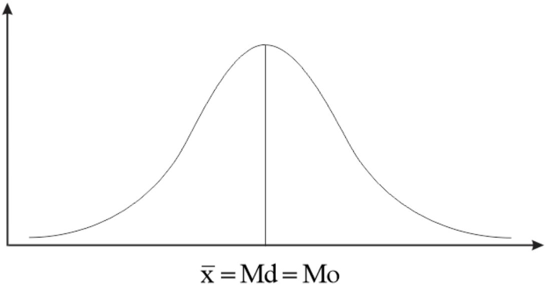


Figura 3.4 – Distribuição simétrica

É aquela que apresenta $\bar{x} \equiv Mo \equiv Md$ e os quartís Q_1 e Q_3 são eqüidistantes do Q_2. Nesse caso tem-se 50% dos elementos antes do valor central e 50% após esse valor.

3.5.2 Distribuição assimétrica

a) Assimétrica à direita, ou positiva

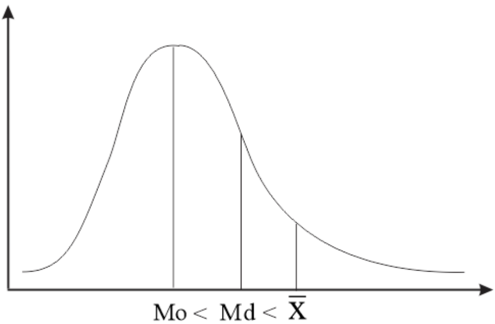


Figura 3.5 – Distribuição assimétrica positiva

b) Assimétrica à esquerda, ou negativa

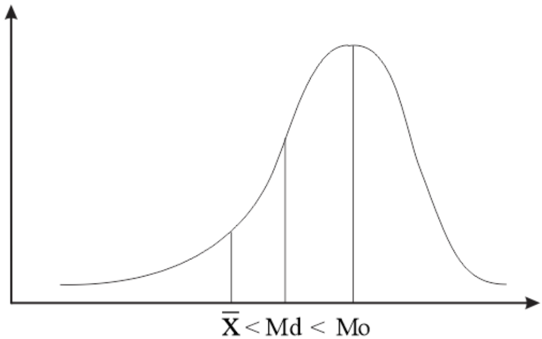


Figura 3.6 – Distribuição assimétrica negativa

Pode-se medir a assimetria de uma distribuição calculando-se os coeficientes de assimetria, sendo o mais utilizado o coeficiente de assimetria de Pearson.

3.5.3 Coeficiente de assimetria de Pearson

$$As = \frac{\overline{X} - Mo}{S} \; .$$

a) Intensidade do coeficiente de assimetria

Interpretação:

As < 0	$\Rightarrow$	distribuição Assimétrica Negativa;
As > 0	$\Rightarrow$	distribuição Assimétrica Positiva;
As = 0	$\Rightarrow$	distribuição Simétrica.

Pode-se calcular o coeficiente de assimetria utilizando somente os quartis (Q_i):

$$As = \frac{Q_3 + Q_1 - 2Md}{Q_3 - Q_1}\text{, lembrado que a } Md = Q_2.$$

3.5.4 Coeficiente momento de assimetria

Este coeficiente mostra a força ou intensidade da assimetria, sendo calculado a seguir:

α_3 corresponde ao terceiro momento abstrato,

$$\alpha_3 = \frac{M_3}{S^3},$$

onde: M_3 = momento centrado de terceira ordem;
S = desvio padrão.

O campo de variação do coeficiente momento de assimetria é:

$-1 \leq \alpha_3 \leq +1$

a) Intensidade do coeficiente momento de assimetria

Interpretação:

$\lvert\alpha_3\rvert < 0,2$	$\Rightarrow$	Simetria;
$0,2 < \lvert\alpha_3\rvert < 1,0$	$\Rightarrow$	Assimetria fraca;
$\lvert\alpha_3\rvert > 1,0$	$\Rightarrow$	Assimetria forte.

3.6 MEDIDAS DE CURTOSE

Já foram mostradas as medidas de tendência central, de dispersão e de assimetria. Examinar-se-á mais uma das medidas de uso comum em Estatística. As medidas de curtose, ou de achatamento, têm por finalidade nos mostrar até que ponto a curva representativa de uma distribuição é mais aguda ou mais achatada, do que uma curva normal, de altura média.

Curtose: grau de achatamento de uma distribuição em curva.

- Curva mesocúrtica

É considerada a curva normal ou padrão (simétrica).

- Curva leptocúrtica

É uma curva mais alta do que a curva normal. Apresenta o topo relativamente alto, cujos valores se apresentam mais agrupado em torno da moda.

- Curva platicúrtica

É uma curva mais baixa do que a curva normal. Apresenta o topo achatado, significando que várias classes apresentam frequências quase iguais.

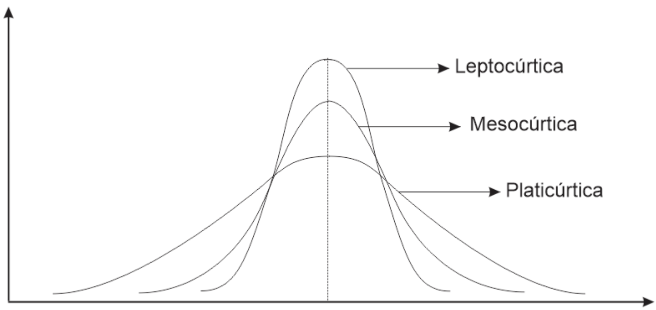


Figura 3.7 – Comparação das curvas quanto a curtose

3.6.1 Coeficiente de curtose

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}\ .$$

onde, Q_1 e Q_3 são o 1º e 3º quartil e P_{10} e P_{90} são o 10º e 90º percentil.

a) Intensidade do coeficiente de curtose

Interpretação:

$K = 0{,}263 \quad \Rightarrow \quad$ distribuição Mesocúrtica;
$K < 0{,}263 \quad \Rightarrow \quad$ distribuição Leptocúrtica;
$K > 0{,}263 \quad \Rightarrow \quad$ distribuição Platicúrtica.

3.6.2 Coeficiente momento de curtose

Serve para medir o grau de achatamento da curva.

$$\alpha_4 = \frac{M_4}{S^4},$$

onde: α_4 corresponde ao momento abstrato de quarta ordem,
M_4 = momento centrado de quarta ordem.

Intensidade do coeficiente momento de curtose

Interpretação:

$\alpha_4 < 3$ $\Rightarrow$ curva Platicúrtica;
$\alpha_4 = 3$ $\Rightarrow$ curva Mesocúrtica;
$\alpha_4 > 3$ $\Rightarrow$ curva Leptocúrtica.

4 PROBABILIDADE E VARIÁVEIS ALEATÓRIAS

4.1 MODELOS MATEMÁTICOS

Podemos definir dois tipos de modelos matemáticos:

4.1.1 Modelos Determinísticos

São modelos que estipulam as condições sob as quais um experimento seja executado, ou seja, as condições *determinam* o resultado do experimento.

Em outras palavras, um modelo determinístico emprega considerações físicas para prever resultados,

4.1.2 Modelos Não Determinísticos ou Probabilísticos

São modelos que informam com que chance ou probabilidade os acontecimentos podem ocorrer. Determina o grau de credibilidade dos acontecimentos (modelos estocásticos).

Em outras palavras, um modelo probabilístico emprega uma mesma espécie de consideração para especificar uma distribuição de probabilidade.

4.2 CONCEITOS EM PROBABILIDADE

Os conceitos fundamentais em probabilidade são experimento aleatório, espaço amostral e evento.

4.2.1 Experimento aleatório (Ω)

Qualquer processo aleatório, capaz de produzir observações, na qual os resultados surgem ao acaso, podendo admitir repetições no futuro. Um experimento aleatório apresenta as seguintes características:

a) os resultados podem repetir-se n vezes ($n \rightarrow \propto$);
b) embora não se possa prever que resultados ocorrerão, pode-se descrever o conjunto de resultados possíveis;

c) a medida que se aumenta o número de repetições, aparece uma certa regularidade nos resultados.

4.2.2 Espaço Amostral (S)

É o conjunto de todos os possíveis resultados de um experimento aleatório. Quanto ao número de elementos pode ser:

Finito
Número limitado de elementos;
Ex.: S = {1, 2, 3, 4, 5, 6}

Infinito
Número ilimitado de elementos, pode ser sub-dividido em:

a - Enumerável
Quando os possíveis resultados puderem ser postos em concordância biunívoca com o conjunto dos números naturais (N) (caso das variáveis aleatórias discretas).
Ex.: Números Naturais = IN

b - Não Enumerável
Quando os possíveis resultados não puderem ser postos em concordância biunívoca com o conjunto dos números naturais (caso das variáveis aleatórias contínuas).
Ex.: Números Reais = IR

4.2.3 Evento (E)

Um evento é qualquer subconjunto de um espaço amostral (S).
Podem-se ter operações entre eventos da mesma forma que operação com conjuntos, como mostra a seguir.

4.2.4 Operações com Eventos

A união B

Símbolo utilizado "$\cup$", é o evento que ocorrerá se, e somente se, A ou B ou ambos ocorrerem;

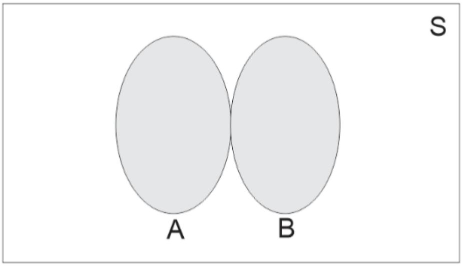


Figura 4.1 - Evento A união B

A interseção B

Símbolo utilizado "$\cap$", é o evento que ocorrerá se, e somente se, A e B ocorrem simultaneamente.

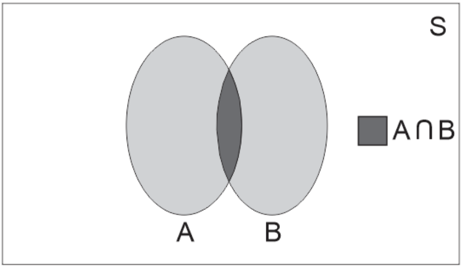


Figura 4.2 - Evento A interseção B

Complementar de A

Simbologia " $\overline{A}$ ", é o evento que ocorrerá se, e somente se A não ocorrer.

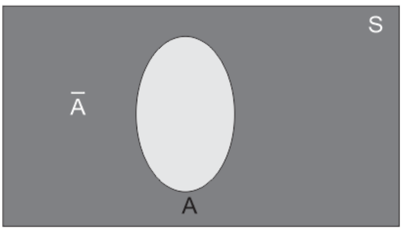


Figura 4.3 - Evento complementar de A $\left(S - A = \overline{A}\right)$

4.2.5 Tipos de eventos

Eventos Mutuamente Excludentes

São ditos eventos mutuamente excludentes, quando a ocorrência de um implica ou não na ocorrência de outro, isto é, não podem ocorrer juntos, e conseqüentemente, $A \cap B$ é o conjunto vazio ($\varnothing$).

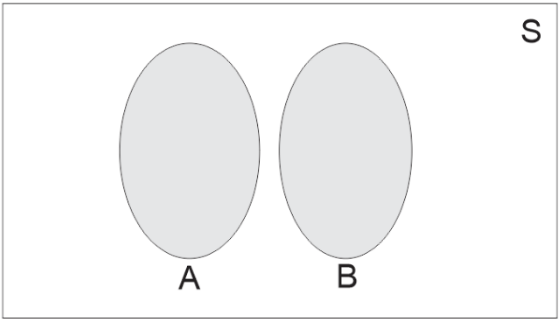


Figura 4.4 - Eventos mutuamente excludentes

Eventos Não Excludentes ou Quaisquer

São ditos eventos não excludentes quando a ocorrência de um implica na ocorrência do outro, isto é, são aqueles que ocorrem ao mesmo tempo, $A \cap B \neq \varnothing$.

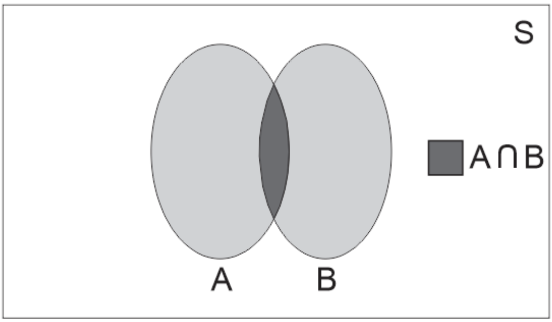


Figura 4.5 - Evento não excludentes

Eventos Independentes

São aqueles cuja ocorrência de um evento, não possui efeito algum na probabilidade de ocorrência do outro.

$A \cap B \neq \varnothing$, se A e B forem Quaisquer;

$A \cap B = \varnothing$, se A e B forem Mutuamente Excludentes.

logo,

$P(A \cap B) = P(A) \, . \, P(B)$.

Ex.: A e B eventos Quaisquer

S = { 1, 2, 3, 4 }   A = {1, 2 }   B = { 2, 4 } $A \cap B$ = { 2 }

$$P(A \cap B) = P(A) \, . \, P(B)$$

$$P(A) = \frac{2}{4} \qquad P(B) = \frac{2}{4} \qquad P(A \cap B) = \frac{1}{4}$$

Eventos Dependentes ou Condicionados

Existem varias situações onde a ocorrência de um evento pode influenciar fortemente na ocorrência de outro.

Assim, se (A) e (B) são eventos, deseja-se definir uma quantidade denominada probabilidade condicional do evento (A) dado que o evento (B) ocorre, ou sob a forma simbólica $P\left(A/B\right)$.

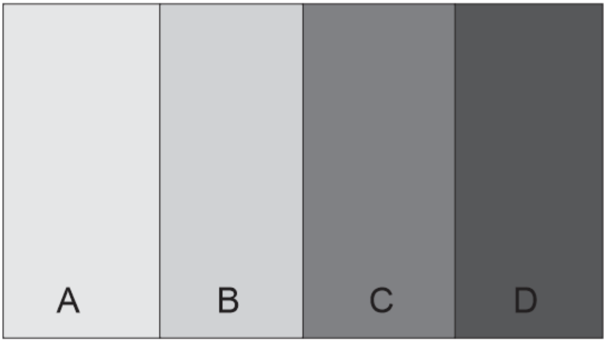
Assim, dá-se a seguinte definição:

$$P\left(A/B\right) = \frac{P(A \cap B)}{P(B)},$$

onde $P(B) > 0$. Se $P(B) = 0$, tem-se que $P\left(A/B\right)$ não é definida.

Eventos Coletivamente Exaustivos

São aqueles que ocorrem se nenhum outro ocorrer.



$$A \cap B \cap C \cap D = \varnothing$$

Figura 4.6 - Evento coletivamente exaustivos

4.3 CONCEITOS DE PROBABILIDADE

4.3.1 Conceito Empírico de Probabilidade

O problema fundamental da probabilidade consiste em atribuir um número a cada evento (E), o qual avaliará quão possível será a ocorrência de "E", quando o experimento for realizado.
Uma possível maneira de tratar a questão seria determinar a frequência relativa do evento E (f_r(E)),

$$f_r(E) = \frac{\text{número de ocorrências do evento}(E)}{\text{número de repetições do experimento}(\Omega)}.$$

Surgem, no entanto, dois problemas:
a - Qual deve ser o número de repetições do experimento (Ω);
b - A sorte ou habilidade do experimentador poderá influir nos resultados, de forma tal que a probabilidade é definida como sendo:

$$P(E) = \lim_{n\to\infty} f_r(E),$$

onde "n" é o número de repetições do experimento Ω.

4.3.2 Definição Clássica ou Enfoque "A priori" de Probabilidade

Se existe "a" resultados possíveis favoráveis a ocorrência de um evento "E" e "b" resultados possíveis não favoráveis, sendo os mesmos mutuamente excludentes, então:

$$P(E) = \frac{a}{a+b},$$

onde os resultados devem ser verossímeis (possível e verdadeiro) e permitem a observação dos valores da probabilidade antes de ser observada qualquer amostra do evento (E).

4.3.3 Definição Axiomática

Seja (Ω) um experimento, seja (S) um espaço amostral associado a (Ω). A cada evento (E) associa-se um número real representado por P(E) e denominaremos de probabilidade de E, satisfazendo as seguintes propriedades:

a - $0 \leq P(E) \leq 1$;

b - $P(S) = 1$;

c - Se A e B são eventos mutuamente excludentes, então:

$$P(A \cup B) = P(A) + P(B).$$

d - Se $A_1, A_2, ..., A_n$ são eventos mutuamente excludentes dois a dois, então:

$$P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n).$$

As propriedades anteriores são conhecidas como axiomas da teoria da probabilidade. Os axiomas, muitas vezes, se inspiram em resultados experimentais e que, assim, definem a probabilidade de forma que possa ser confirmada experimentalmente.

4.3.4 Teoremas Fundamentais

Teorema 1 - Se $\varnothing$ for evento vazio, então $P(\varnothing) = 0$.

Teorema 2 - Se o evento $\overline{A}$ for o evento complementar de A, então

$P(\overline{A})=1 - P(A)$.

Teorema 3 - Se A e B são eventos quaisquer, então:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

RESUMO

$$\begin{array}{c} \cup \\ +\,(\text{OU}) \end{array} \left\{ \begin{array}{c} \text{Mutuamente Excludentes} \Rightarrow P(A \cup B) = P(A) + P(B) \\ A \cap B = \varnothing \\ \text{Quaisquer} \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B) \\ A \cap B \neq \varnothing \end{array} \right.$$

$$\begin{array}{c} \cap \\ \text{x (E)} \end{array} \left\{ \begin{array}{c} \text{Independentes} \Rightarrow P(A \cap B) = P(A) \text{ x } P(B) \\ A \cap B \neq \varnothing \\ \text{Condicionados} \Rightarrow P(A \cap B) = P(A) \text{ x } P(B/A) \end{array} \right.$$

Obs.: Os eventos mutuamente excludentes são dependentes e os eventos quaisquer são independentes;
Os eventos dependentes podem ser quaisquer ou mutuamente excludentes.

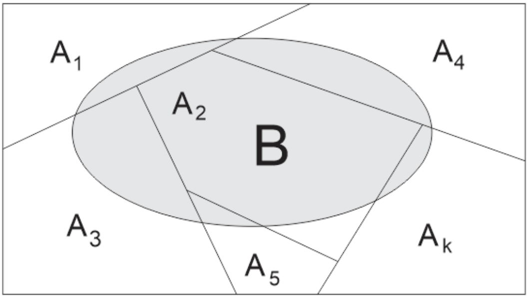
4.3.5 Teorema de Bayes

Definição: Seja S um espaço amostral e A_1, A_2, ... A_k, k eventos. Diz-se que A_1, A_2, ... A_k formam uma partição de S se:

$A_i \neq \varnothing,\ i = 1,\ 2,\ \ldots,\ k$;

$A_1 \cup A_2 \cup \ldots \cup A_k = S$;

$A_i \cap A_j = \varnothing, i \neq j$.



A_1, A_2, ... A_k formam uma partição de S.

Figura 4.7 - Diagrama representativo do Teorema de Bayes

Seja B um evento qualquer de S, onde:

$$B=(B\cap A_1)\cup(B\cap A_2)\cup\ldots\cup(B\cap A_k)$$

$$P(B)=\sum_{i=1}^{k}P(A_i).P\left(B/A_i\right) \text{, i = 1, 2, ..., k} \qquad (4.1)$$

$$P(B\cap A_i)=P(A_i).P\left(B/A_i\right) \qquad (4.2)$$

como

$$P\left(A_i/B\right)=\frac{P(B\cap A_i)}{P(B)} \qquad (4.3)$$

substituindo as equações (4.1) e (4.2) na equação (4.3) temos:

$$P\left(A_i/B\right)=\frac{P(A_i).P\left(B/A_i\right)}{\sum_{i=1}^{k}P(A_i).P\left(B/A_i\right)} \text{, i = 1, 2, ..., k .}$$

4.4 VARIÁVEIS ALEATÓRIAS

Ao descrever um espaço amostral (S) associado a um experimento (Ω) especifica-se que um resultado individual necessariamente, seja um número. Contudo, em muitas situações experimentais, estaremos interessados na mensuração de alguma coisa e no seu registro como um número.

Definição: Seja (Ω) um experimento aleatório e seja (S) um espaço amostral associado ao experimento. Uma função de X, que associe a cada elemento $s \in S$ um número real x(s), é denominada variável aleatória.

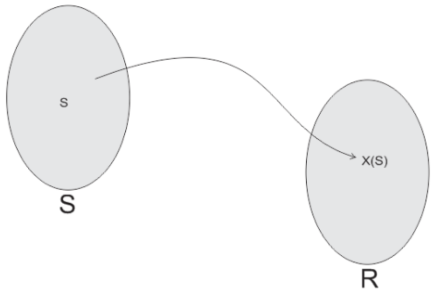


Figura 4.8 - Representação de uma variável aleatória

Uma variável X será aleatória discreta (V.A.D.) se o número de valores de x(s) for finito ou infinito numerável. Caso encontrarmos x(s) em forma de intervalo ou um conjunto de intervalos, teremos uma variável aleatória contínua (V.A.C.).

4.5 FUNÇÃO DE PROBABILIDADE

A probabilidade de que uma variável aleatória "X" assuma o valor "x" é uma função de probabilidade, representada por P(X = x) ou P(x).

4.5.1 Função de Probabilidade de uma V.A.D.

A função de probabilidade para uma variável aleatória discreta é chamada de função de probabilidade no ponto, ou seja, é o conjunto de pares $(x_i \,;\, P(x_i))$, $i = 1, 2, ..., n, ...$, conforme mostra a Figura 4.9.

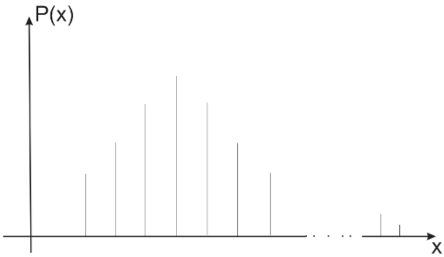


Figura 4.9 - Distribuição de probabilidade de uma V.A.D.

Para cada possível resultado de x teremos:

$0 \le P(X) \le 1$;

$$\sum_{i=1}^{\infty} P(X_i) = 1.$$

4.5.2 Função Repartição de uma V.A.D.

Seja X uma variável aleatória discreta.

Define-se Função Repartição da Variável Aleatória X, no ponto x_i, como sendo a probabilidade de que X assuma um valor menor ou igual a x_i, isto é:

$$F(X) = P(X \le x_i)$$

Propriedades:

1ª) $F(X) = \sum_{x_i \le x} P(x_i)$;

2ª) $F(-\infty) = 0$;

3ª) $F(+\infty) = 1$;

4ª) $P\left(a < x \leq b\right) = F(b) - F(a)$;

5ª) $P\left(a \leq x \leq b\right) = F(b) - F(a) + P(X = a)$;

6ª) $P\left(a < x < b\right) = F(b) - F(a) - P(X = b)$;

7ª) $F(X)$ é contínua à direita $\Rightarrow \lim\limits_{x \to x_0} F(X) = F\left(X_o\right)$. ;

8ª) $F(X)$ é descontínua à esquerda, nos pontos em que a probabilidade é diferente de zero;

$$\lim_{x \to x_0} F(X) \neq F\left(X_o\right), \text{ para } P\left(X = x_0\right) \neq 0$$

9ª) A função não é decrescente, isto é, $F(b) \geq F(a)$ para $b > a$.

4.5.3 Esperança Matemática de uma V.A.D.

Definição: Seja X uma V.A.D., com valores possíveis x_1, x_2 ,..., x_n, ...; Seja $P(x_i) = P(X = x_i)$, $i = 1, 2, ..., n, ...$. Então, o valor esperado de X (ou Esperança Matemática de X), denotado por $E(X)$ é definido como:

$$E(X) = \sum_{i=1}^{\infty} x_i \; P(x_i)$$

se a série $E(X) = \sum\limits_{i=1}^{\infty} x_i \; P(x_i)$ convergir absolutamente, isto é, se $\sum\limits_{i=1}^{\infty} |x_i| \; P(x_i) < \infty$, este número é também denominado o valor médio de X, ou expectância de X.

4.5.4 Variância de uma V.A.D.

Definição: Seja X uma V.A.D.. Define-se a variância de X, denotada por V(X) ou σ_x^2, da seguinte maneira:

$$V(X) = \sum_{i=1}^{\infty} \left(x_i - E(X)\right)^2 . P(x_i) \text{ ou } V(X) = E(X^2) - \left[E(X)\right]^2,$$

onde $E(X^2) = \sum\limits_{i=1}^{\infty} x_i^2 \; P(x_i)$ e a raiz quadrada positiva de V(X) é denominada o desvio-padrão de X, e denotado por σ_X.

4.5.5 Função de Probabilidade de uma V.A.C.

No instante em que X é definida sobre um espaço amostral contínuo, a função de probabilidade será contínua, onde a curva limitada pela área em relação aos valores de x será igual a 1.

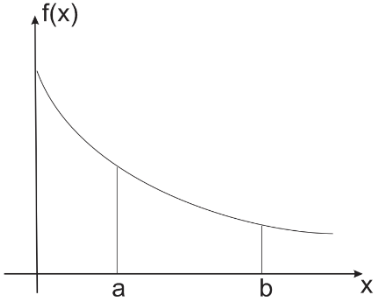


Figura 4.10 - Distribuição de probabilidade de uma V.A.C.

Se quisermos calcular a probabilidade de X assumir um valor x entre "a" e "b" devemos calcular:

$$P(a \le x \le b) = \int_a^b f(x)\, dx \cdot$$

Pelo fato de que a área representa probabilidade, e a mesma tem valores numéricos positivos, logo a função precisa estar inteiramente acima do eixo das abscissas (x).

Definição: A função f(x) é uma Função Densidade de Probabilidade (f.d.p.) para uma V.A.C. X, definida nos reais quando:

$$f(x) \ge 0;$$

$$\int_{-\infty}^{+\infty} f(x)\, dx = 1;$$

$$P(a \le x \le b) = \int_a^b f(x)\, dx.$$

4.5.6 Função Repartição de uma V.A.C.

Seja X uma variável aleatória contínua.

Define-se Função Repartição da Variável Aleatória X, no ponto x_i, como sendo:

$$F(X) = \int_{-\infty}^{x} f(x)\ dx\ .$$

Lembrete: Para variáveis aleatórias contínuas a probabilidade num intervalo entre a e b pode ser expressão como:

$$P(a \le x \le b) = P(a < x < b) = P(a < x \le b) = P(a \le x < b)$$

4.5.7 Esperança Matemática de uma V.A.C.

Definição: Seja X uma V.A.C. com f.d.p. f(x). O valor esperado de X é definido como:

$$E(X) = \int_{-\infty}^{+\infty} x.f(x)\ dx,$$

pode acontecer que esta integral imprópria não convirja.

Conseqüentemente, diremos que E(X) existirá se, e somente se,

$\int_{-\infty}^{+\infty} |\ x\ |\ f(x)$ for finita.

4.5.8 Variância de uma V.A.C.

Definição: Seja X uma V.A.C. de uma função distribuição de probabilidade (f.d.p.). A variância de X é:

$V(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x)\ dx$ ou $V(X) = E(X^2) - [E(X)]^2$.

onde:

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 \; f(x) \; dx \cdot$$

4.6 VARIÁVEL ALEATÓRIA BIDIMENSIONAL

Sejam E um experimento aleatório e S um espaço amostral associado a E. Sejam x = x(s) e y = y(s) duas funções, cada uma associando um numero real a cada resultado s $\in$ S, denomina-se (X,Y) uma variável aleatória bidimensional (V.A.B.).

Uma variável aleatória bidimensional (VAB) (X,Y) pode ser discreta ou contínua. Uma VAB será discreta (V.A.B.D.) se o número de valores de x(s) e y(s) for finito ou infinito numerável e será continua (V.A.B.C.) se encontrarmos x(s) e y(s) em forma de intervalo ou um conjunto de intervalos.

4.6.1 Variáveis Aleatórias Discretas Bidimensionais

- Distribuições conjuntas

Seja (X,Y) uma V.A.D.B., a cada possível valor (x_i, y_i) associaremos um numero real $p(x_i, y_i)$ representado por $p(X = x_i, Y = y_i)$ satisfazendo as condições:

i) $p(x_i, y_i) \geq 0$;

ii) $\sum_{i}^{\infty} \sum_{j}^{\infty} p(x_i, y_i) = 1$ ou $\sum_{i,j}^{\infty} p(x_i, y_i) = 1$.

- Distribuição de Probabilidade Marginal

Dada uma variável aleatória bidimensional e sua distribuição conjunta pode-se determinar a distribuição de X sem considerar Y, ou vice-versa.

Seja (X,Y) uma V.A. Discreta então:

$p(x_i) = P(X = x_i) = \sum_{j=1}^{\infty} (x_i, y_j)$ distribuição marginal de x;

e $q(y_i) = P(Y = y_i) = \sum_{i=1}^{\infty} (x_i, y_j)$ distribuição marginal de y.

- Variáveis Aleatórias Independentes

Seja (X , Y) uma variável aleatória bidimensional discreta, diz-se que X e Y são independentes se, e somente se,

$p(x_i, y_i) = p(x_i).p(y_j)$,

para quaisquer i e j.

- Distribuição de Probabilidade Condicionada

Seja (X,Y) uma variável aleatória discreta bidimensional com função de probabilidade conjunta $p(x_i, y_j)$, sejam $p(x_i)$e $p(y_j)$as funções de probabilidade marginais de x e y respectivamente.

A f.p. de x condicionada a um dado $y = y_i$ é definida por:

$p(x_i / y_j) = \dfrac{p(x_i, y_j)}{q(y_j)}$ se $q(y_j) > 0$.

A função de y condicionada a uma dado valor X=x_i é definida por:

$p(y_j / x_i) = \dfrac{p(x_i, y_j)}{p(x_i)}$ se $p(x_i) > 0$.

Valor Esperado (Função de variáveis aleatórias)

Sejam X e Y duas variáveis aleatórias discretas, a partir desta, pode-se considerar uma variável aleatória Z = X + Y ou uma variável Z = X.Y.

O valor esperado da variável (X + Y) é dado por:

$$E(X + Y) = \sum_x \sum_y (x + y)\, p(x, y) = E(X) + E(Y).$$

O valor esperado do produto XY, será o produto dos valores esperados, sempre que as variáveis forem independentes.

$$E(XY)=\sum_{x}\sum_{y} xy\, p(x,y)=E(X).E(Y).$$

- Covariância

É uma medida de relação linear entre duas variáveis aleatórias.

Definição: Sejam x e y duas variáveis aleatórias, a covariância de x e y é definida por:

$$COV(x,y)=E[(x-E(x))(y-E(y))]=E(xy)-E(x)E(y).$$

Quando a COV(x,y)=0 dizemos que x e y são não correlacionados.

- Correlação

É o grau de associação linear entre x e y.

Definição: Sejam x e y duas variáveis aleatórias, a covariância de x e y é definida por:

$$\rho(xy)=\frac{E(xy)-E(x)E(y)}{\sqrt{V(x)}\sqrt{V(y)}}=\frac{COV(x,y)}{DP(x).DP(y)},$$

onde $-1 \leq \rho(xy) \leq 1$.

4.6.2 Variáveis Aleatórias Contínuas Bidimensionais

- Distribuições conjuntas

Seja (x,y) uma variável aleatória bidimensional contínua. Diz-se que f(x,y) é uma função densidade de probabilidade conjunta se:

1) $f(x,y)) \geq 0;$

2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dx\, dy = 1 \text{ ou } \iint_{R} f(x,y)dx\, dy = 1.$

- Distribuição de Probabilidade Marginal

Dada uma variável aleatória bidimensional e sua distribuição conjunta podemos determinar a distribuição de x sem considerar y, e vice-versa. Tem-se:

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy \;;$$

$$h(x) = \int_{-\infty}^{\infty} f(x, y) dy \,.$$

onde g(x) distribuição marginal de x e h(x) distribuição marginal de y.

- Variáveis Aleatórias Independentes

Seja (x,y) uma variável aleatória contínua bidimensional, diz-se que x e y são independentes se, e somente se, $f(x, y) = g(x)h(x)$ para todo (x,y).

Para verificar se x e y são independentes, deve-se observar se todos os pares e condições são satisfeitas e logo após concluir que *x* e *y* são independentes.

- Distribuição de Probabilidade Condicionada

Seja (x,y) uma V.A.C.B., e a f.p. conjunta f(x,y). Sejam g(x) e h(x) as funções de probabilidade marginais de x e y, respectivamente.

A f.d.p. de x condicionada a um dado Y = y é:

$$g(x / y) = \frac{f(x, y)}{h(y)} \,,\; h(y) > 0 \,.$$

A f.d.p. de y condicionada a um dado X= x é:

$$h(y / x) = \frac{f(x, y)}{g(x)} \,,\; g(x) > 0 \,.$$

- Valor Esperado (Função de Variáveis Aleatórias)

Sejam x, y duas V.A.C.B., a partir desta pode-se considerar uma variável aleatória Z = X + Y ou uma variável Z = X.Y.

O valor esperado da variável (X + Y) é dado por:

$$E(x+y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)f(x,y)dxdy = E(x)+E(y).$$

O valor esperado do produto XY, será o produto dos valores esperados, sempre que as variáveis forem independentes. Logo:

$$E(xy) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)dxdy = E(x)E(y).$$

4.7 COEFICIENTE DE CORRELAÇÃO E COVARIÂNCIA

O grau de dispersão conjunta e associação linear entre duas variáveis aleatórias podem ser avaliados pela covariância e pelo coeficiente de variação, respectivamente.

Define–se a covariância entre x e y como:

$$COV(x,y) = E[(x-E(x))(y-E(y))] = E(xy) - E(x)E(y)$$

Define-se a correlação entre x e y por:

$$\rho_{xy} = \frac{E(xy)-E(x)E(y)}{\sqrt{V(x)}\sqrt{V(y)}}, \; -1 \le \rho_{xy} \le 1.$$

5 DISTRIBUIÇÕES DE PROBABILIDADE

Após termos visto as definições de V.A.D. e V.A.C., citaremos as principais distribuições de probabilidade relacionadas a estas variáveis.

5.1 DISTRIBUIÇÕES DISCRETAS DE PROBABILIDADE

5.1.1 Distribuição Binomial

O termo "Binomial" é utilizado quando uma variável aleatória está agrupada em duas classes ou categorias. As categorias devem ser mutuamente excludentes, de modo a deixar bem claro a qual categoria pertence determinada observação; e as classes devem ser coletivamente exaustivas, de forma que nenhum outro resultado fora dela seja possível.

Sejam, "p" probabilidades de sucesso e "q" probabilidades de falha, ou seja, p + q = 1.

A probabilidade de x sucessos em n tentativas é dado por p^x e de (n - x) falhas em (n - x) tentativas é dado por q^{n-x}, onde o número de vezes em que pode ocorrer x sucessos e (n-x) falhas é dado por:

$C_{n,x} = \binom{n}{x} = \frac{n!}{x!\,(n-x)!}$, x = 0 ,1, 2, ..., n e x! = x . ... . 4 . 3 . 2 . 1 .

logo, a probabilidade de ocorrer x sucessos com n tentativas será:

$$P(X = x) = \binom{n}{x} p^x\, q^{n-x}.$$

Propriedades necessárias para haver uma utilização da Distribuição Binomial:

1ª) Número de tentativas fixas;
2ª) Cada tentativa deve resultar numa falha ou sucesso;
3ª) As probabilidades de sucesso devem ser iguais para todas as tentativas;
4ª) Todas as tentativas devem ser independentes.

- Esperança Matemática de Distribuição Binomial

$E(X) = n \, . \, p$.

- Variância de uma Distribuição Binomial

$V(X) = n. \, p. \, q$.

5.1.2 Distribuição de Poisson

Consiste em uma distribuição discreta de probabilidade útil para estimar o número de ocorrências sobre um intervalo de tempo ou de espaço específicos. Por exemplo, número de chamadas telefônicas recebidas num determinado intervalo de tempo.

Características:

1ª) A probabilidade de ocorrência é a mesma para qualquer intervalo de igual comprimento;
2ª) As ocorrências devem ser aleatórias e independentes;

A função de probabilidade de Poisson é dado por:

$$P(X = x) = \frac{e^{-\lambda} . \lambda^{x}}{x!},$$

onde:

λ é o número médio de eventos ocorrido no intervalo considerado;
x = 0 ,1, 2, ..., n ;
e = 2,7183... ;
x! = x . ... . 4 . 3 . 2 . 1 .

- Esperança Matemática da Distribuição de Poisson

$E(x) = \lambda$.

- Variância da Distribuição de Poisson

$V(x) = \lambda$.

Obs.: A Distribuição Binomial se aproxima da Poisson, com $\lambda = n \cdot p$, quando o tamanho da amostra n é muito grande ($n \to \infty$) e a probabilidade p é pequena ($p \to 0$), na prática quando $n > 30$ e $p<0,05$.

5.2 DISTRIBUIÇÕES CONTÍNUAS DE PROBABILIDADE

5.2.1 Distribuição Uniforme

É uma distribuição de probabilidade usada para variáveis aleatórias contínuas, definida num intervalo $[a, b]$, e sua função densidade de probabilidade é dada por:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{se } a \le x \le b \\ 0 & \text{se } x < a \text{ ou } x > b \end{cases} .$$

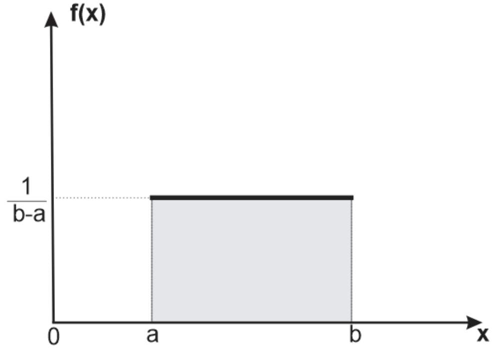


Figura 5.1 - Representação de uma Distribuição Uniforme

- Esperança Matemática da Distribuição Uniforme

$$E(X) = \frac{(b+a)}{2}.$$

- Variância da Distribuição Uniforme

$$V(X) = \frac{(b-a)^2}{12}.$$

5.2.2 Distribuição Normal ou Gaussiana

É um modelo de distribuição contínua de probabilidade, usado tanto para variáveis aleatórias discretas como contínuas.

Uma variável aleatória X, que tome todos os valores reais $-\infty < x < +\infty$ tem distribuição normal quando sua função densidade de probabilidade (f.d.p.) for da forma:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < +\infty \ .$$

Os parâmetros μ e σ seguem as seguintes condições:

$-\infty < \mu < +\infty$ e $\sigma > 0$.

- Propriedades da Distribuição Normal

a) O aspecto gráfico da função f(x) tem:

- Semelhança de um sino, unimodal e simétrico em relação a média μ.

- A especificação da média μ e do desvio padrão σ é completamente evidenciado.

- A área total da curva equivale a 100%.

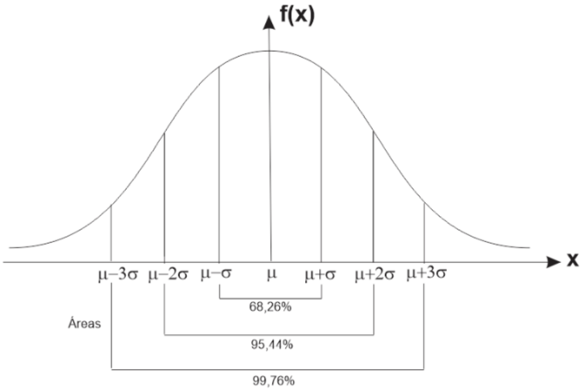


Figura 5.2 - Distribuição Normal em função da μ e σ

- Esperança Matemática da Distribuição Normal

$E(X) = \mu$.

- Variância da Distribuição Normal

$V(X) = \sigma^2$.

Distribuição Normal Padronizada

Tem como objetivo solucionar a complexidade da f(x) através da mudança de variável. f(z).

Fazendo $z = \dfrac{x-\mu}{\sigma}$ e z ~ N(0,1) temos que

$$f(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}},$$

com E(z) = 0 e VAR(z) = 1.

onde:

z = número de desvios padrões a contar da média;
x = valor arbitrário;

μ = média da distribuição normal;
σ = desvio padrão da distribuição normal.

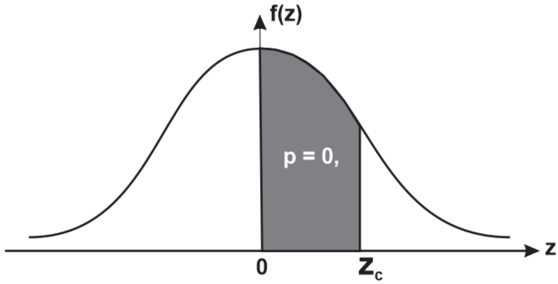


Figura 5.3 – Cálculo da probabilidade “p” na curva normal padrão

Estas probabilidades estão tabeladas e este caso particular é chamado de Forma Padrão da Distribuição Normal (Anexo B).

5.2.3 Distribuição “t” de Student

Trata-se de um modelo de distribuição contínua que se assemelha à distribuição Normal Padrão, N ~ (0,1). Ë utilizada para inferências estatísticas, particularmente, quando se tem amostras com tamanhos inferiores a 30 elementos.

A distribuição t também possui parâmetros denominado "grau de liberdade - φ".

A distribuição t é simétrica em relação a sua média.

$t = \frac{\overline{x} - \mu}{S_{\overline{x}}}$, onde:

$\overline{x}$: corresponde a média da amostra;

μ: corresponde a média populacional;

$S_{\overline{x}} = \frac{S}{\sqrt{n}}$: corresponde ao erro padrão da média.

A função de densidade é dada por:

$$f(t) = K \frac{1}{\left[1 + \frac{t^2}{n-1}\right]^{n/2}} \;.$$

onde: $-\infty < t < +\infty$ e K é a constante para cada valor de n;

Por se tratar de uma f.d.p. a área da curva é 1 e $\varphi = n - 1$ e corresponde aos graus de liberdade da distribuição.

A medida que φ aumenta $t \to Z$, observando que ao ultrapassar 30 graus de liberdade já é possível usar a Distribuição Normal, pois a diferença entre os resultados será bastante pequena.

Genericamente, existe uma família de distribuições "t", cuja forma tende à Distribuição Normal Padrão, à medida que *n* cresce (pois "s" tende a "σ" e, portanto, "t" tende a "Z") (Anexo B).

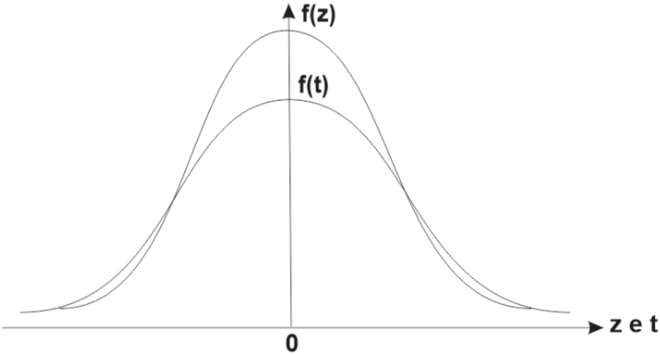


Figura 5.4 – Comparação das distribuições normal e t

Quando se usa a tabela para encontrar os valores das probabilidades, a coluna da esquerda fornece os graus de liberdade, a primeira linha fornece a área e o corpo da tabela fornece os valores de t.

ou seja, $P(|\,t\,| \leq t_0) = 1 - \alpha$.

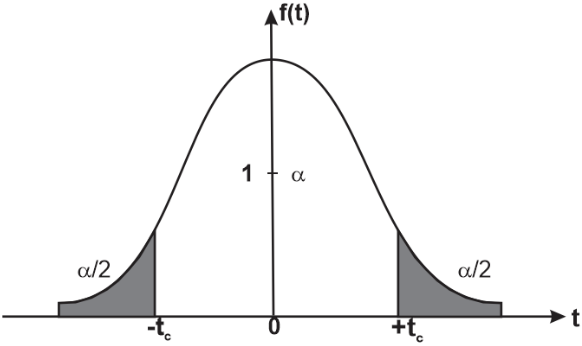


Figura 5.5 – Valores críticos da distribuição t

- Esperança Matemática da Distribuição t de Student

E(X) = 0

- Variância da Distribuição t de Student

$$\mathrm{VAR}\left[t_\varphi\right] = \sigma^2\left(t_\varphi\right) = \frac{\varphi}{\varphi - 2} = \frac{n-1}{n-3}$$, para $\varphi > 2$.

5.2.4 Distribuição Qui-quadrado (χ2)

A Distribuição Qui-quadrado possui numerosas aplicações importantes em inferência estatística. Uma das mais importantes é em testes não paramétricos tais como: testes de aderência e testes de independência.

É um caso particular da distribuição Gama, onde a variável aleatória, com valores positivos, tem uma distribuição qui-quadrado com ϕ graus de liberdade, denotada por $\chi^2(\phi)$, cuja função densidade é dada por:

$$f(x;\phi) = \frac{x^{(\phi/2)-1} . e^{-x/2}}{\Gamma(\phi/2).2^{\phi/2}}$$, x > 0.

Dependendo do número de graus de liberdade, a distribuição χ^2 assume as seguintes formas gráficas:

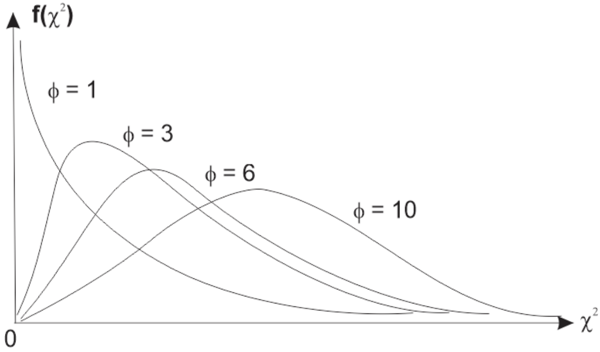


Figura 5.6 – Variações da distribuição qui-quadrado em função de ϕ

1) Para n > 30, a distribuição tende a normal;
2) Para $\phi = 1 \rightarrow \chi_1^2 = Z^2$, isto é, a variável χ^2 é igual ao quadrado de uma normal reduzida;

Sejam x_1, x_2, …, x_n, V.A. independentes, normalmente distribuídas com média zero e variância σ^2. Define-se a V.A. χ^2, com ϕ graus de liberdade como sendo a soma do quadrado de ϕ variáveis normais padronizadas e independentes, isto é:

$$\chi_\phi^2 = \sum_{i=1}^{\phi} z_i^2 = \sum_{i=1}^{\phi}\left(\frac{x_i - \mu}{\sigma}\right)^2 .$$

- Esperança Matemática da Distribuição Qui-quadrado

E(X) = ϕ .

- Variância da Distribuição Qui-quadrado

V(X) = 2 ϕ .

- Propriedade da aditividade:

Sejam duas V.A. $\chi^2_{\phi_1}$ e $\chi^2_{\phi_2}$ **independentes** com distribuição χ^2, com (ϕ_1 + ϕ_2) graus de liberdade, têm-se que:

$$\chi^2_{\phi_1} + \chi^2_{\phi_2} = \chi^2_{\phi_1+\phi_2} \ .$$

Podendo ser generalizada para k variáveis χ^2 independentes.

A Distribuição χ^2 constitui-se de uma família de curvas, onde cada uma é tabelada e caracterizada pelos graus de liberdade ϕ. A curva mais freqüente é a curva unicaudal à direita, isto é:

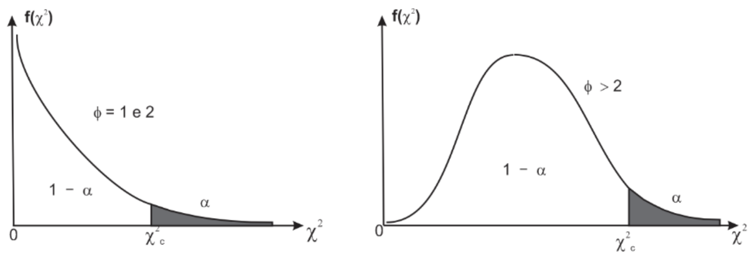


Figura 5.7 – Valores críticos da distribuição qui-quadrado em função de ϕ

Para uma dada probabilidade α e para um dado ϕ, o corpo da tabela fornece o valor de χ^2_0 tal que $P(\chi^2 \geq \chi^2_0) = \alpha$, probabilidade essa representada, na figura, pela região hachurada.

- Diferenças em relação a curva normal:

1) É sempre positiva;
2) É assimétrica;
3) A tabela fornece o valor χ^2 a partir de uma probabilidade α e um certo número de graus de liberdade ϕ.

5.2.5 Distribuição F de Snedecor

A Distribuição **F** é usada em um dos testes mais importantes da estatística, que é a análise de variância.
Suponha-se U e V duas amostras independentes retiradas de populações normais forneçam variâncias amostrais S_u^2 e S_v^2 e que deseja-se conhecer a distribuição amostral do quociente S_u^2 / S_v^2. Isso será possível através da Distribuição *F* de Snedecor, conhecendo-se os graus de liberdade ϕ_u e ϕ_v, cuja distribuição é definida por:

$$W = \frac{U/\phi_u}{V/\phi_v}.$$

$$f(w;\phi_u;\phi_v) = \frac{\Gamma((\phi_u + \phi_v)/2)}{\Gamma(\phi_u/2)\Gamma(\phi_v/2)} \left(\frac{\phi_u}{\phi_v}\right)^{\phi_u/2} \frac{w^{(\phi_u - 2)/2}}{(1+\phi_u f/\phi_v)^{(\phi_u+\phi_v)/2}}, \ w > 0.$$

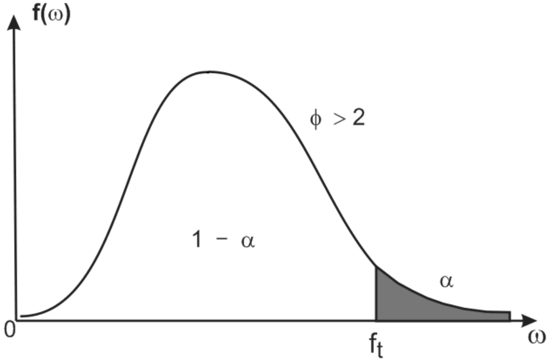


Figura 5.8 – Gráfico da distribuição F

Diz-se que W tem distribuição F de Snedecor, com ϕ_u e ϕ_v graus de liberdade. Para se encontrar o valor da abscissa para outros níveis de probabilidade na cauda à esquerda, ou seja, $F_{1-\alpha}(\phi_u,\phi_v)$ é dado por:

$$F_{1-\alpha}(\phi_u,\phi_v) = \frac{1}{F_\alpha(\phi_v,\phi_u)}.$$

- Esperança Matemática da Distribuição F de Snedecor

$$E(W) = \frac{\phi_v}{\phi_v - 2}.$$

- Variância da Distribuição F de Snedecor

$$V(W) = \frac{2\phi_v^2(\phi_u + \phi_v - 2)}{\phi_u(\phi_v - 2)^2(\phi_v - 2)}.$$

Os valores da estatística **F** apresentam-se tabelados, onde a linha mostra os graus de liberdade do numerador (ϕ_u) e a coluna os graus de liberdade do denominador (ϕ_v). No cruzamento de uma linha com uma coluna aparecem os valores tabelados.

No Quadro 5.1 mostrar-se-ão os principais modelos para variáveis aleatórias contínuas, incluindo a densidade, a média e a variância.

Quadro 5.1 – Modelos para variáveis contínuas

Modelo	f(x)	Média	Variância
Uniforme	$\frac{1}{b-a}, \; a < x < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal	$\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < +\infty$	μ	σ^2
t - Student	$K \frac{1}{\left[1 + \frac{t^2}{n-1}\right]^{n/2}}$	0	$\frac{\varphi}{\varphi - 2}$
Qui-quadrado	$\frac{x^{(\phi/2)-1}.e^{-x/2}}{\Gamma(\phi/2).2^{\phi/2}}$	ϕ	2ϕ
F - Snedecor	$\frac{\Gamma((\phi_u + \phi_v)/2)}{\Gamma(\phi_u/2)\Gamma(\phi_v/2)} \left(\frac{\phi_u}{\phi_v}\right)^{\phi_u/2} \frac{w^{(\frac{\phi_u - 2}{2}}}{(1 + \phi_u f/\phi_v)^{\frac{(\phi_u + \phi_v)}{2}}}$	$\frac{\phi_v}{\phi_v - 2}$	$\frac{2\phi_v^2(\phi_u + \phi_v - 2)}{\phi_u(\phi_v - 2)^2(\phi_v - 2)}$

5.3 ESCORE PADRONIZADO (ESCORE Z)

Sabe-se que determinados tipos de testes padronizados possuem sistemas próprios de pontuação. O SAT (*Scholastic Aptitude Test* - Teste de Aptidão Escolar), utilizado pelo ENEM, é pontuado em uma escala de 800 pontos para cada seção. Quando um candidato recebe suas notas, ele só sabe o quão bem ele marcou para esse teste. Este é o escore bruto, ou escore padrão.

O aluno não vai saber como a sua pontuação se compara com outras instituições nacionais ou estadual, sem outras informações disponíveis para ele.

O escore padronizado (escore z), ao contrário do Coeficiente de Variação, é útil para comparação dos resultados individuais. O escore z é uma pontuação que recai em algum lugar dentro da distribuição normal padrão.

O escore z é uma excelente medida de avaliação de desempenho.

Sejam x_1, x_2, ...x_n os dados observados em uma amostra de tamanho n e $\bar{x}$ a média e S o desvio padrão, então:

$\text{escore } z_i = \dfrac{x_i - \bar{x}}{s}$, i = 1, 2, ..., n , é denominado escore z padronizado.

O escore padrão ou escore z, representa o número de desvios padrões que separa um dado valor x_i da média.

Exemplo:

Um concurso teve pontuação média de 152 pontos e desvio padrão de 7 pontos. Um candidato x_1 tirou 161 pontos e outro x_2 tirou 146 pontos, determine os valores de z para classificar as pontuações em baixa (0 a 33,3%), média (33,3 a 66,7%) e alta (66,7 a 100%) em função do escore padronizado e verifique em que área os candidatos se enquadram.

Observando no Anexo (Tabela 1),

p = 0,33333 (0,5 – 0,33333 = 0,16667), o valor de z = - 0,43σ

p = 0,66667 (0,66667 – 0,5) = 0,16667), z = + 0,43σ

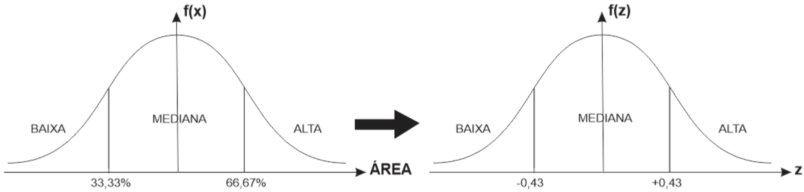


Figura 5.9 – Transformação para o escore z

Para quem teve 161 pontos, qual a sua classificação?

$$\text{escore } z_1 = \frac{x_1 - \bar{x}}{s} = \frac{161 - 152}{7} = 1{,}29\sigma$$

como 1,29 é maior que 0,43, essa pontuação classifica-se em ALTA, por estar acima de 66,67% das pontuações.

Para quem teve 146 pontos, qual a sua classificação?

$$\text{escore } z_2 = \frac{x_2 - \bar{x}}{s} = \frac{146 - 152}{7} = -0{,}86\sigma$$

como -0,86 é um valor menor que -0,43 essa pontuação classifica-se em BAIXA, ou seja, menor que 33,33% das pontuações.

6 AMOSTRAGEM

6.1 CONCEITOS EM AMOSTRAGEM

Inferência Estatística - É o processo de obter informações sobre uma população a partir de resultados observados numa amostra.

Amostragem: É o processo de retirada de informações dos "n" elementos amostrais, na qual deve-se seguir um método adequado (tipos de amostragem).

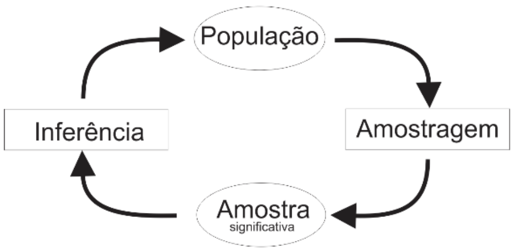


Figura 6.1 – Ciclo da amostragem

A amostragem é uma etapa de grande importância no delineamento da pesquisa capaz de determinar a validade dos dados obtidos. Sua idéia básica refere-se "à coleta de dados relativos a alguns elementos da população e a sua análise, que pode proporcionar informações relevantes sobre toda a população" (MATTAR, p. 128).

Dentre os elementos do planejamento de pesquisa está o plano de amostragem. Segundo Schiffman e Kanuk (2000, p. 26), um plano de amostragem deve responder às seguintes questões: quem pesquisar (unidade de amostragem), quantos pesquisar (o tamanho da amostra) e como selecionar (o procedimento da amostragem).

A decisão de quem pesquisar exige que o universo seja definido de modo que uma amostra adequada possa ser selecionada. As entrevistas realizadas com o público correto são fatores básicos para

a validade do estudo. O tamanho da amostra depende do orçamento disponível e do grau de confiança que a empresa quer alocar aos resultados.
O procedimento de amostragem pode ser realizado por meio de uma amostra probabilística ou não probabilística. No primeiro caso, os resultados podem ser projetáveis para a população total, já no segundo caso, os resultados não podem ser generalizados.

Para a escolha do processo de amostragem, o pesquisador deve levar em conta o tipo de pesquisa, a acessibilidade aos elementos da população, a disponibilidade ou não de ter os elementos da população, a representatividade desejada ou necessária, a oportunidade apresentada pela ocorrência de fatos ou eventos, a disponibilidade de tempo, recursos financeiros e humanos etc. (MATTAR, 1996, p. 133).

6.2 PLANO DE AMOSTRAGEM

1º) Definir os Objetivos da Pesquisa

2º) População a ser Amostrada
- Parâmetros a serem estimados (Objetivos)

3º) Definição da Unidade Amostral
- Seleção dos elementos que farão parte da amostra

4º) Forma de seleção dos elementos da população

$$\text{Tipo de Amostragem Probabilística}\begin{cases}\text{- Aleatória Simples}\\ \text{- Sistemática}\\ \text{- Estratificada}\\ \text{- por Conglomerado}\end{cases}$$

$$\text{Tipo de Amostragem Não probabilística}\begin{cases}\text{- Conveniência}\\ \text{- Intencional}\\ \text{- Quotas}\end{cases}$$

5º) Tamanho da Amostra

Ex.: Moradores de uma Cidade (população alvo)

Objetivo: Tipo de Residência $\begin{cases} \text{própria} \\ \text{alugada} \\ \text{emprestada} \end{cases}$ $\begin{cases} \text{um piso} \\ \text{dois pisos} \\ \text{três ou mais pisos} \end{cases}$

Unidade Amostral: Domicílios (residências)

Elementos da População: Família por domicílio

6.3 TIPOS DE AMOSTRAGEM

- Probabilísticas

É aquela em que cada elemento da população tem uma chance conhecida e diferente de zero de ser selecionado para compor a amostra. As amostragens probabilísticas geram amostras probabilísticas (MATTAR, p. 132).

6.3.1 Amostragem Aleatória Simples ou Ocasional

É o processo mais elementar e freqüentemente utilizado. Todos os elementos da população tem igual probabilidade de serem escolhidos. Para uma população finita o processo deve ser sem reposição. Todos os elementos da população devem ser numerados. Para realizar o sorteio dos elementos da população devemos usar a **Tabela de Números Aleatórios** (Anexo B).

6.3.2 Amostragem Sistemática

Trata-se de uma variação da Amostragem Aleatória Ocasional, conveniente quando a população está naturalmente ordenada, como fichas em um fichário, lista telefônica, etc.

Ex.: N = 5000   n = 50, então $r = \frac{N}{n} = 10$, (P. A. de razão 10)

Sorteia-se usando a **Tabela de Números Aleatórios** um número entre 1 e 10, (x = 3), o número sorteado refere-se ao 1º elemento da amostra, logo os elementos da amostra serão:

3 13 23 33 43 ......

Para determinar qualquer elemento da amostra podemos usar a equação do termo geral de uma Progressão Aritmética (PA).

$$a_n = a_1 + (n-1).r \ .$$

6.3.3 Amostragem Estratificada

É um processo de amostragem usado quando nos depararmos com populações heterogêneas, na qual pode-se distinguir subpopulações mais ou menos homogêneas, denominados estratos.

Após a determinação dos estratos, seleciona-se uma amostra aleatória de cada subpopulação (estrato).

As diversas subamostras retiradas das subpopulações devem ser proporcionais aos respectivos números de elementos dos estratos, e guardarem a proporcionalidade em relação a variabilidade de cada estrato, obtendo-se uma estratificação ótima.

Esta técnica consiste em dividir a população em k subgrupos homogêneos.

$$N = N_1 + N_2 + ... + N_k = \sum_{i=1}^{k} N_i ;$$

$$n = n_1 + n_2 + ... + n_k = \sum_{i=1}^{k} n_i \ .$$

onde k é o número de estratos.

- Estratificada Proporcional

Neste caso a proporcionalidade do tamanho de cada estrato da população é mantida na amostra, ou seja:

$$\frac{n}{N}=\frac{n_i}{N_i}.$$

Este procedimento tenta gerar resultados mais precisos, quando comparados com o processo aleatório simples.

- Estratificada Uniforme

Esta situação é utilizada quando se pretende estimativas separadas para cada estrato, e possuem os tamanhos aproximadamente iguais.

$$N_1 \cong N_2 \cong ... \cong N_k \ ,$$

logo

$$n_1 \cong n_2 \cong ... \cong n_k \, .$$

6.3.4 Amostragem por Conglomerados (ou Agrupamentos)

Algumas populações não permitem, ou tornam extremamente difícil que se identifiquem seus elementos, mas se podem identificar subgrupos da população. Em tais casos, uma amostra aleatória simples desses subgrupos (conglomerados) pode ser escolhida, e uma contagem completa deve ser feita no conglomerado sorteado.

Agregados típicos são: quarteirões, famílias, organizações,
agências, edifícios, etc.

- Não probabilísticas

É aquela em que a seleção dos elementos da população para compor a amostra depende ao menos em parte do julgamento do pesquisador ou do entrevistador no campo (MATTAR, 1996, p. 132).

Uma razão para o uso de amostragem não probabilística pode ser a de não haver outra alternativa viável porque a população não está disponível para ser sorteada. Outra razão é que apesar da amostragem probabilística ser tecnicamente superior na teoria, ocorrem problemas em sua aplicação na prática o que enfraquece essa superioridade.

Segundo Mattar (1996) o resultado de um processo de amostragem probabilístico *“a priori“* pode resultar em um estudo não probabilístico devido a erros que os entrevistadores podem cometer quando não seguem corretamente as instruções. Outro motivo pode ser o e que a obtenção de uma amostra de dados que reflitam precisamente a população não seja o propósito principal da pesquisa.

“Se não houver intenção de generalizar os dados obtidos na amostra para a população, então não haverá preocupações quanto à amostra ser mais ou menos representativa da população. A última razão para usar amostragem não probabilística se refere às limitações de tempo, recursos financeiros, materiais e "pessoas". necessários para a realização de uma pesquisa com amostragem probabilística". (MATTAR, 1996, p. 157).

6.3.5 Amostragem por conveniência

A amostragem por conveniência é formada por indivíduos que o pesquisador reuniu simplesmente porque dispunha deles. Então, se o pesquisador tomar os enfermeiros de uma ala de hospital como amostra de todo hospital, a mesma será uma amostra por conveniência.

Os estatísticos têm muitas restrições ao uso de amostras de conveniência. Mesmo assim, a amostragem por conveniência são comuns nas áreas sociais e humanas e de saúde, em que se fazem pesquisas com indivíduos de uma só organização ou de um só hospital.

Mais ainda, as amostras por conveniência constituem, muitas vezes, a única maneira de estudar determinado problema. De qualquer forma, o pesquisador que utiliza amostras de conveniência precisa de muito senso crítico. Os dados podem ser tendenciosos.

Por exemplo, para estimar a probabilidade de se estar deprimido ou estressado não se deve recorrer aos dados de uma clínica. Como só são internados os casos de depressão ou estresse, é possível que a incidência dos pacientes internados seja bem maior do que entre pacientes não-internados. Conseqüentemente, a amostra por conveniência constituída, nesse exemplo, por pacientes internados na clínica, seria tendenciosa.

Aaker; Kumar e Day (1995, p. 187) comentam que este método também pode ser empregado em pré-testes de questionários. Mattar (1996, p. 133) ilustra os usos de pesquisa com amostras por conveniência nos seguintes casos:

- Solicitar a pessoas que voluntariamente testem um produto e que em seguida respondam a uma entrevista;
- Parar pessoas na rua e colher suas opiniões.
- Colocar linhas de telefone adaptadas para que durante um programa de televisão os telespectadores possam dar suas opiniões.

6.3.6 Amostragem intencional

A seleção de amostras intencionais ou por julgamento são realizadas de acordo com o **julgamento** do pesquisador. Se for adotado um critério razoável de julgamento, pode-se chegar a resultados favoráveis.

É comum a escolha de *experts* (profissionais especializados) quando se trata de amostras por julgamento. Aaker; Kumar e Day (1995, p. 376) argumentam que a escolha de *experts* é uma forma de amostragem por julgamento ou intencional usada para escolher elementos "típicos" e "representativos" para uma amostra.

A abordagem da amostragem por julgamento pode ser útil quando é necessário incluir um pequeno número de unidades na amostra. O método de julgamento é muito utilizado para a escolha de uma localidade "representativa" de um país na qual serão realizadas outras pesquisas, sendo algumas vezes até preferida em relação à seleção de uma localidade por métodos aleatórios.

A amostragem por julgamento também é útil quando é preciso obter uma "amostra deliberadamente enviesada". Aaker; Kumar e Day (1995, p. 376) explicam essa afirmação com o exemplo seguinte:

Quando se quer avaliar uma modificação em um produto ou serviço, pode-se identificar grupos específicos que estariam dispostos a dar sua opinião em relação à modificação. Se o pesquisador avaliar que este grupo não gostou da modificação, então ele poderia assumir que o resto da população também teria uma percepção negativa em relação à mudança. Se o grupo gostou da modificação, então mais pesquisas poderiam ser requeridas para se chegar a uma conclusão a respeito do assunto.

6.3.7 Amostragem por quotas

Segundo Mattar (1995, p. 134) a amostra por quotas constitui um tipo especial de amostra intencional, em que o pesquisador procura obter uma amostra que seja similar à população sob algum aspecto.

A seleção de amostra por quotas é a forma mais usual de amostragem não probabilística. Neste caso, são consideradas várias características da população, como áreas geográficas, sexo, idade, raça e uma medida qualquer de nível econômico onde a amostra pretende incluir proporções similares de pessoas com as mesmas características.

A idéia de amostragem por quotas sugere que se as pessoas são representativas em termos de características, elas também poderão ser representativas em termos da informação procurada pela pesquisa. Depois de serem identificadas as proporções de cada tipo a ser incluído na amostra, o pesquisador estabelece um número ou quota de pessoas que possuem as características determinadas e que serão contatadas pela pesquisa (Curwin e Slater, 1991, p. 8*).*

O entrevistador recebe instruções para continuar a amostragem até que a quota necessária tenha sido atingida em cada estrato. Uma pesquisa com amostragem por quotas poderá ser utilizado e trazer bons resultados quando as características relevantes para controle e delineamento da amostra forem conhecidas, estiverem disponíveis ao pesquisador, estiverem relacionadas ao objeto de estudo e se constituírem em poucas categorias.

As amostras por quotas são muito usadas em pesquisa de opinião eleitoral e pesquisas de mercado. "O processo de quotas produz amostras com tendências, embora esteja freqüentemente de acordo com as amostras pelas probabilidades quando se trata de questões de opinião e pesquisa" (COCHRAN, p. 191).

6.4 DISTRIBUIÇÃO POR AMOSTRAGEM

Consideram-se todas as possíveis amostras de tamanho “n” retiradas da população. Para cada amostra calcula-se a estatística de interesse, obtendo-se, desta maneira, uma distribuição desses resultados, originando uma Distribuição Amostral.

6.4.1 Amostragem "COM" e "SEM" reposição

Seja "N" o número de elementos de uma população, e "n" o número de elementos de uma amostra, então, como já descrito no capítulo 1:

Se o processo de retirada dos elementos for COM reposição (população infinita (fr $\leq$ 5%)), o número de amostras possíveis será:

nº de amostras = N^n

Se o processo de retirada de elementos for SEM reposição (população finita (fr > 5%)), o número de amostras possíveis será:

nº de amostras = $C_{N,n} = \dfrac{N!}{n!(N-n)!}$

Distribuição Amostral das Médias

- (COM reposição)

N = { 1, 2, 3, 4} n = 2 nº de amostras = N^n = $4^2 = 16$

$$
\begin{array}{cccc}
\{1,1\} & \{1,2\} & \{1,3\} & \{1,4\} \\
\{2,1\} & \{2,2\} & \{2,3\} & \{2,4\} \\
\{3,1\} & \{3,2\} & \{3,3\} & \{3,4\} \\
\{4,1\} & \{4,2\} & \{4,3\} & \{4,4\}
\end{array}
$$

- (SEM reposição)

N = { 1, 2, 3, 4} n = 2 nº de amostras = $C_{4,2} = \dfrac{4!}{2!\ 2!} = 6$

$$\{1,2\} \quad \{1,3\} \quad \{1,4\}$$
$$\{2,3\} \quad \{2,4\} \quad \{3,4\}$$

Para ilustrar as estatísticas amostrais usaremos o processo COM reposição.

$$\{1,1\} \Rightarrow \bar{x} = 1{,}0 \quad \{1,2\} \Rightarrow \bar{x} = 1{,}5 \quad \{1,3\} \Rightarrow \bar{x} = 2{,}0 \quad \{1,4\} \Rightarrow \bar{x} = 2{,}5$$
$$\{2,1\} \Rightarrow \bar{x} = 1{,}5 \quad \{2,2\} \Rightarrow \bar{x} = 2{,}0 \quad \{2,3\} \Rightarrow \bar{x} = 2{,}5 \quad \{2,4\} \Rightarrow \bar{x} = 3{,}0$$
$$\{3,1\} \Rightarrow \bar{x} = 2{,}0 \quad \{3,2\} \Rightarrow \bar{x} = 2{,}5 \quad \{3,3\} \Rightarrow \bar{x} = 3{,}0 \quad \{3,4\} \Rightarrow \bar{x} = 3{,}5$$
$$\{4,1\} \Rightarrow \bar{x} = 2{,}5 \quad \{4,2\} \Rightarrow \bar{x} = 3{,}0 \quad \{4,3\} \Rightarrow \bar{x} = 3{,}5 \quad \{4,4\} \Rightarrow \bar{x} = 4{,}0$$

- Representações de uma Distribuição Amostral

- **Tabela**

$\bar{X}_i$	$P(\bar{X} = \bar{x}_i)$
1,0	1/16
1,5	2/16
2,0	3/16
2,5	4/16
3,0	3/16
3,5	2/16
4,0	1/16
Σ	16/16

- **Gráfico**

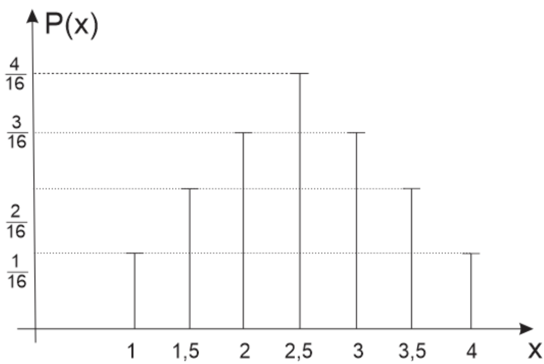


Figura 6.2 – Gráfico das médias e suas respectivas probabilidades

6.5 DISTRIBUIÇÕES AMOSTRAIS DE PROBABILIDADE

6.5.1 Distribuição amostral das médias

Se a variável aleatória "x" segue uma distribuição normal:

$\bar{x} \sim N\left(\mu(\bar{x}); \sigma^2(\bar{x})\right)$, onde $z = \dfrac{\bar{x} - \mu(\bar{x})}{\sigma(\bar{x})}$.

$\mu(\bar{x}) = \mu$ (a média da distribuição amostral é igual a média populacional)

$\sigma(\bar{x}) = \dfrac{\sigma(x)}{\sqrt{n}}$ (Desvio Padrão Amostral).

- Caso COM reposição (pop. infinita)

$\bar{x} \sim N\left[\mu(\bar{x});\ \dfrac{\sigma^2(x)}{n}\right]$.

- Caso SEM reposição (pop. finita)

Quando a amostra for > 5% da população $\left(\dfrac{n}{N}\right)$ devemos usar um fator de correção.

$\bar{x} \sim N\left[\mu(\bar{x});\ \dfrac{\sigma^2(x)}{n}\ \dfrac{N-n}{N-1}\right]$, onde $\dfrac{N-n}{N-1}$ é o fator de correção.

6.5.2 Distribuição amostral das proporções

$p = \bar{p}$ e $\bar{\sigma}_p = \sqrt{\dfrac{p(1-p)}{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}$,

onde $\sqrt{\dfrac{N-n}{N-1}}$ é usado para população finita.

6.5.3 Tipos de erros

Há dois tipos de erro ao testar uma hipótese estatística. Pode-se rejeitar uma hipótese quando ela é, de fato verdadeira, ou aceitar uma hipótese quando ela é, de fato, falsa. A rejeição de uma hipótese verdadeira é chamada "Erro tipo I". A aceitação de uma hipótese falsa constitui um "Erro tipo II".
As probabilidades desses dois tipos de erros são designadas, respectivamente, por α e β.

A probabilidade α do erro do tipo I é denominada "nível de significância" do teste.

Os possíveis erros e acertos de um teste estão sintetizados abaixo:

		Realidade	
		H_0 verdadeira	**H_0 falsa**
Decisão	**Aceitar H_0**	Decisão correta (1 - α)	Erro Tipo II (β)
	Rejeitar H_0	Erro Tipo I (α)	Decisão Correta (1 - β)

observe que o erro tipo I só poderá acontecer se for rejeitado H_0 e o erro tipo II quando for aceito H_0.

P(Erro tipo I) = P(rejeitar H_0 | H_0 é verdadeira) = α

P(Erro tipo II) = P(rejeitar H_0 | H_0 é falsa) = β

6.5.4 Poder do Teste

Definimos poder de um teste estatístico como a probabilidade do teste rejeitar H_0 quando H_0 é realmente falsa, ou seja, o poder de um teste é igual a 1 - β.

O poder do teste dependerá de alguns fatores:

- Do nível de significância α adotado;

- Da distância entre o valor "real" do parâmetro e o considerado verdadeiro em H_0;

- Da variabilidade da população;

- Do tamanho da amostra retirada.

Para o mesmo tamanho de amostra n:

Se o valor considerado como “real” for muito próximo daquele adotado em H_0, o teste terá maior dificuldade para detectar a diferença, logo terá **menor** poder, menor 1 - β, maior β, mas menor gravidade do erro.

Se o valor considerado como “real” for muito distante daquele adotado em H_0, o teste terá maior facilidade para detectar a diferença, logo terá **maior** poder, maior 1 - β, menor β, mas maior gravidade do erro.

Exemplo:

Um concurso teve pontuação média de 152 pontos e desvio padrão de 47 pontos, prova essa aplicada em 150 candidatos. Supondo que a distribuição das médias é normalmente distribuída e que a variância não tenha sido alterada. Use α = 1% e calcule o poder do teste se a verdadeira média fosse:

145	141	143	144	139	144	138	135	143	143	139	138

Média “real” = 141,0

H_0: μ = 152 α = 0,01 n = 150 σ = 47
H_1: μ < 152 $Z_{cal.}$ = - 2,58

$$\overline{x}_{cal.} = \mu - Z_{cal.}\frac{\sigma}{\sqrt{n}} = 152 - 2{,}58\frac{47}{\sqrt{150}} = 152 - 9{,}9 = 142{,}1$$

$$Z_{\beta} = \frac{\overline{x}_{cal.} - \overline{x}_{real}}{\sigma/\sqrt{n}} = \frac{142{,}1 - 141{,}0}{47/\sqrt{150}} = \frac{1{,}1}{3{,}84} = 0{,}29\sigma\ (0{,}5 - 0{,}11409)$$

β = 0,38591 Poder = 0,61409 (1 - β)

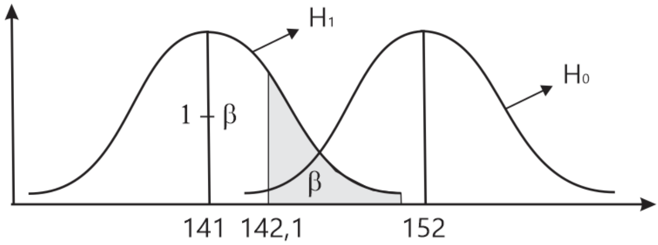


Figura 6.3 – Gráfico do poder do teste

6.6 TAMANHO DA AMOSTRA

6.6.1 Introdução

A técnica da amostragem, devido a sua larga utilização, ainda necessita de alguma didática mais adequada aos pesquisadores iniciantes.

Na teoria da amostragem, são consideradas duas dimensões:

1ª) Dimensionamento da Amostra;

2ª) Composição da Amostra.

6.6.2 Procedimentos para determinar o tamanho da amostra

1º) Analisar o questionário, ou roteiro da entrevista e escolher uma variável que julgue mais importante para o estudo. Se possível mais do que uma;

2º) Verificar o nível de mensuração da variável: nominal, ordinal ou intervalar;

3º) Considerar o tamanho da população: infinita ou finita;

4º) Se a estimativa escolhida for:

- Variância populacional conhecida (variável quantitativa)

População Infinita	População Finita
$n = \left(\dfrac{Z_{\alpha/2} \, . \, \sigma}{e} \right)^2$	$n = \dfrac{Z^2_{\alpha/2} \, . \, \sigma^2 . N}{e^2 (N-1) + Z^2_{\alpha/2} . \sigma^2}$

onde, Z é abscissa da curva normal padrão, fixada a um nível de confiança $(1-\alpha)$

Z = 1,65 $\rightarrow$ (1 - α) = 90%

Z = 1,96 $\rightarrow$ (1 - α) = 95%

Z = 2,0 $\rightarrow$ (1 - α) = 95,5%

Z = 2,33 $\rightarrow$ (1 - α) = 98%

Z = 2,57 $\rightarrow$ (1 - α) = 99%

σ = desvio padrão da população, expresso na unidade variável, onde poderá ser determinado por:

- Especificações Técnicas;
- Resgatar o valor de estudos semelhantes;
- Fazer conjecturas sobre possíveis valores.

e = erro amostral, expresso na unidade da variável. O erro amostral é a máxima diferença que o investigador admite suportar entre μ e $\overline{x}$, isto é: $|\mu - \overline{x}| < e$.

- Variância populacional desconhecida (variável quantitativa)

População Infinita	População Finita
$n = \left(\dfrac{t_{\varphi,\alpha/2} \, . \, s}{e} \right)^2$	$n = \dfrac{(t_{\varphi,\alpha/2})^2 \, . \, s^2 . N}{e^2 (N-1) + (t_{\varphi,\alpha/2})^2 . s^2}$

onde φ = n -1 graus de liberdade.

- Usando a proporção populacional (variável qualitativa)

População Infinita	População Finita
$n = \dfrac{Z^2_{\alpha/2} \, . \, \hat{p} \, . \, \hat{q}}{e^2}$	$n = \dfrac{Z^2_{\alpha/2} \, . \, \hat{p} \, . \, \hat{q} \, . \, N}{e^2 (N-1) + Z^2_{\alpha/2} \, . \, \hat{p} \, . \, \hat{q}}$

onde:

Z = abscissa da normal padrão;
N = tamanho da população;
$\hat{p}$ = estimativa da verdadeira proporção de um dos níveis da variável escolhida. Por exemplo, se a variável escolhida for parte da empresa, $\hat{p}$ poderá ser a estimativa da verdadeira proporção de grandes empresas do setor que está sendo estudado. $\hat{p}$ será expresso em decimais ($\hat{p}$ = 30% $\rightarrow$ $\hat{p}$ = 0,30);
$\hat{q} = 1 - \hat{p}$;
e = erro amostral, expresso em decimais. O erro amostral neste caso será a máxima diferença que o investigador admite suportar entre π e $\hat{p}$, isto é: $|\pi - \hat{p}| < e$, em que π é a verdadeira proporção (frequência relativa do evento a ser calculado a partir da amostra).

Estas fórmulas são básicas para qualquer tipo de composição da amostra; todavia, existem fórmulas específicas segundo o critério de composição da amostra.

Obs.: Quando não se tem condições de prever o possível valor para $\hat{p}$, admita $\hat{p}$ = 0,50, pois, dessa forma, você terá o maior tamanho da amostra, admitindo-se constantes os demais elementos.

6.6.3 Tamanho da amostra em estudos experimentais

O cálculo amostral para comparação de subgrupos (testes de hipóteses) dentro de uma amostra depende do teste estatístico escolhido, das diferenças entre os grupos, da tolerância do pesquisador à detecção de diferenças quando elas não existem (Erro tipo I) ou da falha na detecção de diferenças entre os subgrupos quando elas realmente existem (Erro tipo II). As probabilidades associadas aos erros tipos I e II são convencionadas como α e β, e comumente, são adotados valores de 5% (bilateralmente) e 20%, mas outros valores podem ser utilizados de forma criteriosa (MIOT, 2011).

Uma estratégia que permite a redução da variabilidade das medidas, aumentando a comparabilidade dos indivíduos em uma amostra, e, consequentemente, reduzindo a necessidade numérica amostral para

a detecção de um fenômeno, é o pareamento (ou emparelhamento) das observações (Quadro 6.2).
Isso ocorre quando um mesmo indivíduo é observado em diferentes momentos (estudo longitudinal), desde que se respeitem os limites éticos dessa comparação. Outra forma de emparelhamento mais elaborada é a escolha de indivíduos com as mesmas características: idade, gênero, etnia, classe social, entre outras variáveis que possam controlar a variabilidade individual. Nesses casos, a medida ocorre entre os pares, ao invés da comparação direta dos subgrupos.

Quadro 6.2 - Fórmulas para cálculo do tamanho de amostras para comparação de dois grupos segundo variáveis quantitativas e qualitativas e segundo pareamento dos casos

Variável Quantitativa	
Amostra Não pareada	**Amostra Pareada**
$n = \left(s_1^2 + s_2^2\right).\left(\dfrac{Z_{\alpha/2} + Z_\beta}{d}\right)^2$	$n_p = \left(\dfrac{\left(Z_{\alpha/2} + Z_\beta\right).S_d}{\overline{D}}\right)^2$
Variável Qualitativa	
Amostra Não pareada	**Amostra Pareada**
$n = \dfrac{\left(p_1q_1 + p_2q_2\right).\left(Z_{\alpha/2} + Z_\beta\right)^2}{\left(p_1 - p_2\right)^2}$	$n_p = \dfrac{\left(Z_{\alpha/2} + 2Z_\beta\sqrt{p_1.q_2}\right)^2}{4p_d\left(p_1 - 0{,}5\right)^2}$

onde:

n = tamanho da amostra para cada grupo;
S_1 e S_2 desvio padrão da variável do grupo 1 ou 2;
$Z_{\alpha/2}$ = valor de Z padronizado para α (usualmente 5% = 1,96);
Z_β = valor de Z padronizado para β (usualmente 20% = 0,84);
d = diferença mínima entre as médias.

n_p = número de pares;
S_d = desvio padrão para as diferenças entre os pares;
$\overline{D}$ = média das diferenças entre os pares.

p_1 e p_2 – proporção de resultados favoráveis no subgrupo 1 ou 2;
q_1 e q_2 – proporção de resultados desfavoráveis no subgrupo 1 ou 2;

p_1 – proporção de pares discordantes para grupo 1;
q_2 – proporção de pares concordantes para o grupo 1;
p_d – soma da proporção dos pares discordantes dos dois grupos.

Ex.: Caso se objetivasse comparar dois grupos, sendo que o grupo 1 tem ocorrência de 70% das vezes que o fenômeno é estudado. O grupo 2 a ser estudado tem ocorrência de 80%, o cálculo do tamanho amostral para cada grupo experimental é de:

$$n = \frac{(p_1q_1 + p_2q_2).(Z_{\alpha/2} + Z_\beta)^2}{(p_1 - p_2)^2} = \frac{(0{,}7\,.\,0{,}3 + 0{,}8\,.\,0{,}2)(1{,}96 + 0{,}84)^2}{(0{,}7 - 0{,}8)^2}$$

$$n = \frac{(0{,}37)(7{,}84)}{0{,}01} = 290{,}08 \cong 291 \text{indivíduos por grupo}$$

6.6.4 Tamanho da amostra em estudos correlacionais

O cálculo amostral para estudos que envolvam a estimativa da correlação linear entre duas variáveis quantitativas depende exclusivamente do coeficiente de correlação linear:

$$n_r = 4 + \left(\frac{(Z_{\alpha/2} + Z_\beta)}{0{,}5\ln\left(\frac{1+r}{1-r}\right)} \right)^2 .$$

onde:

n_r = tamanho da amostra para o par de dados;
$Z_{\alpha/2}$ = valor de Z padronizado para α (usualmente 5% = 1,96);
Z_β = valor de Z padronizado para β (usualmente 20% = 0,84);
r = coeficiente de correlação linear (Pearson ou de Spearman).

7 ESTIMAÇÃO DE PARÂMETROS

É um processo de indução, na qual usamos dados extraídos de uma amostra para produzir inferência sobre a população.

- **Tipos de Estimações de Parâmetros**

i) Estimação Pontual
ii) Estimação Intervalar

7.1 ESTIMAÇÃO PONTUAL

É usada quando a partir da amostra procura-se obter um único valor de certo parâmetro populacional, ou seja, obter estimativas a partir dos valores amostrais.

a) Estatísticas

Seja (X_1 ,X_2, ..., X_n) uma amostra aleatória e (x_1 ,x_2, ..., x_n) os valores tomados pela amostra; então y = H(x_1 ,x_2, ..., x_n) é uma estatística.

Principais estatísticas:

- Média Amostral;
- Proporção Amostral;
- Variância Amostral.

7.2 ESTIMAÇÃO INTERVALAR

A maneira mais correta de se calcular um estimativa de um parâmetro desconhecido, é construir um intervalo de confiança para esse parâmetro com uma probabilidade de $1-\alpha$ (nível de confiança) de que o intervalo contenha o verdadeiro parâmetro. Dessa maneira α será o nível de significância, isto é, o erro que se estará cometendo ao afirmar que o parâmetro está entre o limite inferior e o superior calculado.

7.2.1 Intervalo de confiança para a média (μ)

- Variância (σ^2) conhecida

$$P\left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$ (Pop. Infinita);

$$P\left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \le \mu \le \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right) = 1-\alpha$$ (Pop. Finita).

- Variância (σ^2) desconhecida

$$P\left(\bar{x} - t_{\varphi,\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{x} + t_{\varphi,\alpha/2}\frac{S}{\sqrt{n}}\right) = 1-\alpha$$ (Pop. Infinita);

$$P\left(\bar{x} - t_{\varphi,\alpha/2}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \le \mu \le \bar{x} + t_{\varphi,\alpha/2}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right) = 1-\alpha$$ (Pop. Finita).

Obs.:Quando $n > 30$ e σ for desconhecido poderemos usar S como uma boa estimativa de σ. Esta estimação será melhor quanto maior for o tamanho da amostra.

7.2.2 Intervalo de Confiança para a Proporção

Sendo $\hat{p}$ o estimador de π, onde $\hat{p}$ segue uma distribuição normal, logo:

$$\hat{p} \sim N\left(\hat{p}; \frac{\hat{p}\,.\,\hat{q}}{n}\right)$$ (Pop. Infinita);

$$\hat{p} \sim N\left(\hat{p}; \frac{\hat{p}\,.\,\hat{q}}{n}\left(\frac{N-n}{N-1}\right)\right)$$ (Pop. Finita).

$$Z = \frac{\hat{p} - \pi}{\sigma_{\hat{p}}},$$

onde $\begin{cases} \sigma_{\hat{p}} = \sqrt{\dfrac{\hat{p}\,.\,\hat{q}}{n}} \\ \hat{p} = \dfrac{x}{n} = \dfrac{\text{característica}}{\text{número de elementos da amostra}} \end{cases}$

$$P\left(\hat{p} - Z_{\alpha/2}\ \sigma_{\hat{p}} \le \pi \le \hat{p} + Z_{\alpha/2}\ \sigma_{\hat{p}}\right) = 1 - \alpha \text{ (Pop. Infinita)}$$

Para caso de populações finitas usa-se a seguinte fórmula:

$$P\left(\hat{p} - Z_{\alpha/2}\ \sigma_{\hat{p}}\sqrt{\frac{N-n}{N-1}} \le \pi \le \hat{p} + Z_{\alpha/2}\ \sigma_{\hat{p}}\sqrt{\frac{N-n}{N-1}}\right) = 1 - \alpha \text{ (Pop. Finita).}$$

7.2.3 Intervalo de Confiança para a Variância

Como o estimador de σ^2 é s^2 pode-se considerar que $\dfrac{(n-1)s^2}{\sigma^2}$ tem distribuição Qui-quadrado, ou seja:

$$\chi^2_{n-1} \cong \frac{(n-1)s^2}{\sigma^2},$$

logo o intervalo será:

$$P\left(\frac{(n-1)s^2}{\chi^2_{sup.}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{inf.}}\right) = 1 - \alpha.$$

7.2.4 Intervalo de Confiança para a diferença de médias

- Variâncias conhecidas

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm (\text{Erro Padrão} = e_o);$$

$$e_o = Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}};$$

$$P\left[\left(\bar{x}_1-\bar{x}_2\right)-Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\leq\mu_1-\mu_2\leq\left(\bar{x}_1-\bar{x}_2\right)+Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right]=1-\alpha\,.$$

- Variâncias desconhecidas e iguais

Em geral conhecemos duas variâncias populacionais (σ_1^2 e σ_2^2). Se as mesmas são desconhecidas o melhor que se pode fazer é estimá-las por meio de variâncias amostrais s_1^2 e s_2^2.

Como as amostras serão pequenas, introduziremos uma fonte de erro compensada pela distribuição "t":

$$P\left[\left(\bar{x}_1-\bar{x}_2\right)-e_o\leq\mu_1-\mu_2\leq\left(\bar{x}_1-\bar{x}_2\right)+e_o\right]=1-\alpha\,,$$

onde, $\varphi=n_1+n_2-2$.

Obs.: Se as variâncias populacionais são desconhecidas, mas as estimativas são iguais, poderemos usar para o erro padrão (e_o) o seguinte critério:

$$e_o=t_{\varphi,\alpha/2}\,.\,s_c\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$ onde s_c é o desvio padrão comum,

$$s_C=\sqrt{\frac{\left(n_1-1\right)s_1^2+\left(n_2-1\right)s_2^2}{n_1+n_2-2}}\,.$$

- Variâncias desconhecidas e diferentes

$$e_o=t_{\varphi,\alpha/2}.\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\quad\text{com}\quad\varphi=\frac{\left(s_1^2/n_1+s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1+1}+\frac{\left(s_2^2/n_2\right)^2}{n_2+1}}-2\,.$$

7.2.5 Intervalo de Confiança para a diferença de proporções

$$P\left[\left(\hat{p}_1-\hat{p}_2\right)-Z_{\alpha/2}\sqrt{\frac{\hat{p}_1.\hat{q}_1}{n}+\frac{\hat{p}_2.\hat{q}_2}{n}}\le \pi_1-\pi_2\le\left(\hat{p}_1-\hat{p}_2\right)+Z_{\alpha/2}\sqrt{\frac{\hat{p}_1.\hat{q}_1}{n}+\frac{\hat{p}_2.\hat{q}_2}{n}}\right]=1-\alpha.$$

8 TESTES DE HIPÓTESES PARAMÉTRICOS

Trata-se de uma técnica para se fazer inferência estatística. Ou seja, a partir de um teste de hipóteses, realizado com dados amostrais, pode-se inferir sobre a população.

No caso da inferência através de Intervalo de confiança, busca-se "cercar" o parâmetro populacional desconhecido. Aqui formula-se uma hipótese quanto ao valor do parâmetro populacional, e pelos elementos amostrais faz-se um teste que indicará a ACEITAÇÃO ou REJEIÇÃO da hipótese formulada.

8.1 PRINCIPAIS CONCEITOS

8.1.1 Hipótese estatística

Trata-se de uma suposição quanto ao valor de um parâmetro populacional, ou quanto à natureza da distribuição de uma variável populacional.

8.1.2 Teste de hipótese

É uma regra de decisão para aceitar ou rejeitar uma hipótese estatística com base nos elementos amostrais.

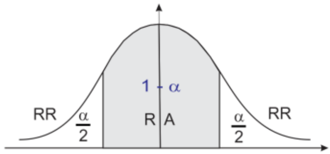
8.1.3 Tipos de hipóteses

Designa-se por H_0, a hipótese nula, ou seja, a hipótese estatística a ser testada, e por H_1 a hipótese alternativa. A hipótese nula expressa uma igualdade, enquanto que a hipótese alternativa é dada por uma desigualdade ($\neq$, $<$, $>$).

Exemplos: Seja Θ um parâmetro e seja θ_0 um estimador qualquer:

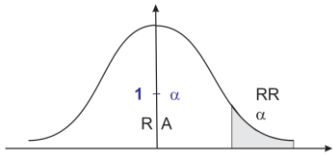
$$\begin{cases} H_0 : \Theta = \theta_0 \\ H_1 : \Theta \neq \theta_0 \end{cases}$$

Teste Bicaudal
ou Bilateral



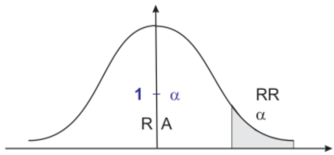
$$\begin{cases} H_0 : \Theta = \theta_0 \\ H_1 : \Theta > \theta_0 \end{cases}$$

Teste Unilateral
Direito



$$\begin{cases} H_0 : \Theta = \theta_0 \\ H_1 : \Theta < \theta_0 \end{cases}$$

Teste Unilateral
Esquerdo



8.2 TESTE DE SIGNIFICÂNCIA

Os testes de significância consideram somente erros do tipo α, pois são os mais usados em pesquisas educacionais ou sócio-econômicas.

O procedimento para realização dos testes de significância é resumido nos seguintes passos:

1º) Enunciar as hipóteses H_0 e H_1;

2º) fixar o limite do erro α, e identificar a variável do teste;

3º) com o auxílio das tabelas estatísticas, considerando α e a variável do teste, determinar as RR (Região de Rejeição) e RA (Região de Aceitação) para H_0;

4º) com os elementos amostrais, calcular o valor da variável do teste;

5º) concluir pela aceitação ou rejeição do H_0 pela comparação do valor calculado no 4º passo com RA e RR.

8.2.1 Teste de significância para a média

1. Enunciar as hipóteses:

$$H_0 : \mu = \mu_0 \qquad\qquad H_1 : \begin{cases} \mu \neq \mu_0 & \text{(a)} \\ \mu > \mu_0 & \text{(b)} \\ \mu < \mu_0 & \text{(c)} \end{cases}$$

2. Fixar α. Admitindo:

- Se a variância populacional σ^2 for conhecida, a variável teste será "Z" (n > 30);
- Se a variância populacional σ^2 for desconhecida, a variável teste será "t" de Student com φ = n -1 (n $\leq$ 30).

3. Com o auxílio das tabelas "Z" e "t" determinar as regiões RA e RR;

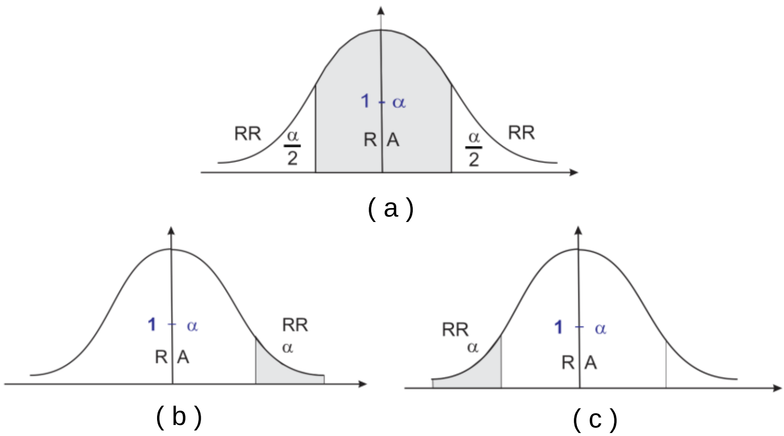


Figura 1.8 – Regiões de aceitação e rejeição do teste para a média

4. Calcular o valor da variável:

$$Z_{cal.} = \frac{\overline{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}} \qquad\qquad t_{cal.} = \frac{\overline{x} - \mu_0}{\dfrac{s}{\sqrt{n}}}$$

onde: $\overline{x}$ = média amostral,

μ_0 = valor da hipótese nula.

5. Conclusão para a situação (a)

- Se $-Z_{\alpha/2} \leq Z_{cal.} \leq Z_{\alpha/2}$ ou $-t_{\alpha/2;\varphi} \leq t_{cal} \leq t_{\alpha/2;\varphi}$, não se pode rejeitar H_0.

- Se $Z_{cal.} < -Z_{\alpha/2}$ ou $Z_{cal.} > Z_{\alpha/2}$ ou $t_{cal.} < -t_{\alpha/2;\varphi}$ ou $t_{cal.} > t_{\alpha/2;\varphi}$, rejeita-se H_0.

Obs.: Para qualquer tipo de teste de significância devemos considerar:
- Se a variável teste (calculada) cair dentro da região de aceitação (RA) não se pode rejeitar H_0;
- Se a variável teste (calculada) cair fora da região de aceitação (RA) rejeita-se H_0.

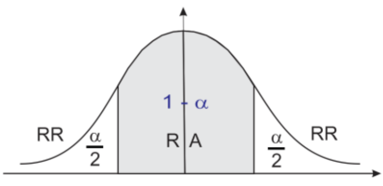
8.2.2 Teste de significância para a proporção

1. Enunciar as hipóteses:

$$H_0 : \pi = \pi_0 \qquad H_1 : \begin{cases} \pi \neq p_0 & \text{(a)} \\ \pi > p_0 & \text{(b)} \\ \pi < p_0 & \text{(c)} \end{cases}$$

2. Fixar α. Escolhendo a variável normal padrão "Z";

3. Com o auxílio da tabela "Z" determinar as regiões RA e RR;



(a)

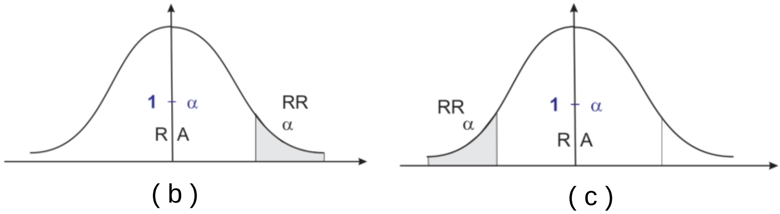


Figura 2.8 – Regiões de aceitação e rejeição do teste para a proporção

4. Calcular o valor da variável:

$$Z_{cal.} = \frac{f - p_0}{\sqrt{\frac{p_0 \cdot q_0}{n}}} \qquad \text{sendo } f = \frac{X}{n},$$

onde: X = característica dentro da amostra;
f = frequência relativa do evento na amostra;
p_0 = valor da hipótese nula.

5. Conclusão para a situação (a)

- Se $-Z_{\alpha/2} \leq Z_{cal.} \leq Z_{\alpha/2}$, não se pode rejeitar H_0.

- Se $Z_{cal.} < -Z_{\alpha/2}$ ou $Z_{cal.} > Z_{\alpha/2}$, rejeita-se H_0.

8.2.3 Teste de significância para a variância

1. Enunciar as hipóteses:

$$H_0 : \sigma^2 = \sigma_0^2 \qquad H_1 : \begin{cases} \sigma^2 \neq \sigma_0^2 & \text{(a)} \\ \sigma^2 > \sigma_0^2 & \text{(b)} \\ \sigma^2 < \sigma_0^2 & \text{(c)} \end{cases}$$

2. Fixar α. Escolhendo a variável qui-quadrado com $\varphi = n - 1$.

3. Com o auxílio da tabela "χ^2" determinar as regiões RA e RR;

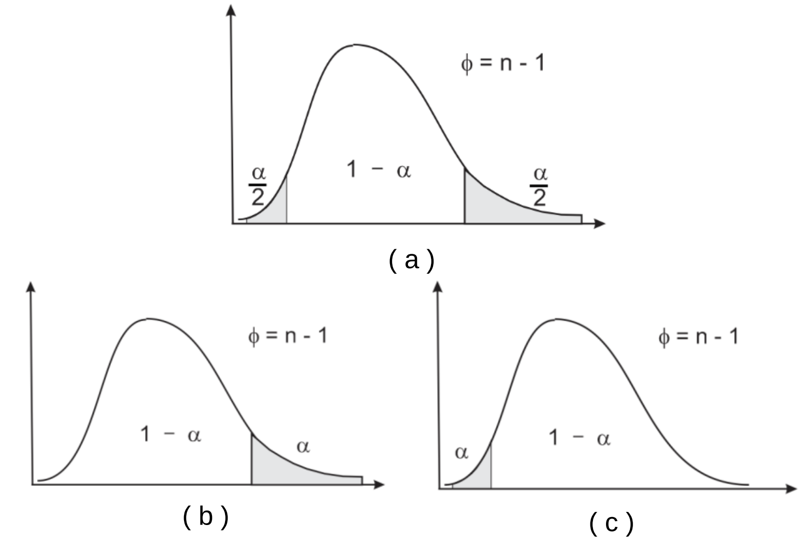


Figura 3.8 – Regiões de aceitação e rejeição do teste da variância

Para (a) temos: $\chi^2_{\inf.} = \chi^2_{1-\alpha/2;\varphi}$ $\chi^2_{\sup.} = \chi^2_{\alpha/2;\varphi}$

Para (b) temos: $\chi^2_{\sup.} = \chi^2_{\alpha;\varphi}$

Para (c) temos: $\chi^2_{\inf.} = \chi^2_{1-\alpha;\varphi}$

4. Calcular o valor da variável:

$$\chi^2_{cal.} = \frac{(n-1)\,s^2}{\sigma_0^2},$$

onde: s^2 = variância amostral;

σ_0^2 = valor da hipótese nula.

5. Conclusão para a situação (a)

- Se $\chi^2_{\inf.} \le \chi^2_{cal.} \le \chi^2_{\sup.}$, não se pode rejeitar H_0.
- Se $\chi^2_{cal.} < \chi^2_{\inf.}$ ou $\chi^2_{cal.} > \chi^2_{\sup.}$, rejeita-se H_0.

8.2.4 Teste de significância para igualdade de duas médias

1º caso) Se as variâncias populacionais σ^2 forem **conhecidas**, independentes e normais, a variável teste será "Z" $(n_1 + n_2 > 30)$;

- Variâncias diferentes ou iguais

1. Enunciar as hipóteses:

$$H_0 : \begin{cases} \mu_1 = \mu_2 \\ \mu_1 - \mu_2 = d \end{cases}$$

$$H_1 : \begin{cases} \mu_1 \neq \mu_2 \\ \mu_1 - \mu_2 \neq d \end{cases} \text{ ou } H_1 : \begin{cases} \mu_1 > \mu_2 \\ \mu_1 - \mu_2 > d \end{cases} \text{ ou } H_1 : \begin{cases} \mu_1 < \mu_2 \\ \mu_1 - \mu_2 < d \end{cases}$$

onde d é a diferença admitida entre as duas médias.

2. Fixar α. Escolhendo a variável normal padrão "Z";

3. Com o auxílio da tabela "Z" determinar as regiões RA e RR;

4. Calcular o valor da variável:

$$Z_{cal.} = \frac{(\bar{x}_1 - \bar{x}_2) - d}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

5. Conclusão:

Optar pela aceitação ou rejeição de H_0.

2º caso) Se as variâncias populacionais σ^2 forem desconhecidas, independentes e normais, a variável teste será "t" $(n_1 + n_2 \leq 30)$ com $\varphi = n_1 + n_2 - 2$;

a) Variâncias iguais

1. Enunciar as hipóteses:

$$H_0: \begin{cases} \mu_1 = \mu_2 \\ \mu_1 - \mu_2 = d \end{cases}$$

$$H_1: \begin{cases} \mu_1 \neq \mu_2 \\ \mu_1 - \mu_2 \neq d \end{cases} \text{ ou } H_1: \begin{cases} \mu_1 > \mu_2 \\ \mu_1 - \mu_2 > d \end{cases} \text{ ou } H_1: \begin{cases} \mu_1 < \mu_2 \\ \mu_1 - \mu_2 < d \end{cases}$$

onde: d é a diferença admitida entre as duas médias.

2. Fixar α. Escolhendo a variável "t" de Student;

3. Com o auxílio da tabela "t" determinar as regiões RA e RR;

4. Calcular o valor da variável:

$$t_{cal.} = \frac{(\bar{x}_1 - \bar{x}_2) - d}{s_c \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \text{onde: } s_c = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 1}}$$

5. Conclusão:

Optar pela aceitação ou rejeição de H_0.

b) Variâncias diferentes

1. Enunciar as hipóteses:

$$H_0: \begin{cases} \mu_1 = \mu_2 \\ \mu_1 - \mu_2 = d \end{cases}$$

$$H_1: \begin{cases} \mu_1 \neq \mu_2 \\ \mu_1 - \mu_2 \neq d \end{cases} \text{ ou } H_1: \begin{cases} \mu_1 > \mu_2 \\ \mu_1 - \mu_2 > d \end{cases} \text{ ou } H_1: \begin{cases} \mu_1 < \mu_2 \\ \mu_1 - \mu_2 < d \end{cases}$$

onde: d é a diferença admitida entre as duas médias.

2. Fixar α. Escolhendo a variável "t" de Student;

3. Com o auxílio da tabela "t" determinar as regiões RA e RR;

4. Calcular o valor da variável:

$$t_{cal.} = \frac{\left(\vec{x}_1 - \bar{x}_2\right) - d}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}} \qquad \text{onde, } \varphi = \frac{\left(S_1^2 / n_1 + S_2^2 / n_2\right)^2}{\dfrac{\left(S_1^2 / n_1\right)^2}{n_1 + 1} + \dfrac{\left(S_2^2 / n_2\right)^2}{n_2 + 1}} - 2.$$

5. Conclusões:

Optar pela aceitação ou rejeição de H_0.

8.2.5 Teste de significância para igualdade de duas proporções

1. Enunciar as hipóteses:

$$H_0 : \pi_1 = \pi_2 \qquad H_1 : \pi_1 \neq \pi_2$$

2. Fixar α. Escolhendo a variável normal padrão "Z";

3. Com o auxílio da tabela "Z" determinar as regiões RA e RR;
4. Calcular o valor da variável:

$$Z_{cal.} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p} \cdot \hat{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \qquad \text{onde } \hat{p}_1 = \frac{x_1}{n_1}, \hat{p}_2 = \frac{x_2}{n_2}, \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}.$$

5. Conclusão:

Optar pela aceitação ou rejeição de H_0.

8.2.6 Teste de significância para a diferença de duas variâncias

1. Enunciar as hipóteses:

$$H_0 : s_1^2 = s_2^2 \qquad H_1 : \begin{cases} s_1^2 \neq s_2^2 \text{ (a)} \\ s_1^2 > s_2^2 \text{ (b)} \\ s_1^2 < s_2^2 \text{ (c)} \end{cases}$$

2. Fixar α, escolhendo a variável F com $\varphi_1 = n_1 - 1$ e $\varphi_2 = n_2 - 1$.

3. Com o auxílio da tabela "F" determinar as regiões RA e RR;

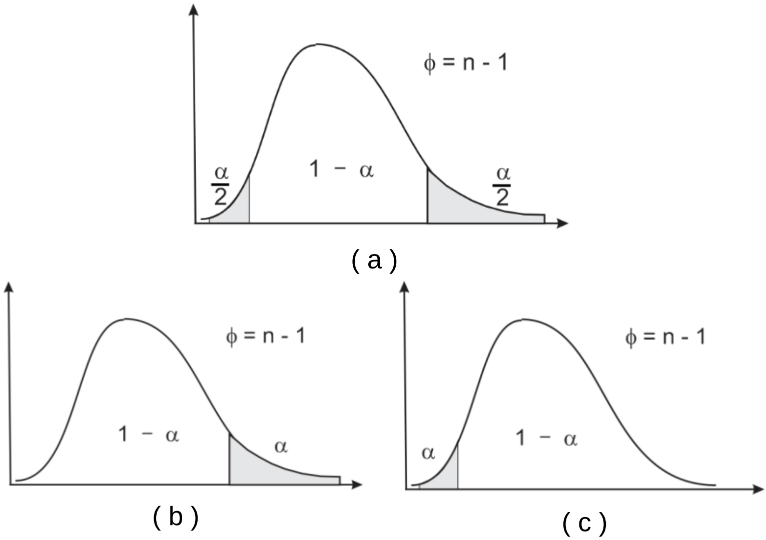


Figura 4.8 – Regiões de aceitação e rejeição do teste para diferença de duas variâncias

Para (a) tem-se: $F_{sup.} = F_{\alpha/2;\varphi_1;\varphi_2}$ e $F_{inf.} = \dfrac{1}{F_{\alpha/2;\varphi_2;\varphi_1}}$;

Para (b) tem-se: $F_{sup.} = F_{\alpha;\varphi_1;\varphi_2}$;

Para (c) tem-se: $F_{inf.} = \dfrac{1}{F_{\alpha;\varphi_2;\varphi_1}}$.

4. Calcular o valor da variável:

$$F_{cal.} = \frac{s_1^2}{s_2^2} \cdot$$

5. Conclusão para a situação (a)

- Se $F_{inf.} \leq F_{cal.} \leq F_{sup.}$, não se pode rejeitar H_0.
- Se $F_{cal.} < F_{inf.}$ ou $F_{cal.} > F_{sup.}$, rejeita-se H_0.

8.2.7 Teste de Homogeneidade

Também conhecido por Teste H (Hartley) onde se deve calcular as variâncias s_i^2 dos grupos a fim de determinar a estatística H dada por:

$$H_{cal.} = \frac{s^2_{máx.}}{s^2_{mín.}} \ ,$$

comparando o seu valor com o valor crítico $H_{tab.}$, que é obtido através da Tabela 8 (Pearson e Hartley, 1970), entrando com os valores i (número de grupos), $\varphi = n - 1$ graus de liberdade e α.

Se $H_{cal.} \geq H_{tab.}$, rejeita-se a hipótese de homogeneidade de variâncias entre os grupos.

Existem outros testes para homogeneidade de variâncias, como o de Cochram e o de Bartley, porem o de Hartley é o de aplicação mais simples.

Ex.: Soma das perguntas de 4 constructos de comportamento ecológico (escala likert de 1 a 6).
Verifique a homogeneidade dos dados para α = 5%.
A = Cuidado com o lixo e consumo de produtos (7 perguntas)
B = Ativismo (4 perguntas)
C = Economia de energia e água (8 perguntas)
D = Meio ambiente (3 perguntas)

H_0: $S^2_A = S^2_B = S^2_C = S^2_D$,
H_1: Pelo menos uma variância difere das demais.

	A		B		C		D	
	17	23	24	11	31	16	3	3
	21	22	16	21	24	32	5	9
	23	21	17	12	26	32	4	6
	13	23	18	16	28	31	3	8
	27	15	22	15	36	19	3	4
Média	20,50		17,20		27,50		4,80	
Variância	4,031		3,970		5,971		2,088	

$$H_{cal.} = \frac{5{,}971}{2{,}088} = 2{,}86$$

$$H_{tab.} = 6{,}31 \text{ (Tabela 8 } (\alpha = 5\%))$$

Como $H_{cal.} < H_{tab.}$, aceita-se a hipótese de homogeneidade das variâncias dos grupos.

8.2.8 Teste t para amostras em par

Os resultados de duas amostras constituem dados emparelhados quando estão relacionados dois a dois, segundo algum critério que introduz uma influência marcante entre os diversos pares, que supõem-se, porém, influir igualmente sobre os valores de cada par.

Deseja-se então calcular as diferenças **"d_i"** correspondentes a cada par de valores, reduzindo, assim, os dados a uma única amostra de "n" diferenças, ou seja, testar simplesmente a hipótese:

$$H_0 : \bar{d} = \Delta \qquad H_1 : \bar{d} \neq \Delta \text{ ou } H_1 : \bar{d} > \Delta \text{ ou } H_1 : \bar{d} < \Delta$$

contra uma alternativa H_1 que poderá corresponder a um teste unilateral ou bilateral, conforme seja de interesse. Logo realiza-se o teste simplesmente através da comparação do t de Student experimental com o valor tabelado obtido em função de α com n – 1 graus de liberdade.

$$t_{cal.} = \frac{\overline{d} - \Delta}{\frac{S_d}{\sqrt{n}}} \text{ ,}$$

onde:

$\overline{d}$ - é a média de amostra das diferenças;

$\triangle$ - o valor testado da média das diferenças populacional (usa-se zero);

$s_d = \sqrt{\frac{\sum\left(d - \overline{d}\right)^2}{n-1}}$ - desvio padrão da amostra das diferenças,

n - o tamanho das diferenças.

Ex.: Análise de um constructo de comportamento ecológico.
C = Economia de energia e água.
Antes e depois de 6 meses de pesquisa e conscientização.

$$H_0 : \overline{d} = 0 \qquad H_1 : \overline{d} \neq 0$$

Antes	31	24	26	28	36	16	32	32	31	19
Depois	36	27	24	32	40	14	38	35	33	20
d_i	5	3	-2	4	4	-2	6	3	2	1

$\overline{d} = 2{,}4$ $\qquad s_d = 2{,}72$ $\qquad n = 10$

$$t_{cal.} = \frac{2{,}4 - 0}{\frac{2{,}72}{\sqrt{10}}} = 2{,}79$$

$$t_{tab.} = \pm 2{,}262 \text{ (Tabela 2 } (\alpha = 5\%))$$

Como $t_{cal.} > +t_{tab.}$, rejeita-se a hipótese de que a diferença média é estatisticamente igual a zero, ou seja, houve efeito da pesquisa de conscientização de economia e consumo de energia e água.

9 TESTES DE HIPÓTESES NÃO-PARAMÉTRICOS

A estatística não-paramétrica não exige suposições quanto a distribuição da população da qual se tenha retirado as amostras para análise. Este tipo de teste independe dos parâmetros populacionais (μ, σ^2, σ, π, ...) e de seus respectivos estimadores ($\overline{x}$, s^2, s, p, ...).

Um dos testes não-paramétricos mais utilizados é o teste qui-quadrado, pois não depende dos parâmetros populacionais, nem de suas respectivas estimativas.

9.1 TESTE DO QUI-QUADRADO x^2

Anteriormente foram testados hipóteses referentes a um parâmetro populacional e a comparação de dois parâmetros, que são os testes paramétricos. Agora será testado aqueles que não dependem de um parâmetro populacional, nem de suas respectivas estimativas. Este teste é chamado de teste do Qui-quadrado.

O teste do Qui-quadrado é usado quando se quer comparar frequências observadas com frequências esperadas. Divide-se em três tipos:

- Teste de adequação do ajustamento;
- Teste de Independência (Tabela de Contingência);
- Teste de Aderência.

9.1.1 Procedimentos para a realização de um teste Qui-quadrado

1º) Determinação das Hipóteses:

$$H_0: F_o = F_e \qquad\qquad H_1: F_o \neq F_e$$

2º) Escolha do Nível de Significância (α)

3º) Estatística Calculada $\chi^2_{cal.} = \sum_{i=1}^{k} \frac{(F_o - F_e)^2}{F_e}$

4º) Estatística Tabelada: $\chi^2_{tab.} = \chi^2_{\varphi,\alpha}$

5º) Comparar o $\chi^2_{cal.}$ com $\chi^2_{tab.}$ e concluir:

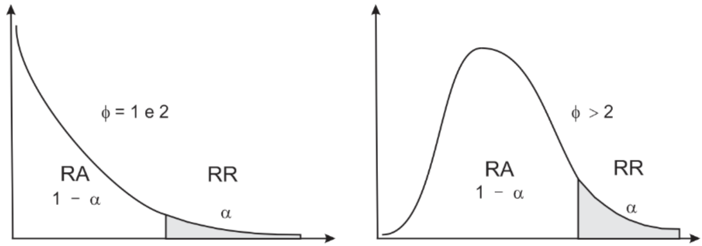


Figura 1.9 - Regiões de aceitabilidade do teste qui-quadrado

6º) Conclusão:

Se $\chi^2_{cal.} \leq \chi^2_{tab.}$ $\rightarrow$ aceita-se H_0

Se $\chi^2_{cal.} > \chi^2_{tab.}$ $\rightarrow$ rejeita-se H_0

9.1.2 Teste de adequação do ajustamento

Suponhamos uma amostra de tamanho n. Sejam E_1, E_2, ..., E_k, um conjunto de eventos possíveis da amostra.

Eventos	Frequência Observada (Fo_i)
E_1	Fo_1
E_2	Fo_2
:	:
E_k	Fo_k
Total	N

Este teste é indicado para verificar se as frequências observadas dos k eventos (k classes em que a variável é dividida) concordam ou não com as frequências teóricas esperadas.

As Frequências Esperadas (Fe_i) são obtidas multiplicando-se o número total de elementos pela proporção teórica da classe i (n . p_i).

Para encontrar o $\chi^2_{cal.}$, necessita-se do nível de significância (α) e dos graus de liberdade (φ), os quais podem ser obtidos da seguinte forma:

1º) $\varphi = k - 1$, quando as frequências esperadas puderem ser calculadas sem que se façam estimativas dos parâmetros populacionais a partir das distribuições amostrais.

2º) $\varphi = k - 1 - m$, quando para a determinação das frequências esperadas, m parâmetros tiverem suas estimativas calculadas a partir das distribuições amostrais. Pearson mostrou que, se o modelo testado for verdadeiro e se as $Fe_i \leq 5$, estas deverão ser fundidas às classes adjacentes.

9.1.3 Teste de independência

Uma importante aplicação do teste χ^2 ocorre quando se quer estudar a relação entre duas ou mais variáveis de classificação. A representação das frequências observadas, nesse caso, pode ser feita por meio de uma tabela de contingência.

H_0: As variáveis são independentes (não estão associadas) ($p > 0{,}05$)

H_1: As variáveis não são independentes (estão associadas) ($p \leq 0{,}05$)

O número de graus de liberdade é dado por: $\varphi = (L - 1)(C - 1)$, onde L é o número de linhas e C o número de colunas da tabela de contingência.

$$\chi^2_{cal} = \sum_{i=1}^{k} \frac{(F_o - F_e)^2}{F_e} \qquad F_e = F_{eij} = \frac{(\text{soma da linha i})(\text{soma da coluna j})}{\text{total de observações}}\ .$$

Obs.: Em tabelas 2 x 2 tem-se 1 grau de liberdade, por isso recomenda-se o teste qui-quadrado com correção de continuidade, conhecido por **Correção de Yates**, no qual para $n \geq 50$ pode ser omitida.

$$\chi^2_{cal.} = \sum_{i=1}^{k} \frac{(|F_o - F_e| - 0{,}5)^2}{F_e}.$$

A correção de Yates produz um teste mais conservador, isto é, um teste com menor probabilidade de rejeição da hipótese de nulidade.

O teste do χ^2 não é indicado em tabelas 2 x 2 nos seguintes casos:

i) quando alguma frequência esperada for menor que 1;
ii) quando a frequência total for menor que 20 ($n < 20$);
iii) quando a frequência total ($20 \leq n \leq 40$) e algumas frequências esperadas for menor que 5. Neste caso aplica-se o Teste Exato de Fischer;
iv) m tabelas c x l, no máximo 20% das F_e podem ser menor que 5 ($F_e < 5$) e nenhuma $F_e < 1$.

9.1.4 Coeficiente de contingência

Uma medida do grau de relacionamento, associação ou dependência das classificações em uma tabela de contingência é dada pelo coeficiente de contingência.

$$C = \sqrt{\frac{\chi^2_{cal.}}{\chi^2_{cal.} + n}} \qquad 0 \leq C \leq 1.$$

Quanto maior o valor de C, maior o grau de associação. O máximo valor de C dependerá do número de linhas e colunas da tabela e pode variar de 0 (independência) a 1 (dependência total).

Obs.: O valor de $C \leq 0{,}35$ indica uma baixa associação e $C > 0{,}65$ uma alta associação.

9.1.5 Teste de aderência

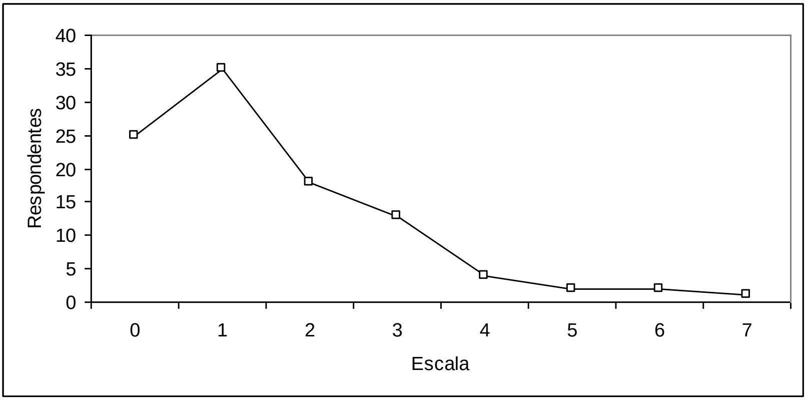
É usada a estatística χ^2 quando deseja-se testar a natureza da distribuição amostral. Por exemplo, quando se quer verificar se a distribuição amostral se ajusta a um determinado modelo de distribuição de probabilidade (Normal, Poisson, Binomial, ...), ou seja, verifica-se a boa ou má aderência dos dados da amostra do modelo.

Ex.: 100 respostas de indivíduos quanto ao valor da saúde utilizando uma escala *likert* de 7 pontos:

1 – discordo completamente 7 – concordo completamente

Pergunta	0	1	2	3	4	5	6	7
Se não tiver saúde não se tem nada	25	35	18	13	4	2	2	1

Verificar se o número de respondente por item da escala segue razoavelmente uma distribuição de Poisson, com α = 5%.



Resolução:

H_0: A distribuição do nº de respostas/item segue uma Poisson
H_1: A distribuição do nº de respostas/item não segue uma Poisson

Valor da Escala	0	1	2	3	4	5	6	7
Nº de Indivíduos	25	35	18	13	4	2	2	1
p_i	0,212	0,329	0,255	0,132	0,051	0,016	0,004	0,001
$Fe_i = n \cdot p_i$	21,2	32,9	25,5	13,2	5,1	1,6*	0,4*	0,1*

* = $F_{ei} \leq$ 5% (0,05), logo deve ser agrupada com a classe adjacente.

Para calcular p_i, temos: $p_i = \dfrac{e^{-\mu}.\mu^i}{i!}$ (Distribuição de Poisson)

$$\mu = \frac{\sum x_i.f_i}{n} = 1{,}55$$

$$p_0 = \frac{e^{-1,55}(1,55)^0}{0!} = 0,212$$

$$p_1 = \frac{e^{-1,55}(1,55)^1}{1!} = 0,329$$

$$p_2 = \frac{e^{-1,55}(1,55)^2}{2!} = 0,255$$

$$p_7 = \frac{e^{-1,55}(1,55)^6}{6!} = 0,001$$

$$p_4 = \frac{e^{-1,55}(1,55)^4}{4!} = 0,051$$

$$p_5 = \frac{e^{-1,55}(1,55)^5}{5!} = 0,016$$

$$p_6 = \frac{e^{-1,55}(1,55)^6}{6!} = 0,004$$

$$\chi^2_{cal.} = \frac{(25-21,2)^2}{21,2} + \frac{(35-32,9)^2}{32,9} + \frac{(18-25,5)^2}{25,5} + \underbrace{\frac{(9-7,2)^2}{7,2}}_{4-5-6-7} = 3,474$$

$$\varphi = k - 1 - m = 5 - 1 - 1 = 3 \quad \begin{cases} k = 5 \text{ (número de classes após agrupamento} \\ m = 1 \text{ (número de estimadores usados } (\mu)) \end{cases}$$

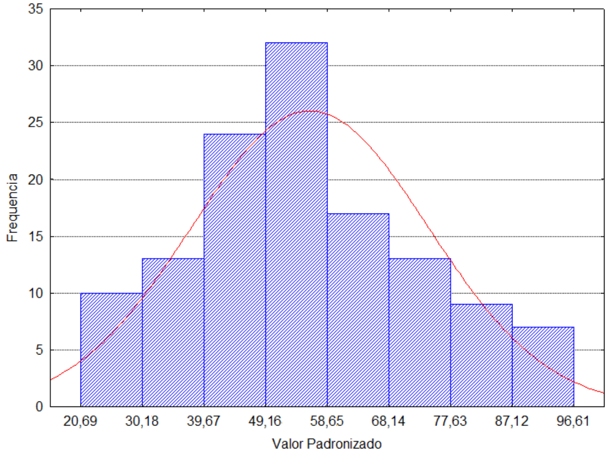
$$\chi^2_{tab.} = \chi^2_{5\%;3} = 7,81$$

Conclusão: Como o $\chi^2_{cal.} \leq \chi^2_{tab.}$, aceita-se H_0, ou seja, para α = 5%, a distribuição do número de respostas/unidade pode seguir uma Distribuição de Poisson.

Ex.: Verificar se os dados da padronização da escala de Engajamento no Trabalho (constructo Dedicação) aplicada em 126 professores da UFSM se aproxima de uma distribuição normal, com α = 5%.

Valor Padronizado			f_i	Transf. Z			p_i	$Fe_i = n \cdot p_i$
20,69	|---	30,18	10	- ∞	|---	-1,43	0,076	9,6
30,18	|---	39,67	13	-1,43	|---	-0,89	0,110	13,9
39,67	|---	49,16	24	-0,89	|---	-0,36	0,173	21,8
49,16	|---	58,65	32	-0,36	|---	0,18	0,212	26,7
58,65	|---	68,14	17	0,18	|---	0,72	0,193	24,3
68,14	|---	77,63	13	0,72	|---	1,25	0,130	16,4
77,63	|---	87,12	10	1,25	|---	1,79	0,068	8,6
87,12	|---	96,61	7	1,79	|---	+ ∞	0,037	4,7*
k = 8			126	k = 7				126,0

* como $Fe_8 \leq$ 5% (0,05), a Fe_8 deve ser agrupada com a classe adjacente.



Resolução:

H_0: A distribuição do constructo dedicação é normal
H_1: A distribuição do constructo dedicação não é normal

Para calcular p_i, temos: $Z_i = \dfrac{x - \bar{x}}{s}$ (Distribuição Normal Padrão)

$$\bar{x} = \frac{\sum x_i . f_i}{n} = 55{,}49 \qquad s = \sqrt{\frac{\sum f_i . (x_i - \bar{x})^2}{n-1}} = 17{,}68$$

$$Z_1 = \frac{30{,}18 - 55{,}49}{17{,}68} = -1{,}43 \qquad Z_2 = \frac{39{,}67 - 55{,}49}{17{,}68} = -0{,}89$$

$$Z_3 = \frac{49{,}16 - 55{,}49}{17{,}68} = -0{,}36 \qquad Z_4 = \frac{58{,}65 - 55{,}49}{17{,}68} = 0{,}18$$

$$Z_5 = \frac{68{,}14 - 55{,}49}{17{,}68} = 0{,}71 \qquad Z_6 = \frac{77{,}63 - 55{,}49}{17{,}68} = 1{,}25$$

$$Z_7 = \frac{87{,}12 - 55{,}49}{17{,}68} = 1{,}79$$

$$\chi^2_{cal} = \frac{(10 - 9{,}6)^2}{9{,}6} + \frac{(13 - 13{,}9)^2}{13{,}9} + \frac{(24 - 21{,}8)^2}{21{,}8} + \frac{(32 - 26{,}7)^2}{26{,}7} + \frac{(17 - 24{,}3)^2}{24{,}3} + \frac{(13 - 16{,}4)^2}{16{,}4} + \underbrace{\frac{(17 - 13{,}3)^2}{13{,}3}}_{7-8} = 5{,}10$$

$$\varphi = k - 1 - m = 7 - 1 - 2 = 4 \quad \begin{cases} k = 7 \text{ (número de classes após agrupamento} \\ m = 2 \text{ (número de estimadores usados } (\bar{x} \text{ e } s)) \end{cases}$$

$$\chi^2_{tab.} = \chi^2_{5\%;4} = 9{,}49$$

Conclusão: Como o $\chi^2_{cal.} \leq \chi^2_{tab.}$, aceita-se H_0, ou seja, para α = 5%, a distribuição a distribuição do constructo dedicação é normal.

9.2 TESTE DE MANN-WHITNEY

O teste de Mann-Whitney serve para testar a hipótese de que duas populações têm a mesma distribuição. Esse teste é, portanto, uma das alternativas para substituir o teste t para amostras independentes. O teste é indicado para amostras pequenas e as pressuposições exigidas pelo teste t estiverem seriamente comprometidas.

Passos para a realização do teste:

$$H_0 : \tau_1 = \tau_2 \qquad H_1 : \tau_1 \neq \tau_2$$

1º Passo: Estabelecer o nível de significância (α).

2º Passo: Passe a chamar os grupos de 1 e 2. O grupo 1 tem n_1 dados e o grupo 2 n_2 dados. Se os grupos tiverem tamanhos diferente $n_1 \neq n_2$, o menor será o grupo 1.

3º Passo: Junte os grupos em um único grupo, onde $n = n_1 + n_2$ dados. Ordene e atribua postos aos dados: o valor mais baixo receberá posto 1, o próximo 2, e assim sucessivamente. Se os dados tiverem valores iguais calcule a média dos postos dos empates, e cada um ficará com o posto igual a média calculada.

4º Passo: Some os postos do grupo 1, isto é, calcule o $\sum R_1$.

5º Passo: Calcular a estatística de Mann-Whitney (U).

$$U = \sum R_1 - \left[\frac{n_1(n_1 + 1)}{2} \right].$$

6º Passo: Calcule da estatística do teste.

$$Z_{cal.} = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}; \qquad (9.1)$$

$$Z_{cal.} = \frac{U + 0{,}5\frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}. \qquad (9.2)$$

ambas as equações seguem a distribuição normal padrão, alguns autores recomendam a equação (9.2), que vem a ser uma correção de continuidade, tornando o teste mais conservador.

7º Passo: Comparar o valor calculado com o valor tabelado ($Z_{tab.}$) da curva normal padrão para o nível de significância estipulado no primeiro passo.

Exemplo: Análise do constructo de comportamento ecológico.
B = Ativismo
Entre homens e mulheres

Dados		**Postos**	
Homens	**Mulheres**	**Homens**	**Mulheres**
24	22	17	18
16	21	15	6,5
17	20	13,5	9,5
11	16	6,5	1
12	15	3,5	2
16	16	6,5	6,5
15	21	16	3,5
19	20	13,5	11,5
19	17	9,5	11,5
---		$\sum R_1$=101	$\sum R_2$=70

$$U = \sum R_1 - \left[\frac{n_1(n_1+1)}{2}\right] = 101 - \frac{9(9+1)}{2} = 56$$

$$Z_{cal.} = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}} = \frac{56 - \frac{9.9}{2}}{\sqrt{\frac{9.9(9+9+1)}{12}}} = \frac{15{,}5}{11{,}32} = 1{,}37$$

$Z_{tab.} = Z_{2,5\%} = \pm 1{,}96$

como - $Z_{tab.} < Z_{cal.} < + Z_{tab.}$, para o constructo Ativismo a diferença entre os postos dos homens e das mulheres não é significativa, para um nível de significância de 5%.

9.3 TESTE DE WILCOXON

O teste de Wilcoxon é um teste para dados pareados, que vem a ser uma alternativa para o teste t pareado. Este teste só deve ser aplicado quando as pressuposições do teste t estiverem seriamente comprometidas.

Passos para a realização do teste:

$$H_0 : \tau_{antes} = \tau_{depois} \qquad H_1 : \tau_{antes} \neq \tau_{depois}$$

1º Passo: Estabelecer o nível de significância (α).

2º Passo: Calcular a diferença entre a primeira e a segunda medidas feita no mesmo par, para todos os pares da amostra.

3º Passo: Sem olhar o sinal atribuir um posto a cada diferença. Procedendo assim, atribui-se à menor diferença (em valor absoluto) o posto 1 e assim sucessivamente, caso ocorram valores iguais, fazer a média dos postos.

4º Passo: Colocar sinais nos postos, respeitando os sinais das diferenças. Postos atribuídos às diferenças que tinham sinal negativo, receberão sinal negativo; postos atribuídos às diferenças que tinham sinal positivo, receberão sinal positivo. Os postos assinalados são indicados por R.

5º Passo: Somam-se os postos positivos e somam-se os postos negativos. Sob a hipótese de nulidade, medidas feitas no mesmo par são iguais. Se a soma dos postos positivos for muito diferente da soma dos postos negativos, é razoável rejeitar a hipótese nula, e concluir que as medidas feitas no mesmo par são diferentes.

6º Passo: Calcule da estatística do teste.

$$Z_{cal.} = \frac{\sum R}{\sqrt{\sum R^2}}.$$

a equação segue a distribuição normal padrão.

7º Passo: Comparar o valor calculado com o valor tabelado (Z_{tab}) da curva normal padrão para o nível de significância estipulado no primeiro passo. Para amostras pequenas (n < 30) deve-se substituir a distribuição Z pela t.

Ex.: Análise de um constructo de comportamento ecológico.
C = Economia de energia e água.
Antes e depois de 6 meses de pesquisa e conscientização.

Dados		**Cálculos**			
Antes	**Depois**	**Diferença**	**Posto**	**Posto assinalado (R)**	**R^2**
31	36	5	9	9	81,00
24	27	3	5,5	5,5	30,25
26	24	-2	1,5	-1,5	2,25
28	32	4	7,5	7,5	56,25
36	40	4	7,5	7,5	56,25
16	14	-2	1,5	-1,5	2,25
32	38	6	10	10	100,00
32	35	3	5,5	5,5	30,25
31	33	2	4	4	16,00
19	20	1	3	3	9,00
---	---	---	---	49,0	383,50

$$t_{cal.} = \frac{\sum R}{\sqrt{\sum R^2}} = \frac{49}{\sqrt{383{,}5}} = 2{,}502$$

$$t_{tab.} = t_{9;2,5\%} = \pm 2{,}262$$

Como $t_{cal.} > + t_{tab.}$, rejeita-se a hipótese nula, ou seja, houve efeito da pesquisa de conscientização de economia e consumo de energia e água.

9.4 TESTE DE KRUSKAL-WALLIS

O teste de Kruskal-Wallis serve para testar a hipótese de que várias populações têm a mesma distribuição. É portanto uma alternativa para a análise de variância com uma classificação.

O teste deve ser aplicado se a amostra for pequena e as pressuposições, exigidas para proceder à análise de variância, estiverem seriamente comprometidas.

Passos para a realização do teste:

H_0: $\tau_1 = \tau_2 = \tau_3 = ... = \tau_n$,

H_1: Pelo menos um grupo difere dos demais.

1º Passo: Estabelecer o nível de significância (α).

2º Passo: Junte os grupos $n_1 + n_2 + ... + n_k = n$. dados em um só conjunto. Atribua um posto a cada dado, como foi feito no teste de Mann-Whitney.

3º Passo: Calcule o somatórios dos postos de cada grupo, isto é, $\sum R_1$, $\sum R_2$, ... $\sum R_k$.

4º Passo: Calcular a estatística de Kruskal-Wallis (H).

$$H_{cal.} = \frac{12}{n(n+1)}\left[\frac{\left(\sum R_1\right)^2}{n_1} + \frac{\left(\sum R_2\right)^2}{n_2} + ... + \frac{\left(\sum R_k\right)^2}{n_k}\right] - 3(n+1).$$

5º Passo: Comparar o valor calculado ($H_{cal.}$) com o valor tabelado ($\chi^2_{tab.}$) a o nível de significância estipulado no primeiro passo.

Ex.: Análise de um constructo de comportamento ecológico entre faixas de idade.
C = Economia de energia e água.

Faixas de Idade (anos)				Postos			
18 a 25	26 a 35	36 a 40	Mais de 41	18 a 25	26 a 35	36 a 40	Mais de 41
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
31	34	24	35	12	20	3,5	24
26	31	24	36	7	12	3,5	27,5
36	39	25	34	27,5	29	5,5	20
32	32	23	25	15	15	2	5,5
35	28	20	35	24	9	1	24
35	34	---	31	24	20	---	12
32	35	---	---	15	24	---	---
30	33	---	---	10	17,5	---	---
28	33	---	---	8	17,5	---	---
ΣR				143,0	163,5	15,5	113,0
$\overline{R}$				15,9	18,2	3,1	18,8

$$H_{cal.} = \frac{12}{29(29+1)}\left[\frac{(143)^2}{9} + \frac{(163{,}5)^2}{9} + \frac{(15{,}5)^2}{5} + \frac{(113)^2}{6}\right] - 3(29+1) = 12{,}33$$

φ = nº de grupos – 1 = 4 – 1 = 3 $\chi^2_{tab.} = \chi^2_{5\%;3} = 7{,}81$.

Como o $H_{cal.}$ é maior que o $\chi^2_{tab.}$, a diferença entre os grupos é significativa, para um nível de significância de 5%.

A hipótese de não haver diferença entre os k grupos foi testada e rejeitada ao nível de significância α. Uma alternativa descrita em Siegel e Castellan (1988) é testar a significância dos pares de diferenças através da seguinte desigualdade,

Para verificar qual o grupo difere dos demais utiliza-se o Teste de Bonferroni (Método de Dunn).

$$\left|\overline{R}_{i.} - \overline{R}_{j.}\right| \geq Z_{\left(\frac{\alpha}{k(k-1)}\right)} \sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}, \quad (9.3)$$

onde:

n_i e n_j são os tamanhos da amostra dos grupos i e j respectivamente;

$N = n_1 + n_2 + ... + n_k$, o número total de elementos considerados em todas as amostras;

R_i e R_j é o efeito dos postos (*ranks*) dos grupos i e j respectivamente;

$\left|\overline{R}_{i.} - \overline{R}_{j.}\right|$ é o valor absoluto da diferença observada;

$Z_{\left(\frac{\alpha}{k(k-1)}\right)}\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$ é a diferença crítica.

Assim, se (9.3) ocorre podemos rejeitar a hipótese $H_0 : \tau_i = \tau_j$ e concluir que $H_1 : \tau_i \neq \tau_j$. Vale lembrar que neste teste de comparações múltiplas, se temos k grupos, então o número de comparações é de $\frac{k(k-1)}{2}$. Agora, vamos aplicar os conceitos no exemplo.

1º Passo: Diferenças observada:

Comparação	$\overline{R}_{i.}$	$\overline{R}_{j.}$	$\left\|\overline{R}_{i.} - \overline{R}_{j.}\right\|$
1 - 2	15,9	18,2	2,3
1 - 3	15,9	3,1	12,8
1 - 4	15,9	18,8	2,9
2 - 3	18,2	3,1	15,1
2 - 4	18,2	18,8	0,6
3 - 4	3,1	18,8	15,7

2º Passo: Consultar a tabela normal padrão:

$$Z_{\left(\frac{\alpha}{k(k-1)}\right)} = Z_{0,05/9} = Z_{0,001} = 2{,}576$$

3º Passo: Comparar o valor calculado t_{ij} com o valor tabelado $t_{tab.}$ ao nível de significância predeterminado (α*).

5º Passo: Calcular as diferenças críticas:

Comparação	$Z_{\left(\frac{\alpha}{k(k-1)}\right)}$	$\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$	Diferença Crítica
1 - 2	2,576	4,014	10,34

1 - 3	2,576	4,749	12,23
1 - 4	2,576	4,488	11,56
2 - 3	2,576	4,749	12,23
2 - 4	2,576	4,488	11,56
3 - 4	2,576	5,156	13,28

5º Passo: Comparar a diferença observada com a diferença crítica:

Comparação	Diferença Observada	Diferença Crítica	Diferença
1 - 2	2,30	10,34	Não
1 - 3	12,80	12,23	Sim
1 - 4	2,90	11,56	Não
2 - 3	15,10	12,23	Sim
2 - 4	0,60	11,56	Não
3 - 4	15,7	13,28	Sim

Conclui-se que existe diferença significativa entre as faixas de idade (1 = 18 a 25), (2 = 26 a 35) e (4 = Mais de 41) em relação a faixa de (3 = 36 a 40) anos, para um nível de significância de 5%.

Pode-se verificar que a faixa de idade 3 (36 a 40 anos) tem menos consciência no que se refere a economia de energia e água que as demais faixas de idade.

9.5 TESTE DE SHAPIRO WILK

O Teste de Shapiro-Wilk (1965), também conhecido por Teste W, vem a ser um procedimento eficiente para avaliar a suposição de normalidade contra um amplo espectro de alternativas não normal, principalmente se é dado um número relativamente pequeno de observação. O teste W é recomendado para amostras com menos de 2.000 observações, acima de 2.000 observações aconselha-se o Teste K (Kolmogorov Smirnov).

Passos para a realização do teste W:

1º) Colocar em ordem crescente as observações da amostra com a menor variância (construir um rol);

$x_1, x_2, ..., x_n$ onde $x_1 \leq x_2 \leq .... \leq x_n$.

2º) Calcular a variância amostral dos dados;

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$, onde $\bar{x}$ é a média amostral.

3º) Se n for par, atribuímos $k = \frac{n}{2}$; se n for ímpar $k = \frac{n+1}{2}$;

$$b = a_n(x_n - x_1) + a_{n-1}(x_{n-1} - x_2) + ... + a_{n-k-1}(x_{n-k-1} - x_k)$$
$$= \sum_{i=1}^{k} a_{n-i+1}(x_{n-i+1} - x_i),$$

onde os valores de a_{n-i+1} são tabelados (Tabela 9 e 10) e quando n for ímpar x_{k-1} não entra no cálculo.

4º) Calculo da estatística teste (W_c);

$$W_{calc.} = \frac{b^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

5º) Comparar o $W_{cal.}$ com o valor $W_{tab.}$ (Anexo B);

Conclusão: Se $W_{cal.} > W_{tab.}$ aceita-se H_0 (normalidade) para um α pré estabelecido.
Se $W_{cal.} < W_{tab.}$ rejeita-se H_0 (normalidade) para um α pré estabelecido.

Ex.: Escore padronizado do constructo de engajamento (Vigor) segue uma distribuição normal? Para α = 5%.

88,57	85,71	91,43	97,14	85,71	74,29	80,00	88,57	82,86

H_0: Os dados estão normalizados

H_1: Os dados não estão normalizados

Rol:

74,29	80,00	82,86	85,71	85,71	88,57	88,57	91,43	97,14

Cálculo da Média e Variância amostral:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 86{,}03 \qquad S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = 43{,}75$$

Visto que n = 9 e k = 10/2 = 5, conforme Tabela 9

k	n = 9
1	a_9 = 0,5888
2	a_8 = 0,3244
3	a_7 = 0,1976
4	a_6 = 0,0947
5	a_5 = 0,0000

b = 0,5888 (97,14–74,29) + 0,3244 (91,43–80,00) + 0,1976 (88,57–82,66) +
+ 0,0947 (88,57-85,71) + 0,0000

b = 18,6 $\qquad$ b^2 = 345,96

$$W_{cal.} = \frac{b^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{345{,}96}{350{,}01} = 0{,}988 \qquad W_{tab.} = 0{,}829\,(\alpha = 5\%)\,(\text{Tabela } 10)\,)$$

Conclusão: Como $W_{cal.} > W_{tab.}$, aceita-se H_0, ou seja, o constructo Vigor segue uma curva normal, para um nível de significância de 5%.

10 REGRESSÃO E CORRELAÇÃO

10.1 INTRODUÇÃO

Muitas vezes é de interesse estudar-se um elemento em relação a dois ou mais atributos ou variáveis simultaneamente.

Nesses casos presume-se que pelo menos duas observações são feitas sobre cada elemento da amostra. A amostra consistirá, então, de pares de valores, um valor para cada uma das variáveis, designadas, X e Y. Um valor "i" qualquer apresenta o par de valores (X_i ; Y_i).

O objetivo da regressão é registrar pares de valores (observações) de uma amostra, e estudar as relações entre as variáveis X e Y.

Para a análise de regressão interessam principalmente os casos em que a variação de um atributo é sensivelmente dependente do outro atributo.

O problema consiste em estabelecer a função matemática que melhor explique a relação existente entre as duas variáveis. Simbolicamente, a relação é expressa por uma equação de regressão e graficamente por uma curva de regressão.

10.2 DEFINIÇÃO

Constitui uma tentativa de estabelecer uma equação matemática linear que descreva o relacionamento entre duas variáveis (uma dependente e outra independente).

A equação de regressão tem por finalidade **ESTIMAR** valores de uma variável, com base em valores conhecidos da outra.

Ex.: Peso x Idade; Vendas x Lucro; Dedicação x Envolvimento.

10.3 MODELO DE REGRESSÃO

$$y = \alpha x + \beta + \varepsilon_i$$

y = variável dependente;
x = variável independente;
α = coeficiente de regressão (coeficiente angular);
β = coeficiente linear;
ε_i = erro aleatório.

10.3.1 Pressuposições

Na regressão, os valores de "y" são estimados com base em valores dados ou conhecidos de "x", logo a variável "y" é chamada variável dependente, e "x" variável independente.

a) A relação entre X e Y é linear (os acréscimos em X produzem acréscimos proporcionais em Y e a razão de crescimento é constante);
b) Y é uma variável aleatória que depende entre outras coisas dos valores de X;
c) ε_i é o erro aleatório, portanto uma variável aleatória com distribuição normal, com média zero e variância σ^2. [$\varepsilon_i \sim N(0, \sigma^2)$]. ε_i representa a variação de Y que não é explicada pela variável independente X;
d) Os erros são considerados independentes.

Ex.: X = Valor padronozado do constructo Dedicação (DE) de Engajamento no Trabalho
Y = Valor padronizado do constructo Envolvimento Pessoal no Trabalho (EPT) de Síndrome de Burnout

Obs.	1	2	3	4	5	6	7	8
DE	62,07	48,57	89,66	86,21	79,31	82,76	86,21	96,55
EPT	68,57	40,00	74,29	85,71	88,57	82,86	82,86	91,43

10.3.2 Gráfico (Diagrama de Dispersão)

Tem como finalidade ajudar na decisão se uma reta descreve adequadamente ou não o conjunto de dados.

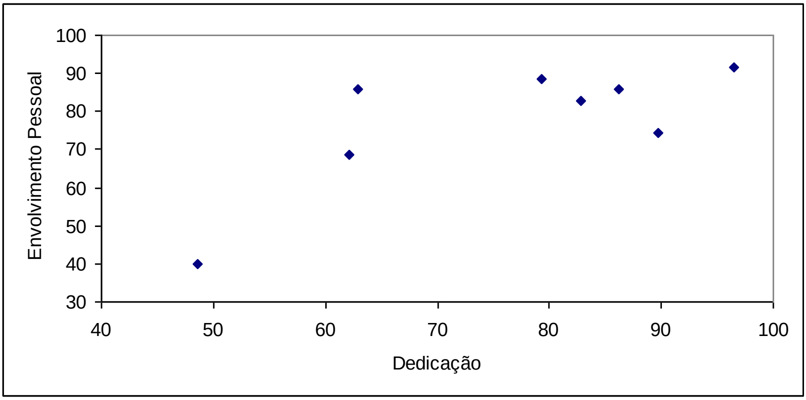


Figura 10.1 –Diagrama de dispersão entre dedicação e envolviemento pessoal

Pelo gráfico pode-se observar que a possível reta de regressão terá um coeficiente de regressão (coeficiente angular) positivo.

10.4 MÉTODO PARA ESTIMAÇÃO DOS PARÂMETROS α E β

As estimativas dos parâmetros α e β dadas por “a” e “b”, serão obtidas a partir de uma amostra de n pares de valores (x_i, y_i) que correspondem a n pontos no diagrama de dispersão.

O método mais usado para ajustar uma linha reta para um conjunto de pontos $(x_1, y_1), ..., (x_n, y_n)$ é o **Método de Mínimos Quadrados.**

O Método dos Mínimos Quadrados consiste em adotar como estimativa dos parâmetros os valores que minimizem a soma dos quadrados dos desvios.

10.4.1 Características

1ª) A soma dos desvios verticais dos pontos em relação a reta é zero;

2ª) A soma dos quadrados desses desvios é mínima.

10.4.2 Estimação dos parâmetros da reta de regressão

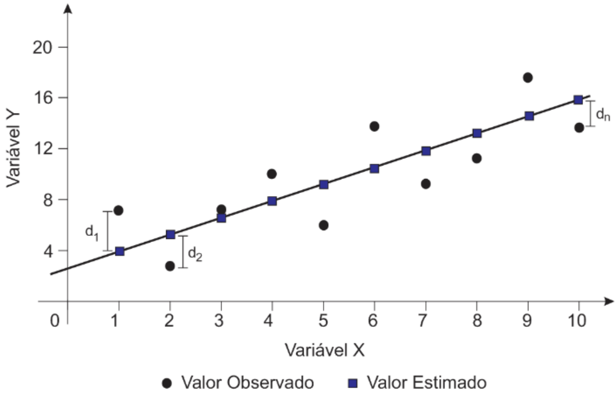


Figura 10.2 – Estimação da reta de regressão

$$S = \sum_{i=1}^{n} d_i^2 = d_1^2 + d_2^2 + \ldots + d_n^2 = \text{soma mínima;}$$

$$S = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (y - \hat{y})^2 = \sum_{i=1}^{n} (y - (ax + b))^2 \ .$$

onde "a" é o estimador de α e "b" é o estimador de β.

Esta soma será mínima se as derivadas parciais em relação a "a" e "b" forem nulas.

$$\frac{\partial S}{\partial a} = -2x \sum_{i=1}^{n} (y - (ax + b)) = 0$$

$$\sum_{i=1}^{n} (xy - bx - ax^2) = 0$$

$$\sum_{i=1}^{n} xy - b\sum_{i=1}^{n} x - a\sum_{i=1}^{n} x^2 = 0$$

$$\frac{\partial S}{\partial b} = -2 \sum_{i=1}^{n} (y - (ax + b)) = 0$$

$$\sum_{i=1}^{n} (y - b - ax) = 0$$

$$\sum_{i=1}^{n} y - nb - a\sum_{i=1}^{n} x = 0$$

$$a = \frac{\sum_{i=1}^{n} xy - \frac{\sum_{i=1}^{n} x \sum_{i=1}^{n} y}{n}}{\sum_{i=1}^{n} x^2 - \frac{\left(\sum_{i=1}^{n} x\right)^2}{n}} \qquad \Bigg| \qquad b = \frac{\sum_{i=1}^{n} y}{n} - a\,\frac{\sum_{i=1}^{n} x}{n}$$

os valores de "a" e "b" da reta de regressão $\hat{y} = a\,x + b$ serão:

$$a = \frac{n\sum_{i=1}^{n} xy - \sum_{i=1}^{n} x \sum_{i=1}^{n} y}{n\sum_{i=1}^{n} x^2 - \left(\sum_{i=1}^{n} x\right)^2} = \frac{S_{xy}}{S_{xx}} \qquad \Bigg| \qquad b = \bar{y} - a\,\bar{x}\,.$$

onde:

$$S_{xx} = \sum_{i=1}^{n} x^2 - \frac{\left(\sum_{i=1}^{n} x\right)^2}{n} = \sum_{i=1}^{n} (x - \bar{x})^2;$$

$$S_{xy} = \sum_{i=1}^{n} xy - \frac{\sum_{i=1}^{n} x \sum_{i=1}^{n} y}{n} = \sum_{i=1}^{n} (x - \bar{x})(y - \bar{y}),$$ (covariância de x e y)

Para cada par de valores (x_i , y_i) podemos estabelecer o desvio:

$e_i = y_i - \hat{y}_i = y_i - (a + bx_i)$.

Para facilitar os cálculos da reta de regressão, acrescentamos três novas colunas na tabela original.

Obs.	DE	EPT	DE x EPT	DE^2	EPT^2
1	62,07	68,57	4.256,16	3.852,56	4.702,04
2	48,57	40,00	1.942,86	2.359,18	1.600,00
3	89,66	74,29	6.660,10	8.038,05	5.518,37
4	86,21	85,71	7.389,16	7.431,63	7.346,94
5	79,31	88,57	7.024,63	6.290,13	7.844,90
6	62,86	85,71	5.387,76	3.951,02	7.346,94
7	82,76	82,86	6.857,14	6.848,99	6.865,31
8	96,55	91,43	8.827,59	9.322,24	8.359,18
Σ	607,98	617,14	48.345,39	48.093,80	49.583,67

$$a = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{8(48.345,39) - (607,98)(617,14)}{8(48.093,80) - (607,98)^2} = 0,7645$$

$$b = \bar{y} - a\,\bar{x} = 77,14 - (0,7645)(76,00) = 19,038$$

$$EPT = 0,7645DE + 19,038$$

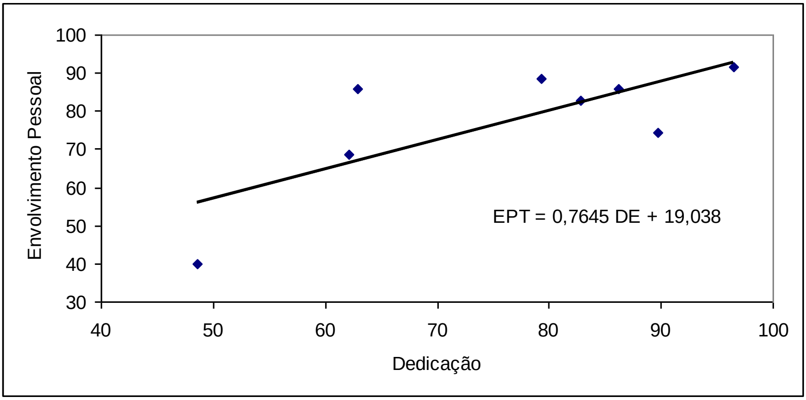


Figura 10.3 – Linha de tendência da reta de regressão

Partindo da reta de regressão pode-se afirmar que para um funcionário que tenha um escore para Dedicação de 55,17 terá um escore de Envolvimento Pessoal de:

$$EPT = 0,7645(55,17) + 19,038 = 61,20$$

Obs.: Para qualquer tipo de equação de regressão deve-se ter muito cuidado para não extrapolar valores para fora do âmbito dos dados. O perigo da extrapolação para fora dos dados amostrais, é que a mesma relação possa não mais ser verificada.

10.5 DECOMPOSIÇÃO DA VARIÂNCIA TOTAL

A dispersão da variação aleatória "y" pode ser medida através da soma dos quadrados dos desvios em relação a sua média $\overline{y}$. Essa soma de quadrados será denominada Soma de Quadrados Total (SQTotal).

$$SQ_{Tot.} = \sum_{i=1}^{n} (y_i - \overline{y})^2 .$$

A $SQ_{Tot.}$ pode ser decomposta da seguinte forma:

$$\sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 .$$

Essa relação mostra que a variação dos valores de Y em torno de sua média pode ser dividida em duas partes: uma $\sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2$ que é explicada pela regressão e outra $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$, que é explicada devido ao fato de que nem todos os pontos estão sobre a reta de regressão, que é a parte "não explicada" pela regressão ou variação residual.

Assim:

$SQ_{Tot.} = SQ_{Reg.} + SQ_{Res.}$.

A estatística definida por $r^2 = \dfrac{SQ_{Reg.}}{SQ_{Tot.}}$, e denominada coeficiente de determinação, indica a proporção ou percentagem da variação de Y que é "explicada" pela regressão.

Note que $0 \leq r^2 \leq 1$.

Fórmulas para cálculo:

$$SQ_{Tot.} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n},$$

com (n - 1) graus de liberdade.

$$SQ_{Reg.} = a\left(\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right),$$

com 1 grau de liberdade.

10.6 ANÁLISE DE VARIÂNCIA DA REGRESSÃO

A Soma dos Quadrados da Regressão (SQ Regressão), segue uma distribuição de χ^2 (Qui-quadrado) com (1) grau de liberdade, enquanto que a Soma dos Quadrados do Resíduo (SQ Resíduos) segue a mesma distribuição, porém com (n - 2) graus de liberdade. Portanto, o quociente:

$$\frac{SQ_{\ Reg.}/1}{SQ_{\ Res.}/n-2} = \frac{QM_{Reg.}}{QM_{Res.}},$$

segue uma distribuição F de Snedecor com 1 e (n - 2) graus de liberdade.

Esse fato nos permite empregar a distribuição F de Snedecor para testar a significância da regressão, através da chamada Análise de Variância, sintetizada no quadro abaixo:

- Análise de Variância do modelo de regressão

Causas de Variação	G. L.	S. Q.	Q. M.	$F_{cal.}$
Regressão	1	$SQ_{Reg.}$	$\frac{SQ_{Reg.}}{1}$	$\frac{QM_{Reg.}}{QM_{Res.}}$
Resíduo	n - 2	$SQ_{Res.}$	$\frac{SQ_{Res.}}{n-2}$	
Total	n - 1	$SQ_{Tot.}$	---	---

onde, QM representa Quadrado Médio que é obtido pela divisão da SM (Soma de Quadrados) pelos respectivos GL (Graus de Liberdade).

10.7 TESTE PARA VERIFICAR A SIGNIFICÂNCIA DA REGRESSÃO

Para testar a significância da regressão, formulam-se as seguintes hipóteses:

H_0: $\alpha = 0$
H_1: $\alpha \neq 0$

onde α representa o coeficiente de regressão paramétrico.

Se o valor de F, calculado a partir do quadro de análise de variância, superar o valor teórico de F com 1 e (n - 2) graus de liberdade, para o nível de significância α, rejeita-se H_0 e conclui-se que a regressão é significativa.

Se $F_{cal.} > F_{\alpha\,[1,(n-2)]}$, rejeita-se H_0.

Para o exemplo anterior:

Obs.	DE	EPT	DE x EPT	DE^2	EPT^2
1	62,07	68,57	4.256,16	3.852,56	4.702,04
2	48,57	40,00	1.942,86	2.359,18	1.600,00
3	89,66	74,29	6.660,10	8.038,05	5.518,37
4	86,21	85,71	7.389,16	7.431,63	7.346,94
5	79,31	88,57	7.024,63	6.290,13	7.844,90
6	62,86	85,71	5.387,76	3.951,02	7.346,94
7	82,76	82,86	6.857,14	6.848,99	6.865,31
8	96,55	91,43	8.827,59	9.322,24	8.359,18
Σ	607,98	617,14	48.345,39	48.093,80	49.583,67

$$EPT = 0{,}7645DE + 19{,}038$$

$$SQ_{Reg.} = a\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right]$$

$$SQ_{Reg.} = 0,7645\left[(48.345,39) - \frac{(607,98)(617,14)}{8}\right] = 1.103,98$$

$$SQ_{Tot.} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

$$SQ_{Tot.} = (49.583,67) - \frac{(617,14)^2}{8} = 1.975,51$$

Causas de Variação	G. L.	S. Q.	Q. M.	$F_{cal.}$
Regressão	1	1.103,98	1.103,98	7,60
Resíduo	6	6.975,75	145,26	
Total	7	1.975,51	---	---

H_0: $\alpha = 0$ $\qquad$ H_1: $\alpha \neq 0$

Comparando o $F_{cal.} = 7,60$ com o $F_{tab.} = F_{0,05\ (1,6)} = 5,99$

Conclui-se que a regressão de y sobre x segundo o modelo $EPT = 0,7645DE + 19,038$ é significativa ao nível de significância de 5%. Uma vez estabelecida e testada a equação de regressão, a mesma pode ser usada para explicar o relacionamento entre as variáveis e também para fazer previsões dos valores de Y para valores fixados de X.

10.8 TESTE PARA O COEFICIENTE DE REGRESSÃO

Para testar a significância da regressão, formulam-se as seguintes hipóteses:

H_0: $\alpha = \alpha_o$
H_1: $\alpha \neq \alpha_o$

onde α_o é uma constante qualquer, no caso, $\alpha_0 = 0$.

Para testar a significância do coeficiente de regressão deve-se calcular:

1º) Desvio padrão residual

$$S_r = \sqrt{\frac{SQ_{Res.}}{n-2}} = \sqrt{QM_{Res.}} \ .$$

2º) Estimativa do desvio padrão do coeficiente de correlação

$$S_a = S_r \sqrt{\frac{1}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}} \ .$$

3º) Determinar o valor do teste calculado (teste t).

$$t_{cal.} = \frac{a - \alpha_o}{S_a} \ ,$$

o valor do $t_{tab.}$ é definido com um nível de significância α com n – 2 graus de liberdades.

Se $t_{cal.} > t_{tab.}$, rejeita-se H_0, ou seja, o coeficiente de correlação é significativo.

Para um nível de confiança α, com n – 2 graus de liberdade construir o intervalo de confiança para o coeficiente de regressão (usar teste unilateral).

$$IC(\alpha) = a \pm t_{\alpha/2} \, . S_a \ .$$

10.9 Teste para o coeficiente linear (intercepto)

Para testar a significância do intercepto, formulam-se as seguintes hipóteses:

H_0: $\beta = \beta_0$
H_1: $\beta \neq \beta_0$

onde β_o é uma constante qualquer, no caso, $\beta_0 = 0$.

Para testar a significância do intercepto deve-se calcular:

1º) Estimativa do desvio padrão do intercepto

$$S_b = S_r . \sqrt{\frac{1}{n} + \frac{\left(\overline{X}\right)^2}{\sum_{i=1}^{n} x_1^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}} .$$

2º) Determinar o valor do teste calculado (teste t).

$$t_{cal.} = \frac{b - \beta_0}{S_b} \ ,$$

o valor do $t_{tab.}$ é definido com um nível de significância α e n – 2 graus de liberdades.

Se $t_{cal.} > t_{tab.}$, rejeita-se H_0, ou seja, o intercepto é significativo.

Para um nível de confiança α, com n – 2 graus de liberdade construir o intervalo de confiança para o intercepto (usar teste unilateral).

$$IC(\beta) = b \pm t_{\alpha/2} . S_b \ .$$

Para calcular a significância dos parâmetros pode-se construir a seguinte tabela:

Parâmetros	G. L.	Erro Padrão	$t_{cal.}$	$t_{tab.\,(5\%)}$
Intercepto	6	21,501	0,886	2,447
Coef. de Regressão	6	0,277	2,777	2,447

Analisando o coeficiente de regressão, o $t_{cal.} > t_{tab.}$, logo o parâmetro é significativo para um nível de significância de 5%.

10.10 COEFICIENTE DE DETERMINAÇÃO (R²)

É o grau em que as predições baseadas na equação de regressão superam as predições baseadas em $\overline{y}$, ou ainda é a proporção entre a variância explicada pela variância total.

Variância Total = soma dos desvios ao quadrado,

$$VT = SQTotal = n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2 .$$

Variância Não-explicada = soma de quadrados dos desvios em relação a reta $\hat{y}$.

Para facilitar os cálculos usar-se-á:

$$r^2 = \frac{\left(n\sum_{i=1}^{n} xy - \sum_{i=1}^{n} x \sum_{i=1}^{n} y\right)^2}{\left[n\sum_{i=1}^{n} x^2 - \left(\sum_{i=1}^{n} x\right)^2\right]\left[n\sum_{i=1}^{n} y^2 - \left(\sum_{i=1}^{n} y\right)^2\right]} = \frac{COV_{xy}}{S_{xx}.S_{yy}}.$$

$$r^2 = \frac{[8(48.345,39) - (607,98)(617,14)]^2}{[8(48.093,8) - (607,98)^2][8(49.583,67) - (617,14)^2]} = 0,5589.$$

O valor de r² varia de 0 a 1, logo o fato de r² = 0,5589 (no exemplo), indica que aproximadamente 56% do Envolvimento Pessoal Dedicação estão relacionados com a Dedicação, em outras palavras 44% da variação do Envolvimento Pessoal não são explicados pela Dedicação.

10.11 COEFICIENTE DE CORRELAÇÃO DE PEARSON (R OU ρ)

Tem como objetivo encontrar o grau de relação entre duas variáveis, ou seja, um coeficiente de correlação.

Esta é a forma mais comum de análise, envolvendo dados contínuos, conhecido como "**r de Pearson**".

10.11.1 Características do "r"

- Pode assumir valores positivos (+) como negativos (-), é semelhante ao coeficiente de regressão de uma reta ajustada num diagrama de dispersão;

- A magnitude de r indica quão próximos da "reta" estão os pontos individuais;

- quando o r se aproxima de +1 indica pouca dispersão, e uma correlação muito forte e positiva;

- quando o r se aproxima de "zero" indica muita dispersão, e uma ausência de relacionamento;

- quando o r se aproxima de -1 indica pouca dispersão, e uma correlação muito forte e negativa.

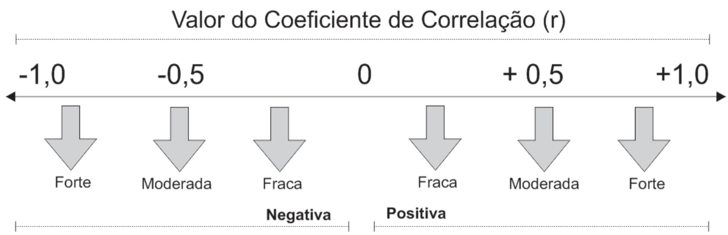


Figura 10.4 – Fiabilidade do coeficiente de correlação

- Fiabilidade do coeficiente de correlação de Pearson:

Valor de r (+ ou -)	Interpretação*
0,00	Nula
0,01 a 0,20	Ínfima fraca
0,21 a 0,40	Fraca
0,41 a 0,60	Moderada
0,61 a 0,80	Forte
0,81 a 0,99	Ínfima Forte
1,00	Perfeita

* a classificação só será válida se o valor da correlação for significativo, p < 0,05

Obs.: Para valores de r ≠ 0 e p > 0,05 entende-se que o valor de r é estatisticamente = 0 em função da interpretação do teste de significância para o coeficiente de correlação ρ (item 10.12);

10.11.2 Medidas de Correlação

- Tratamento Qualitativo (correlação momento produto)

Relação entre as variáveis, mediante a observação do diagrama de dispersão.

- Tratamento Quantitativo

É o estabelecimento das medidas de correlação.

O valor de "r" pode ser enganoso, na realidade, uma estatística mais significante é o r² (coeficiente de determinação), que dá o valor percentual da variação de uma variável explicativa em relação a outra variável.

$$r = \frac{\left(n\sum_{i=1}^{n} xy - \sum_{i=1}^{n} x \sum_{i=1}^{n} y\right)}{\sqrt{\left[n\sum_{i=1}^{n} x^2 - \left(\sum_{i=1}^{n} x\right)^2\right]}\sqrt{\left[n\sum_{i=1}^{n} y^2 - \left(\sum_{i=1}^{n} y\right)^2\right]}} = \frac{S_{xy}}{\sqrt{S_{xx}.S_{yy}}}.$$

Logo pode-se observar que o coeficiente de determinação nos dá uma base intuitiva para a análise de correlação.

No exemplo tem-se: r = + 0,7476

Obs.: O sinal + significa dizer que quanto maior for a Dedicação maior será o Envolvimento Pessoal, apresentando uma correlação forte.

10.11.3 Teste para o coeficiente de correlação de Pearson

Para testar a significância da regressão, formulam-se as seguintes hipóteses.

H_0: $\rho = 0$,
H_1: $\rho \neq 0$.

O teste utilizado é o teste t.

$$t_{cal.} = \sqrt{\frac{r^2(n-2)}{1-r^2}} = \frac{a}{S_a} \ ,$$

o valor do $t_{tab.}$ é definido com um nível de significância α com n – 2 graus de liberdades (usar teste bilateral).

$$t_{cal.} = \sqrt{\frac{0{,}5589.(8-2)}{1-0{,}5589}} = 2{,}7572 \qquad t_{tab.} = t_{5\%;6} = 1{,}943 \ ,$$

observando a Figura 10.3, tem-se que a correlação é forte e como $t_{cal.} > t_{tab.}$, rejeita-se H_0, ou seja, o coeficiente de correlação é significativamente diferente de zero para um nível de significância de 5%.

10.12 CORRELAÇÃO DE SPEARMAN

Conhecida por coeficiente de correlação de postos, nos casos em que os dados não formam uma nuvem comportada, com alguns pontos bem distantes dos demais, ou em que parece existir uma relação crescente ou descrescente num formato de curva, quando os dados não pertencem à uma escala de medida padrão, mas existe uma ordenação clara.

É um método não-paramétrico que usa somente os postos, e não faz quaisquer suposições. Uma equação que é relativamente fácil de usar é:

$$\rho_s = 1 - \frac{6\sum_{i=1}^{n} d_1^2}{n^3 - n},$$

onde,

$d_i^2 =$ (posto de x_i dentre os valor de x) - (posto de y_i dentre os valores de y) .

Os dados a seguir referem-se a valores padronizados de dois constructos de Engajamento no Trabalho:

Ex.: X = Valor padronozado do constructo Vigor (VI);
Y = Valor padronozado do constructo Dedicação (DE).

Obs.	1	2	3	4	5	6	7	8
VI	91,43	85,71	80,00	48,57	51,43	88,57	62,86	77,14
DE	96,55	89,66	89,66	58,62	58,62	93,10	62,07	93,10

Obs.	VI	DE	Posto (VI)	Posto (DE)	d_i	d_i^2
1	91,43	96,55	8	8	0	0
2	85,71	89,66	6	5,5	0,5	0,25
3	80,00	89,66	5	5,5	-0,5	0,25
4	48,57	58,62	1	1,5	-0,5	0,25
5	51,43	58,62	2	1,5	0,5	0,25
6	88,57	93,10	7	7,5	-0,5	0,25
7	62,86	62,07	3	4	-1	1
8	77,14	93,10	4	7,5	-3,5	12,25
Σ	---	---	---	---	---	14,5

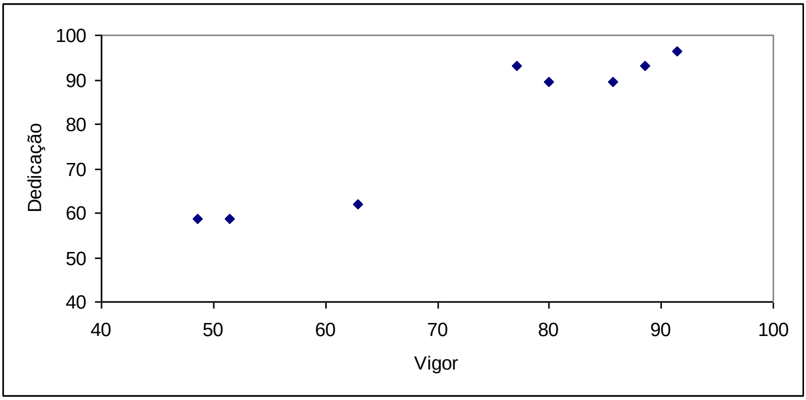


Figura 10.5 – Diagrama de dispersão entre vigor e dedicação

$$r_s = 1 - \left(\frac{6 \;.\; 14{,}5}{512 - 8} \right) = 1 - 0{,}1726 = +0{,}8274$$

O sinal + significa dizer que quanto maior for o Vigor maior será a Dedicação, apresentando uma correlação ínfima forte, como já mencionado na fiabilidade da correlação de Pearson.

10.12.1 Teste para o coeficiente de correlação de Spearman

Para testar a significância da regressão, formulam-se as seguintes hipóteses.

H_0: $\rho_s = 0$,
H_1: $\rho_s \neq 0$.

O teste utilizado é o teste t.

$$t_{cal.} = \sqrt{\frac{r_s^2 (n - 2)}{1 - r_s^2}} \quad ,$$

o valor do $t_{tab.}$ é definido com um nível de significância α com n – 2 graus de liberdades (usar teste bilateral).

$$t_{cal.} = \sqrt{\frac{0{,}6846.(8-2)}{1-0{,}6846}} = 3{,}61 \qquad t_{tab.} = t_{5\%;6} = 1{,}943.$$

Observa-se que a correlação é infimamente forte e como $t_{cal.} > t_{tab.}$, rejeita-se H_0, ou seja, o coeficiente de correlação é significativamente diferente de zero para um nível de significância de 5%.

11 ANÁLISE DE VARIÂNCIA

11.1 INTRODUÇÃO

A análise de variância foi desenvolvida por Fisher, como instrumento para a análise de experimentos agrícolas.

A idéia, na análise de variância, é comparar a **variância devida aos tratamentos** com a **variação devida ao acaso,** ou resíduo. Para fazer uma análise de variância, é preciso proceder a uma série de cálculos. Mas, a aplicação das fórmulas, exige conhecimento da notação, conforme será mostrado posteriormente.

O objetivo da análise de variância é analisar as diferenças entre as médias aritméticas dos grupos, a partir de uma análise na variação dos dados, entre os grupos. Na realidade, toma-se a variação total e subdivide-se-a em **variação entre os grupos** e a **variação dentro do grupo**, a qual se considera como um **erro experimental,** mas se a variação ocorrer **entre os grupos** ela é atribuída ao efeito do tratamento recebido.

11.2 PRESSUPOSIÇÕES BÁSICAS À APLICAÇÃO DA ANOVA

As k populações em estudo tenham a mesma variância, ou seja, sejam homocedásticas.

A variável de estudo seja normalmente distribuída em todas as populações.

11.3 EXPERIMENTO INTEIRAMENTE AO ACASO

- Mesmo número de repetições

Em experimentos completamente casualizados, com um fator fixo, tem-se interesse em verificar a influência dos k níveis desse fator, em uma variável dependente Y, em estudo.

Hipóteses a serem testadas:

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$
H_1: existe pelo menos uma média que difere das demais

- Experimento inteiramente ao acaso

	Tratamento					**Total**
	1	**2**	**3**	**...**	**k**	
	y_{11} y_{12} y_{13} . . . y_{1r}	y_{21} y_{22} y_{23} . . . y_{2r}	y_{31} y_{32} y_{33} . . . y_{3r}	...	y_{k1} y_{k2} y_{k3} . . . y_{kr}	
Total	T_1	T_2	T_3	...	T_k	$\Sigma T = \Sigma y$
N. de Repet.	r	r	r	...	r	n = kr
Média	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	...	$\bar{y}_k$	

Para se fazer a análise de variância de um experimento ao acaso é preciso calcular as seguintes quantidades:

a) os graus de liberdade:

tratamentos : k - 1;
total : (k . r) - 1;
resíduo : k . (r - 1);

b) O valor de C, conhecido como fator de correção:

$$C = \frac{\left(\sum y\right)^2}{N}.$$

c) a soma de quadrados total:

$$SQ_{Tot.} = \sum y^2 - C.$$

d) a soma de quadrados dos tratamentos:

$$SQ_{Trat.} = \frac{\sum T^2}{r} - C.$$

e) a soma de quadrados dos resíduos:

$$SQ_{Res.} = SQ_{Tot.} - SQ_{Trat.}.$$

f) o quadrado médio dos tratamentos:

$$QM_{Tra.t} = \frac{SQ_{Trat.}}{k-1}.$$

g) o quadrado médio de resíduo:

$$QM_{Res.} = \frac{SQ_{Res.}}{k(r-1)}.$$

h) o valor de F:

$$F = \frac{QM_{Trat.}}{QM_{Res.}}.$$

Nota-se que os quadrados médios são obtidos dividindo-se as somas de quadrados pelos respectivos graus de liberdade, obtendo-se, assim, as três variâncias, ou termos quadráticos médios.

Uma vez que a variância é calculada dividindo-se a soma das diferenças ao quadrado por seus graus de liberdade apropriados, todos dos termos quadráticos de médias são variâncias.

Todas as quantidades calculadas são apresentadas em uma tabela de análise de variância.

- Quadro de análise de variância de um experimento inteiramente ao acaso

Fontes de Variação	**S. Q.**	**G. L.**	**Q. M.**	**$F_{cal.}$**
Tratamentos	$SQ_{Trat.}$	k - 1	$QM_{Trat.}$	$F_{calc.} = \frac{QM_{Trat.}}{QM_{Res.}}$
Resíduo	$SQ_{Res.}$	k (r - 1)	$QM_{Res.}$	
Total	$SQ_{Tot.}$	(k .r) - 1		

Para se testar as hipóteses é utilizada a estatística F, de Snedecor, com (k - 1) graus de liberdade no numerador e k (r - 1) graus de liberdade no denominador. Se $F_{cal.} > F_{\alpha, \delta_1, \delta_2}$, rejeita-se H_0 e conclui-se que existe, pelo menos, uma média que difere de outra.

Se $F_{cal.} > F_{tab.}$, rejeitar H_0. Nesse caso, diz-se que existem diferenças estatisticamente significativas entre as médias.

Se $F_{cal.} \leq F_{tab.}$, não rejeitar H_0. Quando isso ocorre, diz-se que não existem evidências estatísticas de que as médias sejam diferentes.

- Diferente número de repetições

A análise estatística de um experimento inteiramente ao acaso, com número diferente de repetições, não apresenta maior dificuldade. Todos os cálculos são feitos da maneira já apresentada antes, com exceção da soma de quadrados de tratamentos. A soma de quadrados de tratamentos é dada pela fórmula:

$$SQ_{Trat.} = \sum \frac{T_i^2}{r_i} - C = \frac{T_1^2}{r_1} + \frac{T_2^2}{r_2} + ... + \frac{T_k^2}{r_k} - C.$$

11.4 EXPERIMENTO EM BLOCOS AO ACASO

A análise de variância de classificação dupla serve para testar, simultaneamente, diferenças entre médias, levando em consideração dois tratamentos ou fatores. Dessa forma, a observação admite a influência dos seguintes elementos:

$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$.

onde:

μ = representa o efeito, devido a média da população que pertence;
α_i = é o efeito do i-ésimo tratamento;
β_j = é o efeito do j-ésimo tratamento;
ε_{ij} = representa o erro residual, que se pressupõem ter distribuição normal com média zero e variância constante.

As hipóteses a serem testadas são as seguintes:

H_{01}: Não existe diferença significativa entre as médias dos tratamentos (colunas);

H_{02}: Não existe diferença significativa entre as médias dos blocos (linhas);

Para entender como se faz a análise de variância de um experimento em blocos ao acaso, primeiro observe a Tabela 10.3. Nessa tabela estão indicados os dados de um experimento em blocos ao acaso, com *k* tratamentos e *r* blocos. O total de cada tratamento é dado pela soma das *r* unidades submetidas a esse tratamento; o total do bloco é dado pela soma das *k* unidades do bloco.

- Experimento em blocos ao acaso

Blocos	**Tratamento**					
	1	**2**	**3**	**...**	K	Total
1	y_{11}	y_{21}	Y_{31}	...	y_{k1}	B_1
2	y_{12}	y_{22}	Y_{32}		y_{k2}	B_2
3	y_{13}	y_{23}	Y_{33}		y_{k3}	B_3
.	.	.	.		.	.
.	.	.	.		.	.
.	.	.	.		.	.
r	y_{1r}	y_{2r}	y_{3r}		y_{kr}	B_r
Total	T_1	T_2	T_3	...	T_k	$\Sigma T = \Sigma B = \Sigma y$
Número de Rep.	r	r	r	...	r	$n = k \cdot r$
Média	$\bar{y}_1$	$\bar{y}_2$	$\bar{y}_3$	...	$\bar{y}_k$	

Para fazer a análise de variância, de um experimento em blocos ao acaso, é preciso calcular:
a) os graus de liberdade:

do total:	k r – 1
dos tratamentos:	k – 1
dos blocos:	r – 1
dos resíduos:	(k r – 1) – (k – 1) – (r – 1) = (k – 1) (r – 1)

b) o valor de C, que é dado pelo total geral elevado ao quadrado e dividido pelo número de observações:

$$C = \frac{\left(\sum y\right)^2}{kr}.$$

c) a soma de quadrados total:

$$SQ_{Tot.} = \sum y^2 - C.$$

d) a soma de quadrados dos tratamentos:

$$SQ_{Trat.} = \frac{\sum T^2}{r} - C.$$

e) a soma de quadrados de blocos:

$$SQ_{Blo.} = \frac{\sum B^2}{k} - C.$$

f) a soma de quadrados dos resíduos:

$$SQ_{Res.} = SQ_{Tot.} - SQ_{Trat.} - SQ_{Blo.}$$

As somas de quadrados são apresentadas na tabela de análise de variância. Para calcular os quadrados médios, basta dividir cada soma de quadrados pelos respectivos graus de liberdade. O valor de $F_{cal.}$, para tratamentos, é dado pelo quociente entre o quadrado médio dos tratamentos e o quadrado médio dos resíduos; o valor de $F_{cal.}$, para os blocos, é dado pelo quociente entre o quadrado médio dos blocos e o quadrado médio dos resíduos.

- Análise de variância de um experimento em blocos ao acaso

Fontes de Variação	S. Q.	G. L.	Q. M.	$F_{cal.}$
Tratamentos	$SQ_{Trat.}$	$k - 1$	$QM_{Trat.}$	$F_{Trat.} = \frac{QM_{Trat.}}{QM_{Res.}}$
Blocos	$SQ_{Blo.}$	$r - 1$	$QM_{Blo.}$	$F_{Blo.} = \frac{QM_{Blo.}}{QM_{Res.}}$
Resíduos	$SQ_{Res.}$	$(k - 1)(r - 1)$	$QM_{Res.}$	---
Total	$SQ_{Tot.}$	$k\,r - 1$	---	---

11.5 TESTE PARA COMPARAÇÃO DE MÉDIAS

O objetivo principal da ANOVA é apontar se um grupo é estatisticamente diferente do outro, ou não. Logo, se a hipótese nula é rejeitada, a um determinado nível de significância, sabe-se, então, que existe, pelo menos, uma das médias de um tratamento que difere de outra.

- Amostra de mesmo tamanho

Para se determinar qual ou quais tratamentos não são estatisticamente iguais, utiliza-se uma diferença mínima significativa (**d.m.s.**), que é utilizada para comparar as médias dos tratamentos.

Nada impede que a hipótese nula (H_0) seja aceita, isto é, que as médias dos tratamentos sejam consideradas iguais e que uma investigação seja conduzida, neste caso, o método de comparação das médias é dito **não-protegido.** Caso H_0 seja rejeitada e uma investigação seja conduzida, então o método é dito **protegido**.

Os testes utilizados, para se encontrar a d.m.s., são: teste "t" de Student; teste de Tukey; teste de Dunnett. Existem outros testes mas não serão abordados.

a) Teste de Tukey

Encontra-se:

$$\text{dms} = q_{\alpha;(\delta,k)} \sqrt{\frac{QM_{Res.}}{r}} \quad ,$$

onde:

q = é o valor tabelado, levando-se em consideração os graus de liberdade do resíduo (δ), o número de tratamentos (k) e o nível de significância (α);
$QM_{Res.}$ = é o quadrado médio dos resíduos;
r = é o número de repetições de cada tratamento.

Os valores para a estatística "q" apresentam-se tabelados.

b) Teste de Dunnett

Este teste deve ser aplicado toda vez que se pretenda comparar as médias dos tratamentos apenas com a média controle.

Encontra-se:

$$\text{dms} = d_{\alpha;(\delta,T)} \sqrt{\frac{2\,QM_{Res.}}{r}} \quad ,$$

onde:

d = é o valor tabelado ao nível de significância estabelecido (α), grau de liberdade do resíduo (δ) e o número de grupos tratados (T);
$QM_{Res.}$ = é o quadrado médio dos resíduos;
r = é o número de repetições de cada tratamento.

Os valores da estatística "d" encontram-se tabelados.

- Amostras de tamanho diferentes

O método, para o cálculo da **dms**, é semelhante ao exposto anteriormente, apenas com o diferencial de que o número de repetições, em cada tratamento, deve ser levado em consideração e que a **dms** deve ser calculada a cada diferença que se queira investigar. Logo, apresenta-se uma tabela com o resumo das formulações.

Teste	Fórmulas
Teste de Tukey	$\mathrm{dms} = q_{\alpha,(\delta,k)} \cdot \sqrt{\left(\frac{1}{r_i} + \frac{1}{r_j}\right)\frac{QM_{Res.}}{2}}$
Teste de Dunnett	$\mathrm{dms} = d_{\alpha,(\delta,T)} \cdot \sqrt{\left(\frac{1}{r_t} + \frac{1}{r_c}\right)\frac{QM_{Res.}}{2}}$

onde r_i e r_j são os números de repetições de cada tratamento. Aqui também o teste de Dunnett é usado para comparar o grupo tratado com o grupo controle, e r_t e r_c representam os números de repetições de cada grupo, respectivamente.

12 NÚMEROS ÍNDICES

12.1 INTRODUÇÃO

Um número índice é uma medida estatística idealizada para mostrar as variações de uma variável, ou um grupo de variáveis, correlacionadas ao tempo, à localização geográfica, ou a outras características como rendimento, profissão etc. Uma coleção de números-índices de diversos anos, localidades etc é freqüentemente denominada série de índices.

A quantidade total de dinheiro gasto cada ano, em relação a certo ano-base, varia de um ano para outro devido as variações no número de unidades compradas dos diferentes artigos e igualmente devido a mudanças nos preços unitários de tais artigos. Temos, portanto, três variáveis em jogo: ***preço***, ***quantidade*** e ***valor***, sendo este último o resultado do produto do preço pela quantidade.

12.2 RELATIVO DE PREÇO

Trata-se do número índice mais simples. Relacionando-se o preço de um produto numa época *t* (chamada época atual) com o de uma época *o* (chamada básica ou simplesmente base) teremos um relativo de preço. Fazendo-se p_t igual ao preço numa época atual e p_o igual ao preço na época-base, definimos relativo de preço pela seguinte quantidade:

$$p_{(o,t)} = \frac{p_t}{p_o} \quad ,$$

se quiser expressar em termos percentuais o relativo de preço, basta multiplicar o quociente acima por 100.

Observação: O simbolo $\boldsymbol{P_{(o,t)}}$ pode ser escrito também por $\boldsymbol{P_{o,t}}$.

Exemplo: O preço de determinado artigo em 2015 foi de R$ 12,00 e em 2016 subiu para R$ 13,80. Tomando-se como base o ano de 2015, determinar o preço relativo em 2016.

Resolução:

O ano tomado como base corresponderá sempre ao índice igual a 100. Os demais apresentarão, portanto, valores que flutuarão em torno de 100.
Então:

$$p_{(2015,2016)} = \frac{p_{2016}}{p_{2015}} = \frac{13,8}{12,0} = 1,15$$, ou 115% ou simplesmente 115.

Esse resultado indica que em 2016 houve um aumento de 15% no preço do artigo com relação ao preço do mesmo artigo em 2015.

Se tiver P_{2016}=11,20 (reais) e P_{2015}=12,00, o relativo de preço seria:

$$p_{(2015,2016)} = \frac{p_{2016}}{p_{2015}} = \frac{11,2}{12,0} = 0,93$$ ou 93% ou 93.

Em 2016 o artigo em questão apresentou um preço 7% inferior ao de 2015.

12.3 RELATIVO DE QUANTIDADE

Assim como se podem comparar os preços de bens, pode-se também fazer em relação a quantidades, quer seja, elas produzidas, vendidas ou consumidas, se utizar q_t igual a quantidade de um produto na época atual (época *t*) e q_o igual a quantidade desse mesmo produto na época *0* (básica), a quantidade relativa será o seguinte quociente:

$$q_{(o,t)} = \frac{q_t}{q_o}$$,

que representa a variação da quantidade na época *t* com relação a uma época *0* (base).

Exemplo: Uma empresa produziu 45 toneladas de aço em 2015 e 68 toneladas em 2016. A quantidade relativa será, tomando-se o ano de 2015 como base:

Resolução:

$q_{(2015,2016)} = \dfrac{q_{2016}}{q_{2015}} = \dfrac{68}{45} = 1{,}51$ ou 151% ou 151.

No ano de 2016 esta empresa aumentou sua produção de 51% acima, em relação a 2015.

12.4 RELATIVO DE VALOR

Se *p* for o preço de determinado artigo em certa época e *q* a quantidade produzida ou consumida desse mesmo artigo na mesma época, então, o produto *p.q* é denominado valor total de produção ou de consumo. Sendo p_t e q_t respectivamente o preço e a quantidade de um artigo na época atual (*t*) e p_o e q_o, o preço e a quantidade do mesmo artigo na época base (*o*), , definimos como valor total relativo ou simplesmente valor relativo o quociente:

$$v_{(o,t)} = \frac{v_t}{v_o} = \frac{p_t . q_t}{p_o . q_o} = p_{o,t} . q_{o,t} \quad .$$

O fato de $v_{(o,t)} = p_{o,t} . q_{o,t}$ é conhecido como propriedade da ***reversibilidade dos fatores,*** ou como, ***critério da decomposição das causas***.

Exemplo: Uma empresa vendeu, em 2015, 1000 unidades de um artigo ao preço unitário de R$ 500,00. Em 2016 vendeu 2000 unidades do mesmo artigo ao preço unitário de R$ 600,00. O valor relativo de venda em 2016 foi:

Resolução:

$$v_{(2015,2016)} = \frac{v_{2016}}{v_{2015}} = \frac{p_{2016} . q_{2016}}{p_{2015} . q_{2015}} = \frac{600x2000}{500x1000} = 2{,}4 = 240\%$$

Em 2016, o valor das vendas foi 140% superior ao de 2015.

12.5 ELOS RELATIVOS E RELATIVOS EM CADEIA

Considere uma seqüência de preços relativos do tipo $p_{1,2}$, $p_{2,3}$, $p_{3,4}$, ..., cujos preços relativos sejam considerados em intervalos sucessivos de tempo. Esta seqüência constitui o que se denomina elos relativos. Suponha-se, para exemplificar, que certa utilidade apresentou os seguintes preços no período de 2013 a 2016: 80, 120, 150 e 180. Os elos relativos serão: 150, 125, 120, os quais foram determinados da seguinte maneira:

$$p_{2013,2014} = \frac{120}{80} x100 = 150\%$$
$$p_{2014,2015} = \frac{150}{120} x100 = 125\%$$
$$p_{2015,2016} = \frac{180}{150} x100 = 120\%$$

Os elos relativos permitem estabelecer comparações para períodos próximos. Se for desejado determinar o preço relativo de 2016 em relação ao ano básico de 2013, basta aplicar propriedade circular modificada, obtendo as comparações através de um processo de encadeamento de comparações binárias. O relativo (índice) assim construído é denominado índice de cadeia, sendo calculado como segue:

$$p_{2013,2016} = p_{2013,2014} \cdot p_{2014,2015} \cdot p_{2015,2016}$$

$$p_{2013,2014} = 1{,}50 \, x \, 1{,}25 \, x \, 1{,}20 = 2{,}25 \text{ ou } 225\%$$

Quando se tiver uma seqüência de preços relativos onde o período básico é fixo, a seqüência será denominada, às vezes, relativos em cadeia referidos a essa base. Os relativos em cadeia da série acima referidos ao ano básico de 2013 serão 150, 187,5 e 225, ou seja:

$$p_{2013,2014} = \frac{120}{80} x100 = 150\%$$
$$p_{2014,2015} = \frac{150}{80} x100 = 187{,}5\%$$

$$p_{2015,2016} = \frac{180}{80} x100 = 225\%$$

Obs.: O que foi afirmado referente aos preços relativos é extensivo às quantidades e valores relativos.

12.6 ÍNDICES AGREGATIVOS SIMPLES

Os índices relativos de preços, quantidade ou valores informam sobre a evolução das variáveis preço, quantidade e valor de uma mercadoria. Entretanto, na maioria das vezes, o que procuramos é saber como se comporta o preço, quantidade ou valor de uma cesta de produtos, já que as famílias adquirem vários produtos para satisfazer suas necessidades. Elas estão interessadas então, no comportamento do preço, na quantidade e no valor desse grupo de itens. Da mesma forma, quando na empresa procuramos construir um índice para descrever a evolução do custo de um produto, devemos considerar os vários insumos envolvidos em sua produção, cada um deles com sua própria evolução de preços.

Uma primeira tentativa para resolver o problema da evolução dos índices relativos a uma cesta de produtos é tomar os relativos da soma dos preços ou das quantidades dos produtos da cesta:

$$p_{o,t} = \frac{\sum p_t}{\sum p_o} \qquad q_{o,t} = \frac{\sum q_t}{\sum q_o}$$

ou ainda pela média de preços:

$$p_{o,t} = \frac{\frac{\sum p_t}{n}}{\frac{\sum p_o}{n}} = \frac{\bar{p}_t}{\bar{p}_o} \qquad q_{o,t} = \frac{\frac{\sum q_t}{n}}{\frac{\sum q_o}{n}} = \frac{\bar{q}_t}{\bar{q}_o},$$

outra maneira é construir o índice com base em uma soma de relativos:

$$p_{o,t} = \sum \frac{p_t}{p_o} \qquad q_{o,t} = \sum \frac{q_t}{q_o}$$

ou ainda a média aritmética:

$$p_{o,t} = \frac{\sum \frac{p_t}{p_o}}{n} \qquad q_{o,t} = \frac{\sum \frac{q_t}{q_o}}{n}.$$

Entretanto , os índices assim obtidos têm várias limitações:

⇒ Apresentam problemas de homogeneidade nas unidades, isto é, não há uma única unidade para expressar as quantidades dos produtos (kg, pacotes de 5kg, fardos, litros, dúzias etc.);

⇒ Não levam em consideração as quantidades relativas adquiridas para satisfazer as necessidades das famílias ou das empresas;

⇒ Não apresentam as propriedades da decomposição, nem a das causas, nem a circular.

Pode-se contornar as duas primeiras dificuldades considerando, além dos preços dos produtos, também as quantidades adquiridas, estabelecendo o índice através da relação entre valores da cesta de produtos nas duas datas, o que constitui os índices agregativos ponderados. A terceira dificuldade será tratada com base na modificação desses mesmos índices.

12.7 ÍNDICES AGREGATIVOS PONDERADOS

Neste ítem serão expressados os índices ponderados mais importantes: Laspeyres, Paasche e Fisher. Esses tipos de índices são estudados em virtude das desvantagens apresentadas pelos índices simples, especialmente no que se refere à inexistência de pesos diferentes para cada utilidade que os compões de acordo com a sua importância relativa.

A ponderação proposta basear-se-á, por conseguinte, na participação de cada bem no valor transacionado total e será feita segundo dois critérios: fixa na época-base ou variável na época atual.

12.7.1 Índice de Laspeyres ou método da época-base

Esse índice constitui uma média aritmética ponderada dos relativos, sendo a ponderação feita utilizando-se os preços ou as quantidades referidas à época básica. Seja, então w_o^i os pesos ou fatores de ponderação dos *n* bens considerados. Simbolizando o índice de preços de Laspeyres por $L_{o,t}$ teremos:

$$L_{o,t} = \frac{\sum \frac{p_t^i}{p_o^i} w_o^i}{\sum w_o^i} \text{ sendo } w_o^i = \frac{p_o^i . q_o^i}{\sum p_o^i . q_o^i} .$$

Pode-se notar que o numerador de w_o^i representa o valor de um dado bem *i* e o denominador a soma dos valores de todos os bens que compõem o índice. Portanto, w_o^i indica a participação relativa do valor do bem *i*, em relação ao valor de todos os bens transacionados, tendo a época-base com referência. Levando-se a expressão de w_o^i na fórmula de $L_{o,t}$ e fazendo-se as simplificações necessárias, chegaremos a uma expressão mais compacta para o índice de Laspeyres, que é um índice de preço agregativo ponderado em relação às quantidades do ano-base, então:

$$L_{o,t} = \frac{\sum p_t^i . q_o^i}{\sum p_o^i . q_o^i} ,$$

o índice de Laspeyres de quantidade $L'_{o,t}$ é obtido substituindo-se o *p* pelo *q* na fórmula anterior:

$$L'_{o,t} = \frac{\sum q_t^i . p_o^i}{\sum q_o^i . p_o^i} .$$

12.7.2 Índice de Paasche ou método da época-atual

O índice proposto por Paasche é uma média harmônica ponderada dos relativos, sendo os pesos dados por:

$$w_t^i = \frac{p_t^i.q_t^i}{\sum p_t^i.q_t^i}$$ (ponderação pela época-atual)

simbolizando-se por $P_{o,t}$ o índice de preços de Paasche, teremos:

$$P_{o,t} = \frac{\sum w_t^i}{\sum \frac{p_o^i}{p_t^i} w_t^i} \quad \text{ou} \quad P_{o,t} = \frac{\sum p_t^i.q_t^i}{\sum p_o^i.q_t^i}$$

o índice de Paasche de quantidade $P'_{o,t}$ será dado por:

$$P'_{o,t} = \frac{\sum q_t^i.p_t^i}{\sum q_o^i.p_t^i}\ .$$

12.7.3 Índice de Fisher (fórmula ideal)

O índice de Fisher é a média geométrica dos números-índices de Laspeyres e Paasche. Se representa por $F_{o,t}$ o índice de Fisher para as variações de preços, teremos:

$$F_{o,t} = \sqrt{L_{o,t}.P_{o,t}}\ .$$

Para as variações de quantidades o índice de Fisher será:

$$F'_{o,t} = \sqrt{L'_{o,t}.P'_{o,t}}\ .$$

12.8 EXEMPLOS DE ÍNDICES PONDERADOS

1) A tabela 12.1 apresenta os preços médios por atacado e as quantidades vendidas dos seguintes produtos: leite, queijo e manteiga, durante os anos de 2000, 2001 e 2015. Calcular o índice de preços e de quantidades, pelo critério de Laspeyres, para o ano de 2015 tomando-se como base:

a) 2000;
b) 2000 e 2001.

Preços dos produtos:

Produtos	2000		2001		2015	
	Preço	**Quant.**	**Preço**	**Quant.**	**Preço**	**Quant.**
Queijo (kg)	8,25	20	9,40	25	15,20	28
Manteiga (500g)	5,60	12	6,95	18	9,30	12
Leite (litro)	1,15	800	1,35	1000	2,65	1200

Resolução:

a) Aplicando-se a fórmula de índice de preços de Laspeyres:

$$L_{o,t} = \frac{\sum p_t^i . q_o^i}{\sum p_o^i . q_o^i} = L_{2000,2015} \frac{\sum p_{2015}^i . q_{2000}^i}{\sum p_{2000}^i . q_{2000}^i}$$

Substituindo-se os símbolos pelos valores constantes da tabela 12.1, tem-se:

$$L_{2000,2015} = \frac{(15{,}20x20) + (9{,}30x12) + (2{,}65x800)}{(8{,}25x20) + (5{,}60x12) + (1{,}15x800)} = \frac{2535{,}6}{1152{,}2} = 2{,}201 = 220{,}2\%$$

Índice de quantidade:

$$L'_{2000,2015} = \frac{\sum q_{2015}^i . p_{2000}^i}{\sum q_{2000}^i . p_{2000}^i}$$

$$L'_{2000,2015} = \frac{(28x8{,}25) + (12x5{,}60) + (1200x1{,}15)}{(20x8{,}25) + (12x5{,}60) + (800x1{,}15)} = \frac{1678{,}2}{1152{,}2} = 1{,}457 = 145{,}7\%$$

b) Para resolver o item b, cuja base é representada pelos anos de 2000 e 2001, calcula-se, inicialmente, a média aritmética dos preços e das quantidades destes dois anos, de forma a trabalhar com um único valor para o preço e para a quantidade de cada item.

Médias dos preços e das quantidades (2000-2001)

Produtos	P_{2000}	Q_{2000}	P_{2001}	Q_{2001}	$\overline{P}$	$\overline{Q}$
Leite	8,25	20	9,40	25	8,83	22,50
Manteiga	5,60	12	6,95	18	6,28	15,00
Queijo	1,15	800	1,35	1000	1,25	900,00

O índice de preços de Laspeyres para o ano de 1990, tomando-se por base os anos de 2000–2001, será:

$$L_{2015/2000,2001} = \frac{\sum p^i_{2015} . q^i_{2000/2001}}{\sum p^i_{2000/2001} . q^i_{2000/2001}}$$

$$L_{2015/2000,2001} = \frac{(15{,}2x22{,}5) + (9{,}30x15) + (2{,}65x900)}{(8{,}83x22{,}5) + (6{,}28x15) + (1{,}25x900)} = \frac{2866{,}5}{1417{,}88} = 2{,}022 = 202{,}2\%$$

portanto, com o aumento de 202,2% em relação aos anos de 2000/2001. O cálculo do índice de quantidades de Laspeyres será feito, como vimos, permutando-se os preços e as quantidades, o que dá origem à fórmula:

$$L'_{o,t} = \frac{\sum q^i_t . p^i_o}{\sum q^i_o . p^i_o} \ .$$

Os dados constantes da Tabela, permite resolver a segunda parte do problema.

$$L'_{2015/2000,2001} = \frac{\sum q^i_{2015} . p^i_{2000/2001}}{\sum q^i_{2000/2001} . p^i_{2000/2001}}$$

$$L'_{2015/2000,2001} = \frac{(28x8{,}83) + (12x6{,}28) + (1200x1{,}25)}{(22{,}5x8{,}83) + (15x6{,}28) + (900x1{,}25)} = \frac{1822{,}6}{1417{,}9} = 1{,}285 = 128{,}5\%$$

2) Com os dados da Tabela, determinar o índice de Paasche de preço e de quantidade para o ano de 2015, tomando por base o ano de 2000.

Resolução:

Índice de preços de Paasche:

$$P_{o,t} = \frac{\sum p_t^i.q_t^i}{\sum p_o^i.q_t^i} \quad .$$

Substituindo os valores da tabela 1 na fórmula acima, teremos:

$$P_{2000,2015} = \frac{\sum p_{2015}^i.q_{2015}^i}{\sum p_{2000}^i.q_{2015}^i}$$

$$P_{2000,2015} = \frac{(15{,}2x28) + (9{,}3x12) + (2{,}65x1200)}{(8{,}25x28) + (5{,}6x12) + (1{,}15x1200)} = \frac{3717{,}2}{1678{,}2} = 2{,}215 = 221{,}5$$

Índice de quantidade de Paasche:

$$P'_{o,t} = \frac{\sum q_t^i.p_t^i}{\sum q_o^i.p_t^i} \, ,$$

então:

$$P'_{2000,2015} = \frac{\sum q_{2015}^i.p_{20015}^i}{\sum q_{2000}^i.p_{2015}^i}$$

$$P'_{2000,2015} = \frac{(28x15{,}2) + (12x9{,}3) + (1200x2{,}65)}{(20x15{,}2) + (12x9{,}3) + (800x2{,}65)} = \frac{3717{,}2}{2536{,}6} = 1{,}466 = 146{,}6\%$$

A Tabela a seguir mostra, a título de ilustração, os índices de preços e de quantidades de Laspeyres e Paasche para o ano de 2015, quando o ano base for 2000.

Índices de Laspeyres e Paasche em 2000-2015

Índice	Preços	Quantidades
Laspeyres	$L_{2000,2015} = 220,2$	$L'_{2000,2015} = 145,7$
Paasche	$P_{2000,2015} = 221,5$	$P'_{2000,2015} = 146,6$

3) Com base nos resultados obtidos na Tabela, determinar o índice de Fisher para as variações de preços e para as variações de quantidades.

Resolução:

Sabe-se que:

$$F_{o,t} = \sqrt{L_{o,t}.P_{o,t}} \ ,$$

para as variações de preços, então,

$$F_{2000,2015} = \sqrt{L_{2000,2015}.P_{2000,2015}} = \sqrt{2,202x2,215} = \sqrt{4,8774} = 2,208 = 220,8\%$$

como $F'_{o,t} = \sqrt{L'_{o,t}.P'_{o,t}}$ representa o índice de Fisher para as quantidades, teremos:

$$F'_{2000,2015} = \sqrt{L'_{2000,2015}.P'_{2000,2015}} = \sqrt{1,457x1,466} = \sqrt{2,953} = 1,461 = 146,1\%$$

Índices de Laspeyres e Paasche e Fisher em 2000-2015

Índice	Preços	Quantidades
Laspeyres	$L_{2000,2015} = 220,2$	$L'_{2000,2015} = 145,7$
Paasche	$P_{2000,2015} = 221,5$	$P'_{2000,2015} = 146,6$
Fisher	$F_{2000,2015} = 220,8$	$F'_{2000,2015} = 146,1$

12.9 ALGUMAS CONSIDERAÇÕES SOBRE OS ÍNDICES DE PAASCHE E DE LASPEYRES

Os índices de Laspeyres e de Paasche, embora visando medir a mesma variação, conduzem-nos a resultados diferentes; assim , o índice de Laspeyres é uma média aritmética ponderada de relativos de preços, sendo a ponderação feita com base na participação relativa (percentual) de cada bem no valor total dos bens, considerados na época-base, enquanto que o índice de Paasche é calculado através de uma média harmônica dos relativos de preços, sendo o peso de cada bem considerado com sua participação relativa no valor total dos bens da época atual.

Os resultados de cada um dos índices, de Laspeyres ou de Paasche, são geralmente diferentes, quando aplicados aos mesmos dados. Os resultados seriam iguais somente se os preços ou as quantidades de todos os bens que compõem o índice variarem na mesma proporção.

Na prática nenhum desses elementos varia na mesma proporção, e a relação entre os índices irá depender da corelação entre suas variações. Assim, supondo que as variações dos preços e das quantidades no tempo, resultassem principalmente de mudanças das condições de oferta, espera-se então uma correlação negativa entre essas variáveis.

Supondo $X = \frac{p_t^i}{p_0^i}$ a variação relativa dos preços de cada mercadoria e $Y = \frac{q_t^i}{q_0^i}$ a variação relativa das quantidades de cada mercadoria, $f = p_0^i.q_0^i$ e $\sum f = n$, logo o coeficiente de correlação é dado por:

$$r_{XY} = \frac{\sum XYf}{N} - \frac{\sum Xf}{N}.\frac{\sum Yf}{N}$$

$$r_{XY} = \frac{\sum p_t^i.q_t^i}{\sum p_0^i.q_0^i} - \frac{\sum p_t^i.q_0^i}{\sum p_0^i.q_0^i}.\frac{\sum p_0^i.q_t^i}{\sum p_0^i.q_0^i} = V_{0,t} - L_{0,t}.L'_{0,t} = P_{0,t}.L'_{0,t} - L_{0,t}.L'_{0,t}$$

se algum dos fatores r_{XY} for nulo, $0 = P_{0,t}.L'_{0,t} - L_{0,t}L'_{0,t}$, logo $L_{0,t} = P_{0,t}$

Assim conclui-se que os índices de Laspeyre e Paasche são iguais quando não houver correlação linear entre os relativos de preço e de quantidade ($r_{XY} = 0$), ou quando todos os preços ou todas as quantidades se alterarem na mesma proporção, inexistindo, conseqüentemente, dispersão de variação dos preços o u das quantidades ($S_X = 0$ ou $S_Y = 0$)

12.10 ÍNDICADORES DE PRODUTIVIDADE

A produção e a produtividade podem parecer sinônimas para alguns administradores de produção, mas tem funções diferentes no âmbito organizacional.

A produção é apenas uma medida de resultados, ou seja, um dado sobre o que foi produzido em determinada organização num dado período de tempo. Estas informações são importantes para o Planejamento e Controle de Produção (PCP), que controla a atividade de decidir sobre o melhor uso dos recursos de produção.

A produtividade é a capacidade de se produzir mais utilizando cada vez menos em menos tempo. “Produtividade é minimizar cientificamente o uso de recursos materiais, mão-de-obra, máquinas, equipamentos, pessoal, etc.” (*Japan Productivity Center for Social – Economics Development*).

$$\text{Prod}_t = \frac{Q_t}{I_t},$$

ou seja, produtividade é a quantidade produzida num dado período (Q_t) em relação aos seus insumos (I_t) disponíveis.

A produção e produtividade são estudos diferentes, mas de grande valia um para o outro, atuando em cooperação. Portanto nunca se esqueça de um deles quando for analisar e organizar seu negócio.

Indicadores de produtividade é uma forma mais nítida de divulgar os resultados, afinal, tem-se metas claras estabelecidas, permitindo que todos as entendam e trazendo mais transparência às ações adotadas nas melhorias dos processos, podendo então se tornar uma medida de produção comparável.

$$I_t = \frac{Prod_t}{Prod_i}.100,$$

é a proditividade no período t ($Prod_t$) em relação a produtividade no período base ($Prod_i$).

A taxa de variação de um indicador no instante (i + 1) em relação a um indicador (i), podendo ser interpretado como uma forma de medir "quão rápido" o indicador (i + 1) está mudando à medida em que o indicador (i) muda.

$$Taxa_{\,i+1} = \frac{Índice_{i+1} - Índice_i}{Índice_i}\,.$$

Como a taxa de variação poderá assumir valores positivos e negativos ao longo de um período, a Taxa Média de Variação, servirá como um indicador com tendência positiva ou negativa para o período, podendo ser calculado pelo método aritmético ou geométrico (o mais aconselhado).

$$TV_{\,Aritm.} = \frac{\sum Taxa_{\,i}}{n};$$

$$Taxa_{\,(trans.)_i} = TV_{Arit.} + 1;$$

$$TV_{\,Geom.} = 1 - \sqrt[n]{\Pi \; Taxa_{\,(trans.)_i}}\,.$$

Ex.: Calcular os indicadores de produtividade a seguir:

Mês	Vendas	Funcion.	Prod. (i)	Ind. (j)	TV Arit.	TV Trans.
JAN	7.520	230	32,70	100,00	---	---
FEV	7.821	233	33,57	102,66	0,03	1,03
MAR	8.422	189	44,56	136,27	0,33	1,33
ABR	7.294	201	36,29	110,98	-0,19	0,81
MAI	8.122	206	39,43	120,58	0,09	1,09
JUN	8.648	210	41,18	125,93	0,04	1,04
JUL	9.099	212	42,92	131,25	0,04	1,04
AGO	9.250	200	46,25	141,44	0,08	1,08
SET	8.874	215	41,27	126,21	-0,11	0,89
OUT	8.949	220	40,68	124,40	-0,01	0,99
NOV	8.723	222	39,29	120,15	-0,03	0,97
DEZ	8.573	235	36,48	111,56	-0,07	0,93

Q = Vendas; I = Funcionários

Produção é a quantidade vendida em relação ao número de funcionários.

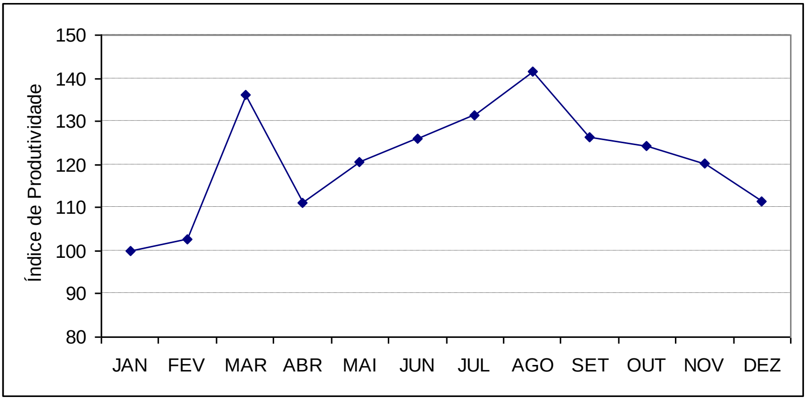


Figura 12.1 – Índice de produtividade de janeiro a dezembro

$$TV_{Aritm.} = \frac{\sum Taxa_i}{n} = \frac{0,19}{11} = 0,017 = 1,7\%$$

$$TV_{Geom.} = 1 - \sqrt[11]{1,116} = 1\text{-}1,010 = -0,01 = -1\%$$

Observa-se que a taxa de variação geométrica para o período ficou em torno de -1%, ou seja, a empresa esta atuando com produtividade decrescente.

13 VALIDAÇÃO DE INSTRUMENTOS

13.1 INTRODUÇÃO

Este capítulo tem como finalidade mostrar as principais técnicas estatísticas utilizadas para a validação de instrumentos de pesquisas utilizando escalas Likert (*checklist*).

Inicialmente para a elaboração do instrumento de pesquisa utilizando-se de escalas com as suas respectivas questões, bem como seus fatores ou constructos, deve-se realizar uma pesquisa bibliográfica com base em material já elaborado por autores consagrados em relação ao tema escolhido, constituído principalmente de livros e artigos científicos e/ou instrumentos já validados.

A validação normalmente requer uma investigação empírica e podem ser utilizadas dois tipos: validação de conteúdo e validação de constructo.

- **Validação de conteúdo**

A validade de conteúdo representa a garantia do planejamento e da construção das assertivas. Normalmente, envolve o conhecimento e familiaridade do construtor do instrumento, com o assunto abordado ou objetivo do questionário.

Sugere-se que a validade de conteúdo também está associada à sensibilidade dos respondentes quanto à facilidade do uso.

- **Validação dos constructos**

Os constructos ou fatores são conceitos teóricos e abstratos que ajudam a explicar e organizar as idéias e representam também a hipótese de que um tipo de comportamento irá se correlacionar com o outro, afetando o tratamento do experimento. Levando em consideração que os constructos são teoricamente abstratos, eles não podem ser diretamente observados. Assim eles devem ser indiretamente definidos, por meio do seu relacionamento, com manifestações observáveis. Após esses relacionamentos serem estabelecidos é possível investigar a relação entre os fatores, ou examinar o relacionamento entre um grupo de fatores e outro processo se o pesquisador estiver interessado.

Antes da aplicação definitiva do instrumento elaborado é fundamental passar por duas etapas pré-aplicação:

- **Pré-teste**

O instrumento deverá ser minuciosamente avaliado por especialistas do assunto (no mínimo 2), onde terão plena liberdade de dar opiniões e sugestões de mudanças no instrumento.

- **Pesquisa piloto**

Após o pré-teste sugere-se realizar uma **calibração do instrumento**, através de uma **pesquisa piloto**, utilizando-se de uma amostragem não probabilística por **conveniência** ($10 \leq n \leq 15$).

> Uma amostragem por **conveniência** nos dará informações sobre as **tendências e resultados** que poderá ser encontrar ao utilizar uma amostra probabilística. Esta informação pode ser usada para modificar o planejamento do estudo antes de definir uma técnica de amostragem mais eficiente (MATTAR, 1996).

- Vantagem e inconveniência

A **principal vantagem** da amostragem por conveniência é sua conveniência, por ser simples, econômica e rápida. Oferecendo informações valiosas em inúmeras circunstâncias, especialmente quando não existem razões fundamentais que diferenciem os indivíduos acessíveis que formam o total da população.

O **principal inconveniente,** a falta de representatividade, impossibilita a realização de declarações sobre os resultados sem correr nenhum risco devido ao critério de amostra aplicado. No pior dos casos, a amostragem por conveniência pode representar um desvio sistemático em relação à população total, produzindo resultados distorcidos.

- **Tema de Pesquisa**

Qualidade da estrutura física e pessoal do Curso de Administração da UFSM - 2016

- **Instrumento de coleta proposto**

Escala utilizada: Escala Likert de 5 pontos

1 - Muito Ruim	2 - Ruim	3 - Regular	4 - Bom	5 - Muito Bom

Fatores e questionamentos

Quadro 1.13 - Estrutura inicial do instrumento de pesquisa

Fatores / Questionamentos		**ESCALA**				
		1	**2**	**3**	**4**	**5**
Coordenação	Nível de relacionamento com os alunos					
	Habilidade em dar informações					
	Divulgação das atividades relacionadas ao curso (site, eventos, palestras, ...)					
	Presteza no atendimento ao aluno					
Infraestrutura	Organização geral do curso					
	Condições gerais do prédio					
	Condições das salas de aula					
	Condições dos banheiros					
	Condições dos laboratórios					
	Condições da biblioteca					
	Localização da biblioteca					
	Qualidade dos equipamentos utilizados					
	Acesso a Internet					
	Eficiência do pessoal de apoio					
Fatores / Questionamentos		**ESCALA**				
		1	**2**	**3**	**4**	**5**
Professores	Conhecimento teórico do assunto					
	Conhecimento prático do assunto					
	Clareza de explicação					
	Facilidade de comunicação e de relacionamento com a turma					
	Pontualidade					
	Capacidade de incentivar a troca de experiências e conhecimentos					
	Atendimento e esclarecimento de dúvidas individuais					
	Coerência entre o programa de curso e a discussão feita em sala de aula					

O instrumento foi aplicado em 145 alunos do curso de Administração da UFSM - 2016.

- **Avaliação do instrumento de pesquisa**

Nos itens a seguir apresentar-se-á as principais medidas estatísticas necessárias para a análise e validação de um instrumento de pesquisa.

13.2 MEDIDAS DESCRITIVAS

As medidas descritivas aconselhadas para descrever os valores encontrados por questão envolvendo escala likert são:

- **Tabelas de frequências (frequência absoluta e percentual)**

Essas duas medidas informarão a intensidade quão foram respondidas uma determinada questão em relação a variação das respostas dada pelo números de pontos definidos na escala likert adotada.

Tabela 13.1 - Indicador de qualidade dos equipamentos utilizados pelos Alunos do Curso de Administração (Questão 12)

Escala	f_i	%
1	0	0,0
2	8	5,5
3	32	22,1
4	78	53,8
5	27	18,6
Σ	145	100,0

Fonte: Curso de Administração (2016)

No que se refere a qualidade dos equipamentos utilizados pelos alunos do curso, 53,8% deles estão satisfeitos com os equipamentos oferecidos pelo curso, O grau de satisfação é de 72,4% (53,8% + 18,6%). O grau de insatisfação dos alunos é de 5,5% (0,0% + 5,5%).

- **Medidas de Posição e Variabilidade**

Para esse tipo de questão aconselha-se utilizar a média aritmética, o desvio padrão e o coeficiente de variação afim de avaliar a representatividade da média. Neste caso a média é usada como um indicador, podendo ser comparada com as demais questões do instrumento ou do constructo.

Como o indicador esta numa escala de 1 a 5 e a média que será apresentada esta numa escala de 1 a 10, ou seja 2x o valor da escala original, o valor da média foi multiplicada por 2 e em decorrência a variância é multiplicada por 4, assim ficando:

Tabela 13.2 - Notas dadas pelos alunos para o constructo - Infraestrutura

Indicadores	n = 145		Infraestrutura	
	Média*	**D. Padrão**	**Média***	**D. Padrão**
Organização geral do curso	7,97	1,447	7,32	2,129
Condições gerais do prédio	8,46	1,435		
Condições das salas de aula	8,64	1,296		
Condições dos banheiros	7,03	2,193		
Condições dos laboratórios	8,46	1,566		
Condições da biblioteca	6,87	2,282		
Localização da biblioteca	5,44	2,394		
Qualidade dos equipamentos	7,71	1,558		
Acesso a Internet	5,31	2,036		
Eficiência do pessoal de apoio	7,26	1,454		

* valores transformados para uma escala de 0 a 10

A média dada pelos estudantes em relação ao fator infraestrutura do curso foi de 7,32 (2,129).

13.3 CONFIABILIDADE DAS RESPOSTAS

A análise da consistência interna de uma medida psicológica é uma medida aceita e muito utilizada em pesquisas envolvendo escala likert pela comunidade científica. Entre os diferentes métodos que nos fornecem estimativas do grau de consistência de uma medida salienta-se o Alfa de Cronbach (α) que tem por finalidade medir a confiabilidade das respostas dos entrevistados na pesquisa.
Os utilizadores deste método indicam a utilização do coeficiente, especialmente para os casos em que os itens da escala são heterogêneos, são dicotômicos ou definem estruturas multi-fatoriais.

A fórmula utilizada para calcular o Alfa de Cronbach é dada por:

$$\alpha = \frac{k}{(k-1)}\left[1 - \frac{\sum_{j=1}^{n} S_j^2}{S_T^2}\right],$$

onde:

k = é o número de itens do instrumento ou constructo;
$S_j^2 =$ é a variância de cada item, j = 1, 2, ... n.
$S_T^2 =$ é a variância total de todos os itens.

- **Fiabilidade do α de Cronbach**

Tabela 13.3 - Fiabilidade dos Valores de α

Valor de α	**Fiabilidade**
< 0,60	Inaceitável
> 0,61	Aceitável
0,61 a 0,79	Fraco
0,80 a 0,90	Moderado
0,91 a 1,00	Forte

Segundo Hair Jr. et al. (2005, p. 90) em caso de pesquisas exploratórias o alfa de Cronbach é aceitável para valores acima de 0,6.

O cálculo do α de Cronbach permite que este assuma valores negativos quando as correlações inter-itens são, elas próprias, negativas. Um alfa negativo reflete normalmente um erro sério na codificação dos pontos dos itens e a solução passa pela recodificação (inversão) dos pontos de forma a assegurar que todos os itens estão codificados na mesma direção conceitual.

Um alfa muito baixo pode refletir a codificação errada de itens ou a mistura de itens de dimensões diferentes exigindo a reavaliação da base teórica que motivou a construção da escala.

Para a pesquisa de qualidade do Curso de Administração .

Tabela 13.4 - Cálculo do alfa de cronbach

Fator	**Alfa de Cronbach**	**Classificação**
Coordenação (4 questões)	0,798	Moderado
Infraestrutura (10 questões)	0,626	Fraco
Professores (8 questões)	0,857	Moderado
Geral	**0,832**	Moderado

13.4 ANÁLISE FATORIAL

Analise fatorial é um dos procedimentos psicométricos mais frequentemente utilizados na construção, revisão e avaliação de instrumentos psicológicos, assim como no desenvolvimento de teorias psicológicas.

A analise fatorial e particularmente útil quando aplicada a escalas que consistem de uma grande quantidade de itens utilizados para medir personalidade, estilos de comportamento ou atitudes.

O ponto central de partida na analise fatorial e o principio da parcimônia: um grande numero de variáveis observadas pode ser explicado por um número menor de variáveis hipotéticas, não-observadas. Estas variáveis hipotéticas, também chamadas de fatores, são responsáveis pelo fato de as variáveis observadas se correlacionarem entre si.

Um outro uso fundamental da analise fatorial ocorre no processo da validação de instrumentos psicológicos. Neste processo, a técnica de análise fatorial e imprescindível. A grande maioria dos pesquisadores em ciências comportamental e social reconhece o relacionamento próximo entre analise fatorial e questões de **validade**.

A fim de avaliar construtos psicológicos, a analise fatorial utiliza dois tipos de uso exploratório que estão alinhados com os objetivos de explicação e redução dos dados.

O primeiro uso, relacionado a explicação, consiste em identificar as dimensões subjacentes de um determinado domínio, no caso da pesquisa de qualidade (curso, coordenação, infraestrutura, docentes, metodologia, conteúdo programático, discente).

Neste caso, a analise fatorial é levada a efeito, a fim de identificar as dimensões subjacentes que representam os construtos teóricos do instrumento. Este procedimento é exploratório porque, presumivelmente, o investigador não tem qualquer expectativa firme *a priori* baseada em teoria ou pesquisa anterior acerca da composição de subescalas.

A analise fatorial e utilizada para descobrir se as variáveis latentes que estão subjacentes a escala. Para alcançar este fim, a análise fatorial usa uma matriz de correlações entre as variáveis mensuradas.

O segundo uso de analise fatorial exploratória está relacionado a redução de dados. O objetivo da redução de dados é descobrir ponderações ótimas para as variáveis mensuradas, de forma que um grande conjunto de variáveis possa ser reduzido a um conjunto menor de índices sumários (fatores) que tenham máxima variabilidade e fidedignidade.

A meta da redução de dados é tipicamente atingida pelo uso da Análise dos Componentes Principais (ACP) e não pelo uso da Analise Fatorial Comum (AFC). Uma diferença fundamental entre os dois métodos é que a ACP trabalha com a variância total observada, enquanto a AFC trabalha somente com a variância partilhada dos itens (variância erro e variância única são excluídas). Em outras palavras, a ACP tem como ponto de partida a variância observada das variáveis e a AFC a covariância observada entre as variáveis.

Na AFC, os fatores são estimados para explicar as covariâncias entre as variáveis observadas, portanto os fatores são considerados como as causas das variáveis observadas. Em contraste, na ACP os componentes são estimados para representar a variância das variáveis observadas de uma maneira tão econômica quanto possível. Os componentes principais são somas otimamente ponderadas das variáveis observadas. Neste sentido, as variáveis observadas são consideradas as causas dos componentes principais.

13.4.1 Decisão quanto ao tamanho da amostra

Numa pesquisa, dificilmente se realiza uma análise fatorial com uma amostra com menos de 50 respostas, e de preferência o tamanho da amostra deve ser superior a 100.

Segundo Hair Jr. (2005, p. 98), como regra geral, o mínimo que deve-se ter como tamanho da amostra é de cinco vezes o número de questões a serem analisadas, e o tamanho aceitável teria uma proporção de 10 para 1.
No caso da pesquisa de qualidade tem-se 22 variáveis, portanto seriam necessários 110 respondentes para validar o instrumento.

13.4.2 Suposições da análise fatorial

As suposições críticas da análise fatorial de um ponto de vista estatístico seriam:

- normalidade dos dados;
- linearidade dos dados (correlação de Pearson) ou (Teste KMO).
- homocedasticidade das Variâncias e;

A normalidade é necessária somente quando algum teste estatístico é feito para verificar a significância dos fatores, mas isso raramente é feito.

Espera-se que exista entre as variáveis um pouco de multicolineariedade, pois deseja-se trabalhar com variáveis inter-relacionadas, isso pode-se ser verificado pela matriz de correlação entre as variáveis e pela significância das correlações.

Quadro 13.1 - Matriz das correlações (Pearson) e suas significâncias

		SIGNIFICÂNCIAS																					
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20	Q21	Q22
CORRELAÇÕES	Q1	1,0000	p=,000	p=,000	p=,000	p=,000	p=,001	p=,042	p=,112	p=,000	p=,043	p=,330	p=,017	p=,840	p=,000	p=,000	p=,008	p=,014	p=,055	p=,000	p=,000	p=,000	p=,002
	Q2	,5820	1,0000	p=,000	p=,000	p=,000	p=,014	p=,525	p=,651	p=,027	p=,123	p=,815	p=,110	p=,125	p=,003	p=,000	p=,002	p=,030	p=,049	p=,005	p=,000	p=,000	p=,010
	Q3	,3927	,5125	1,0000	p=,000	p=,000	p=,013	p=,050	p=,309	p=,008	p=,061	p=,232	p=,014	p=,050	p=,019	p=,025	p=,113	p=,057	p=,013	p=,505	p=,005	p=,015	p=,258
	Q4	,5308	,5967	,4391	1,0000	p=,000	p=,000	p=,008	p=,026	p=,000	p=,102	p=,528	p=,003	p=,058	p=,000	p=,001	p=,001	p=,002	p=,000	p=,008	p=,000	p=,000	p=,000
	Q5	,4082	,4193	,3887	,4750	1,0000	p=,000	p=,000	p=,021	p=,000	p=,000	p=,900	p=,000	p=,021	p=,000	p=,000	p=,000	p=,000	p=,003	p=,031	p=,000	p=,001	p=,000
	Q6	,2868	,2122	,2158	,3559	,4563	1,0000	p=,000	p=,000	p=,000	p=,001	p=,437	p=,000	p=,086	p=,000	p=,000	p=,069	p=,000	p=,006	p=,030	p=,008	p=,004	p=,001
	Q7	,1770	,0556	,1705	,2305	,3297	,6638	1,0000	p=,000	p=,000	p=,002	p=,177	p=,000	p=,015	p=,000	p=,003	p=,168	p=,012	p=,002	p=,004	p=,036	p=,005	p=,025
	Q8	,1383	,0395	-,0889	,1924	,2007	,4458	,4347	1,0000	p=,050	p=,000	p=,244	p=,005	p=,003	p=,008	p=,017	p=,000	p=,000	p=,296	p=,048	p=,034	p=,000	p=,005
	Q9	,3173	,1915	,2282	,3294	,3006	,4317	,5293	,1703	1,0000	p=,015	p=,039	p=,000	p=,355	p=,000	p=,000	p=,450	p=,657	p=,838	p=,479	p=,745	p=,007	p=,310
	Q10	,1759	,1345	,1626	,1422	,3178	,2897	,2640	,3985	,2104	1,0000	p=,053	p=,000	p=,007	p=,000	p=,000	p=,000	p=,003	p=,171	p=,002	p=,000	p=,000	p=,000
	Q11	,0850	,0205	,1044	-,0552	-,0110	,0679	,1178	,1017	-,1794	,1681	1,0000	p=,202	p=,427	p=,269	p=,871	p=,214	p=,329	p=,514	p=,264	p=,838	p=,697	p=,211
	Q12	,2058	,1390	,2134	,2534	,3816	,3133	,4738	,2444	,3547	,3374	,1112	1,0000	p=,000	p=,000	p=,000	p=,000	p=,000	p=,000	p=,079	p=,000	p=,000	p=,000
	Q13	,0177	,1338	,1702	,1651	,2000	,1492	,2101	,2533	,0809	,2321	,0694	,3250	1,0000	p=,002	p=,102	p=,000	p=,008	p=,015	p=,069	p=,086	p=,013	p=,152
	Q14	,3626	,2578	,2032	,2543	,4013	,3235	,3145	,2299	,3874	,3896	,0966	,4951	,2709	1,0000	p=,000	p=,000	p=,000	p=,010	p=,022	p=,000	p=,000	p=,000
	Q15	,3522	,3425	,1938	,2924	,3945	,4363	,2550	,2062	,3315	,2729	-,0142	,4020	,1426	,3897	1,0000	p=,000	p=,000	p=,000	p=,014	p=,000	p=,000	p=,000
	Q16	,2306	,2602	,1381	,2930	,4087	,1580	,1204	,2715	-,0660	,2779	,1084	,3047	,3249	,3610	,4216	1,0000	p=,000	p=,000	p=,000	p=,000	p=,000	p=,000
	Q17	,2136	,1887	,1657	,2629	,3795	,3011	,2165	,2647	-,0388	,2541	,0854	,3422	,2306	,3683	,4887	,5060	1,0000	p=,000	p=,000	p=,000	p=,000	p=,000
	Q18	,1670	,1713	,2149	,3213	,2570	,2367	,2607	,0913	,0179	,1195	,0570	,3230	,2097	,2235	,3573	,4281	,5831	1,0000	p=,000	p=,000	p=,000	p=,000
	Q19	,3492	,2410	,0583	,2278	,1869	,1880	,2462	,1718	,0619	,2718	,0976	,1527	,1582	,1980	,2137	,3008	,3493	,3446	1,0000	p=,000	p=,000	p=,000
	Q20	,3364	,2991	,2415	,3699	,4065	,2299	,1818	,1839	,0285	,3324	,0179	,3656	,1493	,3737	0	,5585	,5571	,4682	,3436	1,0000	p=,000	p=0,00
	Q21	,4115	,3626	,2102	,4445	,2940	,2469	,2400	,3024	,2313	,3395	-,0340	,4172	,2145	,4133	,3313	,3313	,4611	,3773	,3656	,5450	1,0000	p=,000
	Q22	,2603	,2234	,0987	,4153	,3295	,2758	,1941	,2407	,0887	,3651	,1091	,3358	,1249	,4104	,3867	,4213	,5513	,4678	,3996	,6791	,6056	1,0000

Observa-se que a grande maioria das correlações são significativas (p < 0,05)

p = 0,0556	Exemplo de correlação não significativa (p > 0,05)

Uma outra forma de detectar a adequação da análise fatorial é analisar a matriz de correlação interna (teste de esfericidade de Bartlett e Teste Kaiser-Meyer-Olkin (KMO).

- **Teste de Esfericidade de Bartlett**

Tem por finalidade testar a hipótese de que as variáveis não sejam correlacionadas (Ho) (p > 0,05). A hipótese básica diz que a matriz de correlação é uma matriz identidade, indicando que o modelo fatorial é inadequado. A estatística teste é dada por:
;

$$\chi_c^2 = -\left[(n-1) - \frac{2p+5}{6}\right].\ln|R|,$$

onde:

n = tamanho da amostra;
p = número de variáveis; e
|R| = determinante da matriz de correlação.

Para a pesquisa de qualidade o χ^2_c= 1261,7 (p = 0,00001).

- **Teste de Kaiser-Meyer-Olkin (KMO) (Medida de Adequacidade)**

Essa medida avalia a adequacidade da análise fatorial, calculada por:

$$KMO = \frac{\sum\sum_{j \neq k} r_{jk}^2}{\sum\sum_{j \neq k} r_{jk}^2 + \sum\sum_{j \neq k} q_{jk}^2} = \frac{r_1^2 + r_2^2 + \ldots + r_j^2}{\left(r_1^2 + r_2^2 + \ldots + r_j^2\right) + \left(q_{11}^2 + q_{12}^2 + \ldots + q_{jk}^2\right)},$$

tem distribuição qui-quadrado com $\upsilon = \frac{p(p-1)}{2}$ graus de liberdade.

onde:

p = número de variáveis;
$r_{jk}^2 =$ quadrado dos elementos matriz de correlação original fora da diagonal; e
$q_{jk}^2 =$ quadrado das correlações parciais entre as variáveis.

Tabela 13.5 - Fiabilidade dos Valores de KMO

Valor do KMO	Fiabilidade
< 0,5	Inaceitável
≥ 0,51	Aceitável
0,51 a 0,60	Ruim
0,61 a 0,70	Razoável
0,71 a 0,80	Médio
0,81 a 0,90	Bom
0,91 a 1,00	Muito bom

Os valores do índice de KMO, segundo Hair Jr. et al. (2005), são valores aceitáveis entre 0,5 e 1,0 , portanto abaixo de 0,5 indica que a análise fatorial é inaceitável. Kaiser e Rice (1977) indicam que o KMO aceitável seja acima de 0,8.

Para os dados da pesquisa de qualidade, o valor encontrado foi, KMO = 0,832 (aceitável e bom).

13.4.3 Critério quanto ao número de fatores a extrair

1) Critério da raiz latente

É a técnica mais comumente usada, é uma técnica simples, onde se afirma quanto ao número de fatores nada mais é que a quantidade de autovalores maiores que **1,000**.

Na pesquisa de qualidade com 22 questões:

Tabela 13.6 - Número de autovalores a serem extraídos

Fatores	Autovalores Iniciais			Extraídos		
	Total	% da Variância	Acumulado %	Total	% da Variância	Acumulado %
1	**7,012**	31,873	31,873	7,012	31,873	31,873
2	**2,097**	9,533	41,406	2,097	9,533	41,406
3	**1,935**	8,796	50,202	1,935	8,796	50,202
4	**1,209**	5,494	55,696	1,209	5,494	55,696
5	**1,120**	5,089	60,784	1,120	5,089	60,784
6	**1,029**	4,678	65,463	1,029	4,678	65,463
7	0,945	4,294	69,756			
8	0,894	4,066	73,822			
9	0,762	3,462	77,283			
10	0,677	3,079	80,363			

2) Critério a **priori**

É quando o pesquisador sabe quantos fatores devem ser extraídos antes de empreender a análise fatorial. Na verdade o computador é instruído a parar a análise quando atingir o número de fatores desejados. Esse método é usando quando se tenta repetir o trabalho de outro pesquisador e extrair o mesmo número de fatores encontrados pelo pesquisador.

3) Critério do teste *Screen Plot* (Teste de Kattell)

Tal procedimento consiste na observação do gráfico dos autovalores, no qual é apresentado o número de dimensões (eixo "x") e seus autovalores correspondentes (eixo "y"). Por meio da análise do gráfico, é possível observar quais fatores apresentam maiores *autovalores*, sendo, portanto, responsáveis por uma maior variância explicada. O objetivo é encontrar o ponto (comumente chamado de cotovelo) onde os autovalores apresentam uma tendência descente linear.

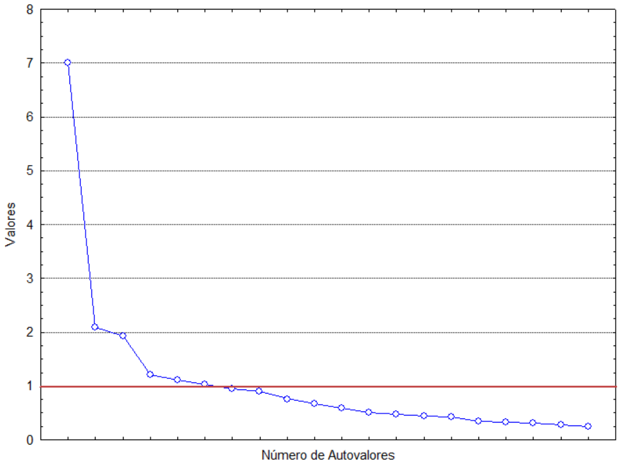


Figura 13.1 - Critério de Kattell (*Screen Plot*)

13.4.4 Comunalidade

As comunalidades representam a proporção da variância para cada variável incluída na análise que é explicada pelos componentes extraídos. Usualmente o valor mínimo aceitável é de 0,50. Logo, caso seja encontrado alguma comunalidade abaixo desse patamar a variável deve ser excluída e a análise fatorial deve ser realizada novamente.

No caso da pesquisa de qualidade:

Tabela 13.7 - Matriz das Comunalidades

Questão	Inicial	Valor Extraído	Questão	Inicial	Valor Extraído
Q1	1,000	0,706	Q12	1,000	0,603
Q2	1,000	0,740	Q13	1,000	0,528
Q3	1,000	0,690	Q14	1,000	0,601
Q4	1,000	0,641	Q15	1,000	0,490
Q5	1,000	0,540	Q16	1,000	0,603
Q6	1,000	0,762	Q17	1,000	0,710
Q7	1,000	0,807	Q18	1,000	0,715
Q8	1,000	0,571	Q19	1,000	0,513
Q9	1,000	0,771	Q20	1,000	0,685
Q10	1,000	0,623	Q21	1,000	0,651
Q11	1,000	0,747	Q22	1,000	0,705

Observa-se na tabela a existência de um valor muito próximo de 0,5 (variável 15), por ser um valor muito próximo de 0,5, deixaremos para a tomada de decisão da exclusão ou não da variável pela carga fatorial.

13.4.5 Carga fatorial

As cargas fatoriais são o meio de interpretar o papel que cada variável tem na definição de cada fator, sendo as cargas maiores fazem a variável tornar-se representativa no fator.

A carga fatorial é a correlação da variável com o fator, indicando o grau de correspondência entre a variável e o fator. Um fator não rotacionado pode não fornecer um padrão significativo de cargas fatoriais, portanto espera-se avaliar as cargas fatoriais após a rotação dos fatores, melhorando sua interpretação, reduzindo as ambiguidades que frequentemente ocorrem antes da rotação.

Num próximo passo, o pesquisador deve avaliar a necessidade de reespecificar o modelo fatorial devido à necessidade de eliminação de variáveis em função da carga fatorial ser abaixo do desejado. Lembrando que as extrações devem ser repetidas a cada eliminação de uma variável com carga fatorial menor que a desejada, definida pelo critério do tamanho da amostra. A Tabela a seguir apresenta uma orientação para a identificação da significância das cargas fatoriais em função do tamanho da amostra.

Significância das cargas fatoriais em função do tamanho da amostra

Tabela 13.8 - Cargas fatoriais x tamanho da amostra

Tamanho necessário da amostra	Carga fatorial
350	0,30
250	0,35
200	0,40
150	0,45
120	0,50
100	0,55
85	0,60
70	0,65
60	0,70
50	0,75

Fonte: Hair Jr. et al. (2005, p. 107)

No caso da pesquisa de qualidade para **n = 149** a carga fatorial mínima deve ser **0,45**.

Quadro 13.2 - Planejamento da análise fatorial em três estágios

Procedimento	O que deve ser observado
Verificar a adequabilidade da base de dados	Nível de mensuração das variáveis, tamanho da amostra, razão entre o número de casos e a quantidade de variáveis e o padrão de correlação entre as variáveis.
Determinar a técnica de extração e o número de fatores a serem extraídos	O tipo de extração **- Componentes principais**; - Máxima verossimilhança.
Decidir o tipo de rotação dos fatores	Se for ortogonal **- Varimax**; - Quartimax; - Equamax. Se for oblíqua - Direct oblimin; - Promax.

Usando a técnica de análise componentes principais (ACP) e uma rotação Varimax, para os dados da pesquisa de qualidade, as cargas fatoriais foram:

Tabela 13.9 - Cargas fatoriais após rotação do modelo

Pergunta	Fatores*					
	F1	F2	F3	F4	F5	F6
Q1	0,118	**0,754**	0,135	0,314	-0,080	0,004
Q2	0,133	**0,842**	-0,036	0,096	0,052	0,015
Q3	0,057	**0,731**	0,075	-0,226	0,276	0,137
Q4	0,279	**0,706**	0,200	0,110	0,005	-0,115
Q5	0,291	**0,498**	0,302	0,042	0,334	-0,065
Q6	0,192	0,212	**0,817**	0,087	0,052	0,043
Q7	0,118	0,049	**0,867**	0,094	0,157	0,072
Q8	0,114	-0,118	0,489	**0,507**	0,132	0,174
Q9	-0,206	0,313	**0,614**	0,156	0,186	-0,442
Q10	0,078	0,076	0,170	**0,636**	0,395	0,148
Q11	-0,001	0,049	0,059	0,106	0,110	**0,848**
Q12	0,271	0,099	0,355	0,158	0,597	-0,111
Q13	0,146	0,025	0,060	0,029	**0,687**	0,169
Q14	0,193	0,239	0,193	0,398	**0,537**	-0,150
Q15	**0,465**	0,260	0,320	0,075	0,234	-0,207
Q16	**0,626**	0,146	-0,065	0,154	0,380	0,129
Q17	**0,801**	0,067	0,138	0,097	0,179	0,057
Q18	**0,798**	0,104	0,185	-0,155	0,070	0,056
Q19	0,430	0,189	0,135	0,430	-0,210	0,211
Q20	**0,700**	0,247	-0,029	0,318	0,159	-0,078
Q21	**0,504**	0,305	0,076	0,435	0,122	-0,204
Q22	**0,672**	0,150	0,062	0,472	0,043	-0,061

* Fatores após rotação Varimax

Pergunta	Fatores*					
	F1	F2	F3	F4	F5	F6
Q1	0,130	**0,754**	0,128	0,260	-0,121	-0,001
Q2	0,139	**0,842**	-0,037	0,075	0,040	0,006
Q3	0,049	**0,733**	0,086	-0,179	0,294	0,132
Q4	0,290	**0,705**	0,197	0,087	-0,023	-0,114
Q5	0,306	**0,506**	0,323	0,107	0,236	-0,030
Q6	0,195	0,215	**0,824**	0,108	-0,023	0,072
Q7	0,103	0,047	**0,860**	0,104	0,178	0,046
Q8	0,152	-0,109	0,479	**0,538**	0,006	0,199
Q9	-0,198	0,314	**0,610**	0,204	0,153	-0,457
Q10	0,126	0,091	0,169	**0,696**	0,246	0,167
Q11	0,004	0,057	0,064	0,126	0,061	**0,874**
Q12	0,292	0,108	0,368	0,258	**0,532**	-0,110
Q13	0,135	0,026	0,048	0,099	**0,804**	0,084
Q14	0,235	0,253	0,206	**0,491**	0,402	-0,128
Q15	**0,483**	0,267	0,345	0,114	0,124	-0,162
Q16	**0,644**	0,154	-0,049	0,181	0,319	0,146
Q17	**0,811**	0,070	0,154	0,081	0,128	0,081
Q18	**0,774**	0,096	0,191	-0,217	0,151	0,031
Q20	**0,735**	0,254	-0,016	0,306	0,044	-0,040
Q21	**0,508**	0,309	0,065	0,404	0,045	-0,208
Q22	**0,710**	0,156	0,066	0,432	-0,080	-0,027

* Fatores após rotação Varimax e com a exclusão da pergunta 19 (Pontualidade) (carga fatorial < 0,5)

Quadro 13.3 - Nova estrutura dos fatores após a análise fatorial

Coordenação	Nível de relacionamento com os alunos
	Habilidade em dar informações
	Divulgação das atividades relacionadas ao curso (site, eventos, palestras, ...)
	Presteza no atendimento ao aluno
	Organização geral do curso
Infraestrutura (espaço físico)	Condições gerais do prédio
	Condições das salas de aula
	Condições dos laboratórios
Infraestrutura (apoio)	Condições da biblioteca
	Condições dos banheiros
	Eficiência do pessoal de apoio
Infraestrutura (informática)	Qualidade dos equipamentos utilizados
	Acesso a Internet
Professores	Conhecimento teórico do assunto
	Conhecimento prático do assunto
	Clareza de explicação
	Facilidade de comunicação e de relacionamento com a turma
	Capacidade de incentivar a troca de experiências e conhecimentos
	Atendimento e esclarecimento de dúvidas individuais
	Coerência entre o programa de curso e a discussão feita em sala de aula

Infraestrutura	Q11 - Localização da biblioteca

Após passar por todos os pressupostos da Análise Fatorial, optou-se em eliminar a questão 11 (Localização da biblioteca) por ter ficado isolada em único fator, essa questão não trará ônus para a pesquisa, pois no novo fator infraestrutura (apoio) possui uma questão avaliativa para a biblioteca. Portanto o instrumento final ficará composto por 21 questões, o fator infraestrutura foi subdividido em 3 novos fatores (espaço físico, apoio e informática), ficando então 5 novos fatores, preservando os fatores coordenação e professores.

13.4.6 As estatísticas-chaves associadas à análise fatorial

As estatísticas existentes e utilizadas regularmente no processo de análise fatorial segundo são:

- **Teste de Esfericidade de Bartlett**

Estatística teste usada para examinar a hipótese de que as variáveis não sejam correlacionadas. Em outras palavras, testa se a matriz de correlação é uma matriz identidade (Ho); cada variável se correlaciona perfeitamente com ela própria (r = 1), mas não apresenta correlação com as outras variáveis (r = 0).

- **Matriz de correlação**

O triângulo inferior da matriz exibe as correlações simples, r, entre todos os pares possíveis de variáveis incluídas na análise. Os elementos da diagonal, que são todos iguais a 1, em geral são omitidos.

- **Comunalidade**

Porção da variância que uma variável compartilha com todas as outras variáveis consideradas. É também a proporção de variância explicada pelos fatores comuns.

- **Autovalor**

Representa a variância total explicada por cada fator.

- **Cargas fatoriais**

Correlação simples entre as variáveis e os fatores.

- **Matriz de fatores ou matriz principal**

Contém as cargas fatoriais de todos as variáveis em todos os fatores extraídos.

- **Escores fatoriais**

Escores compostos estimados para cada entrevistado nos fatores derivados.

- **Medida de adequacidade da amostra de Kaiser-Meyer-Olkin (KMO)**

índice usado para avaliar a adequacidade da análise fatorial. Valores altos (entre 0,5 e 1,0) indicam que a análise fatorial é apropriada. Valores abaixo de 0,5 indicam que a análise fatorial pode ser inadequada.

- **Porcentagem de variância**

Porcentagem da variância total atribuída a cada fator.

- **Resíduos**

Diferenças entre as correlações observadas, dadas na matriz de correlação de entrada (*input*) e as correlações reproduzidas, conforme estimadas pela matriz de fatores.

- ***Screen plot***

Gráfico dos autovalores versus número de fatores por ordem de extração.

14 ANÁLISE DE DADOS QUANTITATIVOS

14.1 CRITÉRIOS DE ESCOLHA

Dentre os inúmeros testes e técnicas estatísticos que se apresentam no contexto de uma trabalho de pesquisa, é natural um certo grau de desorientação inicial quanto à identificação daqueles que são ou não aplicáveis a cada situação. Para realizar a escolha adequada, é importante considerar alguns parâmetros básicos dos dados a serem analisados, tais como:

- **Nº de Amostras**: O número de grupos distintos sendo analisados (Um ou mais).

- **Relações entre Amostras**: Refere-se a duas ou mais amostras consistirem ou não de múltiplas medidas das mesmas entidades ou de entidades relacionadas (serem ou não Pareadas ou Casadas).

- **Escala Numérica**: A forma na qual os dados foram registrados (Escala Nominal, Ordinal, Intervalar ou de Razão).

- **Distribuição**: A densidade de probabilidade (distribuição de probabilidade) dos dados (Normal ou Não-Normal).

- **Dependência entre Variáveis**: O conhecimento de uma variável contribuir ou não para o conhecimento de outras (respectivamente, serem Associadas ou Independentes entre si).

São estes os fatores que determinam quais os procedimentos gráficos e analíticos possíveis para cada combinação de número de amostras e tipos de dados.

14.2 ESTATÍSTICA DESCRITIVA

O Quadro a seguir indica os tipos de técnicas estatísticas que podem ser aplicadas para a descrição de conjuntos de dados para se obter um resumo ou descrição geral deles.

Quadro 14.1 - Estatística descritiva e gráficos

Nº de Amostras	**Escala Numérica**	**Distribuição**	**Análises Aplicáveis**	**Gráficos Aplicáveis**
Uma ou Mais	Ordinal, Intervalar ou Razão	Normal	Média, Moda, Desvio Padrão, Coeficiente de Variação, Intervalo de Confiança, Mínimo, Primeiro Quartil, Mediana, Terceiro Quartil, Máximo, Série Temporal*.	Histograma, *Box & Whiskers*, Gráfico de Séries, Ogiva (Função de Distribuição).
	Ordinal, Intervalar ou Razão	Não-normal	Média, Moda, Desvio Padrão, Coeficiente de Variação, Intervalo de Confiança, Mínimo, Primeiro Quartil, Mediana, Terceiro Quartil, Máximo, Série Temporal*.	Histograma, *Box & Whiskers*, Gráfico de Séries, Ogiva (Função de Distribuição).
	Nominal	Não-normal	Frequências, Série Temporal*.	Pictograma, Gráfico de Séries.

* quando uma das variáveis registradas for o tempo.

14.3 COMPARAÇÕES ENTRE AMOSTRAS

O quadro a seguir indica as técnicas estatísticas que podem ser aplicadas para comparar parâmetros de dois ou mais grupos de dados.

Quadro 14.2 - Testes estatísticos comparáveis

Nº de Amostras	**Tipo de Relação**	**Distribuição**	**Escala Numérica**	**Análises Aplicáveis**
Duas Amostras	Pareadas	Normal	Intervalar ou Razão	Teste *t* de *Student* Pareado
Duas Amostras	Pareadas	Não-Normal	Ordinal, Intervalar ou Razão	Teste de Friedman, Sign-Test, Teste de Wilcoxon
Duas Amostras	Pareadas	Não-Normal	Nominal Dicotômica*	Teste de McNemar
Duas Amostras	Não-Pareadas	Normal	Intervalar ou Razão	Teste *t* de *Student*
Duas Amostras	Não-Pareadas	Não-Normal	Ordinal, Intervalar ou Razão	Teste Mann-Whitney U, Teste de Wald-Wolfowitz, Teste de Kolmogorov-Smirnov
Duas Amostras	Não-Pareadas	Não-Normal	Nominal	Teste de Qui-quadrado (Homogeneidade)
Três ou Mais Amostras	Pareadas	Normal	Intervalar ou Razão	ANOVA c/ Medidas Repetidas
Três ou Mais Amostras	Pareadas	Não-Normal	Ordinal, Intervalar ou Razão	ANOVA de Friedman
Três ou Mais Amostras	Pareadas	Não-Normal	Nominal	Teste Q de Cochran
Três ou Mais Amostras	Não-Pareadas	Normal	Intervalar ou Razão	ANOVA c/ Grupos Independentes
Três ou Mais Amostras	Não-Pareadas	Não-Normal	Ordinal, Intervalar ou Razão	ANOVA de Kruskal-Wallis
Três ou Mais Amostras	Não-Pareadas	Não-Normal	Nominal	Teste de Qui-Quadrado

* Variável com apenas dois valores ou duas categorias (variável binária).

14.4 RELAÇÃO ENTRE VARIÁVEIS

O Quadro a seguir mostra as técnicas analíticas e procedimentos gráficos aplicáveis quando se quer verificar a existência e/ou caracterizar a relação entre duas ou mais variáveis.

Quadro 14.3 - Relação entre variáveis

Nº de Variáveis	**Escala Numérica das Variáveis**	**Distribuição**	**Análises Aplicáveis**	**Gráficos Aplicáveis**
Duas	Intervalar e/ou Razão	Normal e/ou Não-Normal	Correlação de Pearson, Regressão Linear Simples.	Diagrama de Dispersão (X,Y).
Duas	Ordinal e/ou Intervalar e/ou Razão	Normal e/ou Não-Normal	Correlação de Pearson e/ou Spearman.	Diagrama de Dispersão (X,Y).
Duas	Nominal	Não-Normal	Teste de Qui-Quadrado.	---
Três ou Mais	Intervalar e/ou Razão	Normal	Regressão Múltipla	Diagrama Previsão vs. Observação
Três ou Mais	Ordinal e/ou Intervalar e/ou Razão	Não-Normal	Correlação *Partial Rank* de Kendall	Diagrama de Dispersão (X,Y).
Três ou Mais	Nominal	Não-Normal	Análise Discriminante	---
Três ou Mais	Intervalar e/ou Razão	Normal e/ou Não-Normal	Regressão Linear Múltipla, Regressão Não-Linear	---
Três ou Mais	Nominal Dicotômica* (Variável-Resposta) e/ou Nominal e/ou Ordinal e/ou Intervalar e/ou Razão	Normal e/ou Não-Normal	Regressão Logística	---

* Variável com apenas dois valores ou duas categorias (variável binária).

14.5 COMENTÁRIOS

Os quadros anteriores apontam as análises de dados possíveis nas diversas situações de pesquisa, porém, não indicam exatamente os procedimentos a serem adotados em cada situação. Isso ocorre devido ao fato de que a decisão final depende não apenas das restrições matemáticas, mas também dos objetivos do estudo e da própria natureza dos achados que vão sendo produzidos. É importante, contudo, ter em mente que as tabulações apresentadas constituem um mapa de referência que deixa claro espaço de ações dentro do qual pode se manifestar a liberdade do pensador analítico.

15 REFERÊNCIAS

AAKER, D.; KUMAR, V.; DAY, G. **Marketing research**. New York: John Wiley & Sons, Inc. 1995.

ALPERT, M. Identification of determinant attributes: a comparison of methods. **Journal of Marketing Research**, v. 8, n. 2, p. 184-191, 1971.

ANDERSON, L. W. **Guttman scales**. In: WALBERG H. J.; HAERTEL G. D. (Eds.), The International Encyclopedia of Educational Evaluation (pp 333-334). Oxford: Pergamon. 1990.

ANDRICH, D. **Thurstone scales**. In: WALBERG H. J.; HAERTEL G. D. (Eds.),The International Encyclopedia of Educational Evaluation (pp 329-332). Oxford: Pergamon. 1990.

BARBETTA, P. A.; REIS, M. C.; BORNIA, A. C. **Estatística para cursos de engenharia e informática**. São Paulo: Atlas. 2004.

BUSSAB, W. B.; MORETTIN, P. A. **Estatística básica**. São Paulo: Saraiva, 5. ed., 2002.

COCHRAN, W. G. **Técnicas de amostragem**. Rio de Janeiro: Fundo de Cultura. Rio de Janeiro, 1965.

COSTA NETO, P. L. O.; CYMBALISTA, M. **Probabilidades**. São Paulo: Edgard Blucher. 1994.

CURWIN, J.; SLATER, R. **Quantitative methods for business decisions**. 3. ed. United Kingdom: Chapman & Hall. 1991.

FIGUERA, K. K. Competências gerenciais: um estudo em empresas STARTUPS brasileiras. **Dissertação (Mestrado em Administração)** – Programa de Pós-Graduação em Administração. Universidade Federal de Santa Maria, 2016.

FONSECA, J. S.; MARTINS, G. A. **Curso de estatística**. 4. ed., São Paulo: Atlas. 1993.

GUTTMAN, L. A basis for scaling qualitative data. **Annual Meeting of the American Sociological Society**. New York, NY, 38, 1943.

HAIR, Jr., J.; BABIN, B.; MONEY, A. H.; SAMOUEL, P. **Fundamentos de Métodos de Pesquisa em Administração.** Porto Alegre: Bookman, 2005.

LEVINE, J.; FOX, J. A. **Estatística para ciências humanas**. 9. ed., São Paulo: Prentice Hall. 2004.

LEVINE, M. D; BERENSON, L. M.; ESTEPAN, D. **Estatística:** teoria e Aplicações. Rio de Janeiro: LTC - Livros técnicos e científicos. 2000.

LIKERT, R. **A technique for the measurement of attitudes**. New York: Archives of Psichology. n. 140, 1932.

MACRAE, A. W. Measurement scales and statistics: What can significance tests tell us about the world? **British Journal of Psychology**, 79, 161-171. 1988.

MATTAR, F. **Pesquisa de marketing**. São Paulo: Atlas. 1996.

MEYER, P. L. **Probabilidade – Aplicações à Estatística**, 2. ed., Rio de Janeiro: LTC - Livros Técnicos e Científicos, 1983.

MIOT, H. A. Tamanho da amostra e, estudos clínicos e experimentais. **J Vasc Bras**. v. 10, n. 4, p. 275-278. 2011.

MONTGOMERY, D. C.; RUNGEL, G. C. Estatística aplicada e probabilidade para engenheiros, 5. ed., Rio de Janeiro: LTC - Livros Técnicos e Científicos, 2012.

MORETTIN, L. G. **Estatística Básica – inferência**. São Paulo: Makron Books, v. 2., 2000.

MORETTIN, P. A .; TOLOI, C. M. C. **Métodos Quantitativos:** Séries Temporais. São Paulo: Atual Editora Ltda. 1986.

OBREGON, S. L. Síndrome deburnout e engajamento no trabalho: percepção dos servidores de uma instituição pública de ensino. **Dissertação (Mestrado em Administração)** – Programa de Pós-Graduação em Administração. Universidade Federal de Santa Maria, 2016.

OLIVEIRA, S. L. **Tratado de metodologia científica.** São Paulo: Pioneira, 2002.

OSGOOD, C. E.; SUCI, G. J.; TANNENBAUM. The Measurement of Meaning. **Linguistic Society of America**. 35 (1), 58-77, 1959.

ROBERTS, J. S. **Comparative validity of the Likert and Thurstone approaches to attitudes measurement**. In: Annual meeting of the American Educational Research Association. ERIC. 1997.

RON, L.; FARBER, E. **Estatística Aplicada**. São Paulo: Prentice Hall, 2004.

SCHIFFMAN, L.; KANUK, L. **Comportamento do consumidor**. 6. ed. São Paulo: Editora LTC. 2000.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete sample). **Biometrika**, 52. 59, 1965.

SIEGEL, S.; CASTELLAN, J. **Nonparametric statistics for the behavioral sciences**. New York: McGraw-Hill.1988.

SOARES, F. J.; FARIAS, A. A.; CESAR, C. C. **Introdução à estatística**. Rio de Janeiro: Guanabara & Koogan. 1991.

STAPEL, J. What is Job-satisfaction? **The Public Opinion Quarterly**. v. 14, p. 551 - 554. 1950.

THURSTONE, L. Attitudes can be measured. **American Journal of Sociology**, v. 33, p. 529-554. 1928.

TRIOLA, M. F. **Introdução à Estatística**, 7. ed., Rio de Janeiro: LTC - Livros Técnicos e Científicos, 1999.

ANEXO A - EQUAÇÕES ESTATÍSTICAS

Distribuições de Freqüências

Amplitude total:	$H = X_{max.} - X_{min}$
Número de classes:	$n(k) = 1 + 3{,}3 \log n$
Amplitude de classes:	$h = \dfrac{H}{n(k)}$

Medidas Descritivas

Média aritmética: (dados não tabelados)	$\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$
Média aritmética: (dados tabelados)	$\overline{X} = \dfrac{\sum_{i=1}^{k} X_i.f_i}{\sum_{i=1}^{k} f_i}$

Moda de Czuber:
(dados tabelados)

$$Mo_c = l_{Mo} + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right).h \Rightarrow \begin{cases} \Delta_1 = f_{Mo} - f_{ant} \\ \Delta_2 = f_{Mo} - f_{pos} \end{cases}$$

Separatrizes: (dados não tabelados)	Posição $= \dfrac{i(n+1)}{S}$

se for **mediana** $\begin{cases} i = 1 \\ S = 2 \end{cases}$ se for **quartis** $\begin{cases} 1 \le i \le 3 \\ S = 4 \end{cases}$

se for **decis** $\begin{cases} 1 \le i \le 9 \\ S = 10 \end{cases}$ se for **centis** $\begin{cases} 1 \le i \le 99 \\ S = 100 \end{cases}$

Separatrizes: (dados não tabelados)	Posição $= \dfrac{i(n)}{S}$

$$S_i = l_{S_i} + \frac{\left(\frac{i.n}{S} - F_{ant}\right).h}{f_{S_i}}$$

Desvio padrão: (dados não tabelados) (populacional)	$\sigma = \sqrt{\dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n}}$
Desvio padrão: (dados tabelados) (amostral)	$S = \sqrt{\dfrac{\sum_{i=1}^{k} f_i(X_i - \overline{X})^2}{\sum_{i=1}^{k} f_i - 1}}$
Coeficiente de Variação:	$C.V. = \dfrac{S}{\overline{X}} x100$
Coeficiente de Assimetria:	$As = \dfrac{\overline{X} - Mo}{S}$
Coeficiente de Curtose:	$K = \dfrac{Q_3 - Q_1}{2(P_{90} - P_{10})}$

Probabilidade

$$P(E) = \frac{\text{número de ocorrências do evento}(E)}{\text{número de repetições do experimento}(\Omega)}$$

$$\begin{matrix} \cup \\ +(OU) \end{matrix} \begin{cases} \text{Mutuamente Excludentes} \Rightarrow P(A \cup B) = P(A) + P(B) \\ A \cap B = \varnothing \\ \text{Quaisquer} \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B) \\ A \cap B \neq \varnothing \end{cases}$$

$$\begin{matrix} \cap \\ x\ (E) \end{matrix} \begin{cases} \text{Independentes} \Rightarrow P(A \cap B) = P(A) x\ P(B) \\ A \cap B \neq \varnothing \\ \text{Dependentes} \Rightarrow P(A \cap B) = P(A) x\ P(B/A) \end{cases}$$

Teorema de Bayes:

$$P\left(A_i / B\right) = \frac{P(A_i).P\left(B / A_i\right)}{\sum_{i=1}^{k} P(A_i).P\left(B / A_i\right)}$$

Variáveis Aleatórias Discretas

Esperança matemática:	$E(X) = \sum_{i=1}^{\infty} x_i\ P(x_i)$
Variância:	$V(X) = E(X^2) - [E(X)]^2$
onde:	$E(X^2) = \sum_{i=1}^{\infty} x_i^2\ P(x_i)$

Variáveis Aleatórias Contínuas

Esperança matemática:	$E(X) = \sum_{i=1}^{\infty} x_i\ P(x_i)$
Variância:	$V(X) = E(X^2) - [E(X)]^2$
onde:	$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\ dx$

Distribuições de Probabilidade

Binomial:	$P(X = x) = \binom{n}{x} p^x\ q^{n-x}$
Poisson:	$P(X = x) = \dfrac{e^{-\lambda} \lambda^x}{x!}$
Normal Padrão:	$z = \dfrac{x - \mu}{\sigma}$

Amostragem

Sistemática:	$r = \dfrac{N}{n}$
Estratificada:	$N = \sum_{i=1}^{k} N_i$ e $n = \sum_{i=1}^{k} n_i$

Tamanho da Amostra

- Variância populacional conhecida

População Infinita	População Finita
$n = \left(\dfrac{Z_{\alpha/2} \cdot \sigma}{e}\right)^2$	$n = \dfrac{Z^2_{\alpha/2} \cdot \sigma^2 N}{e^2(N-1) + Z^2_{\alpha/2} \cdot S^2}$

- Variância populacional desconhecida

População Infinita	População Finita
$n = \left(\dfrac{t_{\varphi,\alpha/2} \cdot S}{e}\right)^2$	$n = \dfrac{(t_{\varphi,\alpha/2})^2 \cdot S^2 N}{e^2(N-1) + (t_{\varphi,\alpha/2})^2 \cdot S^2}$

- Usando a proporção populacional

População Infinita	População Finita
$n = \dfrac{Z^2_{\alpha/2} \cdot \hat{p} \cdot \hat{q}}{e^2}$	$n = \dfrac{Z^2_{\alpha/2} \cdot \hat{p} \cdot \hat{q} \cdot N}{e^2(N-1) + Z^2_{\alpha/2} \cdot \hat{p} \cdot \hat{q}}$

Estimação de Parâmetros

- Intervalo de confiança para a média:

$$P(\overline{X} - e_0 \leq \mu \leq \overline{X} + e_0) = 1 - \alpha$$

Variância conhecida	Variância desconhecida
$e_0 = Z_{\alpha/2} \dfrac{\sigma}{\sqrt{n}} \cdot \sqrt{f.c.}$	$e_0 = t_{\varphi,\alpha/2} \dfrac{\sigma}{\sqrt{n}} \cdot \sqrt{f.c.}$

- Intervalo de confiança para a proporção:

$$P(\hat{p} - e_0 \leq \mu \leq \hat{p} + e_0) = 1 - \alpha$$

$\hat{p} = \dfrac{x}{n}$	$e_0 = Z_{\alpha/2} \sqrt{\dfrac{\hat{p} \cdot (1-\hat{p})}{n}} \cdot \sqrt{f.c.}$

O fator de correção $\dfrac{N-n}{N-1}$ é usado se $\dfrac{n}{N} > 5\%$ (pop. finita)

- Intervalo de confiança para a variância:

$$P\left[\frac{(n-1)S^2}{X^2_{sup}} \leq \sigma^2 \leq \frac{(n-1)S^2}{X^2_{inf}}\right] = 1 - \alpha$$

- Intervalo de confiança: diferença entre médias

$$P\left((\overline{X}_1 - \overline{X}_2) - e_0 \leq \mu_1 - \mu_2 \leq (\overline{X}_1 - \overline{X}_2) + e_0\right) = 1 - \alpha$$

Variâncias conhecidas
$e_0 = Z_{\alpha/2} \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

Variâncias desconhecidas	
Iguais	Diferentes
$e_0 = t_{\varphi,\alpha/2} S_c \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$	$e_0 = t_{\varphi,\alpha/2} \sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}$
$\varphi = n_1 + n_2 - 2$	$\varphi = \dfrac{(S_1^2/n_1 + S_2^2/n_2)^2}{\dfrac{(S_1^2/n_1)^2}{n_1+1} + \dfrac{(S_2^2/n_2)^2}{n_2+1}} - 2$

- Intervalo de confiança: diferença entre proporções

$$P[(\hat{p}_1 - \hat{p}_2) - e_0 \leq \pi_1 - \pi_2 \leq (\hat{p}_1 - \hat{p}_2) + e_0] = 1 - \alpha$$

$$e_0 = Z_{\alpha/2} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n} + \frac{\hat{p}_2 \hat{q}_2}{n}}$$

Teste de Hipótese Paramétrico

- Teste para a média

Variância conhecida	Variância desconhecida
$Z_{cal} = \dfrac{\overline{x} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$	$t_{cal} = \dfrac{\overline{x} - \mu_0}{\dfrac{S}{\sqrt{n}}}$

- Teste para a proporção

$$Z_{cal} = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 \cdot (1-p_0)}{n}}}$$

- Teste para a diferença entre duas médias

Variâncias conhecidas
$Z_{cal} = \dfrac{(\overline{X}_1 - \overline{X}_2) - d}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

Variâncias desconhecidas	
Iguais	Diferentes
$t_{cal} = \dfrac{(\overline{X}_1 - \overline{X}_2) - d}{S_c \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ $S_c = \sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}}$	$t_{cal} = \dfrac{(\overline{X}_1 - \overline{X}_2) - d}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$
$\varphi = n_1 + n_2 - 2$	$\varphi = \dfrac{(S_1^2/n_1 + S_2^2/n_2)^2}{\dfrac{(S_1^2/n_1)^2}{n_1+1} + \dfrac{(S_2^2/n_2)^2}{n_2+1}} - 2$

- Teste para a diferença entre duas proporções

$$Z_{cal} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

$$\hat{p}_1 = \frac{X_1}{n_1}, \hat{p}_2 = \frac{X_2}{n_2}, \hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$$

- Teste t pareado

$$t = \frac{\overline{d} - \Delta}{\dfrac{S}{\sqrt{n}}} \qquad \varphi = n - 1$$

Teste de Hipótese Não Paramétrico

- Teste Qui-quadrado (Adequação):

$$\chi^2_{cal} = \sum_{i=1}^{k} \frac{(F_o - F_e)^2}{F_e}, \quad \varphi = k - 1$$

- Teste Qui-quadrado (Adequação):

$$\chi^2_{cal} = \sum_{i=1}^{h}\sum_{j=1}^{k} \frac{(Fo_{ij} - Fe_{ij})^2}{Fe_{ij}}, \quad \varphi = (h-1)(k-1)$$

- Teste de Mann-Withney:

$$Z_{cal} = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$

- Teste de Wilcoxon:

$$Z_{cal} = \frac{\sum R}{\sqrt{\sum R^2}}$$

- Teste de Kruskall-Wallis:

$$H = \frac{12}{n(n+1)}\left[\frac{(\sum R_1)^2}{n_1} + \frac{(\sum R_2)^2}{n_2} + \dots + \frac{(\sum R_k)^2}{n_k}\right] - 3(n+1)$$

Regresão Linear Simples e Correlação

- Estimadores: $\hat{y} = ax + b$

$$a = \frac{n\sum_{i=1}^{n} xy - \sum_{i=1}^{n} x \sum_{i=1}^{n} y}{n\sum_{i=1}^{n} x^2 - \left(\sum_{i=1}^{n} x\right)^2} = \frac{S_{xy}}{S_{xx}} \qquad b = \bar{y} - a\,\bar{x}$$

- Teste para o coeficiente de regressão:

$$t_c = \frac{a - \alpha_o}{S_a} \text{ com } \varphi = n - 2$$

$$S_a = S_e \sqrt{\frac{1}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} \qquad S_e = \sqrt{\frac{SQR}{n-2}} = \sqrt{QMR}$$

- Teste para o coeficiente linear:

$$t_c = \frac{b - \beta_o}{S_b} \text{ com } \varphi = n - 2$$

$$S_b = S_e \sqrt{\frac{1}{n} + \frac{(\bar{X})^2}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} \qquad S_e = \sqrt{\frac{SQR}{n-2}} = \sqrt{QMR}$$

- Coeficiente de determinação:

$$r^2 = \frac{\left(n\sum_{i=1}^{n} xy - \sum_{i=1}^{n} x \sum_{i=1}^{n} y\right)^2}{\left[n\sum_{i=1}^{n} x^2 - \left(\sum_{i=1}^{n} x\right)^2\right]\left[n\sum_{i=1}^{n} y^2 - \left(\sum_{i=1}^{n} y\right)^2\right]} = \frac{COV_{xy}}{S_{xx} \cdot S_{yy}}$$

- Coeficiente de correlação:

$$r = \frac{\left(n\sum_{i=1}^{n} xy - \sum_{i=1}^{n} x \sum_{i=1}^{n} y\right)}{\sqrt{\left[n\sum_{i=1}^{n} x^2 - \left(\sum_{i=1}^{n} x\right)^2\right]}\sqrt{\left[n\sum_{i=1}^{n} y^2 - \left(\sum_{i=1}^{n} y\right)^2\right]}} = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$$

- Teste para o coeficiente de correlação:

$$t_c = \sqrt{\frac{r^2 (n-2)}{1 - r^2}} = \frac{a}{S_a} \text{ com } \varphi = n - 2$$

Análise de Variância - ANOVA

Fator de correção:	$C = \frac{(\sum y)^2}{N}$
Soma de quadrados:	
Total:	$SQ_{Total} = \sum y^2 - C$
Tratamentos:	
Amostras de mesmo tamanho:	$SQ_{Trat.} = \frac{\sum T^2}{r} - C$
Amostras de tamanhos diferentes:	$SQTr = \sum \frac{T_i^2}{r_i} - C$
Resíduos:	$SQ_{Res.} = SQ_{Total} - SQ_{Trat}$
Quadrado médio:	
Tratamentos:	$QM_{Trat} = \frac{SQ_{Trat}}{k-1}$
Resíduos:	$QM_{Res} = \frac{SQ_{Res}}{k(r-1)}$
F calculado:	$F = \frac{QM_{Trat}}{QM_{Res}}$
Teste de Tukey:	
Amostras de mesmo tamanho:	$dms = q_{\alpha;(\delta,k)} \sqrt{\frac{QMR}{r}}$
Amostras de tamanhos diferentes:	$dms = q_{\alpha;(\delta,k)} \cdot \sqrt{\left(\frac{1}{r_i} + \frac{1}{r_j}\right)\frac{QMR}{2}}$
Teste de Dunnett:	
Amostras de mesmo tamanho:	$dms = d_{\alpha;(\delta,T)} \sqrt{\frac{2QMR}{r}}$
Amostras de tamanhos diferentes:	$dms = d_{\alpha;(\delta,T)} \cdot \sqrt{\left(\frac{1}{r_i} + \frac{1}{r_c}\right)\frac{QMR}{2}}$

NÚMEROS ÍNDICES

- Relativo de Preço:

$$p_{(o,t)} = \frac{p_t}{p_o}$$

- Relativo de Quantidade:

$$q_{(o,t)} = \frac{q_t}{q_o}$$

- Relativo de Valor:

$$v_{(o,t)} = \frac{v_t}{v_o} = \frac{p_t \cdot q_t}{p_o \cdot q_o} = p_{o,t} \cdot q_{o,t}$$

- Índice Agregativo Simples:

$$p_{o,t} = \frac{\sum p_t}{\sum p_o} \qquad q_{o,t} = \frac{\sum q_t}{\sum q_o}$$

- Índice Agregativo Simples (pela média):

$$p_{o,t} = \frac{\frac{\sum p_t}{n}}{\frac{\sum p_o}{n}} = \frac{\bar{p}_t}{\bar{p}_o} \qquad q_{o,t} = \frac{\frac{\sum q_t}{n}}{\frac{\sum q_o}{n}} = \frac{\bar{q}_t}{\bar{q}_o}$$

- Índice de Laspeyre:

Preço	Quantidade
$L_{o,t} = \frac{\sum p_t^i \cdot q_o^i}{\sum p_o^i \cdot q_o^i}$	$L'_{o,t} = \frac{\sum q_t^i \cdot p_o^i}{\sum q_o^i \cdot p_o^i}$

- Índice de Paasche:

Preço	Quantidade
$P_{o,t} = \frac{\sum p_t^i \cdot q_t^i}{\sum p_o^i \cdot q_t^i}$	$P'_{o,t} = \frac{\sum q_t^i \cdot p_t^i}{\sum q_o^i \cdot p_t^i}$

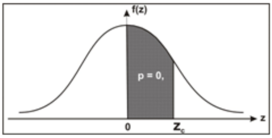
- Índice de Paasche:

Preço	Quantidade
$I_{o,t} = \sqrt{L_{o,t} \cdot P_{o,t}}$	$I'_{o,t} = \sqrt{L'_{o,t} \cdot P'_{o,t}}$

ANEXO B – TABELAS ESTATÍSTICAS

TABELA 1 – DISTRIBUIÇÃO NORMAL PADRÃO

PROBABILIDADES p TAIS QUE $p = (0 < Z < Z_c)$

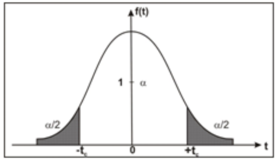


Parte inteira e primeira decimal de Z_c	SEGUNDA DECIMAL DE Z_c										Parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586	0,0
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535	0,1
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409	0,2
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173	0,3
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793	0,4
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240	0,5
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490	0,6
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524	0,7
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327	0,8
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891	0,9
1,0	0,34134	0,34375	0,34614	0,34850	0,35083	0,35314	0,35543	0,35769	0,35993	0,36214	1,0
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37695	0,37900	0,38100	0,38298	1,1
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147	1,2
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41309	0,41466	0,41621	0,41774	1,3
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42786	0,42922	0,43056	0,43189	1,4
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408	1,5
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449	1,6
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327	1,7
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062	1,8
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670	1,9
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169	2,0
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574	2,1
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899	2,2
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158	2,3
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361	2,4
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49492	0,49506	0,49520	2,5
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643	2,6
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736	2,7
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807	2,8
2,9	0,49813	0,49819	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861	2,9
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49897	0,49900	3,0
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929	3,1
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950	3,2
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965	3,3
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49970	3,4
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983	3,5
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989	3,6
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992	3,7
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995	3,8
3,9	0,49995	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997	3,9
4,0	0,49997	0,49997	0,49997	0,49997	0,49997	0,49997	0,49998	0,49998	0,49998	0,49998	4,0
4,5	0,49999	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	0,50000	4,5

Parte inteira e primeira decimal de Z_c	SEGUNDA E TERCEIRA DECIMAL DE Z_c										Parte inteira e primeira decimal de Z_c
	05	15	25	35	45	55	65	75	85	95	
0,0	0,00199	0,00598	0,00997	0,01396	0,01795	0,02193	0,02591	0,02989	0,03387	0,03784	0,0
0,1	0,04181	0,04578	0,04974	0,05369	0,05764	0,06159	0,06553	0,06946	0,07339	0,07730	0,1
0,2	0,08121	0,08512	0,08901	0,09290	0,09677	0,10064	0,10450	0,10834	0,11218	0,11600	0,2
0,3	0,11982	0,12362	0,12741	0,13119	0,13495	0,13871	0,14244	0,14617	0,14988	0,15358	0,3
0,4	0,15726	0,16093	0,16458	0,16822	0,17184	0,17545	0,17903	0,18261	0,18500	0,18970	0,4
0,5	0,19322	0,19672	0,20021	0,20368	0,20712	0,21055	0,21396	0,21735	0,22073	0,22408	0,5
0,6	0,22741	0,23072	0,23401	0,23729	0,24054	0,24377	0,24697	0,25016	0,25333	0,25647	0,6
0,7	0,25959	0,26270	0,26544	0,26883	0,27186	0,27488	0,27786	0,28083	0,28377	0,28669	0,7
0,8	0,28959	0,29246	0,29531	0,29814	0,30094	0,30372	0,30648	0,30921	0,31192	0,31461	0,8
0,9	0,31727	0,31990	0,32252	0,32511	0,32767	0,33021	0,33273	0,33522	0,33769	0,34013	0,9

TABELA 2 – DISTRIBUIÇÃO t DE STUDENT

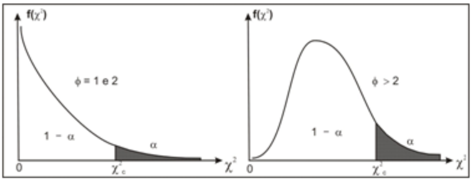
VALORES CRÍTICOS DE t TAIS QUE P $(-t_c < t < +t_c) = 1 - p$



φ	p = 90%	80%	70%	60%	50%	40%	30%	20%	10%	5%	4%	2%	1%	0,20%	0,10%	φ
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	15,894	31,821	63,657	318,309	636,619	**1**
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	4,849	6,965	9,925	22,327	31,598	**2**
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	3,482	4,541	5,841	10,214	12,924	**3**
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	2,998	3,747	4,604	7,173	8,610	**4**
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	2,756	3,365	4,032	5,893	6,869	**5**
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	2,612	3,143	3,707	5,208	5,959	**6**
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,517	2,998	3,499	4,785	5,408	**7**
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,449	2,896	3,355	4,501	5,041	**8**
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,398	2,821	3,250	4,297	4,781	**9**
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,359	2,764	3,169	4,144	4,587	**10**
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,328	2,718	3,106	3,025	4,437	**11**
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,303	2,681	3,055	3,930	4,318	**12**
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,282	2,650	3,012	3,852	4,221	**13**
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,264	2,624	2,977	3,787	4,140	**14**
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,248	2,602	2,947	3,733	4,073	**15**
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,235	2,583	2,921	3,686	4,015	**16**
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,224	2,567	2,898	3,646	3,965	**17**
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,214	2,552	2,878	3,610	3,922	**18**
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,205	2,539	2,810	3,579	3,883	**19**
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,197	2,528	2,845	3,552	3,850	**20**
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,189	2,518	2,831	3,527	3,819	**21**
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,183	2,508	2,819	3,505	3,792	**22**
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,177	2,500	2,807	3,485	3,768	**23**
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,172	2,492	2,797	3,467	3,745	**24**
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,165	2,485	2,787	3,450	3,725	**25**
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,162	2,479	2,779	3,435	3,707	**26**
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,158	2,473	2,771	3,421	3,690	**27**
28	0,127	0,256	0,389	0,530	0,684	0,855	1,056	1,313	1,701	2,048	2,154	2,467	2,763	3,408	3,674	**28**
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,150	2,462	2,756	3,396	3,659	**29**
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,147	2,457	2,750	3,385	3,646	**30**
35	0,126	0,255	0,388	0,529	0,682	0,852	1,052	1,306	1,690	2,030	2,133	2,438	2,724	3,340	3,591	**35**
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,123	2,423	2,704	3,307	3,551	**40**
50	0,126	0,254	0,387	0,528	0,679	0,849	1,047	1,299	1,676	2,009	2,109	2,403	2,678	3,261	3,496	**50**
60	0,126	0,254	0,387	0,527	0,679	0,848	1,045	1,296	1,671	2,000	2,099	2,390	2,660	3,232	3,460	**60**
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,076	2,358	2,617	3,160	3,373	**120**
$Z \cong \infty$	**0,126**	**0,253**	**0,385**	**0,524**	**0,674**	**0,842**	**1,036**	**1,282**	**1,645**	**1,960**	**2,054**	**2,326**	**2,576**	**3,090**	**3,291**	**$Z \cong \infty$**

TABELA 3 – DISTRIBUIÇAO QUI-QUADRADO

VALORES CRÍTICOS DE χ^2



α / φ	99,50%	99,00%	97,50%	95,00%	90,00%	50,00%	10,00%	5,00%	2,50%	1,00%	0,50%	α / φ
1	0,00004	0,00016	0,00098	0,00393	0,0158	0,455	2,71	3,84	5,02	6,63	7,88	**1**
2	0,0100	0,0201	0,0506	0,103	0,211	1,386	4,61	5,99	7,38	9,21	10,60	**2**
3	0,072	0,115	0,216	0,352	0,584	2,366	6,25	7,81	9,35	11,34	12,84	**3**
4	0,207	0,297	0,484	0,711	1,064	3,357	7,78	9,49	11,14	13,28	14,86	**4**
5	0,412	0,554	0,831	1,145	1,61	4,251	9,24	11,07	12,83	15,09	16,75	**5**
6	0,676	0,872	1,24	1,64	2,20	5,35	10,64	12,59	14,45	16,81	18,55	**6**
7	0,989	1,24	1,69	2,17	2,83	6,35	12,02	14,07	16,01	18,48	20,23	**7**
8	1,34	1,65	2,18	2,73	3,49	7,34	13,36	15,51	17,53	20,09	21,96	**8**
9	1,73	2,09	2,70	3,33	4,17	8,34	14,68	16,92	19,02	21,67	23,59	**9**
10	2,16	2,56	3,25	3,94	4,87	9,34	15,99	18,31	20,48	23,21	25,19	**10**
11	2,60	3,05	3,82	4,57	5,58	10,34	17,28	19,63	21,92	24,72	26,76	**11**
12	3,07	3,57	4,40	5,23	6,30	11,34	18,55	21,03	23,34	26,22	28,30	**12**
13	3,57	4,11	5,01	5,89	7,04	12,34	19,81	22,36	24,74	27,69	29,82	**13**
14	4,07	4,66	5,63	6,57	7,79	13,34	21,06	23,68	26,12	29,14	31,32	**14**
15	4,60	5,23	6,26	7,26	8,55	14,34	22,31	25,00	27,49	30,58	32,80	**15**
16	5,14	5,81	6,91	7,96	9,31	15,34	23,54	26,30	28,85	32,00	34,27	**16**
17	5,70	6,41	7,56	8,67	10,09	16,34	24,77	27,59	30,19	33,41	35,72	**17**
18	6,26	7,01	8,23	9,39	10,86	17,34	25,99	28,87	31,53	34,81	37,16	**18**
19	6,84	7,63	8,91	10,12	11,65	18,34	27,20	30,14	32,85	36,19	38,58	**19**
20	7,43	8,26	9,59	10,85	12,44	19,34	28,41	31,41	34,17	37,57	40,00	**20**
21	8,03	8,90	10,28	11,59	13,24	20,34	29,62	32,67	35,48	38,93	41,40	**21**
22	8,64	9,54	10,98	12,34	14,04	21,34	30,81	33,92	36,78	40,29	42,80	**22**
23	9,26	10,20	11,69	13,09	14,85	22,34	32,01	35,17	38,08	41,64	44,13	**23**
24	9,89	10,36	12,40	13,85	15,66	23,34	33,20	36,42	39,36	42,98	45,56	**24**
25	10,52	11,52	13,12	14,61	16,47	24,34	34,38	37,65	40,65	44,31	46,93	**25**
26	11,16	12,20	13,84	15,38	17,29	25,34	35,56	38,89	41,92	45,64	48,29	**26**
27	11,81	12,83	14,57	16,15	18,11	26,34	36,74	40,11	43,19	46,96	49,64	**27**
28	12,46	13,56	15,31	16,93	18,94	27,34	37,92	41,34	44,46	38,28	50,99	**28**
29	13,12	14,26	16,05	17,71	19,77	28,34	39,09	42,56	45,72	49,59	52,34	**29**
30	13,79	14,95	16,79	18,49	20,60	29,34	40,26	43,77	46,98	50,89	53,67	**30**
40	20,71	22,16	24,43	26,51	29,05	39,34	51,80	55,76	59,34	63,69	66,77	**40**
50	27,99	29,71	32,36	34,76	37,69	49,33	63,17	67,50	71,42	76,15	79,49	**50**
60	35,53	37,43	40,48	43,19	46,46	59,33	74,40	79,08	83,30	83,38	91,95	**60**
70	43,28	45,44	48,76	51,74	55,33	69,33	85,53	90,53	95,02	100,4	104,22	**70**
80	57,17	53,54	51,17	60,39	64,28	79,33	98,58	101,9	106,6	112,3	116,32	**80**
90	59,20	61,75	65,65	69,13	73,29	89,33	107,6	113,1	118,1	124,1	123,3	**90**
100	67,33	70,06	74,22	77,93	82,36	99,33	113,5	124,3	129,6	135,3	140,2	**100**

TABELA 4 – VALORES PERCENTUAIS (5%) DA DISTRIBUIÇÃO F DE SNEDECOR

ϕ_1 / ϕ_2	1	2	3	4	5	6	7	8	9	10	20	30	120	∞	ϕ_1 / ϕ_2
1	161,4	199,5	215,7	224,6	230,2	234	236,8	238,9	240,5	241,9	248	250,1	253,3	254,3	**1**
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,45	19,46	19,49	19,50	**2**
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,66	8,62	8,55	8,53	**3**
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,80	5,75	5,66	5,63	**4**
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,56	4,50	4,40	4,36	**5**
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,87	3,81	3,70	3,67	**6**
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,44	3,38	3,27	2,23	**7**
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,15	3,08	2,97	2,92	**8**
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	2,94	2,86	2,75	2,71	**9**
10	4,96	4,1	3,71	3,48	3,33	3,22	3,14	4,07	3,02	2,98	2,77	2,70	2,58	2,54	**10**
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,65	2,57	2,45	2,40	**11**
12	4,75	3,89	3,49	3,26	3,11	3,00	2,90	2,85	2,80	2,75	2,54	2,47	2,34	2,30	**12**
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,46	2,38	2,25	2,21	**13**
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,39	3,31	2,18	2,13	**14**
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,33	2,25	2,11	2,07	**15**
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,28	2,19	2,06	2,01	**16**
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,23	2,15	2,01	1,96	**17**
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,19	2,11	1,97	1,92	**18**
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,16	2,07	1,93	1,88	**19**
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,12	2,04	1,90	1,84	**20**
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,1	2,01	1,87	1,81	**21**
22	4,3	3,44	4,05	2,82	2,66	2,55	2,46	2,4	2,34	2,3	2,07	1,98	1,84	1,78	**22**
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,05	1,96	1,81	1,76	**23**
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,3	2,25	2,03	1,94	1,79	1,73	**24**
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	1,93	1,84	1,68	1,62	**30**
40	4,08	3,23	2,84	2,61	2,45	2,34	2,21	2,18	2,12	2,08	1,84	1,74	1,58	1,51	**40**
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,75	1,65	1,47	1,39	**60**
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,66	1,55	1,35	1,25	**120**
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,57	1,46	1,22	1,00	∞

TABELA 5 – VALORES PERCENTUAIS (2,5%) DA DISTRIBUIÇÃO F DE SNEDECOR

ϕ_2 \ ϕ_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	ϕ_1 / ϕ_2
1	647,8	799,5	864,2	899,6	921,8	937,1	948,2	956,7	963,3	968,6	976,7	984,9	993,1	997,2	1001	1006	1010	1014	1018	**1**
2	38,51	39,00	39,17	39,25	39,30	39,33	39,36	39,37	39,39	39,40	39,41	39,43	39,45	39,46	39,46	39,47	39,48	39,49	39,50	**2**
3	17,44	16,04	15,44	15,1	14,88	14,73	14,62	14,54	14,47	14,42	14,34	14,25	14,17	14,12	14,08	14,04	13,99	13,95	13,90	**3**
4	12,22	10,65	9,98	9,60	9,36	9,20	9,07	8,98	8,90	8,84	8,75	8,66	8,56	8,51	8,46	8,41	8,36	8,31	8,26	**4**
5	10,01	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68	6,62	6,52	6,43	6,33	6,28	6,23	6,18	6,12	6,07	6,02	**5**
6	8,81	7,26	6,6	6,23	5,99	5,82	5,7	5,6	5,52	5,46	5,37	5,27	5,17	5,12	5,07	5,01	4,96	4,90	4,85	**6**
7	8,07	6,54	5,89	5,52	5,29	5,21	4,99	4,90	4,82	4,76	4,67	4,57	4,47	4,42	4,36	4,31	4,25	4,20	4,14	**7**
8	7,57	6,06	5,42	5,05	4,82	4,65	4,53	4,43	4,36	4,30	4,20	4,10	4,00	3,95	3,89	3,84	3,78	3,73	3,67	**8**
9	7,21	5,71	5,08	4,72	4,48	4,32	4,20	4,10	4,03	3,96	3,87	3,77	3,67	3,61	3,56	3,51	3,45	3,39	3,33	**9**
10	6,94	5,46	4,83	4,47	4,24	4,07	3,95	3,85	3,78	3,72	3,62	3,52	3,42	3,37	3,31	3,26	3,20	3,14	3,08	**10**
11	6,72	5,26	4,63	4,28	4,04	3,88	3,76	3,66	3,59	3,53	3,43	3,33	3,23	3,17	3,12	3,06	3,00	2,94	2,88	**11**
12	6,55	5,1	4,47	4,12	3,89	3,73	3,61	3,51	3,44	3,37	3,28	3,18	3,07	3,02	2,96	2,91	2,85	2,79	2,72	**12**
13	6,41	4,97	4,35	4,00	3,77	3,60	3,48	3,39	3,31	3,25	3,15	3,05	2,95	2,89	2,84	2,78	2,72	2,66	2,60	**13**
14	6,3	4,86	4,24	3,89	3,66	3,50	3,38	3,29	3,21	3,15	3,05	2,95	2,84	2,79	2,73	2,67	2,61	2,55	2,49	**14**
15	6,20	4,77	4,15	3,80	3,58	3,41	3,29	3,20	3,12	3,06	2,96	2,86	2,76	2,70	2,64	2,59	2,52	2,46	2,40	**15**
16	6,12	4,69	4,08	3,73	3,50	3,34	3,22	3,12	3,05	2,99	2,89	2,79	2,68	2,63	2,57	2,51	2,45	2,38	2,32	**16**
17	6,04	4,62	4,01	3,66	3,44	3,28	3,16	3,06	2,98	2,92	2,82	2,72	2,62	2,66	2,50	2,44	2,38	2,32	2,25	**17**
18	5,98	4,56	3,95	3,61	3,38	3,22	3,10	3,01	2,93	2,87	2,77	2,67	2,56	2,50	2,44	2,38	2,32	2,26	2,19	**18**
19	5,92	4,51	3,90	3,56	3,33	3,17	3,05	2,96	2,88	2,82	2,72	2,62	2,51	2,45	2,39	2,33	2,27	2,20	2,13	**19**
20	5,87	4,46	3,86	3,51	3,29	3,13	3,01	2,91	2,84	2,77	2,68	2,57	2,46	2,41	2,35	2,29	2,22	2,16	2,09	**20**
21	5,83	4,42	3,82	3,48	3,25	3,09	2,97	2,87	2,8	2,73	2,64	2,53	2,42	2,37	2,31	2,25	2,18	2,11	2,04	**21**
22	5,79	4,38	3,78	3,44	3,22	3,05	2,93	2,84	2,76	2,7	2,60	2,50	2,39	2,33	2,27	2,21	2,14	2,08	2,00	**22**
23	5,75	4,35	3,75	3,41	3,18	3,02	2,90	2,81	2,73	2,67	2,57	2,47	2,36	2,30	2,24	2,18	2,11	2,04	1,97	**23**
24	5,72	4,32	3,72	3,38	3,15	2,99	2,87	2,78	2,70	2,64	2,54	2,44	2,33	2,27	2,21	2,15	2,08	2,01	1,94	**24**
25	5,69	4,29	3,69	3,35	3,13	2,97	2,85	2,75	2,68	2,61	2,51	2,41	2,3	2,24	2,18	2,12	2,05	1,98	1,91	**25**
26	5,66	4,27	3,67	3,33	3,10	2,94	2,82	2,73	2,65	2,59	2,49	2,39	2,28	2,22	2,16	2,09	2,03	1,95	1,88	**26**
27	5,63	4,24	3,65	3,31	3,08	2,92	2,80	2,71	2,63	2,57	2,47	2,36	2,25	2,19	2,13	2,07	2,00	1,93	1,85	**27**
28	5,61	4,22	3,63	3,29	3,06	2,90	2,78	2,69	2,61	2,55	2,45	2,34	2,23	2,17	2,11	2,05	1,98	1,91	1,83	**28**
29	5,59	4,20	3,61	3,27	3,04	2,88	2,76	2,67	2,59	2,53	2,43	2,32	2,21	2,15	2,09	2,03	1,96	1,89	1,81	**29**
30	5,57	4,18	3,59	3,25	3,03	2,87	2,75	2,65	2,57	2,51	2,41	2,31	2,20	2,14	2,07	2,01	1,94	1,87	1,79	**30**
40	5,42	4,05	3,46	3,13	2,90	2,74	2,62	2,53	2,45	2,39	2,29	2,18	2,07	2,01	1,94	1,88	1,8	1,72	1,64	**40**
60	5,29	3,93	3,34	3,01	2,79	2,63	2,51	2,41	2,33	2,27	2,17	2,06	1,94	1,88	1,82	1,74	1,67	1,58	1,48	**60**
120	5,15	3,80	3,23	2,89	2,67	2,52	2,39	2,30	2,22	2,16	2,05	1,94	1,82	1,76	1,69	1,61	1,53	1,43	1,31	**120**
∞	5,02	3,69	3,12	2,79	2,57	2,41	2,29	2,19	2,11	2,05	1,94	1,83	1,71	1,64	1,57	1,48	1,39	1,27	1,00	∞

TABELA 6 – VALORES PERCENTUAIS (1%) DA DISTRIBUIÇÃO F DE SNEDECOR

ϕ_2 \ ϕ_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	ϕ_1 / ϕ_2
1	4052	4999,5	8403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	**1**
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	**2**
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	**3**
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	**4**
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,72	9,05	9,47	9,38	9,29	9,20	**5**
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	**6**
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,34	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	**7**
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	**8**
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	**9**
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	**10**
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	4,00	3,86	3,78	**11**
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	**12**
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	**13**
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	**14**
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	**15**
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	**16**
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	**17**
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	**18**
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	**19**
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,63	**20**
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	**21**
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	**22**
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	**23**
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	**24**
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	**25**
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	**26**
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	**27**
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	**28**
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	**29**
30	7,56	5,39	4,51	4,02	3,7	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	**30**
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	**40**
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	**60**
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	**120**
∞	6,63	4,61	3,78	3,32	3,02	2,8	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	**∞**

TABELA 7 – NUMEROS ALEATORIOS

03991	10461	93716	16894	98953	73213	39528	72484	82474	25593
38555	95554	32886	59780	09958	18065	81616	18711	53342	44276
17546	73704	92052	46215	15917	06253	07586	16120	82641	22820
32643	52861	95819	06831	19640	99413	90767	04235	13574	17200
69572	68777	39510	35905	85244	35159	40188	28193	29593	88627
24122	66591	27699	06494	03152	19121	34414	82157	86887	55087
61196	30231	92692	61773	22109	78508	63439	75363	44989	16822
30532	21704	10274	12202	94205	20380	67049	09070	93399	45547
03788	97599	75867	20717	82037	10268	79495	04146	52162	90286
48228	63379	85783	47619	87481	37220	91704	30552	04737	21031
88618	19161	41290	67312	74857	15957	48545	35247	18619	13674
71299	23853	05870	01119	92784	26340	75122	11724	74627	73707
27954	58909	82444	99005	04921	73701	92904	13141	32392	19763
80863	00514	20247	81759	45197	25332	69902	63742	78464	22501
33564	60780	48460	85558	15191	18782	94972	11598	62095	36787
90899	75754	60833	25983	01291	41349	19152	00023	12302	80783
78038	70267	43529	06318	38384	74761	36024	00867	76378	41605
55986	66485	88722	56736	66164	49431	94458	74284	05041	49807
87539	08823	94813	31900	54155	83436	54158	34243	46978	35482
16818	60311	74457	90561	722848	11834	75051	93029	47665	64382
34677	58300	74910	64345	19325	81540	60365	94653	35075	33949
45305	07521	61318	31855	14413	70951	83799	42402	56623	34442
59747	67277	76503	34513	39663	77544	32960	07405	36409	83232
16520	69676	11654	99893	02181	68161	19322	53845	57620	52606
68652	27376	92852	55866	88448	03584	11220	94747	07399	37408
79375	95220	01159	63267	10622	48391	31751	57260	68980	05339
33521	26665	55823	47641	86225	31704	88492	99382	14454	04504
59589	49067	66821	41575	49767	04037	30934	47744	07481	83828
20554	91409	96277	48257	50816	97616	22888	48893	27499	98748
59404	72059	43947	51680	43852	59693	78212	16993	35902	91386
42614	29297	01918	28316	25163	01889	70014	15021	68971	11403
34994	41374	70071	14736	65251	07629	37329	33295	18477	65622
99385	41600	11133	07586	36815	43625	18637	37509	14707	93997
66497	68646	78138	66559	64397	11692	05327	82162	83745	22567
48509	23929	27482	45476	94515	25624	95096	67946	16930	33361
15470	48355	88651	22596	83761	60873	43253	84145	20368	07126
20094	98977	74843	93413	14387	06345	80854	09279	41196	37480
73788	06533	28597	20405	51321	92246	80088	77074	66919	31678
60530	45128	74022	84617	72472	00008	80890	18002	35352	54131
44372	15486	65741	14014	05466	55306	93128	18464	79982	68416
18611	19241	66083	24653	84609	58232	41849	84547	46850	52323
58319	15997	08355	60860	29735	47762	46352	33049	69248	93460
18627	90872	00911	98936	76355	93779	52701	08337	56303	87315
00441	58997	14060	40619	29549	69616	57275	36898	81304	48585
32624	68691	14845	46672	61958	77100	20857	73156	70284	24326
65961	73488	41839	55382	17267	70943	15633	84924	90415	93614
20288	34060	39685	23309	10061	68829	92694	48297	39904	02115
59362	95938	74416	53166	35208	33374	77613	19019	88152	00080
99782	93478	53152	67433	35663	52972	38688	32486	45134	63545
73625	48824	30016	10122	14567	81234	67812	12876	56712	28190
66124	43101	23251	32145	88374	53167	68901	43658	44563	23341
28373	56712	83109	56722	12577	47891	50012	76802	12893	58900
34321	72913	42191	98634	32145	43567	56928	44123	42298	54439
12823	32190	72019	23173	57912	78924	46734	32675	44566	78231

TABELA 8 – VALORES CRÍTICOS DA ESTATÍSTICA H

i = número de grupos
φ = n - 1 = graus de liberdade de cada grupo

α = 5%

φ	i										
	2	3	4	5	6	7	8	9	10	11	12
2	39,00	87,50	142,00	202,00	266,00	333,00	403,00	475,00	550,00	626,00	704,00
3	15,40	27,80	39,20	50,70	62,00	72,90	83,50	93,90	104,00	114,00	124,00
4	9,60	15,50	20,60	25,20	29,50	33,60	37,50	41,10	44,60	48,00	51,40
5	7,15	10,80	13,70	16,30	18,70	20,80	22,90	24,70	26,50	28,20	29,90
6	5,82	8,38	10,40	12,10	13,70	15,00	16,30	17,50	18,60	19,70	20,70
7	4,99	6,94	8,44	9,70	10,80	11,80	12,70	13,50	14,30	15,10	15,80
8	4,43	6,00	7,18	8,12	9,03	9,78	10,50	11,10	11,70	12,20	12,70
9	4,03	5,34	6,31	7,11	7,80	8,41	8,95	9,45	9,91	10,30	10,70
10	3,72	4,85	5,67	6,34	6,92	7,42	7,87	8,28	8,66	9,01	9,34
12	3,28	4,16	4,79	5,30	5,72	6,09	6,42	6,72	7,00	7,25	7,48
15	2,86	3,54	4,01	4,37	4,68	4,95	5,19	5,40	5,59	5,77	5,93
20	2,46	2,95	3,29	3,54	3,76	3,94	4,10	4,24	4,37	4,49	4,59
30	2,07	2,40	2,61	2,78	2,91	3,02	3,12	3,21	3,29	3,36	3,39
60	1,67	1,85	1,96	2,04	2,11	2,17	2,22	2,26	2,30	2,33	2,36
∞	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

α = 1%

φ	i										
	2	3	4	5	6	7	8	9	10	11	12
2	199,00	448,00	729,00	1036,0	1362,0	1705,0	2063,0	2432,0	2813,0	3204,0	3605,0
3	47,50	85,00	120,00	151,00	184,00	216,00	249,00	281,00	310,00	337,00	361,00
4	23,20	37,00	49,00	59,00	69,00	79,00	89,00	97,00	106,00	113,00	120,00
5	14,90	22,00	28,00	33,00	38,00	42,00	46,00	50,00	54,00	57,00	60,00
6	11,10	15,50	19,10	22,00	25,00	27,00	30,00	32,00	34,00	36,00	37,00
7	8,89	12,10	14,50	16,50	18,40	20,00	22,00	23,00	24,00	26,00	27,00
8	7,50	9,90	11,70	13,20	14,50	15,80	16,90	17,90	18,90	19,80	21,00
9	6,54	8,50	9,90	11,10	12,10	13,10	13,90	14,70	15,30	16,00	16,60
10	5,85	7,40	8,60	9,60	10,40	11,10	11,80	12,40	12,90	13,40	13,90
12	4,91	6,10	6,90	7,60	8,20	8,70	9,10	9,50	9,90	10,20	10,60
15	4,07	4,90	5,50	6,00	6,40	6,70	7,10	7,30	7,50	7,80	8,00
20	3,32	3,80	4,30	4,60	4,90	5,10	5,30	5,50	5,60	5,80	5,90
30	2,63	3,00	3,30	3,40	3,60	3,70	3,80	3,90	4,00	4,10	4,20
60	1,96	2,20	2,30	2,40	2,40	2,50	2,50	2,60	2,60	2,70	2,70
∞	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

TABELA 9 – COEFICIENTES a_{n-k+1} PARA O TESTE W - NORMALIDADE

k	n									
	1	2	3	4	5	6	7	8	9	10
1	---	0,7071	0,7071	0,6872	0,6646	0,6431	0,6233	0,6052	0,5888	0,5739
2	---	---	0,0000	0,1677	0,2413	0,2806	0,3031	0,3164	0,3244	0,3291
3	---	---	---	---	0,0000	0,0885	0,1401	0,0174	0,1976	0,2141
4	---	---	---	---	---	---	0,0000	0,0561	0,0947	0,1224
5	---	---	---	---	---	---	---	---	0,0000	0,0399

k	N									
	11	12	13	14	15	16	17	18	19	20
1	0,5601	0,5475	0,5359	0,5251	0,5150	0,5056	0,4968	0,4886	0,4808	0,4734
2	0,3315	0,3325	0,3325	0,3318	0,3306	0,3290	0,3273	0,3253	0,3232	0,3211
3	0,2260	0,2347	0,2412	0,2460	0,2495	0,2521	0,2540	0,2553	0,2561	0,2565
4	0,1429	0,1586	0,1707	0,1802	0,1878	0,1939	0,1988	0,2027	0,2059	0,2085
5	0,0695	0,0927	0,1099	0,1240	0,1353	0,1447	0,1109	0,1587	0,1641	0,1686
6	0,0000	0,0303	0,0539	0,0727	0,0880	0,1005	0,0725	0,1197	0,1271	0,1334
7	---	---	0,0000	0,0240	0,0433	0,0593	0,3590	0,0837	0,0932	0,1013
8	---	---	---	---	0,0000	0,0196	0,0000	0,0496	0,0612	0,0711
9	---	---	---	---	---	---	---	0,0163	0,0303	0,0422
10	---	---	---	---	---	---	---	---	0,0000	0,0140

k	n									
	21	22	23	24	25	26	27	28	29	30
1	0,4643	0,4590	0,4542	0,4493	0,4450	0,4407	0,4366	0,4328	0,4291	0,4245
2	0,3185	0,3156	0,3126	0,3098	0,3069	0,3043	0,3018	0,2992	0,2968	0,2944
3	0,2578	0,2571	0,2563	0,2554	0,2543	0,2533	0,2522	0,2510	0,2499	0,2487
4	0,2119	0,2131	0,2139	0,2145	0,2148	0,2151	0,2152	0,2151	0,2150	0,2148
5	0,1736	0,1764	0,1787	0,1807	0,1822	0,1836	0,1848	0,1857	0,1864	0,1870
6	0,1399	0,1443	0,1480	0,1512	0,1539	0,1563	0,1584	0,1601	0,1616	0,1630
7	0,1092	0,1150	0,1201	0,1245	0,1283	0,1316	0,1346	0,1372	0,1395	0,1415
8	0,0804	0,0878	0,0941	0,0997	0,1046	0,1089	0,1128	0,1162	0,1192	0,1219
9	0,0530	0,0618	0,0696	0,0764	0,0823	0,0876	0,0923	0,0965	0,1002	0,1036
10	0,0263	0,0368	0,0459	0,0539	0,0610	0,0672	0,0728	0,0778	0,0822	0,0862
11	0,0000	0,0122	0,0228	0,0321	0,0403	0,0476	0,0540	0,0598	0,0650	0,0697
12	---	---	0,0000	0,0107	0,0200	0,0284	0,0358	0,0424	0,0483	0,0537

TABELA 10 – PONTOS PERCENTUAIS DO TESTE W PARA n = 3 a 50

n	Nível								
	1%	2%	5%	10%	50%	90%	95%	98%	99%
3	0,753	0,756	0,767	0,789	0,959	0,998	0,999	1,000	1,000
4	0,687	0,707	0,748	0,792	0,935	0,987	0,992	0,996	0,997
5	0,686	0,715	0,762	0,806	0,927	0,979	0,986	0,991	0,993
6	0,713	0,743	0,788	0,826	0,927	0,974	0,981	0,986	0,989
7	0,730	0,760	0,803	0,838	0,928	0,972	0,979	0,985	0,988
8	0,749	0,778	0,818	0,851	0,932	0,972	0,978	0,984	0,987
9	0,764	0,791	0,829	0,859	0,935	0,972	0,978	0,984	0,986
10	0,781	0,806	0,842	0,869	0,938	0,972	0,978	0,983	0,986
11	0,792	0,817	0,850	0,876	0,940	0,973	0,979	0,984	0,986
12	0,805	0,828	0,859	0,883	0,943	0,973	0,979	0,984	0,986
13	0,814	0,837	0,866	0,889	0,945	0,974	0,979	0,984	0,986
14	0,825	0,846	0,874	0,895	0,947	0,975	0,980	0,984	0,986
15	0,835	0,855	0,881	0,901	0,950	0,975	0,980	0,984	0,987
16	0,844	0,863	0,887	0,906	0,952	0,976	0,981	0,985	0,987
17	0,851	0,869	0,892	0,910	0,954	0,977	0,981	0,985	0,987
18	0,858	0,874	0,897	0,914	0,956	0,978	0,982	0,986	0,988
19	0,863	0,879	0,901	0,917	0,957	0,978	0,982	0,986	0,988
20	0,868	0,884	0,905	0,920	0,959	0,979	0,983	0,986	0,988
21	0,873	0,888	0,908	0,923	0,960	0,980	0,983	0,987	0,989
22	0,878	0,892	0,911	0,926	0,961	0,980	0,984	0,987	0,989
23	0,881	0,895	0,914	0,928	0,962	0,981	0,984	0,987	0,989
24	0,884	0,898	0,916	0,930	0,963	0,981	0,984	0,987	0,989
25	0,888	0,901	0,918	0,931	0,964	0,981	0,985	0,988	0,989
26	0,891	0,904	0,920	0,933	0,965	0,982	0,985	0,988	0,989
27	0,894	0,906	0,923	0,935	0,965	0,982	0,985	0,988	0,990
28	0,896	0,908	0,924	0,936	0,966	0,982	0,985	0,988	0,990
29	0,898	0,910	0,926	0,937	0,966	0,982	0,985	0,988	0,990
30	0,900	0,912	0,927	0,939	0,967	0,983	0,985	0,988	0,990
31	0,902	0,914	0,929	0,940	0,967	0,983	0,986	0,988	0,990
32	0,904	0,915	0,930	0,941	0,968	0,983	0,986	0,988	0,990
33	0,906	0,917	0,931	0,942	0,968	0,983	0,986	0,989	0,990
34	0,908	0,919	0,933	0,943	0,969	0,983	0,986	0,989	0,990
35	0,910	0,920	0,934	0,944	0,969	0,984	0,986	0,989	0,990
36	0,912	0,922	0,935	0,945	0,970	0,984	0,986	0,989	0,990
37	0,914	0,924	0,936	0,946	0,970	0,984	0,987	0,989	0,990
38	0,916	0,925	0,938	0,947	0,971	0,984	0,987	0,989	0,990
39	0,917	0,927	0,939	0,948	0,971	0,984	0,987	0,989	0,991
40	0,919	0,928	0,940	0,949	0,972	0,985	0,987	0,989	0,991
41	0,920	0,929	0,941	0,950	0,972	0,985	0,987	0,989	0,991
42	0,922	0,930	0,942	0,951	0,972	0,985	0,987	0,989	0,991
43	0,923	0,932	0,943	0,951	0,973	0,985	0,987	0,990	0,991
44	0,924	0,933	0,944	0,952	0,973	0,985	0,987	0,990	0,991
45	0,926	0,934	0,945	0,953	0,973	0,985	0,988	0,990	0,991
46	0,927	0,935	0,945	0,953	0,974	0,985	0,988	0,990	0,991
47	0,928	0,936	0,946	0,954	0,974	0,985	0,988	0,990	0,991
48	0,929	0,937	0,947	0,954	0,974	0,985	0,988	0,990	0,991
49	0,929	0,937	0,947	0,955	0,974	0,985	0,988	0,990	0,991
50	0,930	0,938	0,947	0,955	0,974	0,985	0,988	0,990	0,991

Este livro tem por finalidade despertar o interesse pelos métodos quantitativos e seu conteúdo é direcionado a auxiliar pesquisadores das Ciências Sociais e Humanas e da Saúde. A Estatística é uma ciência aplicável a qualquer ramo do conhecimento em que se manipulem dados resultantes de uma pesquisa. Assim, a Administração, a Engenharia, a Enfermagem, a Medicina, a Biologia, a Economia, as Ciências Agrárias etc, tendem cada vez mais a servir-se desta técnica como ferramenta de estudo, daí surge a crescente importância desta ciência. O principio deste livro é apresentar os principais conceitos utilizados em métodos quantitativos e suas técnicas fundamentais de análise de dados.

EDITORA Voix

9788594414076