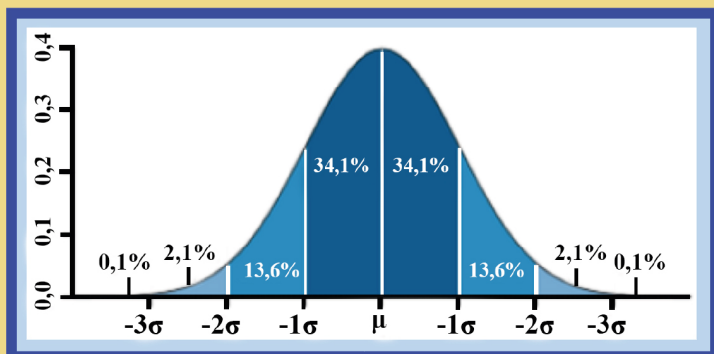


INTERVALOS DE CONFIANÇA E TESTE DE HIPÓTESES APLICADOS EM PSICOLOGIA

Teoria e aplicação no SPSS e R



Edwirde Luiz Silva Camêlo
Dalila Camêlo Aguiar
Ivan Olier
Ramón Gutiérrez Sánchez

EDWIRDE LUIZ SILVA CAMÊLO
DALILA CAMÊLO AGUIAR
IVAN OLIER
RAMÓN GUTIÉRREZ SÁNCHEZ

**Intervalos de confiança e teste de hipóteses aplicados
em psicologia**

Teoria e aplicação no SPSS e R



Campina Grande-PB | 2025



Universidade Estadual da Paraíba

Prof^a. Célia Regina Diniz | *Reitora*

Prof^a. Ivonildes da Silva Fonseca | *Vice-Reitora*



Editora da Universidade Estadual da Paraíba

Cidoval Morais de Sousa | *Diretor*

Conselho Editorial

Alessandra Ximenes da Silva (UEPB)

Alberto Soares de Melo (UEPB)

Antonio Roberto Faustino da Costa (UEPB)

José Etham de Lucena Barbosa (UEPB)

José Luciano Albino Barbosa (UEPB)

Melânia Nóbrega Pereira de Farias (UEPB)

Patrícia Cristina de Aragão (UEPB)



Editora indexada no SciELO desde 2012



Editora filiada a ABEU

EDITORA DA UNIVERSIDADE ESTADUAL DA PARAÍBA

Rua Baraúnas, 351 - Bairro Universitário - Campina Grande-PB - CEP 58429-500
Fone: (83) 3315-3381 - <http://eduepb.uepb.edu.br> - email: eduepb@uepb.edu.br

Expediente EDUEPB***Design Gráfico e Editoração***

Erick Ferreira Cabral

Jefferson Ricardo Lima A. Nunes

Leonardo Ramos Araujo

Revisão Linguística e Normalização

Antonio de Brito Freire

Elizete Amaral de Medeiros

Assessoria Editorial

Eli Brandão da Silva

Assessoria Técnica

Thaise Cabral Arruda

Divulgação

Danielle Correia Gomes

Comunicação

Efigênio Moura

Depósito legal na Câmara Brasileira do Livro - CDL

I61 Intervalos de confiança e teste de hipóteses aplicados em psicologia [recurso eletrônico] : teoria e aplicação no SPSS e R / Edwirde Luiz Silva Camêlo ... [et al.]. – Campina Grande : EDUEPB, 2025.
205 p. : il. color. ; 15 x 21 cm.

ISBN: 978-65-5221-100-2 (Impresso)

ISBN: 978-65-5221-103-3 (PDF)

ISBN: 978-65-5221-102-6 (Epub)

1. Estatística Aplicada à Psicologia. 2. Intervalo de confiança. 3. Teste de Confiança. 4. Programa SPSS. 5. Programa R. I. Camêlo, Edwirde Luiz Silva. II. Aguiar, Dalila Camêlo. III. Olier, Ivan. IV. Sánchez, Ramón Gutiérrez. V. Título.

21. ed. CDD 519.53

Ficha catalográfica elaborada por Fernanda Mirelle de Almeida Silva - CRB - 15/483

Copyright © **EDUEPB**

A reprodução não-autorizada desta publicação, por qualquer meio, seja total ou parcial, constitui violação da Lei nº 9.610/98.

AGRADECIMENTOS

Gostaríamos de agradecer à EDUEPB, que tem apoiado e estimulado a divulgação dos trabalhos dos professores da UEPB. Devemos também agradecer ao trabalho profissional do Prof. XXXX, que pacientemente abrilhantou este trabalho realizando as devidas correções gramaticais. Finalmente, nós autores não poderíamos deixar de agradecer aos alunos que nos permitiram a experiência para elaboração deste livro, aos colegas que contribuíram para a realização desta compilação e em especial ao **Eterno, Criador dos céus e da terra** que nos outorgou vida para concluir esta obra.

Lista de Tabelas

1.1	Gastos com limpezas em clínicas psicológicas . . .	37
1.2	Três níveis de confiança mais usados	38
1.3	tempo de recuperação de um transtorno afetivo bi- polar	43
1.4	Idade de 40 idosos de uma clínica	44
1.5	Idade dos 20 jovens	45
1.6	Tempo de reação de um novo medicamento contra ansiedade	45
1.7	Uma amostra de 10 níveis de ansiedade	51
1.8		64
1.9	Uma amostra de 10 indivíduos com dificuldades educativas	65
1.10	Uma amostra de 12 indivíduos com abstinência do consumo de álcool	65
1.11	I.C. para distribuição de Poisson	121
1.12	I.C. para distribuição de Bernoulli	122
1.13	I.C. para os métodos Tchebyshev, T.C.L. e M.V. . .	123
1.14	I.C. para os métodos Tchebyshev, T.C.L. e M.V. . .	124

LISTA DE TABELAS

2.1		173
2.2	Número de homens e mulheres	177
2.3	Valores calculados	188

Lista de Figuras

1.1	Normal padranizada	28
1.2	Distribuição da variável aleatória normal padronizada	29
1.3	Z para uma área de 0,34 é 0,99	30
1.4	Distribuição Z padronizado	33
1.5	Intervalo de confiança e outras estatísticas no SPSS	40
1.6	Output do SPSS	41
1.7	Região de confiança entre C_1 e C_2	47
1.8	Distribuição qui-quadrada com n-1 grau de liberdade	53
1.9	Passos para encontrar I.C. em teste paramétricos . .	66
1.10	Passos para encontrar I.C. em teste não paramétricos	66
1.11	Distribuição F de Sneedecor	75
1.12	Área que tem de manter-se constante quando se des- locam (z_1, z_2) para esquerda ou para direita.	104
1.13	Distribuição Z padronizado com $Z_{cal} = 2,18$	106
1.14	Distribuição Normal com as caudas C_1 e C_2	112
1.15	Intervalo da coordenada $\theta^- = a_n$	114
1.16	Intervalo de confiança para θ	116
1.17	Intervalo de confiança para θ	118

LISTA DE FIGURAS

1.18	Intervalo com limite superior limitado	119
2.1	Possíveis erros de acertos de uma decisão	131
2.2	Distribuição t de Student unicaudal à esquerda com nível de significância 5%	132
2.3	Distribuição t de Student unicaudal à direita com nível de significância 5%	133
2.4	População (N) e amostra (n)	134
2.5	Distribuição t de Student bicaudal informando as áreas de aceitação e rejeição com nível de signifi- cância 5%	136
2.6	Opções para o teste T para uma amostra	139
2.7	Menu: teste T para uma amostra	139
2.8	Output do teste T para uma amostra	140
2.9	Limite crítico para aceitar ou rejeitar a hipótese nula	141
2.10	Limite para aceitar ou rejeitar a hipótese nula . . .	142
2.11	Limite para aceitar ou rejeitar a hipótese nula . . .	143
2.12	Analisar - descritiva e explorar	147
2.13	Verificando a normalidade da variável Idade	151
2.14	Output do teste de Shapiro Wilks	151
2.15	Analisar - descritiva e explorar	153
2.16	Explorar - Estatísticas	154
2.17	Explorar - Gráficos	154
2.18	Valores ausentes	155
2.19	Principais testes de hipóteses para duas amostras .	156
2.20	Parte da Tabela t de Student	158
2.21	Teste para duas amostras independentes com vari- âncias conhecidas	159
2.22	Saída do SPSS para o teste para duas amostras in- dependentes com variâncias conhecidas	159

LISTA DE FIGURAS

2.23	Teste para duas amostras independentes com variâncias desiguais	162
2.24	Qui-Quadrado unicaudal e bicaudal	164
2.25	Distribuição t de Student unilateral	170
2.26	1Passos no SPSS	172
2.27	Passos no SPSS continuação	172
2.28	Output do SPSS	173
2.29	Distribuição normal padronizada bicaudal	178
2.30	Tabela t de Student com v grau de liberdade	185
2.31	Tabela T de Student bicaudal com 8 graus de liberdade	185
2.32	Distribuição t de Student bicaudal informando as áreas de aceitação e rejeição com nível de significância 5%	188
2.33	Distribuição t de Student com 29,74 graus de liberdade, informando as áreas de aceitação e rejeição	189
2.34	Limites unilaterais da distribuição F Fisher-Snedecor aos níveis 2,5% e 5%	191
2.35	Erros tipo I e erro tipo II	195

Sumário

Prefácio	15
1 Intervalo de confiança	17
1.1 Introdução ao intervalo de confiança	17
1.1.1 Noções fundamentais sobre a estimação de intervalo de confiança	17
1.1.2 Algumas distribuições de probabilidades teó- ricas	19
1.1.3 Distribuições discretas	20
1.1.4 Distribuições contínuas	20
1.2 Construção da região de confiança	20
1.3 Características de uma distribuição amostral	26
1.3.1 Erro padrão da média	27
1.4 I.C. para Média de uma distribuição Normal com variância conhecida	31
1.4.1 No SPSS	40
1.5 I.C. para Média de uma distribuição Normal com variância desconhecida	41
1.5.1 Região de confiança para a média μ e a va- riância σ^2 de uma normal	45

1.6	Intervalo de confiança para a variância	47
1.6.1	Intervalo unilateral à esquerda a um nível de confiança de $(1 - \alpha)100\%$	49
1.6.2	Intervalo de confiança unilateral à direita a um nível de confiança de $(1 - \alpha)100\%$	49
1.7	Intervalo de confiança para o desvio padrão	54
1.8	Intervalos de confiança para duas populações normais	55
1.8.1	Intervalos de confiança para diferença de medias	56
1.8.2	Intervalo de confiança para quando as vari- âncias são desconhecidas e distintas	60
1.9	Intervalo de confiança para a diferença de média em amostras emparelhadas	62
1.10	Intervalo para os parâmetros de duas populações normais independentes	67
1.11	Intervalos de confiança para o quociente de variâncias	70
1.11.1	Com médias μ_x y μ_y desconhecidas	70
1.11.2	Com médias μ_x y μ_y conhecidas	73
1.12	I.C. para diferença de médias sendo as variâncias desconhecidas e iguais	77
1.13	I.C. para diferença de médias sendo desconhecidas as variâncias amostrais	78
1.14	Outros método para a construção de I.C.	78
1.15	Intervalos de confiança para proporções	84
1.15.1	Intervalos de confiança para a proporção sem utilizar o dobro da aproximação	91
1.15.2	Intervalos de confiança para a proporção uti- lizando o dobro da aproximação	92

1.15.3	Determinação do tamanho amostral para proporção	93
1.16	I.C. para diferença de duas proporções dicotômicas	94
1.17	Intervalo de média para diferença de médias de duas populações normais de variâncias desconhecidas não iguais	99
1.18	I.C. para diferença de duas proporções populacionais	102
1.19	I.C. para populações não normais	108
1.19.1	Caso geral	108
1.20	I.C. assintótico para qualquer parâmetro	120
1.21	Intervalo de confiança para o correlação entre duas populações normais	124
1.21.1	Pequenas amostras para correlação	124
1.21.2	Grandes amostras para correlação	124
2	Teste de hipóteses	128
2.1	Relação entre testes de hipóteses e intervalos de confiança	128
2.2	Tipos de erros	129
2.2.1	Erro tipo I e erro tipo II	130
2.3	Teste de hipóteses através do valor p	131
2.3.1	O teste de hipótese para uma média	133
2.3.2	Teste T para uma média no SPSS	137
2.3.3	Resumo de teste de hipótese para uma média	142
2.4	Exploração dos dados: teste de normalidade e igualdade de variâncias	145
2.5	Normalidade univariada	147
2.5.1	Dispersão por nível com o teste de Levene	152
2.6	Teste de hipóteses para duas amostras	155
2.7	Teste de hipóteses para duas amostras independentes	156

2.8	Teste de hipóteses para diferenças de médias populacionais com variâncias populacionais desconhecidas	160
2.9	Teste de hipótese para variância populacional	163
2.10	Teste de hipótese para T pareado	167
2.10.1	Resumo de teste de hipóteses para a diferenças de médias populacionais relacionadas	174
2.11	Teste para uma proporção populacional	176
2.12	Teste T para duas amostras independentes	179
2.12.1	Uso do SPSS para um teste T para duas amostras independentes	181
2.13	Teste com amostras independentes e variâncias iguais	183
2.14	Teste com amostras independentes com variâncias desconhecidas e diferentes	186
2.15	Teste de hipóteses para a comparação de variâncias com médias populacionais conhecidas e desconhecidas	189

Prefácio

A estatística básica e sua prática constitui uma introdução à estatística para estudantes de estatística e psicologia I e II e também os tópicos especiais em estatística para o mestrado de Pós-Graduação em Psicologia da Saúde da Universidade Estadual da Paraíba. Neste prefácio, dou uma descrição do livro para permitir aos professores julgarem se ele é adequado aos seus alunos de graduação e pós-graduação.

Em linhas gerais esse livro dar ênfase ao pensamento estatístico aplicado a psicologia; mais dados e conceitos, menos teoprias e menos receitas; promover o ensino ativo que é caracterizado como métodos de ensino que fortalecem a formação de vínculos democráticos na relação entre professor, alunos e conteúdo.

O livro é elementar quanto ao nível da matemática exigida e aos processos estatísticos apresentados, o livro tem como objetivo dar ao estudante não só a compreensão das idéias fundamentais da estatística e psicologia como a habilidade necessária para lidar também com data-frame. Os exemplos e exercícios foram aplicados em psicologia com objetivo de proporcionar fundamento suficiente. Com frequência, faço conclusões que são mais do que um simples número (ou “o intervalo variada de tanto a tanto”). Alguns exercícios exigem julgamento, além de cálculos e conclusões do tipo “certo ou errado”. Tanto o estudante quanto o professor devem reconhecer que nem toda parte do exercício admite uma única resposta correta.

Os Capítulos 1 apresenta uma breve introdução de intervalo de confiança. O aluno tem a oportunidade de apreciar as distribuições discretas e contínuas. No capítulo 2 apresenta a construção do intervalo de confiança. No capítulo 3 apresenta algumas medidas de posição e dispersão. No capítulo 4 apresenta uma introdução de in-

tervalo de confiança no SPSS. No capítulo 5 apresenta os intervalos de confiança. Os demais capítulos trata-se de diversos intervalo de confiança.

Queremos ainda mostrar aos graduandos e pós-graduandos como incorporar os resultados das suas análises e como interpretar os resultados nos artigos. Tentamos simplificar conceitos complexos, e, algumas vezes, bastantes complexos. Entretanto, ao facilitar existe uma perda de acurácia. Nos exercícios propostos após exemplos de cálculos ajudarão a resolver os problemas de intervalo de confiança. Os conjuntos de dados para exemplos e exercícios constam no próprio texto, de onde podem ser lidos para qualquer programa estatístico além do SPSS, como por exemplo, o programa R.

Sobre o autor

Edwirde Luiz Silva Camêlo é professor associado da Universidade Estadual da Paraíba, campus I em Campina Grande. Possui graduação em Matemática pela Universidade Rural de Pernambuco, mestrado em Biometria e doutorado (2007) em Estadística e Investigación Operativa pela Universidad de Granada (UGR) - España. Tem experiência na área de estatística univariada e multivariada aplicado a psicologia da saúde, experiência em Redes neurais artificiais. Possui vários trabalhos publicados na área. Estou hebraico bíblico e hebraico moderno na UGR.

Capítulo 1

Intervalo de confiança

1.1 Introdução ao intervalo de confiança

Um intervalo de confiança (I.C.) faz referência a probabilidade de que o I.C. está contido no parâmetro. Extraem-se amostras aleatórias independentes da população, 95% dessas proporcionarão um intervalo que contém o verdadeiro valor do parâmetro, mas o estimado estimado a partir de uma realização amostral concreta, pertence a esse 95% que no caso contém o parâmetro com probabilidade 1, ou pertence a 5% do restante, em cujo caso contém o parâmetro com probabilidade 0.

1.1.1 Noções fundamentais sobre a estimação de intervalo de confiança

O intervalo de confiança descreve a variabilidade entre a medida obtida em um estudo e a medida real da população (o valor verdadeiro). Corresponde a um intervalo de valores, cuja distribui-

ção é normal e em que se encontra o valor real de uma determinada variável, com elevada probabilidade.

Definição 1.1.1. *Seja duas estatísticas $a(X_1, \dots, X_n) < b(X_1, \dots, X_n)$ com $a(X_1, \dots, X_n) < b(X_1, \dots, X_n)$ e um vetor $\alpha \in (0, 1)$. Suponha que se verifica $P[a(X_1, \dots, X_n) < \theta < b(X_1, \dots, X_n)] = 1 - \alpha$ para todo $\theta \in \Theta$. Então, para uma realização específica da amostra $x_1, \dots, x_n$ diz-se que $[a(x_1, \dots, x_n), b(x_1, \dots, x_n)]$ é um I.C. para θ com nível de confiança $1 - \alpha$. O que se denota $I.C_{1-\alpha}(\theta)$*

Começamos com uma definição formal do que procuramos: consideramos uma amostra $X_1, \dots, X_n$ de uma variável aleatória com função de distribuição $F(\cdot; \theta)$, sendo $\theta \in \Theta \subset \mathbb{R}$ um parâmetro desconhecido.

Exemplo 1.1.1. *Por exemplo, num estudo sobre a eficácia de um novo tratamento contra a ansiedade, os investigadores poderiam recolher dados de uma amostra de pacientes e calcular de alguma forma esse nível de redução média da ansiedade no grupo de tratamento. Um intervalo de confiança de 95% para a redução média da ansiedade seria um intervalo de valores dentro do qual se espera que a verdadeira redução média da ansiedade na população esteja com uma probabilidade de 95%. Isto significa que se o mesmo estudo for realizado 100 vezes, aproximadamente 95 estudos terão a verdadeira média populacional dentro do seu intervalo de confiança de 95%.*

Existem duas formas básicas de inferência, teste de hipótese e intervalos de confiança. Neste livro estudaremos apenas intervalo de confiança.

O objetivo do intervalo de confiança é encontrar duas funções de amostra (dependendo da amostra), θ_1^* e θ_2^* , para que possamos

afirmar que o verdadeiro valor do parâmetro (θ) se encontra entre esses parâmetros com uma certa distribuição de probabilidade.

A probabilidade de que o valor verdadeiro do parâmetro esteja no intervalo com *nível de confiança*, denotado por $1 - \alpha$, é

$$P[\theta_1^* < \theta < \theta_2^*] = 1 - \alpha$$

O valor de α é chamado de nível de significância ou risco assumido, sendo $\alpha \in [0, 1]$.

Com tudo isso, o intervalo formado por θ_1^* e θ_2^* , é chamado de **Intervalo de confiança no nível $1 - \alpha$ do parâmetro θ** . Ressaltamos que quanto maior o risco, α , maior é a largura do intervalo.

Existem duas formas básicas de inferência estatística: intervalo de confiança (I.C.) e no teste de hipótese. No I.C. existe a estimação pontual e a estimação intervalar.

Devido a importância que tem as populações normais, dedicou-se esse livro especialmente no estudo dos intervalos de confiança para seus parâmetros. A facilidade de cálculo de quantidade pivotal (fundamental) que apresentam essas populações fazem recomendável a obtenção destes intervalos de confiança através do método pivotal.

1.1.2 Algumas distribuições de probabilidades teóricas

Sabe-se que as distribuições de probabilidades teóricas dependem de uma série de parâmetros. Vejamos alguns deles:

1.1.3 Distribuições discretas

Lei de probabilidade	Função de massa de probabilidade	Parâmetros
Bernouilli	$P[X = x] = p^x q^{1-x}$	p
$\beta(p)$	$x=0,1; p+q=1$	$0 < p < 1$
Poisson	$P[X = x] = \exp -\lambda \frac{\lambda^x}{x!}$	λ
$P(\lambda)$	$x = 0, 1, 2, \dots$	$\lambda > 0$
Binomial	$P[X = x] = \binom{n}{x} p^x q^{n-x}$	n,p
B(n,p)	$x = 0, 1, 2, \dots, n; p + q = 1$	$n \in N, 0 < p < 1$
Hipergeométrica	$P[X = x] = \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}}$	N,D,n
H(N,D,n)	$\max(0, n - N + D) \leq x \leq \min(D, n)$	N,D, $n \in N$

1.1.4 Distribuições contínuas

Vamos a estudar as seguites distribuições contínuas:

Lei de probabilidade	Função de probabilidade	Parâmetros
Normal	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x-\mu}{2\sigma^2}\right)$	μ, σ^2
$N(\mu, \sigma^2)$	$-\infty \leq x + \infty$	$\mu \in R, \sigma > 0$
t Student	$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$	n
t_n	$x \in R$	$n \in N$
Qui-quadrada	$f(x) = \frac{1}{\Gamma(n/2)2^{n/2}} x^{n/2-1} \exp -x/2$	α, β
$\Gamma(\alpha, \beta) : \chi_n^2 \sim \Gamma(n/2, 2)$	$x > 0$	$\alpha, \beta > 0$

1.2 Construção da região de confiança

Seja $X_1, \dots, X_n$ uma amostra procedente de uma população cuja distribuição depende de um certo parâmetro θ , e seja um estatístico $t = T(X_1, \dots, X_n)$ cuja função de distribuição de probabilidade é representada por $F(t, \theta)$. Vamos definir o que é uma estatística.

Em muitas ocasiões interessa manejar uma função da amostra aleatória no lugar de utilizar a própria amostra.

Definição 1.2.1. Dado $T(X_1, \dots, X_n)$ uma massa aleatória de uma

variável aleatória com distribuição F_θ , denomina-se estatística a qualquer função T das variáveis aleatórias da amostra que não contém nenhum valor desconhecido ($T : R^n \rightarrow R^k, k$), sendo $k =$ dimensão da estatística. Uma estatística não é nada mais que uma variável aleatória que se obtém a partir de uma massa aleatória. Um valor específico desta nova variável aleatória quando se dispõe de uma realização $(x_1, \dots, x_n)$ da amostra. Quem é uma estatística S_1 ou S_2 ?

$$S_1 = \frac{1}{\mu} \left(\sum_{i=1}^4 x_i^2 - 6\bar{x}^2 \right) = \frac{1}{\mu} (10 - 6) = \frac{4}{\mu}$$

$$S_2 = \frac{1}{4} \left(\sum_{i=1}^4 x_i^2 - 6\bar{x}^2 \right) = \frac{1}{4} (10 - 6) = \frac{4}{4} = 1$$

Claramente se observa que S_2 é uma estatística, já S_1 não é uma estatística.

Definição 1.2.2. Quando se dispõe de uma massa aleatória com valores no espaço amostral (α^n, B^n) , a estatística $T(X_1, \dots, X_n)$ com $T : (\alpha^n, B^n) \sim (R^k, B^k)$ é uma nova variável aleatória. Isso faz que se deva considerar sua distribuição de probabilidade, distribuição que indique com que probabilidade pode tomar essa estatística cada valor possível.

Em geral, sempre que t seja variável contínua $F(t, \theta)$ é uma variável aleatória com distribuição uniforme no intervalo $(0, 1)$. Chamamos $Z = h(t, \theta)$, Z_o o quantil $\alpha/2$ e Z_1 o quantil $1 - \frac{\alpha}{2}$. Evidentemente, se cumpre que para $\forall \theta$:

$$P_o(Z_o < h(t, \theta) < Z_1) = 1 - \alpha$$

Define-se:

$$I(\theta, \alpha) = x/Z_o < h(x, \theta) < Z_1 R(t, \theta) = y/Z_o < h(t, y) < Z_1$$

Se θ_o representa o verdadeiro e desconhecido valor do parâmetro então:

$$t \in I(\theta_o, \alpha) \leftrightarrow \theta_o \in R(t, \alpha)$$

Já que t e θ_o se pertencem, verifica-se que: $Z_o < h(t, \theta_o) < Z_1$. Por tanto,

$$P(t \in I(\theta_o, \alpha)) = P(\theta_o \in R(t, \alpha)) = P(Z_o < h(t, \theta_o) < Z_1) = 1 - \alpha$$

Como consequência $R(t, \alpha)$ é uma região aleatória (que depende do valor que se obtém de t segundo a amostra), que tem probabilidade $1 - \alpha$ de “acampar” (juntar) no seu interior o verdadeiro valor do parâmetro, assim constitui uma região de confiança. Se a função $h(t, \alpha)$ é monótona em θ para todo t , então a região $R(t, \alpha)$ é um intervalo de confiança.

Exemplo 1.2.1. *Suponha que o tempo que demora para uma assistência psicológica eficaz seja uma variável aleatória com distribuição exponencial. Calcular um intervalo de confiança do tempo médio para essa assistência psicológica.*

O tempo médio da assistência é o parâmetro θ da distribuição exponencial. Dada uma amostra de tamanho n a estatística $t = \sum_{i=1}^n X_i$ tem uma distribuição gamma com parâmetro n , conhecido, que pertence a θ . A variável aleatória será: $Z = \frac{2t}{\theta}; \forall \theta > 0$ tem distribuição χ^2 com $2n$ graus de liberdade (gamma de parâmetros

n e 2). Os valores Z_o e Z_1 são os quantis $\alpha/2$ e $1 - \alpha/2$ da dita distribuição. Então:

$$I(\theta_o, \alpha) = \left(x/Z_o < \frac{2x}{\theta} < Z_1 \right)$$

$$R(t, \alpha) = \left(y/Z_o < \frac{2t}{y} < Z_1 \right)$$

É evidente que:

$$R(t, \alpha) = \left(y/\frac{2t}{Z-1} < \theta_o < \frac{2t}{Z_o} \right) = 1 - \alpha$$

E, portanto, temos:

$$P(\theta_o \in R(t, \alpha)) = P\left(\frac{2t}{Z_1} < \theta_o < \frac{2t}{Z_o}\right) = 1 - \alpha$$

Exemplo 1.2.2. Se a média de uma amostra de tamanho 10 tem uma determinada média, então o I.C. da média populacional com nível de confiança de 0,90 com $Z_0 = 10,85$ e $Z_1 = 31,42$. será:

$$I.C. = \left[\frac{2.65}{31,42} < \theta_o < \frac{2.65}{10,85} \right] = [4,14 < \theta_o < 11,96]$$

Um estimador tem maior significação se vem acompanhado de uma medida de precisão. Uma das medidas tem suposição de probabilidade do tipo:

$$P_\theta[|V - \theta| \leq \varepsilon]$$

Sendo $|V - \theta|$ o erro de estimação, e ε é uma cota superior para este erro. Interessa também considerar uma cota inferior para essa probabilidade. Assim, o erro ε tem a seguinte forma:

$$P_{\theta}[|V - \theta| \leq \varepsilon] \geq 1 - \alpha, \quad \forall \theta \in \Theta$$

Podemos ainda reescrever da seguinte forma:

$$P_{\theta}[V - \varepsilon \leq \theta \leq V + \varepsilon] \geq 1 - \alpha, \quad \forall \theta \in \Theta$$

Pergunta-se: qual é o intervalo que contém o verdadeiro valor θ com probabilidade ao menos $1 - \alpha$? Sabe-se que tal intervalo é $[V - \varepsilon, V + \varepsilon]$.

Exemplo 1.2.3. *Seja X uma variável aleatória com distribuição uniforme $U(0, \theta)$, com $\theta > 0$. Sendo $x = 1.80$ encontre um I.C. ao nível de $1 - \alpha$ para θ baseando-se em uma observação apenas. Faça como ponto de partida $T(X) = X$, cuja função de densidade dada por: $f(x, \theta) = \frac{1}{\theta}$, $0 \leq x \leq \theta$ simétrica em relação a $E_{\theta}[X] = \frac{\theta}{2}$.*

Para encontrar o I.C. temos que:

$$P_{\alpha} \left[\left| X - \frac{\theta}{2} \right| \leq k \right] = 1 - \alpha, \quad \forall \theta \in \Theta$$

Subtrai k na inequação acima, fica então:

$$P_{\alpha} \left[\left| \frac{\theta}{2} - k \leq X \leq \frac{\theta}{2} + k \right| \right] = 1 - \alpha, \quad \forall \theta \in \Theta$$

O I.C. que produz a estimação mais precisa para θ será aquele que fica entre $[-\frac{\theta}{2}; +\frac{\theta}{2}]$, ou seja:

$$\frac{\alpha}{2} = P_{\alpha} \left[X < \frac{\theta}{2} - k \right] = \int_0^{\theta/2 - k} \frac{1}{\theta} dx$$

$$\frac{\alpha}{2} = P_{\alpha} \left[X > \frac{\theta}{2} + k \right] = \int_{\theta/2+k}^0 \frac{1}{\theta} dx$$

Resolvendo qualquer uma das integrais acima, encontramos $k = \frac{\theta(1-\alpha)}{2}$. Substituindo em 1.2, temos:

$$P \left[\frac{\theta}{2} - \frac{\theta(1-\alpha)}{2} \leq X \leq \frac{\theta}{2} + \frac{\theta(1-\alpha)}{2} \right] = 1 - \alpha, \quad \forall \theta \in \Theta$$

Resolvendo a inequação para θ , fica:

$$P_{\alpha} \left[\frac{2X}{2-\alpha} \leq \theta \leq \frac{2X}{\alpha} \right] = 1 - \alpha$$

Assim, o I.C. para θ ao nível de $1 - \alpha$ é:

$$\left[\frac{2X}{2-\alpha}, \frac{2X}{\alpha} \right]$$

Exemplo 1.2.4. Para uma amostral $x = 0,80$. Qual será o I.C. a 95%?

$$\left[\frac{2X}{2-\alpha}, \frac{2X}{\alpha} \right] = \left[\frac{2.0,80}{2-0,05}, \frac{2.0,80}{0,05} \right] = [0,82; 35]$$

Assim, o I.C. para a observação amostral $x = 0,80$, o intervalo de confiança ao nível de 95% é igual a $[0,82; 35]$.

Exemplo 1.2.5. Para $x = 0,86$. Qual será o I.C. a 90%?

$$\left[\frac{2X}{2-\alpha}, \frac{2X}{\alpha} \right] = \left[\frac{2.0,86}{2-0,10}, \frac{2.0,86}{0,10} \right] = [0,90; 17,2]$$

Depois de considerar o raciocínio por trás dos intervalos de confiança, você aprenderá a construir esses intervalos para a média da população quando soubermos ou não o valor de α . Inicialmente, vamos trabalhar com intervalo de confiança.

O objetivo do intervalo de confiança é encontrar duas funções de amostra (dependendo da amostra), θ_1^* e θ_2^* , para que possamos afirmar que o verdadeiro valor do parâmetro (θ) se encontra entre esses parâmetros com uma certa distribuição de probabilidade.

A probabilidade de que o valor verdadeiro do parâmetro esteja no intervalo com *nível de confiança*, denotado por $1 - \alpha$, é

$$P[\theta_1^2 < \theta < \theta_2^2] = 1 - \alpha$$

O valor de α é chamado de nível de significância ou risco assumido, sendo $\alpha \in [0, 1]$.

Com tudo isso, o intervalo formado por θ_1^2 e θ_2^2 , é chamado de **Intervalo de confiança no nível $1 - \alpha$ do parâmetro θ** .

Nota: Ressaltamos que quanto maior o risco, α , maior é a largura do intervalo.

Exemplo 1.2.6. Para $x = 0,78$. Qual será o I.C. a 95%?

1.3 Características de uma distribuição amostral

- Aproxima-se de uma curva normal (desde que o tamanho da amostra seja razoavelmente grande, ou seja, $N > 30$)
- A média de uma distribuição amostral (a média das médias) é igual à verdadeira média populacional (μ)

- O desvio padrão de uma distribuição amostral ($\sigma_{\bar{x}}$) é menor do que o da população (σ). A média amostral é mais estável do que os escores que a compõe.

1.3.1 Erro padrão da média

É definido por ($\sigma_{\bar{x}}$). Desvio padrão das amostras. Medida de quão representativa a amostra poderá ser da população. Na realidade não podemos selecionar centenas de amostras para construir uma distribuição amostral. Técnica para estimar o erro padrão a partir do desvio padrão da amostra(s): Dividir s pela raiz quadrada do tamanho da amostra (N).

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = EP$$

Além de nos fornecer uma ideia da diferença entre a média da amostra (X) e a média populacional (μ). Com ajuda do Erro Padrão da Média podemos estimar a probabilidade de nossa média populacional situar-se realmente dentro de um intervalo de valores médios à Conceito de intervalo de confiança.

Como encontrar um intervalo de confiança? Se selecionarmos 100 amostras aleatória de idosos estressados, calcularmos a média e, depois de determinarmos o intervalo de confiança para aquela média, 95% dos intervalos de confiança conterão o valor real da média populacional.

Tabela da Distribuição Normal Padrão
 $P(Z \leq z)$

z	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964

$z = 1.96$
 1.9 linha
 0.06 coluna

$95\% + 2.5\% = 97.5\%$
 Lembre-se foi 5%
 Para ser bicaudal: $5\%/2$
 $5\%/2 = 2.5\%$
 Assim, $95\% + 2.5\% = 0.9750\%$

Figura 1.1: Normal padranizada

O valor 1,96 é um valor de z importante, pois 95% dos escores de z estão entre -1,96 e 1,96 que tem um desvio à esquerda de $\mu - 2\sigma$ e um desvio à direita de $\mu + 2\sigma$, sendo s o desvio padrão amostral do desvio padrão populacional σ .

$$Z = \frac{x - \bar{x}}{s}$$

Assim, os valores de -1,96 e +1,96 serão:

$$1,96 = \frac{x - \bar{x}}{s} \quad -1,96 = \frac{x - \bar{x}}{s}$$

Multiplicando ambos os lados por s , temos:

$$1,96s = x - \bar{x} \rightarrow 1,96s + \bar{x}$$

Assim, temos:

- L_{inf} = limite inferior do intervalo de confiança: $\bar{x} - 1,96.EP$

- $L_{sup.}$ = Limite superior do intervalo de confiança: $\bar{x} + 1,96.EP$

$$I.C.(95\%) = \bar{x} \pm 1,96.EP \quad I.C.(99\%) = \bar{x} \pm 2,58.EP$$

Distribuição normal reduzida – estude esses percentuais observando os valores na Tabela Z. Uma grande vantagem de assumir uma forma normal para uma distribuição é que muitas grandezas importantes podem ser obtidas de maneira simples a partir de tabelas ou softwares. A Figura 1.2, por exemplo, mostra a posição da média e três desvios-padrão do lado esquerdo e do lado direito da média. A partir das propriedades matemáticas da distribuição normal, sabe-se que 95% da população estará abarcada no intervalo dado pela média ± 2 desvios-padrão, e 99,7% no centro $\pm$ desvios-padrão.

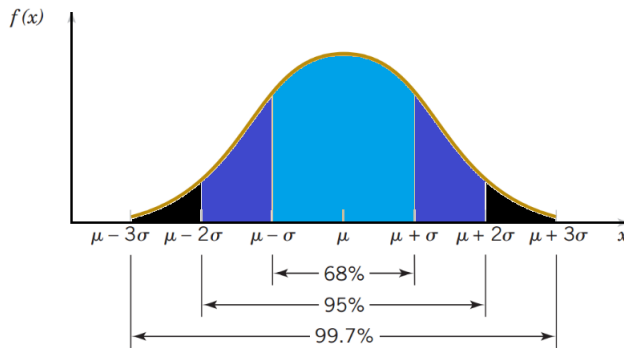


Figura 1.2: Distribuição da variável aleatória normal padronizada

Exemplo 1.3.1. Digamos que tenhamos coletados de forma aleatória uma amostra de tamanho ($N = 100$), com média amostral $\bar{x} = 3800$ e desvio padrão amostral, $s = 1500$. O erro padrão (EP) é dado por: $EP = 1500/\sqrt{100} = 150$. Com esses valores os I.C.

serão:

- Limite inferior do intervalo de confiança = $3800 - (1,96 * 150) = 3506 = L_{inf}$.
- Limite superior do intervalo de confiança = $3800 + (1,96 * 150) = 4094 = L_{sup}$.

Interpretação: Nesse intervalo contém a média populacional (μ) da população com um erro de 5%, podemos dizer que entre o intervalo 3506 e 4094 tem 95% de chance de conter a média populacional, (μ).

Exemplo 1.3.2. *Em um estudo sobre longevidade se analisaram as idades de 489 aposentados. Obtendo-se uma média de idade 72 anos com um desvio padrão de 8,6. Entre que valores se situam 68% próximo da média?*

Se 68% se situam próximo da média isso significa que cada lado se encontra a metade de 0,68 que é 0,34.

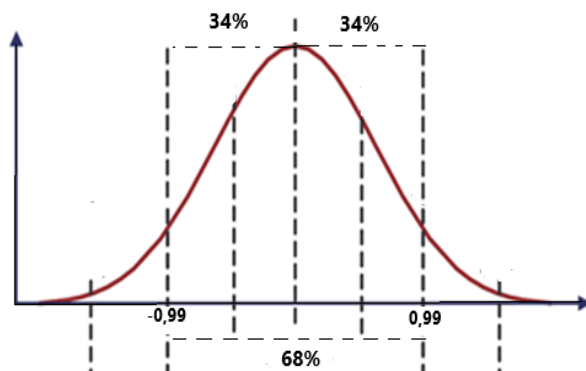


Figura 1.3: Z para uma área de 0,34 é 0,99

Substituindo em $Z = \frac{X - \bar{X}}{S}$, temos: $\pm 0,99 = \frac{X - 72}{8,6}$. Assim,

$$X = 72 \pm (0,99.86) \begin{cases} X_{\text{inf.}} = 72 - (0,99.86) = 63,48 \\ X_{\text{sup.}} = 72 + (0,99.86) = 80,51 \end{cases}$$

Portanto, o valor 68% se situa entre 63,48 e 80,51.

Exercício 1.3.1. *Em um estudo sobre inteligência naturalista (cuidados com os animais, cuidado com as plantas e ciências) se analisaram as idades de 400 estudantes. A média das idades foi 72 anos e o desvio padrão foi de 8,1. Entre que valores se situam 70% próximo da média?*

Exercício 1.3.2. *Em um estudo sobre a inteligência lingüística analisaram as idades de 356 idosos. A média das idades foi 68 anos e um desvio padrão de 7,6. Entre que valores se situam 81% próximo da média?*

1.4 I.C. para Média de uma distribuição Normal com variância conhecida

Sabemos que nas condições iniciais, $\bar{x} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$, portanto:

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

Como Z é $N(0, 1)$, podemos escrever:

$$P \left[-Z_{\frac{\alpha}{2}} < Z < Z_{\frac{\alpha}{2}} \right] = 1 - \alpha$$

$$P \left[-Z_{\frac{\alpha}{2}} < \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\frac{\alpha}{2}} \right] = 1 - \alpha$$

Resolve de modo que o μ fique no meio da inequação.

$$P \left[-Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} - \bar{x} < -\mu < Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} - \bar{x} \right]$$

$$P \left[Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} + \bar{x} > \mu > -Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} + \bar{x} \right]$$

$$P \left[\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

O intervalo de confiança no nível $1 - \alpha$ para a média de uma população normal com variância conhecida é:

$$\left[\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

Assim, o I.C para μ populacional tem a seguinte forma:

$$P \left[\bar{x} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right] = 1 - \alpha$$

Para o caso de populações finitas, usa-se a seguinte fórmula:

$$P \left[\bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}; \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right] = 1 - \alpha$$

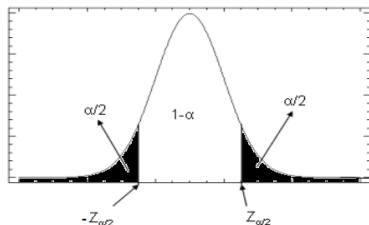


Figura 1.4: Distribuição Z padronizado

Considere ainda a seguinte proposição:

Proposição 1.4.1. *Se $\bar{x}$ é uma estatística com média amostral de uma amostra aleatória de tamanho n de uma variável aleatória com distribuição $N(\mu; \sigma^2)$, a distribuição na amostra dessa estatística é $N(\mu, \sigma^2/2)$.*

Da proposição segue que:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Se considera X uma variável aleatória $X \sim N(\mu, \sigma^2)$, $-\infty < \mu < +\infty$, σ^2 , da qual se extrai uma amostra aleatória de tamanho n , $\mathbf{X} = (X_1, \dots, X_n)$. As estatísticas a utilizar para obter o I.C. para μ serão:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

Caso σ^2 seja conhecido ou desconhecido. Para encontrar o I.C., no caso da variância conhecida precisa-se encontrar:

$$c(\alpha)/P \left[\left| \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right| \leq c(\alpha) \right] = 1 - \alpha$$

Operando no sucesso anterior, chaga-se a:

$$P \left[\bar{X} - c(\alpha) \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + c(\alpha) \frac{\sigma}{\sqrt{n}} \right] = 1 - \alpha$$

Em que o I.C. será dado por:

$$P \left[\bar{X} - c(\alpha) \frac{\sigma}{\sqrt{n}}; \bar{X} + c(\alpha) \frac{\sigma}{\sqrt{n}} \right]$$

Sendo $c(\alpha)$ uma constante cujo valor se determinará considerando a distribuição de $Z \sim N(0, 1)$. Assim, o I.C. para μ conhecida σ^2 :

$$\left[\bar{X} - z_{\sigma/2} \frac{\sigma}{\sqrt{n}}; \bar{X} + z_{\sigma/2} \frac{\sigma}{\sqrt{n}} \right]$$

O valor da abscissa z_α sob uma $N(0, 1)$ deixa a sua direita uma área α .

Fazendo $C_1 = C_2 = z_{\sigma/2}$. A longitude desse intervalo será:

$$D = \left(\bar{X} - C_1 \frac{\sigma}{\sqrt{n}} \right) - \left(\bar{X} - C_2 \frac{\sigma}{\sqrt{n}} \right) = \frac{\sigma}{\sqrt{n}} (C_2 - C_1)$$

Pretende-se encontrar o intervalo 1.4 usando o método dos multiplicadores de Lagrange, sujeita a condição:

$$P[C_1 \leq \mathbf{N}(\mathbf{0}; \mathbf{1}) \leq C_2] = 1 - \alpha$$

Assim, deverá buscar o mínimo de:

$$\mathfrak{K} = \frac{\sigma}{\sqrt{n}} (C_2 - C_1) + \mathfrak{J} \left[\int_{C-1}^{C_2} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}a^2\right) da - (1 - \alpha) \right]$$

Encontrando a primeira derivada de $\frac{\delta \mathfrak{K}}{\delta C_2}$, $\frac{\delta \mathfrak{K}}{\delta C_1}$ e $\frac{\delta \mathfrak{K}}{\delta \mathfrak{J}}$, temos:

$$\frac{\delta \bar{x}}{\delta C_2} = \frac{\sigma}{\sqrt{n}} + \frac{1}{\sqrt{n}} \exp\left(-\frac{1}{2}C_2^2\right) = 0$$

$$\frac{\delta \bar{x}}{\delta C_1} = \frac{\sigma}{\sqrt{n}} + \frac{1}{\sqrt{n}} \exp\left(-\frac{1}{2}C_1^2\right) = 0$$

$$\int_{C_1}^{C_2} \frac{1}{\sqrt{n}} \exp\left(-\frac{1}{2}a^2\right) da - (1 - \alpha) = 0$$

Das duas primeiras derivadas se obtém:

$$e^{-\frac{1}{2}C_2^2} = e^{-\frac{1}{2}C_1^2}$$

Ou seja, $C_1^2 = C_2^2$ as soluções são:

- $C_1 = C_2$, incompatível com a existência do intervalo posto que faz sua longitude nula, $D = 0$
- $C_1 = -C_2$, conduz que o intervalo de longitude mínima seja simétrico em uma distribuição $N(0; 1)$, ou seja, o formato pelos opostos $(-C; C)$.

Assim, o I.C. para a média μ de uma população $N(\mu; \sigma)$ com σ conhecida é o mesmo visto na 1.4:

$$\left[\bar{X} - C \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + C \frac{\sigma}{\sqrt{n}} \right]$$

Exemplo 1.4.1. De uma população $N(\mu; 3)$ se extrai uma amostra aleatória simples de tamanho 25, sendo $\sum_{i=1}^{n=25} X_i = 100$. Determinar o I.C. de 95% para a média da população.

$$\bar{X} = \frac{100}{25} = 4$$

E, por outra parte, em uma $N(0, 1)$ se $P[-C \leq N(0; 1) \leq 1] = 1 - 5\% = 0,95$. Sendo $C = 1,96$; o I.C. para a média μ com uma confiança de 95% será;

$$\left[4 - 1,96 \frac{3}{\sqrt{25}} \leq \mu \leq 4 + C \frac{3}{\sqrt{25}} \right] = [2,82; 5,17]$$

Exemplo 1.4.2. *Determinar a altura média dos valores de 100 estudantes que tem ansiedades (que segue uma distribuição normal), tomou uma amostra aleatória de 10 deles em que tiveram as seguintes medidas em centímetros: 162, 176, 169, 165, 171, 169, 172, 168, 167 e 175. Determinar o valor da altura média da população com uma confiança de 95%, suponha que $\sigma = 4$.*

Um estimador pontual da média da altura μ é a média $\bar{x} = (162 + \dots + 175)/10 = 169,40$, mas não se sabe como de precisa é. Como $n = 10$ e $\alpha = 5\%$. O I.C. para a média populacional, μ , será:

$$\left[169,40 - 1,92 \frac{4}{\sqrt{10}}; 169,40 + 1,92 \frac{4}{\sqrt{10}} \right] = [166,9; 171,8]$$

Assim, com base na amostra, espera-se que este intervalo seja um dos 95 de cada 100 que contém a média μ , ou seja, a altura média é um número entre 166,9 e 171,8 com uma confiança de 95%.

Exemplo 1.4.3. *Um empresário quer testar, com base nos dados do quadro a seguir, e para um nível de significância $\alpha = 5\%$, se o gasto médio de limpeza de cada clínica psicológica está próximo de R\$1055,00. Das várias clínicas foi retirada uma amostra de tamanho 20 de onde obteve os seguintes valores em real de gasto*

com limpeza.

903,88	1036,92	1098,04	1011,26
1020,70	915,38	1014,53	1097,79
934,52	1214,08	993,45	1120,19
860,41	1039,19	950,38	942,83
936,78	1086,98	1144,94	1066,12

Tabela 1.1: Gastos com limpezas em clínicas psicológicas

Construir um I.C. para a média populacional dos gastos. Suponha que os gastos se distribuam numa normal. O grau de liberdade será g.l. = $n - 1 = 20 - 1 = 19$ com um nível de 90% de confiança, ou seja, $(1 - \alpha)100 = (1 - 5\%)100 = 0,95.100 = 95\%$.

$$IC(\mu, 1 - \alpha) = \left(\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}}; \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right)$$

$$IC(\mu, 1 - \alpha) = \left(1019,37 - 2,093 \frac{91,37}{\sqrt{20}}; 1019,37 + 2,093 \frac{91,37}{\sqrt{20}} \right)$$

Assim, o $IC(\mu, 1 - \alpha) = [976,6; 1062,13]$.

Como: IC de 95% = $\bar{X} \mp 2,093EP$. Observando o valor $t_{20-1,5\%/2} = 2,093$ (bicaudal) (ver na Tabela da distribuição t de Student). Se o tamanho da amostra fosse $n > 30$ usaria a distribuição Z.

- $L_{inf.} = 1019,3885 - 2,0938 * 20,43065 = 976,60$.
- $L_{sup.} = 1019,3885 + 2,0938 * 20,43065 = 1062,13$.

Assim, $P(976,60 \leq \mu \leq 1062,13) = 1 - 10\% = 0,95 = 95\%$, ou seja, a média populacional com 95% de acerto está neste intervalo.

Aplica-se para grandes amostras, ou seja, $n \geq 30$. Lembre-se que a estimativa do valor desconhecido μ é $\bar{x}$ e a margem de erro é $z \frac{\sigma}{\sqrt{n}}$.

Na Tabela 1.4 podemos encontrar alguns valores de z .

Nível de confiança	Área da cauda	z
90%	0,05	1,645
95%	0,025	1,960
99%	0,005	2,576

Tabela 1.2: Três níveis de confiança mais usados

Exemplo 1.4.4. *Encontrar os seguintes I.C. para a média populacional com variância conhecida. Suponha que a amostra é extraída de uma população que tem distribuição normal.*

1. Estabelecer um IC de 95% de confiança para a média populacional. Dado $\bar{X} = 141$, $\sigma = 5,73$ e $n = 30$.
2. Estabelecer um IC de 90% de confiança para a média populacional. Dado $\bar{X} = 141$, $\sigma = 5,73$ e $n = 34$.
3. Estabelecer um IC de 92% de confiança para a média populacional. Dado $\bar{X} = 90$, $\sigma = 4,1$ e $n = 38$.
4. Em determinada população, a quantidade de homens com transtorno de ansiedade é distribuído normalmente com um desvio padrão de 16. Uma amostra aleatória simples de 36 homens adultos é sorteada dessa população, obtendo-se uma média amostral de 78,2. Construa um intervalo de confiança para a média da quantidade de homens com transtorno de ansiedade dessa população. Faça $\alpha = 5\%$.

5. Considere os dados que correspondem a quantidade de pessoas que tem o mecanismo de defesa fantasia. Sendo $\bar{x} = 151,9$ e $\sigma = 10$.(Exemplo fictício).

137	154	159	155	167	159	158	159	152	169
154	158	140	149	145	157	160	155	155	143
157	139	159	139	129	162	151	150	134	151

- a) Construa um intervalo de confiança, a 95%, para a média da população de todas as pessoas que usam mecanismo.
- b) Construa um intervalo de confiança, a 90%, para a média da população de todas as pessoas que usam mecanismo.
6. Considere os seguintes dados:

12	10	14	20	21	13	18	14	16	14
10	11	19	31	18	17	18	19	37	18
20	10	19	29	19	21	21	24	24	26
22	9	20	29	17	21	24	24	24	21

- a) Estabelecer um IC de 90% de confiança para a média populacional.
- b) Estabelecer um IC de 95% de confiança para a média populacional.
- c) Estabelecer um IC de 96% de confiança para a média populacional.
- d) Estabelecer um IC de 99% de confiança para a média populacional.
7. Em 30 cidades de um estado foi colhida uma amostra aleatória de jovens ansiosos após pandemia, obtiveram-se os seguintes dados:

10	11	11	11	12	13	18	14	16	14
10	11	19	14	18	17	18	19	13	18
14	14	14	14	14	15	15	15	16	16

- (a) Construir o I.C. para a média, sendo $\alpha = 5\%$.
- (b) Construir o I.C. para a média, sendo $\alpha = 10\%$.

1.4.1 No SPSS

Inserir os dados no SPSS e seguir os passos abaixo.

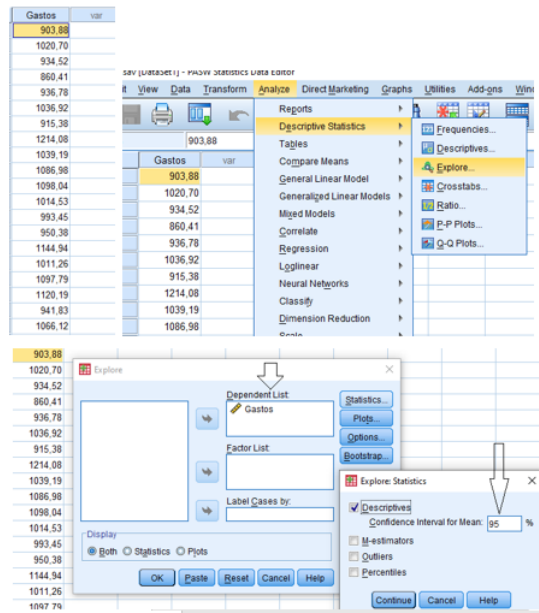


Figura 1.5: Intervalo de confiança e outras estatísticas no SPSS

No output do SPSS temos:

Descriptives			Média amostral
Gastos	Statistic	Std. Error	
Mean	1019,3685	20,43065	Erro padrão EP
95% Confidence Interval for Mean	Lower Bound 976,6067		I.C.
	Upper Bound 1062,1303		
5% Trimmed Mean	1017,3822		
Median	1017,6150		
Variance	8348,227		
Std. Deviation	91,36863		
Minimum	860,41		
Maximum	1214,08		
Range	353,67		amplitude = Valor Max - Valor Min
Interquartile Range	157,05		
Skewness	,239	,512	assimetria
Kurtosis	-,448	,992	

Figura 1.6: Output do SPSS

Para o caso de populações finitas, usa-se a seguinte fórmula:

$$P\left(\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}; \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}\right) = 1 - \alpha$$

Sendo N o número do tamanho populacional e n o número do tamanho amostral.

1.5 I.C. para Média de uma distribuição Normal com variância desconhecida

Se não conhecemos a variância da população, não podemos usar a estatística Z como no caso anterior. Usa-se então a distribuição t de Student (T).

$$T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

Esta estatística é conhecida por seguir uma distribuição t de Student com n-1 graus de liberdade.

Sendo o procedimento de cálculo idêntico ao anterior, resulta o

intervalo de confiança para a média de uma população normal com variância desconhecida, no nível $1 - \alpha/2$, da forma:

$$\left[\bar{X} - t_{(n-1; 1-\frac{\alpha}{2})} \frac{S}{\sqrt{n}}; \bar{X} + t_{(n-1; 1-\frac{\alpha}{2})} \frac{S}{\sqrt{n}} \right]$$

com $t_{n-1; 1-\frac{1-\alpha}{2}}$ o valor x tal que, $P[t_{n-1} < x] = 1 - \frac{\alpha}{2}$

Sendo S é o desvio padrão amostral. Sabendo que $\bar{X}$ e s são independentes, tem-se que:

$$\frac{\bar{X} - \mu}{S} \sqrt{n} \sim t_{n-1} \quad \begin{cases} \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \sim N(0; 1) \\ (n-1) \frac{S^2}{\sigma^2} \sim \chi_{n-1}^2 \end{cases}$$

Sendo t_{n-1} representa a distribuição t de Student com $n-1$ graus de liberdade. Expresando em termos de variância, S^2 , tem-se como se observou acima.

$$IC(\mu, 1 - \alpha) = \left(\bar{X} - t_{(n-1; 1-\alpha/2)} \frac{S}{\sqrt{n}}; \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}} \right)$$

Exemplo 1.5.1. Considere a seguinte amostra: 55,95; 68,24; 52,73; 21,50; 23,78 retirada de uma população normal. Estabeleça um intervalo de 95% para essa amostra $\bar{X}$.

A média e os desvio padrão valem $\bar{x} = 4,44$ e $s = 20,741$, respectivamente. O grau de liberdade é $gl = n - 1 = 5 - 1 = 4$. Para 95% de confiança o valor $t_{tab} = 2,776$. O I.C. será:

$$IC = [\bar{X} \pm t \frac{S}{\sqrt{n}}]$$

$$IC = [44,44 - 2,776 \frac{20,741}{\sqrt{5}}; 44,44 + 2,776 \frac{20,741}{\sqrt{5}}]$$

$$IC = [18,69; 70,19]$$

Exemplo 1.5.2. *Apartir de uma amostra de tempo de recuperação de um transtorno afetivobipolar cujos períodos de duração (em horas) se observa na tabela abaixo. Pede-se um I.C. de 95% para o tempo médio de recuperação desse transtorno.*

386	432	429	378	440	434	503	480	345	427
436	429	451	466	392	422	412	507	480	433

Tabela 1.3: tempo de recuperação de um transtorno afetivo bipolar

Tendo que $\bar{X} = 434,2$, $S = 40,63$ e que para $\alpha = 0,05$ e $n = 20$ é $t_{n-1; 1-\frac{\alpha}{2}} = 2,093$. Assim, o I.C. será:

$$IC = [\bar{X} \pm t \frac{S}{\sqrt{n}}]$$

$$IC = [434,2 \pm 2,093 \frac{40,63}{\sqrt{20}}]$$

$$IC = [415,2; 453,21]$$

Exercício 1.5.1. *Numa população obteve-se desvio padrão de 1,30 com uma amostra aleatória de 120 elementos. Sabendo que para essa mesma amostra obteve-se uma média de 6, determine o valor mais provável para a média dos dados. (Resp. $\bar{X} = 6 \pm 0,11$)*

Exercício 1.5.2. *Como interpretar um intervalo de confiança de 95% para a média populacional?*

Exercício 1.5.3. *Uma amostra aleatória 9,8,12,7,9,6,11,6,10,9 foi extraída de uma população normal. Construir um I.C. para a média ao nível populacional com 95% de confiança. (Resp. $P(7,27 \leq \mu \leq 10,13) = 95\%$)*

Exercício 1.5.4.

Uma amostra de tamanho 9, extraída de uma população normal. Teve uma média, $\bar{X} = 1$ e $S = 0,264$. Construir um I.C. de 98 e 95% de confiança para a média populacional.

Exercício 1.5.5. *Uma amostra de tamanho 16, extraída de uma população normal. Teve uma $\bar{X} = 1,2$ e $S = 0,364$. Construir o IC de 97 e 95% de confiança para a média populacional.*

Exercício 1.5.6. *Dado a idade de 29 idosos. Encontre o I.C. para média populacional considerando que a idade dos idosos se distribui numa normal.*

78	99	87	89	69	87	89	87	68	87
81	82	92	93	99	87	87	94	99	89
96	89	97	98	94	91	89	89	87	

Tabela 1.4: Idade de 40 idosos de uma clínica

Exercício 1.5.7. *Dado a idade de 20 jovens que usa o mecanismo de defesa intelectualização. Encontre o I.C. para média populacional considerando que a idade dos jovens se distribui numa normal.*

23	25	18	21	24	23	24	25	23	21
24	30	25	21	25	28	26	24	28	26

Tabela 1.5: Idade dos 20 jovens

Exercício 1.5.8. *O tempo de reação de um novo medicamento contra ansiedade pode ser considerado como tendo distribuição normal. Pretende-se fazer inferência sobre a média que é desconhecida obtendo um intervalo de confiança para média populacional. Vinte pacientes foram sorteados (amostras aleatórias) e tiveram seu tempo de reação anotado. Os dados foram os seguintes (em minutos):*

2,9	3,4	3,5	4,1	4,6	4,7	4,5	3,8	5,3	4,9
4,8	5,7	5,8	5	3,4	5,9	6,3	4,6	5,5	6,2

Tabela 1.6: Tempo de reação de um novo medicamento contra ansiedade

Exercício 1.5.9. *Considere o grau de transtorno (suponha que varia de 1 a 20) de 100 jovens. Um histograma obtido tem forma aproximadamente Normal. A média amostral e o desvio padrão amostral foram $s = 74$ e $\bar{x} = 2,34$, respectivamente. Construir um intervalo de confiança para esse grau de transtorno médio populacional. Adote $\alpha = 5\%$. Resp. $IC = [73,54; 74,4586]$.*

1.5.1 Região de confiança para a média μ e a variância σ^2 de uma normal

Seja uma população $N(\mu; \sigma)$ e retira-se uma amostra aleatória simples de tamanho n e considera-se os estimadores de máxima verossimilhança: $\mu = \bar{x}$ e $\hat{\sigma}^2 = s^2$. A região de confiança é:

$$p[(\mu; \sigma^2) \in R(\bar{x}; s^2)] = 1 - \alpha$$

Pelo lema de Fisher-Cochran as estatísticas $\bar{X}$ e nS^2 são variáveis aleatórias independentes, usando essa circunstância determina-se a região de confiança:

$$P[(\mu; \sigma^2) \in R(\bar{X}; s^2)] = 1 - \alpha = P[\mu \in R_1(\bar{X})]P[\sigma^2 \in R_2(nS^2)]$$

$$P\left(\bar{X} - C \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + C \frac{\sigma}{\sqrt{n}}\right)$$

$$P\left(\frac{nS^2}{\sqrt{C_2}} \leq \sigma^2 \leq \frac{nS^2}{C_1}\right)$$

Em que R_1 e R_2 representam os correspondentes I.C. para os parâmetros μ e σ^2 , respectivamente.

Para determinar as constantes tabuladas C , C_1 e C_2 se considera o seguinte critério:

$$P\left(\bar{X} - C \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + C \frac{\sigma}{\sqrt{n}}\right) = \sqrt{1 - \alpha}$$

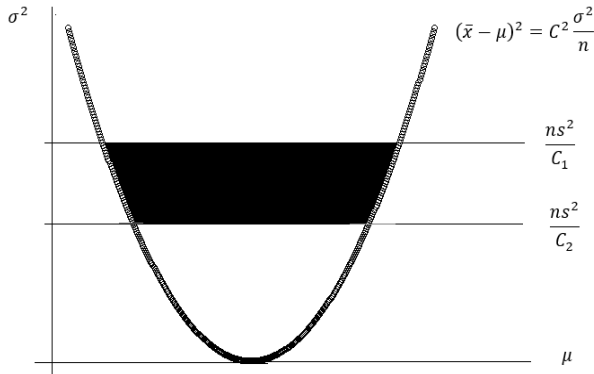
$$P\left(\frac{nS^2}{C_2} \leq \sigma^2 \leq \frac{nS^2}{C_1}\right) = \sqrt{1 - \alpha}$$

Sendo:

$$P[-C \leq N(0; 1) \leq C] = \sqrt{1 - \alpha}$$

$$P[C_1 \leq \chi^2(n-1) \leq C_2] = \sqrt{1 - \alpha}$$

A região de confiança obtida graficamente é:

Figura 1.7: Região de confiança entre C_1 e C_2

1.6 Intervalo de confiança para a variância

O I.C. para a variância de uma população normal $(X_1, \dots, X_n)$ uma variável aleatória de uma população $N(\mu, \sigma^2)$ cuja variância σ^2 se pretende estimar. Sabendo que a variância amostral é definida por $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. Sabe-se ainda que $Q = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$, sendo χ_{n-1}^2 a distribuição qui-quadrada de $(n-1)$ graus de liberdade.

O I.C. a um nível de confiança de $(1 - \alpha)100\%$ será:

$[LI, LS] = [C_1, C_2]$ sendo necessário calcular o valor de C_1 e C_2 .
Então:

$$P\left(\frac{C_1}{S^2} \leq \frac{\sigma^2}{S^2} \leq \frac{C_2}{S^2}\right) = 1 - \alpha$$

$$P\left(\frac{(n-1)S^2}{C_2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \frac{(n-1)S^2}{C_1}\right) = 1 - \alpha$$

Sabendo que $Q = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$, temos:

$$P(\chi_{1-\frac{\alpha}{2}}^2 \leq Q \leq \chi_{\frac{\alpha}{2}}^2) = 1 - \alpha$$

Assim, temos que: $P(Q < \chi_{1-\frac{\alpha}{2}}^2) = \frac{\alpha}{2}$ e $P(Q \leq \chi_{\frac{\alpha}{2}}^2) = 1 - \frac{\alpha}{2}$

$$\frac{(n-1)S^2}{C_2} = \chi_{1-\frac{\alpha}{2}}^2 \quad \frac{(n-1)S^2}{C_1} = \chi_{\frac{\alpha}{2}}^2$$

Então os valores de C_2 e C_1 serão:

$$C_2 = \frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2} \quad C_1 = \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2}$$

O intervalo de confiança no nível $1 - \alpha$ para a variância de uma população normal com média desconhecida será:

$$\left[\frac{(n-1)S^2}{\chi_{(n-1, 1-\frac{\alpha}{2})}^2}; \frac{(n-1)S^2}{\chi_{(n-1, \frac{\alpha}{2})}^2} \right]$$

Para facilitar a busca dos valores tabelados para I.C. para a variância, vamos considerar o seguinte I.C. para variância.

$$P\left(\frac{(n-1)S^2}{\chi_{sup}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{inf}^2}\right) = 1 - \alpha$$

A amplitude do intervalo é:

$$A = LS - PI = C_2 - C_1 = \frac{(n-1)S^2}{\chi_{(n-1, 1-\frac{\alpha}{2})}^2} - \frac{(n-1)S^2}{\chi_{(n-1, \frac{\alpha}{2})}^2}$$

$$A = \frac{(n-1)S^2(\chi_{\alpha/2}^2 - \chi_{1-\alpha/2}^2)}{\chi_{1-\alpha}^2 \chi_{\frac{\alpha}{2}}^2}$$

1.6.1 Intervalo unilateral à esquerda a um nível de confiança de $(1 - \alpha)100\%$

O intervalo é $[0, LS] = [0, C]$. Sendo LS o limte superior de confiança. Encontra-se o valor de C da seguinte forma:

$$P\left(\frac{\sigma^2}{s^2} \leq \frac{C}{S^2}\right) = 1 - \alpha$$

$$P\left(\frac{(n-1)S^2}{\sigma^2} \geq \frac{(n-1)S^2}{C}\right) = 1 - \alpha$$

$$P\left(Q \geq \frac{(n-1)S^2}{C}\right) = 1 - \alpha$$

$$P\left(Q < \frac{(n-1)S^2}{C}\right) = \alpha \Leftrightarrow P(Q < \chi_{1-\alpha}^2) = \alpha$$

$$\left(\frac{(n-1)S^2}{C}\right) = \chi_{1-\alpha}^2 \Leftrightarrow C = \frac{(n-1)S^2}{\chi_{1-\alpha}^2}$$

Sendo que $\chi_{1-\alpha}^2$ é o quantil apropriado da distribuição χ_{n-1}^2 ou seja, $P(Q < \chi_{1-\alpha}^2) = \alpha$. Logo, obtem-se o I.C. pretendido:

$$\left[0, \frac{(n-1)S^2}{\chi_{1-\alpha}^2}\right]$$

1.6.2 Intervalo de confiança unilateral à direita a um nível de confiança de $(1 - \alpha)100\%$

O intervalo é $[LI, \infty] = [C, \infty]$. Sendo LI o limte inferior de confiança. Vamos encontrar o valor de C.

$$P\left(\frac{s^2}{\sigma^2} \leq \frac{S^2}{C}\right) = 1 - \alpha$$

$$P\left(\frac{(n-1)S^2}{\sigma^2} \leq \frac{(n-1)S^2}{C}\right) = 1 - \alpha$$

$$P\left(Q \leq \frac{(n-1)S^2}{C}\right) = 1 - \alpha$$

$$P(Q \leq \chi_\alpha^2) = 1 - \alpha$$

$$\frac{(n-1)S^2}{C} = \chi_\alpha^2 \Leftrightarrow C = \frac{(n-1)S^2}{\chi_\alpha^2}$$

Sendo que χ_α^2 é o quantil apropriado da distribuição χ_{n-1}^2 ou seja, $P(Q > \chi_\alpha^2) = \alpha \Leftrightarrow P(Q \leq \chi_\alpha^2) = 1 - \alpha$. Logo, obtem-se o I.C. pretendido:

$$\left[\frac{(n-1)S^2}{\chi_\alpha^2}, \infty \right[$$

Resumindo do I.C. para a variância será:

- I.C. para σ^2 com média populacional μ conhecida:

$$I.C. = \left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n;\alpha/2}^2}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n;1-\alpha/2}^2} \right]$$

O valor de $\chi_{n;\alpha} =$ abscissa que sob uma χ_n^2 deixa à sua direita uma área α .

Exemplo 1.6.1. *Suponha que uma psicóloga coletou os seguintes números:*

3,02 2,99 2,99 3,05 3,00 3,01
3,01 2,99 2,97 2,97 3,02 2,98

Tabela 1.7: Uma amostra de 10 níveis de ansiedade

Sabe-se que a média populacional, $\mu = 3$, coletou-se 12 indivíduos que tem ansiedade aleatoriamente e que se modela mediante uma variável aleatória Normal. Considere o nível de confiança do intervalo $1 - \alpha = 0,99$, onde $\chi_{12}^{1\%/2} = \chi_{12}^{0,005} = 28,3$ (observando o valor na tabela χ^2 com 12 graus de liberdade a 0,005, e $\chi_{12}^{1-1\%/2} = \chi_{12}^{1-0,995} = 3,074$ (observando o valor na tabela χ^2 com 12 graus de liberdade a 0,995.

$$I.C. = \left[\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n;\alpha/2}^2}; \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{n;1-\alpha/2}^2} \right] = \left[\frac{\sum_{i=1}^{12} (X_i - 3)^2}{28,3}; \frac{\sum_{i=1}^{12} (X_i - 3)^2}{3,074} \right]$$

$$I.C. = \left[\sqrt{\frac{0,006}{28,3}}; \sqrt{\frac{0,006}{3,074}} \right] = [0,0146; 0,0442]$$

Pode-se dizer então que, em 99% dos casos, esse I.C. contém o verdadeiro valor de σ .

- Intervalo para σ^2 com média populacional μ desconhecida:

$$I.C. = \left[\frac{(n-1)S^2}{\chi_{(n-1;1-\alpha/2)}^2}; \frac{(n-1)S^2}{\chi_{(n-1;\alpha/2)}^2} \right]$$

Em termos da variância S^2 , tem-se:

$$I.C. = \left[\frac{nS^2}{\chi_{(n-1; 1-\alpha/2)}}; \frac{nS^2}{\chi_{(n-1; \alpha/2)}} \right]$$

Exemplo 1.6.2. Considere o caso de $n = 12$ pessoas com distúrbios de desenvolvimento selecionadas aleatoriamente, mostra uma variância amostral de 0,22. Construir um I.C. para a variância a um nível de confiança de 95%.

Exemplo 1.6.3. Sabe-se que o peso por comprimido de um certo medicamento se distribui segundo uma normal. Com objetivo de estudar a variância da distribuição, extraem-se uma amostra aleatória simples de tamanho 6. Sabendo que a variância amostral é 40, pretende-se estimar a variância populacional mediante um intervalo de confiança de 95%.

Como μ é desconhecido, um I.C. para σ^2 é dado por:

$$I.C_{1-\alpha}(\sigma^2) = \left[\frac{nS^2}{\chi_{(n-1; 1-\alpha/2)}^2}; \frac{nS^2}{\chi_{(n-1; \alpha/2)}^2} \right]$$

Aqui, $\alpha = 10\%$, $n = 6$, $S^2 = 40$. Assim, $\chi_{5;0,95}^2 = 11,07$ e $\chi_{5;0,05}^2 = 1,145$. Assim, o I.C. será:

$$I.C_{0,90}(\sigma^2) = \left[\frac{6 \cdot 40}{11,07}; \frac{6 \cdot 40}{1,145} \right] = [21,68; 209,61]$$

Exemplo 1.6.4. Vamos admitir uma distribuição normal. A idade de 10 pacientes, com variância $S^2 = 25$. Pede-se construir um IC para a variância populacional ao nível de 90%.

Temos: $n = 10$, $S^2 = 5$, $\varphi = (n - 1) = (10 - 1) = 9$ e com nível $(1 - \alpha)\% = 90\%$. Consultando a tabela $\chi_{gl=n-1}^2$ temos:

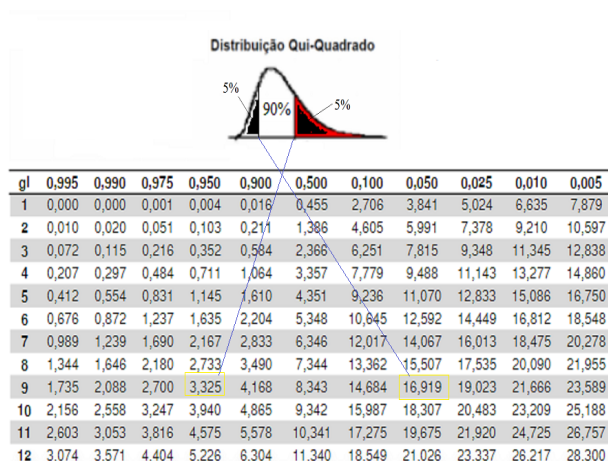


Figura 1.8: Distribuição qui-quadrada com $n-1$ grau de liberdade

$$P\left(\frac{(10-1)5}{16,91} \leq \sigma^2 \leq \frac{(10-1)4}{3,325}\right) = 1 - 10\%$$

$$P(2,66 \leq \sigma^2 \leq 13,5113) = 90\%$$

Assim, podemos afirmar que o intervalo $[2,66; 13,51]$ contém a variância populacional com 90% de confiança.

Exercício 1.6.1. Vamos admitir uma distribuição normal. A idade de 9 idosos com inteligência de parâmetros (refere-se aos ritmos e rotinas da vida, a capacidade de proteger determinadas áreas da vida como forma de preservação de ordem, segurança e bem estar pessoal) tem variância $S^2 = 36$. Pede-se construir um IC para a variância populacional ao nível de 95%.

Exercício 1.6.2. Vamos admitir uma distribuição normal. A idade de $n = 12$ jovens com inteligência interpessoal tem variância $S^2 =$

16. *Pede-se construir um IC para a variância populacional ao nível de 95%.*

Exemplo 1.6.5. *Sabe-se que o peso por comprimido de um certo medicamento se distribui segundo uma normal. Com objetivo de estudar a variância da distribuição, extraem-se uma amostra aleatória simples de tamanho 10. Sabendo que a variância amostral é 49, pretende-se estimar a variância populacional mediante um intervalo de confiança de 90%.*

1.7 Intervalo de confiança para o desvio padrão

Admitindo que a distribuição de probabilidade populacional de onde se extraiu a amostra seja normal, um intervalo de confiança aproximado para o desvio padrão, σ , é dado pela raiz quadrada do I.C. para variância populacional, σ^2 . Assim, o I.C. para o desvio padrão será:

$$P\left(\sqrt{\frac{(n-1)s^2}{\chi_{sup}^2}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi_{inf}^2}}\right) = 1 - \alpha$$

Em que a distribuição qui-quadrada é tomada com $n - 1$ graus de liberdade.

No exemplo anterior vimos que o I.C. para variância populacional foi $[2,66; 13,51]$.

Assim, o I.C. para o desvio padrão populacional será:

$$P(\sqrt{2,66} \leq \sigma \leq \sqrt{13,51}) = 90\%$$

Podemos concluir que esse intervalo contém o desvio padrão populacional com 90% de confiança.

Exemplo 1.7.1. *Sabe-se que o peso por comprimido de um certo medicamento se distribui segundo uma normal. Com objetivo de estudar a variância da distribuição, extraem-se uma amostra aleatória simples de tamanho 10. Sabendo que a variância amostral é 49, pretende-se estimar o desvio padrão populacional mediante um intervalo de confiança de 90%.*

1.8 Intervalos de confiança para duas populações normais

Suponha que X seja uma população Normal com média μ_X e desvio padrão σ_X , ou seja, $X \sim N(\mu_X, \sigma_X)$. Retira-se uma amostra aleatória simples $X_1, \dots, X_{n_1}$ e Y outra população normal com média μ_Y e desvio padrão σ_Y , tal que $Y \sim N(\mu_Y, \sigma_Y)$. Considere uma amostra aleatória de tamanho n_1 da variável aleatória X , $X_1, \dots, X_{n_1}$, e uma amostra aleatória de tamanho n_2 da variável aleatória Y , $Y_1, \dots, Y_{n_2}$. Denota-se $\bar{X}$, $\bar{Y}$, S_1^2 e S_2^2 as correspondentes médias e variâncias amostrais para as duas amostras.

Como $\bar{X} \sim N\left(\mu_X, \frac{\sigma_X}{\sqrt{n_1}}\right)$ e $\bar{Y} \sim N\left(\mu_Y, \frac{\sigma_Y}{\sqrt{n_2}}\right)$ são independentes, sabe-se que:

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y; \frac{\sigma_X}{\sqrt{n_1}} + \frac{\sigma_Y}{\sqrt{n_2}}\right)$$

Usando que $S_{cX}^2 \sim \chi_{n_1-1}^2$ e $S_{cY}^2 \sim \chi_{n_2-1}^2$. Assim,

$$(n_1 - 1) \frac{S_{cX}^2}{\sigma_X^2} + (n_2 - 1) \frac{S_{cY}^2}{\sigma_Y^2} \sim \chi_{n_1+n_2-2}^2$$

Sendo σ_X^2 e σ_Y^2 as variâncias amostrais para as amostras X e Y independentes.

1.8.1 Intervalos de confiança para diferença de médias

Com variâncias σ_X e σ_Y conhecidas

A distribuição da estatística da diferença de médias com variâncias conhecidas será:

$$\bar{X} - \bar{Y} \sim N\left(\mu_X - \mu_Y; \frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}\right)$$

Com o qual se pode tomar como pivô:

$$\frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}}} \sim N(0, 1)$$

E agindo da mesma forma que nos intervalos para a média de um população, chegamos ao intervalo de confiança para o diferença nas médias de uma população normal no nível $1 - \alpha$. Essa diferença é a seguinte:

$$\left[\bar{X} - \bar{Y} \pm Z_{1-\alpha/2} \sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}} \right]$$

Sendo, $\delta = \mu_1 - \mu_2$ tem-se:

$$P \left[(\bar{X} - \bar{Y}) - Z_{(1-\alpha/2)} \sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}} < \delta < (\bar{X} - \bar{Y}) + Z_{(1-\alpha/2)} \sqrt{\frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}} \right]$$

O centro do intervalo é $(\bar{X} - \bar{Y})$ e seu raio será: $r = Z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

O tamanho amostral necessário para alcançar uma cota de erro com um nível de confiança dado será:

$$n = \left(\frac{Z_{1-\alpha/2}}{r} \right)^2 (\sigma_1^2 + \sigma_2^2)$$

O I.C. cobre a diferença $(\mu_1 - \mu_2)$ com probabilidade $1 - \alpha$ e pode partir-se do mesmo para passar a intervalos de confiança a $(1 - \alpha)100\%$ para essa diferença.

Exemplo 1.8.1. *Seja $\bar{X} = 1400$, $\bar{Y} = 1200$, $Z_{1-\alpha/2} = 1,96$ (intervalo de confiança de 95%), $\sigma_X = 120$, $\sigma_Y = 80$, $n_1 = 150$ e $n_2 = 200$. Encontre um I.C. para diferença de médias quando as variâncias populacionais são conhecidas.*

$$\left[1400 - 1200 \pm 1,96 \sqrt{\frac{120^2}{150} + \frac{80^2}{200}} \right] = [177,8; 222,2]$$

Exemplo 1.8.2. *Suponha que o tempo de recuperação de uma crise de ansiedade por um método é uma variável aleatória aproximadamente normal com média μ_1 e variância 49. O tempo de recuperação usando um método alternativo é uma variável aleatória que também pode ser considerada normal com média μ_2 e variância 64. Obtenha o I.C. para a diferença de média. Suponha que o tamanho da amostra seja 194.*

O tamanho amostral necessário para alcançar um erro de 1,5 minutos (no tempo de recuperação) com um nível de confiança de 0,95 será:

$$n = \left(\frac{Z_{\alpha/2}}{r} \right)^2 (\sigma_1^2 + \sigma_2^2) = \left(\frac{1,96}{1,5} \right)^2 (49 + 64) = 192,9$$

Suponha que o tamanho da amostra seja 193 e se obteve $\bar{X} = 25$ e $\bar{Y} = 29$. O I.C. será então:

$$P \left[(-4) - 1,96\sqrt{\frac{49}{194} + \frac{64}{194}} < \delta < (-4) + 1,96\sqrt{\frac{49}{194} + \frac{64}{194}} \right] = 0,95$$

Ou seja, $-5,5 < \mu_1 - \mu = \delta_2 < -2,5$. Observa-se que a diferença entre as duas médias populacionais, δ , com 95% de confiança, admite apenas valores negativos, $\mu_1 - \mu_2 < 0$. Por tanto, esse resultado nos permite afirmar, com um nível de confiança de 0,95, que o primeiro método tem melhor tempo médio (menor média).

Exercício 1.8.1. *Suponha que o tempo de recuperação de uma depressão por um método é uma variável aleatória aproximadamente normal com média μ_1 e variância 64. O tempo de recuperação usando um método alternativo é uma variável aleatória que também pode ser considerada normal com média μ_2 e variância 81. Obtenha o I.C. para a diferença de média. Suponha que o tamanho da amostra seja 225.*

Com variâncias σ_X e σ_Y desconhecidas e iguais

Como S_{cX}^2 e S_{cY}^2 são independentes, sabe-se que:

$$(n_1 - 1) \frac{S_{cX}^2}{\sigma^2} \sim \chi_{n_1-1}^2$$

$$(n_2 - 1) \frac{S_{cY}^2}{\sigma^2} \sim \chi_{n_2-1}^2$$

Logo,

$$\frac{(n_1 - 1)S_{cX}^2 + (n_2 - 1)S_{cY}^2}{\sigma^2} \sim \chi_{n_1+n_2-2}^2$$

Sabendo também que $\bar{X}$ e S_{cX} assim como, $\bar{Y}$ e S_{cY} são independentes, tem-se como pivô:

$$T(X, Y, \mu_1 - \mu_2) = \frac{\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma}}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \sqrt{\frac{(n_1 - 1)S_{cX}^2 + (n_2 - 1)S_{cY}^2}{n_1 + n_2 - 2}}}}$$

Que é igual a:

$$T(X, Y, \mu_1 - \mu_2) = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)S_{cX}^2 + (n_2 - 1)S_{cY}^2}{n_1 + n_2 - 2}}} \left(\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right)^{-1} \sim t_{n_1 + n_2 - 2}$$

Assim, obtém um $IC_{(\mu_1 - \mu_2)}$ para diferença de médias populacionais com variância desconhecida e iguais será:

$$\left[\bar{X} - \bar{Y} \pm t_{(n_1 + n_2 - 2; 1 - \alpha/2)} \sqrt{\frac{(n_1 - 1)S_{cX}^2 + (n_2 - 1)S_{cY}^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right]$$

Expressando em função da variância amostral fica:

$$\left[\bar{X} - \bar{Y} \pm t_{(n_1 + n_2 - 2; 1 - \alpha/2)} \sqrt{\frac{n_1 S_{cX}^2 + n_2 S_{cY}^2}{(n_1 + n_2 - 2)} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right]$$

Exemplo 1.8.3. *Uma população de pessoas que tem uma boa conduta social num clube social que se distribui numa normal $N(\mu_X, \sigma^2)$. Extrai-se uma amostra aleatória simples de tamanho $n_1 = 10$, tal que a média amostral seja 4,1 e a variância amostral seja 6,09. De outra população de outro clube social que tem uma distribuição*

$N(\mu_Y, \sigma^2)$ extra-se uma amostra aleatória simples de tamanho 16 ($n_2 = 16$) e independente do clube anterior, cuja média e variância amostrais são respectivamente 3,87 e 3,60. Pede-se um I.C. de 95% para a diferença de médias populacionais com variâncias iguais.

Para $\alpha = 5\%$ tem-se que $t_{(10+16-2; 1-5\%/2)} = t_{(24; 0,975)} = 2,064$ (Tabela t de Student). Como o I.C. está em função das variâncias amostrais tem-se:

$$\left[4,1 - 3,87 \pm 2,064 \sqrt{\frac{10.6,09 + 16.3,609}{24}} \left(\frac{1}{10} + \frac{1}{16} \right) \right]$$

Portanto, $I.C_{0,95}(\mu_1 - \mu_2) = [-1,62; 2,07]$.

Exemplo 1.8.4. Uma população de pessoas que tem inteligência intuitiva num departamento de uma grande universidade tem uma distribuição normal $N(\mu_X, \sigma^2)$. Extraí-se uma amostra aleatória simples de tamanho $n_1 = 12$, tal que a média amostral seja 6 e a variância amostral seja 5,76. De outra população de outro clube social que tem uma distribuição $N(\mu_Y, \sigma^2)$ extra-se uma amostra aleatória simples de tamanho 20 ($n_2 = 18$) e independente do clube anterior, cuja média e variância amostrais são respectivamente 4,81 e 3,72. Pede-se um I.C. de 90% para a diferença de médias populacionais com variâncias iguais.

1.8.2 Intervalo de confiança para quando as variâncias são desconhecidas e distintas

Sendo $\bar{X}_1$ e S_1^2 , assim como $\bar{X}_2$ e S_2^2 independentes, tem-se que:

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

Assim, como

$$(n_1 - 1) \frac{S_1^2}{\sigma_1^2} + (n_2 - 1) \frac{S_2^2}{\sigma_2^2} \sim \chi_{n_1+n_2-2}^2$$

Pela teoria sabemos que: $\frac{Z}{\chi^2} = t$. Portanto,

$$\frac{\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}}{(n_1 - 1) \frac{S_1^2}{\sigma_1^2} + (n_2 - 1) \frac{S_2^2}{\sigma_2^2}} \sim t_{n_1+n_2-2}$$

Como a estatística depende de σ_1 e σ_2 então recorreremos à aproximação de Welch, em função da qual a estatística:

$$\frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Tem uma distribuição aproximada $t_{(b; 1-\frac{\alpha}{2})}$, sendo b um fator de correção que se calcula considerando o interio mais próximo a:

$$b = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1+1} \left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2+1} \left(\frac{S_2^2}{n_2}\right)^2} - 2$$

O I.C. para $\mu_1 - \mu_2$ vem dado por:

$$I.C_{1-\alpha}(\mu_1 - \mu_2) = \left[\bar{X}_1 - \bar{X}_2 \pm t_{b; 1-\frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right]$$

Neste caso, considera-se o seguinte I.C. para essa diferença de média:

$$I.C. = \left[\bar{X} - \bar{Y} \pm t_{(b; 1 - \frac{\alpha}{2})} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right]$$

Exemplo 1.8.5. *Para realizar um estudo sobre hipertensão em uma pessoa estressada e suas consequências, tomaram-se duas amostras de 13 e 16 pacientes de cidades distintas. Os resultados foram os seguintes: $\bar{X}_1 = 166\text{mm}$, $S_1 = 28\text{mm}$, $\bar{X}_2 = 164,7\text{mm}$ e $S_2 = 7\text{mm}$. Supondo que ambas populações sejam normais e que suas variâncias sejam desconhecidas e distintas. Pede-se um I.C. de 95% de confiança para a diferença de médias.*

1.9 Intervalo de confiança para a diferença de média em amostras emparelhadas

O teste t com amostras emparelhadas (dependente) tenta determinar se existe uma diferença entre as médias de duas variáveis emparelhadas. Uma pergunta relacionada e normalmente mais informativa é: Qual é o tamanho da diferença entre as médias das variáveis emparelhadas?. Essa, diferença pode ser estimada com um I.C..

Esse método se aplica diretamente ao problema de avaliação da diferença média de escores de uma população.

A magnitude dessa diferença média pode ser estimada pela construção do I.C. apropriado nas diferenças de escores da amostra. Ao criar esses intervalos, os pesquisadores em psicologia podem abordar perguntas como Quanto a média mudou? ou qual é o tamanho da diferença entre os resultados médios dos dois grupos de tratamento?. Assim, em vez de fazer a pergunta do tipo um medica-

mento cria uma resposta média maior do que a outra?, como seria feito com um teste t de amostras emparelhadas, o pesquisador pode abordar a questão a resposta média produzida por um medicamento é muito maior do que aquela produzida por outro medicamento?, por meio de um intervalo de confiança.

Seja $X = (X_1, \dots, X_n)$ uma amostra aleatória de uma variável aleatória $X \sim N(\mu_1, \sigma_1^2)$ e $Y = (Y_1, \dots, Y_n)$ uma amostra aleatória de $Y \sim N(\mu_2, \sigma_2^2)$. Então, $d = (X_1 - Y_1, \dots, X_n - Y_n) = (d_1, \dots, d_n)$ é uma amostra aleatória $N(\mu_1 - \mu_2, \sigma^2)$ com σ^2 desconhecida. O I.C. para $\mu_1 - \mu_2$ a um nível de $1 - \alpha$ é:

$$I.C. = \left[\bar{d} - t_{n-1, \alpha/2} \frac{S_d}{\sqrt{n}}; \bar{d} + t_{n-1, \alpha/2} \frac{S_d}{\sqrt{n}} \right]$$

Sendo que $t_{n-1; \alpha}$ é igual a abscissa que sob uma t_{n-1} deixa a sua direita uma área α . Observa-se, como sempre, que para construir um I.C. considera-se:

$$L_i = \bar{d} - t \frac{S_d}{\sqrt{n}} \quad (1.1)$$

$$L_s = \bar{d} + t \frac{S_d}{\sqrt{n}} \quad (1.2)$$

Em que s_d é o desvio padrão amostral da diferença dos escores, n é o número tamanho amostral e t é o valor apropriado observado na tabela t de Student com $n-1$ graus de liberdade.

Exemplo 1.9.1. *Considere as medições sanguíneas sistólicas pré-tratamento e pós-tratamento de um problema de ansiedade.*

Tabela 1.8:

Pré-tratamento	Pós-tratamento	diferenças	d^2
217	209	8	$8^2 = 64$
252	241	11	$11^2 = 121$
229	230	1	$1^2 = 1$
200	208	8	$8^2 = 64$
209	206	3	$3^2 = 9$
213	211	2	$2^2 = 4$

No R seria:

```
A<-c(217,252,229,200,209,213)
D<-c(209,241,230,208,206,211)
d<-A-D; d
n<-length(A)
md<-mean(d); md
s<-sd(d); s
d^2
t<-md/(s/sqrt(6));t
Li<- md-qt(0.95,n-1)*s/sqrt(n); Li
Ls<- md+qt(0.95,n-1)*s/sqrt(n); Ls
```

O I.C. seria:

$$\begin{aligned}
 I.C. &= [L_i; L_s] \\
 &= \left[\bar{d} - t \frac{s_d}{\sqrt{n}}; \bar{d} + t \frac{s_d}{\sqrt{n}} \right] \\
 &= [-3.0245; 8.0245]
 \end{aligned} \tag{1.3}$$

Uma interpretação estatística desse intervalo manteria que podemos ter 95% de confiança que μ_d é maior ou igual a -3,0245 e menor ou igual a 8,0245.

Exercício 1.9.1. *Os dados da tabela abaixo representa a pressão arterial sistólica de 10 alunos com dificuldades educativas. Foram extraídas amostras da pressão arterial no início do semestre e após o semestre (antes e depois). Pede-se o intervalo de confiança para a diferença de média nessas amostras emparelhadas (existe uma dependência entre o antes e depois). Considere $\alpha = 5\%$.*

Ind.	1	2	3	4	5	6	7	8	9	10
Antes	140	171	160	148	175	189	171	170	154	160
Depois	143	156	150	148	173	175	169	169	145	168
Dif.										

Tabela 1.9: Uma amostra de 10 indivíduos com dificuldades educativas

Exercício 1.9.2. *Considere que os dados da tabela abaixo representa a pressão arterial sistólica de 12 alcoólatras antes e depois de meses de abstinência do consumo de álcool. Pede-se o I.C. com um risco de 5%.*

Ind.	1	2	3	4	5	6	7	8	9	10	11	12
Antes	144	168	162	170	177	189	170	174	152	159	178	178
Depois	148	152	151	170	170	175	164	166	145	170	171	188
Dif.												

Tabela 1.10: Uma amostra de 12 indivíduos com abstinência do consumo de álcool

Basicamente no SPSS seria:

1. Criamos duas variáveis, chamadas janeiro e fevereiro, que contêm os dados

2. Depois de prepararmos os dados, selecionamos: Analisar - Comparar médias - Teste T para amostras relacionadas
3. Na caixa de diálogo, as duas variáveis são selecionadas passando para a primeira linha da caixa variável emparelhadas
4. Clique OK. Deve-se obter a saída desejada.

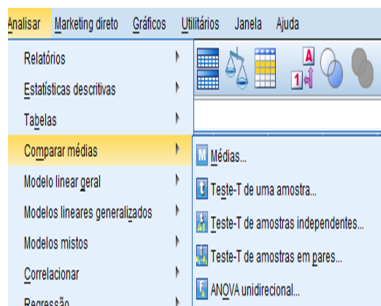


Figura 1.9: Passos para encontrar I.C. em teste paramétricos

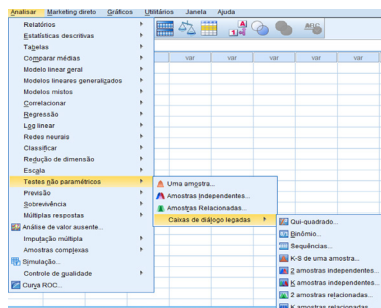


Figura 1.10: Passos para encontrar I.C. em teste não paramétricos

1.10 Intervalo para os parâmetros de duas populações normais independentes

Na seção anterior vimos o caso para duas população dependentes (dados emparelhados). Aqui, veremos o caso para duas populações normais independentes.

A base na médias amostrais e, mais concretamente, na diferença $\bar{X} - \bar{Y}$. O objetivo é conhecer sua distribuição ou alguma função da diferença.

Dado $X_1, \dots, X_{n_1}$ uma amostra aleatória. Sendo $X \sim N(\mu_1, \sigma_1^2)$ e $Y_1, \dots, Y_{n_2}$ uma amostra aleatória de uma variável aleatória $Y \sim N(\mu_2, \sigma_2^2)$, sendo as duas populações independentes. Sendo $\bar{X}, \bar{Y}, S_1^2, S_2^2$ as correspondentes médias e variâncias amostrais para as duas populações. Temos que:

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1) \sim N(0, 1)$$

Para o caso das variâncias serem conhecidas. Se as variâncias são desconhecidas e iguais se considera a seguinte estatística:

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{n_1 + n_2}{n_1 n_2}}} \sim t_{n_1 + n_2 - 2}$$

Sendo S_p , a variância conjunta. A fórmula para encontrar S_p é:

$$S_p^2 = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

Se as variâncias são desconhecidas e distintas a fórmula é:

$$T = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{gl}$$

Sendo gl (grau de liberdade) um número inteiro mais próximo a:

$$gl = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1+1} \left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2+1} \left(\frac{S_2^2}{n_2}\right)^2} - 2$$

- I.C. para $\mu_1 - \mu_2$ com σ_1^2 e σ_2^2 conhecidos é:

$$\left[\bar{X} - \bar{Y} - z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}; \bar{X} - \bar{Y} + z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

z_α é a abscissa que sob uma $N(0, 1)$ deixa a sua direita uma área α .

Exemplo 1.10.1. *Em duas turmas do primeiro período de duas faculdades deseja-se testar a diferença do nível de estresse. A primeira turma teve $\sigma_1 = 1$ e a segunda, $\sigma_2 = 1,5$. Testam-se duas amostras de tamanhos $n_1 = 10$ e $n_2 = 12$, com média $\bar{x}_1 = 87,6$ e $\bar{x}_2 = 74,5$. Assumindo distribuições normais, construa um I.C. para a diferença de médias do nível de estresse ao nível de 90%.*

O nível de confiança do intervalo é $1 - \alpha = 0,90$, logo $z_{\alpha/2} = z_{0,10/2} = z_{0,05} = 1,654$.

$$I.C. = \left[87,6 - 74,5 \pm 1,654 \sqrt{\frac{1^2}{10} + \frac{1,5^2}{12}} \right] \approx [12,2; 13,9]$$

- I.C. para $\mu_1 - \mu_2$ com σ_1^2 e σ_2^2 desconhecidos é:

$$\left[\bar{X} - \bar{Y} \pm t_{(n_1+n_2-2; \alpha/2)} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

$t_{(n_1+n_2; \alpha)}$ é a abscissa que sob uma $t_{(n_1+n_2)}$ deixa a sua direita uma área α .

Exemplo 1.10.2. *Seleciona-se duas amostras aleatórias de tamanhos $n_1 = 25$ e $n_2 = 27$ com variâncias $s_1^2 = 0,42$ e $s_2^2 = 0,51$ e médias $\bar{x}_1 = 9,83$ e $\bar{x}_2 = 8,75$. Assumindo que os dados foram extraídos de distribuições normais com $\sigma_1^2 = \sigma_2^2$. Assumindo distribuições normais, construa um I.C. para a diferença de médias do nível de estresse ao nível de 95%. O nível de confiança do intervalo é $1 - \alpha = 0,95$ de onde temos:*

Pode-se calcular encontrar o valor de t da seguinte forma:

$t_{n_1+n_2-2; \alpha/2} = t_{25+27-2; 5\%/2} = t_{50; 0,05/2} = t_{50; 0,025} = 2,009$ (Veja na tabela t de Student)

Vamos calcular a variância conjunta: S_p e posteriormente o I.C. para $\mu_1 - \mu_2$ ao nível de 95%.

$$S_p^2 = \sqrt{\frac{(24)0,42^2 + (26)0,51^2}{25 + 27 - 2}} = \sqrt{\frac{10,99}{50}} = 0,468$$

$$\left[9,83 - 8,75 \pm 2,0090,468 \sqrt{\frac{1}{25} + \frac{1}{27}} \right] = [0,69; 1,46]$$

No intervalo $\mu_1 - \mu_2 \in [0,69; 1,46]$ contém a verdadeira diferença de média populacional.

$$b = \frac{\left(\sqrt{28^2}13 + \frac{7^2}{16}\right)^2}{\frac{1}{13+1}\left(\frac{28^2}{13}\right)^2 + \frac{1}{16+1}\left(\frac{7^2}{16}\right)^2} - 2 = 13,4$$

Logo, toma-se $b = 13$. O valor de t será: $t_{13;0,975} = 2,16$. Assim, o I.C. será:

$$\begin{aligned} I_{95\%}(\mu_1 - \mu_2) &= [166 - 164,7 \pm 2,16\sqrt{\frac{784}{13} + \frac{49}{16}}] \\ &= [-15,9; 18,5] \end{aligned}$$

Exercício 1.10.1. *Para realizar um estudo sobre hipertensão em jovens com Anorexia e Bulimia e suas consequências, tomaram-se duas amostras de 15 e 18 pacientes de cidades distintas. Os resultados foram os seguintes: $\bar{X}_1 = 170\text{mm}$, $S_1 = 35\text{mm}$, $\bar{X}_2 = 169,8\text{mm}$ e $S_2 = 8\text{mm}$. Supondo que ambas populações sejam normais e que suas variâncias sejam desconhecidas e distintas. Pede-se um I.C. de 90% de confiança para a diferença de médias.*

1.11 Intervalos de confiança para o quociente de variâncias

1.11.1 Com médias μ_x y μ_y desconhecidas

Muitas vezes em psicologia é necessário comparar a variância de duas populações. Por exemplo, um número de pessoas com dois tipos de aprendizagem com condicionamento operante: reforço positivo e reforço negativo. Interessa nesse quociente aquele tipo de aprendizagem que tem uma variabilidade pequena. A comparação de duas variâncias se faz mediante quociente e não com a diferença

como no caso da média.

Dado duas amostras independentes procedentes de duas populações normais $N(\mu_1; \sigma_1^2)$ e $N(\mu_2; \sigma_2^2)$, as variáveis aleatórias $\frac{(n-1)S_1^2}{\sigma_1^2}$ e $\frac{(m-1)S_2^2}{\sigma_2^2}$ tem distribuição χ^2 com $n-1$ e $m-1$ graus de liberdade respectivamente, sendo ambas independentes.

Por outra parte, a variável aleatória $F = \frac{\frac{u}{k_1}}{\frac{v}{k_2}}$ tem distribuição F de Snedecor com k_1 e k_2 graus de liberdade para o numerador e denominador, quando u e v são independentes e com distribuição χ^2 com k_1 e k_2 graus de liberdade, respectivamente. Como consequência:

$$F = \frac{S_1^2}{S_2^2} \frac{\sigma_2^2}{\sigma_1^2}$$

Dada duas amostras $X_1, \dots, X_n$ e $Y_1, \dots, Y_m$, procedentes de duas populações normais, independentes, com médias desconhecidas e variâncias σ_1^2 e σ_2^2 , também desconhecidas, a variável F , definida por:

$$\frac{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}{\frac{1}{m-1} \sum_{i=1}^m (Y_i - \bar{Y})^2} \frac{\sigma_2^2}{\sigma_1^2} = \frac{S_1^2}{S_2^2} \frac{\sigma_2^2}{\sigma_1^2}$$

Tem uma distribuição F de Snedecor com $n-1$ graus de liberdade para o numerador e $m-1$ graus de liberdade para o denominador, a inversa é também F com $m-1$ e $n-1$ graus de liberdade. Se F_α é o quantil $\alpha/2$ e F_1 é o quantil $1 - \alpha/2$ de dita distribuição, então, temos:

$$P\left(F_\alpha < \frac{S_1^2}{S_2^2} \frac{\sigma_2^2}{\sigma_1^2} < F_1\right) = 1 - \alpha$$

Assim , o I.C. será:

$$P\left(F_o \frac{S_2^2}{S_1^2} < \frac{\sigma_2^2}{\sigma_1^2} < F_1 \frac{S_2^2}{S_1^2}\right) = 1 - \alpha$$

Observa que a expressão acima (ou seu equivalente, resultante de um processo similar da variável inversa) obtem-se um intervalo de confiança da seguinte forma:

$$P\left(F_o' \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < F_1' \frac{S_1^2}{S_2^2}\right) = 1 - \alpha$$

Sendo $F_o' = \frac{1}{F_1}$ e $F_1' = \frac{1}{F_o}$, teremos então:

$$P\left(\frac{1}{F_1} \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1}{F_o} \frac{S_1^2}{S_2^2}\right) = 1 - \alpha$$

Observe que o centro e o raio do intervalo serão:

$$\frac{S_1^2}{S_2^2} \left(\frac{F_o + F_1}{2} \right)$$

e

$$\frac{S_1^2}{S_2^2} \left(\frac{F_o - F_1}{2} \right)$$

O centro é um estimador pontual da razão da variância que tende a coincidir com o quociente da variância amostral quando o tamanho de ambas amostras cresce indefinidamente, já que a distribuição F de Snedecor com $n - 1$ e $m - 1$ graus de liberdade tem media que converge a 1 e variância que converge a 0 quando m e n crescem simultaneamente, e como consequência os dois quantis F_o e F_1 tendem ambos ao valor 1. Portanto, é consistente, suficiente, assintoticamente eficiente e assintoticamente inviesado.

1.11.2 Com médias μ_x y μ_y conhecidas

Vamos fazer de forma diferente para encontrar o I.C. para o quociente de variância com as médias μ_x y μ_y conhecidas. Como X e Y seguem uma distribuição normal, $X \sim N(\mu_x, \sigma_x)$ e $Y \sim N(\mu_y, \sigma_y)$, respectivamente.

$$X \sim N(\mu_x, \sigma_x) \text{ e}$$

$$\sum_{i=1}^{n_1} \left(\frac{x_i - \mu_x}{\sigma_x} \right)^2 \sim \chi_{n_1}^2$$

$$Y \sim N(\mu_y, \sigma_y) \text{ e}$$

$$\sum_{i=1}^{n_2} \left(\frac{y_i - \mu_y}{\sigma_y} \right)^2 \sim \chi_{n_2}^2$$

O quociente de duas distribuições χ é uma distribuição F de Sneedecor com $n - 1$ e $m - 1$ graus de liberdade. Assim, teremos:

$$F = \frac{\frac{Q_1}{n_1-1}}{\frac{Q_2}{n_2-1}} = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F_{[(n_1-1), (n_2-1)]}$$

O I.C. a um nível de confiança de $(1 - \alpha)100\%$ será:

$[LI; LS] = [C_1, C_2]$ sendo necessário calcular o valor do limite inferior C_1 e do limite superior C_2 .

$$P \left(C_1 \frac{S_2^2}{S_1^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \frac{S_2^2}{S_1^2} \leq \frac{S_2^2}{S_1^2} \right) = 1 - \alpha$$

$$P \left(\frac{1}{C_2} \frac{S_1^2}{S_2^2} \leq \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \leq \frac{1}{C_1} \frac{S_1^2}{S_2^2} \right) = 1 - \alpha$$

$$P \left(\frac{1}{C_2} \frac{S_1^2}{S_1^2} \leq \frac{\frac{(n_1-1)S_1^2}{\frac{\sigma_1^2}{n_1-1}}}{\frac{(n_2-1)S_2^2}{\frac{\sigma_2^2}{n_2-1}}} \leq \frac{1}{C_1} \frac{S_1^2}{S_2^2} \right) = 1 - \alpha$$

$$P \left(\frac{1}{C_2} \frac{S_1^2}{S_1^2} \leq \frac{\frac{Q_1}{n_1-1}}{\frac{Q_2}{n_2-1}} \leq \frac{1}{C_1} \frac{S_1^2}{S_2^2} \right) = 1 - \alpha$$

Como

$$F = \frac{\frac{Q_1}{n_1-1}}{\frac{Q_2}{n_2-1}}$$

Temos então:

$$P \left(\frac{1}{C_2} \frac{S_1^2}{S_1^2} \leq F \leq \frac{1}{C_1} \frac{S_1^2}{S_2^2} \right) = 1 - \alpha$$

$$P(F_{1-\alpha/2} \leq F \leq F_{\alpha/2}) = 1 - \alpha$$

. Assim:

$$P(F < F_{1-\alpha/2}) = \alpha/2$$

e

$$P(F \leq F_{\alpha/2}) = 1 - \alpha/2$$

. Colocando C_1 e C_2 em evidência temos:

$$\frac{1}{C_2} \frac{S_1^2}{S_2^2} = F_{1-\alpha/2} \Rightarrow C_2 = \frac{1}{F_{1-\alpha/2}} \frac{S_1^2}{S_2^2}$$

$$\frac{1}{C_1} \frac{S_1^2}{S_2^2} = F_{\alpha/2} \Rightarrow C_1 = \frac{1}{F_{\alpha/2}} \frac{S_1^2}{S_2^2}$$

Sendo $F_{\alpha/2}$ e $F_{1-\alpha/2}$ os quantis apropriados da distribuição $F_{[(n_1-1), (n_2-1)]}$,

ou seja,

$$P(F < F_{1-\alpha/2}) = P(F > F_{\alpha/2}) = \alpha/2$$

Assim, o intervalo de confiança será então:

$$I.C. = [LI; LS] = \left[\frac{1}{F_{\alpha/2}} \frac{S_1^2}{S_2^2}, \frac{1}{F_{1-\alpha/2}} \frac{S_1^2}{S_2^2} \right]$$

A amplitude $A = LS - LI$ será:

$$A = \frac{1}{F_{1-\alpha/2}} \frac{S_1^2}{S_2^2} - \frac{1}{F_{\alpha/2}} \frac{S_1^2}{S_2^2} = \frac{S_1^2}{S_2^2} \frac{F_{\alpha/2} - F_{1-\alpha/2}}{F_{1-\alpha/2} F_{\alpha/2}}$$

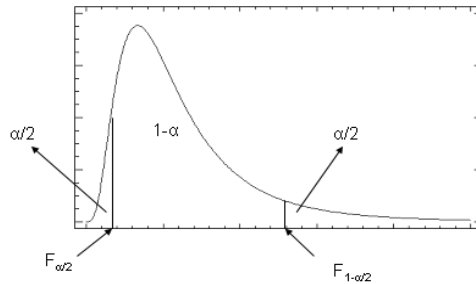


Figura 1.11: Distribuição F de Snedecor

Assim, usando o método de distribuição equitativa do nível de significância tem-se:

$$I.C_{1-\alpha} \left(\frac{\sigma_1^2}{\sigma_2^2} \right) = \left[\frac{1}{F_{(n_1-1, n_2-1, 1-\frac{\alpha}{2})}} \frac{S_1^2}{S_2^2}, \frac{1}{F_{(n_1-1, n_2-1, \frac{\alpha}{2})}} \frac{S_1^2}{S_2^2} \right]$$

Fazendo $S_c^2 = \frac{n}{n-1} S^2$. Foram consideradas as variâncias observadas acima: $S_c^2 = S_1^2$ ou $S_c^2 = S_2^2$ para evitar a escrita de dois subín-

lices nas variâncias.

$$I.C_{1-\alpha} \left(\frac{\sigma_1^2}{\sigma_2^2} \right) = \left[\frac{1}{F_{(n_1-1, n_2-1, 1-\frac{\alpha}{2})}} \frac{n_1}{n_1-1} \frac{S_1^2}{S_2^2}; \frac{1}{F_{(n_1-1, n_2-1, \frac{\alpha}{2})}} \frac{n_2}{n_2-1} \frac{S_1^2}{S_2^2} \right]$$

Exemplo 1.11.1. *A porcentagem de estresse em duas escolas secundárias em algumas turnas é a seguinte:*

<i>Escola(A)</i>	0,7	0,8	0,6	0,95	0,65	1	0,9	0,2	0,55
<i>Escola(B)</i>	0,4	0,7	0,7	0,85	0,9	1	0,95		

Pede-se determinar um I.C. para o quociente de variâncias. Os dados amostras são: $n_A = 9$, $s_A^2 = 0,1986$, $\hat{s}_A^2 = \frac{9}{8} \cdot 0,1986 = 0,2234$. Os dados amostras são: $n_B = 7$, $s_B^2 = 0,0362$, $\hat{s}_B^2 = \frac{7}{6} \cdot 0,0362 = 0,0422$.

Para os graus de liberdade (8,6)temos $F_{(8,6;0,95)} = 4,1468$. Para os graus de liberdade (6,8)temos $F_{(6,8;0,95)} = 3,5806$, de modo que para (8,6) graus de liberdade, $F_{(8,6;0,05)} = \frac{1}{3,5806}$. Assim, o intervalo fica:

$$\left(\frac{s_1^2}{s_2^2} F_{(8,6;0,05)}; \frac{s_1^2}{s_2^2} F_{(8,6;0,95)} \right) = (0,0528; 0,7838)$$

Esse é o intervalo confiança para $\frac{\sigma_2^2}{\sigma_1^2}$ para um nível de confiança de 90%.

Um detalhe muito interessante do intervalo de confiança encontrado é que não contém a unidade, que seria o valor do quociente em caso de igualdade de variâncias. Isso vem determinado pelo fato de que as variâncias amostrais são bastantes dispersas, pelo que não parece muito acertado considerar que as variâncias populacionais possam ser iguais.

1.12 I.C. para diferença de médias sendo as variâncias desconhecidas e iguais

Aqui vamos encontrar o I.C. para diferença de médias, desconhecidas σ_1^2 e σ_2^2 e considerando $\sigma_1^2 = \sigma_2^2 = \sigma^2$. O t de Student será: $t_{n_1+n_2-2; \alpha/2}$ é a abscissa que sob uma $t_{n_1+n_2-2}$ deixa a sua direita uma área α .

E, padronizando, se obtém a distribuição buscada. Se as variâncias são desconhecidas, o resultado anterior não poderia ser usado para cumprir o objetivo. Faz-se necessário encontrar a seguinte solução:

Dado $X_1, \dots, X_n$ uma amostra aleatória, sendo $X \sim N(\mu_1, \sigma_1^2)$ e $Y_1, \dots, Y_n$ uma amostra aleatória de uma variável aleatória $Y \sim N(\mu_2, \sigma_1^2)$. Verifica-se que:

$$I.C. = \left[\bar{X} \pm \bar{Y} \pm t_{n_1+n_2-2; \alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

Exemplo 1.12.1. *Selecionam-se duas amostras aleatórias de tamanhos $n_1 = 25$ e $n_2 = 27$ com variâncias amostrais $s_1^2 = 0,42$ e $s_2^2 = 0,51$ e médias $\bar{x}_1 = 9,82$ e $\bar{x}_2 = 8,75$, respectivamente. Sabendo que os dados foram extraídos de uma população normal encontre uma I.C. para $\mu_1 - \mu_2$, sendo desconhecidas as variâncias populacionais e que as mesmas sejam iguais. Sendo $\alpha = 5\%$.*

O intervalo de confiança será:

$$\begin{aligned} I.C. &= \left[9,82 - 8,75 \pm 2,009 \left[\frac{(24)0,42 + (26)0,51}{50} \right] \sqrt{\frac{1}{25} + \frac{1}{27}} \right] \\ &= [0,70; 1,46] \end{aligned} \tag{1.4}$$

1.13 I.C. para diferença de médias sendo desconhecidas as variâncias amostrais

Aqui vamos encontrar o I.C. para diferença de médias, desconhecidas S_1^2 e S_2^2 e considerando $S_1^2 = S_2^2 = \sigma^2$. Com distribuição t de Student com gl (graus de liberdade). O I.C. é calculado de forma análoga ao anterior, considerando agora o grau de liberdade da distribuição t.

$$I.C. = \left[\bar{X} - \bar{Y} \pm t_{(gl; \alpha/2)} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right]$$

Exercício 1.13.1. *Selecionam-se duas amostras aleatórias de tamanhos $n_1 = 25$ e $n_2 = 27$ com variâncias amostrais $s_1^2 = 0,42$ e $s_2^2 = 0,51$ e médias $\bar{x}_1 = 9,82$ e $\bar{x}_2 = 8,75$, respectivamente. Sabendo que os dados foram extraídos de uma população normal encontre uma I.C. para $\mu_1 - \mu_2$, sendo desconhecidas as variâncias populacionais e que as mesmas sejam iguais.*

Sendo os graus de liberdade (gl) dado por:

$$gl = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{1}{n_1+1} \left(\frac{S_1^2}{n_1} \right)^2 + \frac{1}{n_2+1} \left(\frac{S_2^2}{n_2} \right)^2} - 2$$

1.14 Outros métodos para a construção de I.C.

Nas seções anteriores estudamos o problema da estimação por intervalo de confiança. Nesta seção o objetivo é descrever alguns métodos gerais para a construção de intervalos de confiança. Vamos fazer uma introdução do método de desigualdade de Tchebyshev.

Existe um procedimento alternativo, de uso universalmente aceito, segundo o qual é possível deduzir um I.C. para o parâmetro desconhecido sem ter mais informação que pela média e variância de um estimador de parâmetro. Trata-se do método baseado na desigualdade de Tchebyshev.

Seja X uma variável aleatória tal que exista $E[X^2]$. O valor $P[|X - E[X]|] \leq c$ será então:

$$P[|X - E[X]| \leq c] \geq 1 - \frac{\text{Var}[\hat{\theta}]}{c^2}, \quad \forall c > 0$$

Demonstração 1.14.1. A demonstração se baseia na desigualdade básica aplicada a variável $(X - E[X])^2 \geq 0$ e tomando $\varepsilon = c$.

$$P(|X - E[X]| \leq c) = P((X - E[X])^2 \leq c^2) \leq \frac{E[(X - E[X])^2]}{c^2}$$

Expressões alternativas da desigualdade seria:

$$P[|X - E[X]| \leq c] \geq 1 - \frac{\text{Var}(X)}{c^2}, \quad \forall c > 0$$

Exemplo 1.14.1. Seja X uma variável aleatória com μ e σ . Utilizar a desigualdade de Tchebyshev para estimar $P(\mu - 4\sigma \leq X \leq \mu + 4\sigma)$

Pela desigualdade temos que :

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) \leq 1 - \frac{1}{4^2} = 0,937$$

Exemplo 1.14.2. Seja Z uma variável aleatória normal padronizada com $\mu = 0$ e $\sigma = 1$. Use a desigualdade de Chebyshev para encontrar um valor β tal que $P(-\beta \leq Z \leq \beta) = 0,8$.

$$P(0 - \beta \leq Z \leq \beta) \geq \frac{\sigma^2}{c^2} \quad P(0 - \beta \leq Z \leq \beta) \geq \frac{1}{c^2}$$

$$1 - \frac{1}{c^2} = 0,8 \quad \frac{1}{c^2} = 0,2 \quad c^2 = \frac{1}{0,2} = 5$$

Como $c = \sqrt{5} = \beta$, logo o I.C. será:

$$P(0 - \sqrt{5} \leq Z \leq \sqrt{5}) = \frac{1}{5}$$

Seja $\tilde{\beta}$ um estimador de β com variância finita denominada por $var[\tilde{\beta}] = \sigma^2$, pela desigualdade de Chebyshev, tem-se:

$$P[|\tilde{\beta} - E[\tilde{\beta}]| \leq \varepsilon] \geq 1 - \frac{Var(\tilde{\beta})}{\varepsilon^2} \quad (1.5)$$

Sabe-se que um estimador imparcial é o seu valor esperado seja igual ao parâmetro a estimar, ou seja, $B(X)$ é imparcial para β se $E_{\beta}[B(X)] = \beta$, $\forall \beta$ que pertença ao conjunto de todos os possíveis valores do parâmetro que se denota por Θ e a função de distribuição da variável aleatória ou distribuição da população, denotada por F_{β} . Fazendo $\varepsilon = c\sqrt{var[\tilde{\beta}]} = c\sigma$ e substituindo em 1.5 tem-se:

$$P[|\tilde{\beta} - \beta| \leq c\sigma] \geq 1 - \frac{Var(\tilde{\beta})}{(c\sqrt{var[\tilde{\beta}]})^2}$$

Simplificando, fica então:

$$P[|\tilde{\beta} - \beta| \leq c\sigma] \geq 1 - \frac{1}{c^2}$$

Com objetivo de encontrar um I.C. é justo fazer: $\frac{1}{c^2} = 1 - \alpha$,

colocando c em evidência tem-se: $c = \frac{1}{\sqrt{\beta}}$, assim tem-se:

$$P \left[|\tilde{\beta} - \beta| \leq \frac{\sigma}{\sqrt{c}} \right] \geq 1 - \alpha \quad (1.6)$$

Aplicando a propriedade de uma inequação modular: $|x - a| \leq b \Rightarrow -b \leq x - a \leq b \Rightarrow a - b \leq x \leq a + b$ em 1.6, tem-se:

$$P \left[\beta - \frac{\sigma}{\sqrt{\alpha}} \leq \tilde{\beta} \leq \beta + \frac{\sigma}{\sqrt{\alpha}} \right] \geq 1 - \alpha$$

Assim, o I.C. para β a um nível $1 - \alpha$ será:

$$I.C. = \left[\beta - \frac{\sigma}{\sqrt{\alpha}} \leq \tilde{\beta} \leq \beta + \frac{\sigma}{\sqrt{\alpha}} \right]$$

Sempre que seja possível escrever com sucesso:

$$\left[\beta - \frac{\sigma}{\sqrt{\alpha}} \leq \tilde{\beta} \leq \beta + \frac{\sigma}{\sqrt{\alpha}} \right]$$

Da forma:

$$\left[\beta_{inf}(\tilde{\beta}) \leq \beta \leq \beta_{sup}(\tilde{\beta}) \right]$$

Esse será um I.C. para β ao um nível $1 - \alpha$. Esse I.C. não é para β a um nível de $1 - \alpha$, senão que o nível é ao menos $1 - \alpha$.

Exemplo 1.14.3. *Na população de pessoas com problemas psicológico por causa de política, de uma determinada região, se conhece por estudos anteriores que o desvio padrão é de 60.000, desconhecendo sua média. Seleciona-se uma amostra aleatória simples de 100 indivíduos nessa região resultando numa média de 170000 indivíduos. Pede-se encontrar um I.C. de pelo menos 90% do número de indivíduo médio nessa região.*

Dado que se conhece a distribuição de probabilidade nessa re-

gião utilizaremos a desigualdade de Tchebyshev. Segundo essa desigualdade sabe-se que para uma variável aleatória X , tem-se:

$$P[|X - E[X]| \leq c] \geq 1 - \frac{Var(X)}{nc^2}, \quad \forall c > 0$$

Fazendo $\bar{X} = X$ tem-se que:

$$E(\bar{X}) = E(X) = \mu \quad V(\bar{X}) = \frac{X}{n} = \frac{\sigma^2}{n}$$

A desigualdade será:

$$P[|\bar{X} - \mu| \leq c] \geq 1 - \frac{\sigma^2}{nc^2}$$

Assim,

$$\begin{aligned} P[-c \leq \bar{X} - \mu \leq c] &\geq 1 - \frac{\sigma^2}{nc^2} \\ P[-c - \bar{X} \leq -\mu \leq c - \bar{X}] &\geq 1 - \frac{\sigma^2}{nc^2} \\ P[\bar{X} - c \leq \mu \leq \bar{X} + c] &\geq 1 - \frac{\sigma^2}{nc^2} \end{aligned}$$

Como o intervalo $[\bar{X} - c; \bar{X} + c]$ é um intervalo aleatório com probabilidade de ao menos $1 - \frac{\sigma^2}{nc^2}$. Fazendo $1 - \frac{\sigma^2}{nc^2} = 90\%$ tem-se:

$$1 - \frac{\sigma^2}{nc^2} = 90\%$$

$$1 - 0,90 = \frac{\sigma^2}{nc^2}$$

$$c^2 = \frac{\sigma^2}{0,10n}$$

$$c = \frac{\sigma^2}{\sqrt{100n}}$$

Como $n=100$ e $\sigma = 60000$. O valor de c será: $c = \frac{60000}{\sqrt{10}} = 18974$

Portanto o I.C. será:

$$I.C_{\mu 90\%} = P[\bar{X} - 18974; \bar{X} + 18974]$$

$$= P[170000 - 18974; 170000 + 18974]$$

$$= [151026,3; 188973,7]$$

```
n<-100
s<-60000
k<- s/sqrt(10); k
IC<-c(170000-k, 170000+k); IC
```

Exercício 1.14.1. *Na população de pessoas com problemas trans-torno de déficit de atenção e hiperatividade por causa de estresse, em uma determinada região, se conhece por estudos anteriores que o desvio padrão é de 70.000, desconhecendo sua média. Seleciona-se uma amostra aleatória simples de 91 indivíduos nessa região resultando numa média de 160000 indivíduos. Pede-se encontrar um*

I.C. de pelo menos 95% do número de indivíduo médio nessa região com transtorno de déficit de atenção e hiperatividade.

Exercício 1.14.2. *Suponha que quantidade de pessoas estressadas em várias escolas tem uma variância populacional $\sigma = 9$. Encontre um I.C. para $\tilde{\beta}$ ao nível de 10%.*

Exercício 1.14.3. *Suponha que quantidade de pessoas que tem inteligência espacial em várias universidades tem uma variância populacional $\sigma = 16$. Encontre um I.C. para $\tilde{\beta}$ ao nível de 10%.*

Exemplo 1.14.4. *Seja X uma variável aleatória $U(0, \theta)$. Obtenha um I.C. para θ ao um nível $1 - \alpha$. A função densidade da variável aleatória é:*

$$f(x; \theta) = \frac{1}{\theta}, \quad 0 \leq x \leq \theta$$

Sendo $E_{\theta} = \frac{\theta}{2}$ e $Var_{\theta} = \frac{\theta^2}{12}$. Fazendo $\tilde{\theta} = 4X$ terá:

$$E_{\theta}[\tilde{\theta}] = 4 \frac{\theta}{2} = 2\theta, \quad Var_{\theta}[\tilde{\theta}] = 4^2 \frac{\theta^2}{12} = 16 \frac{\theta^2}{12} = \frac{4}{3} \theta^2$$

Como $\sigma = 9$ e $\alpha = 10\%$, temos que:

$$I.C. = \left[\beta - \frac{9}{\sqrt{0,10}} \leq \tilde{\beta} \leq \beta + \frac{9}{\sqrt{0,10}} \right]$$

1.15 Intervalos de confiança para proporções

Vamos adotar a seguinte convenção: z_p significa o valor da variável aleatória Z que segue uma distribuição normal para o qual $P(z \leq z_p) = p$, ou seja, a área sob a curva normal padrão à esquerda de z_p é igual a p .

Suponhamos que se deseja obter informação sobre o parâmetro p de uma população com distribuição Bernoulli. Obtem-se n testes de Bernoulli independentes e se define a variável aleatória: $X =$ número de sucessos nas n provas. O parâmetro desconhecido apresenta $P[\text{sucesso}] = p$ em cada testes de . Nesta situação considera-se a estatística:

$$\hat{p} = \frac{X}{n}$$

Conhecido como proporção amostral. Esta estatística, tal que $0 \leq \hat{p} \leq 1$, é a estatística a utilizar para obter informação sobre p , proporção populacional.

Proposição 1.15.1. *Verifica-se:*

1. $E[\hat{p}] = p$
2. $Var[\hat{p}] = \frac{p(1-p)}{n}$

Demonstração: Em efeito, tem em conta que a variável aleatória X segue uma distribuição $B(n, p)$,

1. $E[\hat{p}] = E\left[\frac{X}{n}\right] = \frac{E(X)}{n} = \frac{np}{n} = p$
2. $Var[\hat{p}] = \frac{p(1-p)}{n} = Var\left[\frac{X}{n}\right] = \frac{1}{n^2} Var[X] = \frac{np(n-1)}{n^2} = \frac{p(1-p)}{n} =$
 $\frac{pq}{n}$

Do teorema central do limite de Lévy segue-se o seguinte resultado que permite aproximar a distribuição na amostra da proporção amostral.

Proposição 1.15.2. *A variável aleatória $\hat{p}$ padronizada tem distribuição assintótica $N(0, 1)$, ou seja,*

$$\lim_{n \rightarrow \infty} P \left[\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z \right] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

A expressão $x \sim B(n; p)$ e np e $nq > 5 \Rightarrow x \sim N(\mu = np; \sigma = \sqrt{npq})$ indica que uma Binomial x se aproxima de uma Normal sob determinadas condições. Sabendo que $x \in \mu \pm z_\alpha \sigma$ se obtém que $x \in np \pm z_\alpha \sqrt{npq}$ a $(1 - \alpha)100\%$ das vezes. Igualando x aos extremos, colocando a incógnita p e efetuando $(x \pm 0,5)$ obtém-se o seguinte I.C. para p , esse intervalo é o método de Wilson com correção por continuidade.

$$p \in \frac{(x \pm 0,5) + \frac{z_\alpha^2}{2} \pm \sqrt{\frac{\alpha^2}{4} + (x \pm 1/2) \left(1 - \frac{x \pm 0,5}{n}\right)}}{n + z_\alpha^2}$$

O valor de z_α se observa na tabela z padrão. Esse I.C. só será válido se $n\hat{p} = x > 5$ e $n\hat{q} = n - x > 5$ (com $\hat{q} = 1 - \hat{p}$)

Exemplo 1.15.1. *Se de 100 jovens com problemas, 30 foram a favor de um partido político. Que porcentagem de votos obteria dito político nas eleições?*

Observe que $n = 30 > 5$ e $n - x = 100 - 30 = 70 > 5$, e sabendo que $z_\alpha^2 = 1,96$. Com uma confiança de 95%, teremos depois de multiplicar por 100.

$$p \in \frac{(30 \pm 0,5) + \frac{1,96}{2} \pm \sqrt{\frac{1,96}{4} + (30 \pm 1/2) \left(1 - \frac{30 \pm 0,5}{100}\right)}}{100 + 1,96^2}$$

Ou seja, a população a favor do partido político está entre 21,45% e 40,11%.

O I.C. anterior é a melhor solução assintótica possível sob certa condição, mas a fórmula é muito incômoda para efeito de cálculo. Se no I.C. antes era $(x \in np \pm z_\alpha \sqrt{npq})$, substitui $\sqrt{npq}$ por $\sqrt{n\hat{q}\hat{p}}$, esse novo I.C. apenas será válido se $x > 20$ e $n - x > 20$. Logo, o I.C. será:

$$p \in \hat{p} \pm \left[z_\alpha \sqrt{\frac{\hat{p}\hat{q}}{n}} + \frac{1}{2n} \right] = \frac{x \pm \left[z_\alpha \sqrt{\frac{x(n-x)}{n}} + 0,5 \right]}{n}$$

A primeira expressão 1.15 é mais conhecida na literatura, mas a expressão 1.15 é mais cômoda para efeito de cálculo. No exemplo, o I.C. é válido, pois $x = 30 > 20$ e $n - x = 70 > 20$, assim o intervalo será:

$$p \in \frac{30 \pm \left(1,96 \sqrt{\frac{30 \cdot 70}{100}} + 0,5 \right)}{100} = [20,5\%; 39,48\%]$$

Quando não é válido a expressão acima, usa-se a seguinte fórmula geral:

$$p \in \frac{(x+2) \pm \left[z_\alpha \sqrt{(x+2)(n-x+2)} \right]}{n+4}$$

Considerando os valores do exemplo, temos:

$$p \in \frac{(30+2) \pm \left[1,96 \sqrt{(30+2)(100-30+2)} \right]}{100+4} = [21,9\% \leq p \leq 39,4\%]$$

Vamos fazer um resumo dos três métodos visto acima:

- O I.C. de Wilson: $21,4\% \leq p \leq 40,1\%$ proporciona praticamente a mesma solução
- O I.C. do Wald com correção por continuidade: $20,5\% \leq$

$p \leq 39,5\%$ proporciona praticamente a mesma solução.

- O I.C. do Wald ajustado (Agresti and Caffo, 2000): $21,9\% \leq p \leq 39,63\%$.

Exemplo 1.15.2. *Seja uma proporção $B(1; p)$ da qual se extrai uma amostra aleatória simples de tamanho $n = 255$, resultando ser a proporção amostral $0,6$. Construir um I.C. de 95% para o parâmetro p .*

O estimador de máxima verossimilhança de p é $\hat{p} = \frac{\sum_{i=1}^n x_i}{n}$ em que, pelo teorema de Moivre-Laplace, tem-se:

$$\sum_{i=1}^n x_i \sim N[(np; \sqrt{np(1-p)})]$$

Portanto,

$$\frac{\sum_{i=1}^n x_i - np}{\sqrt{np(1-p)}} = \frac{\bar{x} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0; 1)$$

Portanto, em probabilidade tem-se:

$$P \left[-1,96 \leq \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq 1,96 \right] = 0,95$$

Que conduz a:

$$\begin{aligned}
P \left[|\hat{p} - p| \leq 1,96 \frac{\hat{p} - p}{n} \right] &= 0,95 \\
P \left[(\hat{p} - p)^2 \leq \frac{1,96^2}{n} p(1-p) \right] &= 0,95 \\
P \left[\hat{p}^2 - 2\hat{p}p + p^2 - \frac{1,96^2 p}{n} + \frac{1,96^2}{n} p^2 \leq 0 \right] &= 0,95 \\
P \left[\left(1 + \frac{1,96^2}{n} \right) p^2 - \left(2\hat{p} + \frac{1,96^2}{n} \right) p + \hat{p}^2 \leq 0 \right] &= 0,95
\end{aligned}$$

Sendo a solução para p da desigualdade entre colchêta:

$$\begin{aligned}
|p| \leq & \frac{\left(2\hat{p} + \frac{1,96^2}{n} \right) \pm \sqrt{\left(2\hat{p} + \frac{1,96^2}{n} \right)^2 - 4 \left(1 + \frac{1,96^2}{n} \right) \hat{p}^2}}{2 \left(1 + \frac{1,96^2}{n} \right)} = \\
& \frac{\left(2\hat{p} + \frac{1,96^2}{n} \right) \pm \sqrt{4\hat{p}^2 + 4\hat{p}\frac{1,96^2}{n} + \left(\frac{1,96^2}{n} \right)^2 - 4\hat{p}^2 - 4\hat{p}^2\frac{1,96^2}{n}}}{2 \left(1 + \frac{1,96^2}{n} \right)} = \\
& \frac{\left(2\hat{p} + \frac{1,96^2}{n} \right) \pm \sqrt{4\hat{p}\frac{1,96^2}{n}(1-\hat{p}) + \left(\frac{1,96^2}{n} \right)^2}}{2 \left(1 + \frac{1,96^2}{n} \right)} = \\
& \frac{\left(2\hat{p} + \frac{1,96^2}{n} \right) \pm 1,96 \sqrt{\frac{4\hat{p}(1-\hat{p}) + \frac{1,96^2}{n}}{n}}}{2 \left(1 + \frac{1,96^2}{n} \right)} = \\
& \frac{\left(\hat{p} + \frac{1,96^2}{2n} \right) \pm 1,96 \sqrt{\frac{4\hat{p}(1-\hat{p}) + \frac{1,96^2}{n}}{4n}}}{\left(1 + \frac{1,96^2}{n} \right)}
\end{aligned}$$

Pelo que a probabilidade inicial finalmente adota a expressão:

$$P \left[\frac{\left(\hat{p} + \frac{1,96^2}{2n} \right) - 1,96 \sqrt{4\hat{p}(1-\hat{p}) + \frac{1,96^2}{n}}}{1 + \frac{1,96^2}{n}} \leq p \leq \frac{\left(\hat{p} + \frac{1,96^2}{2n} \right) + 1,96 \sqrt{4\hat{p}(1-\hat{p}) + \frac{1,96^2}{n}}}{1 + \frac{1,96^2}{n}} \right]$$

Considerando $n = 225$, $\hat{p} = 0,6$, $\hat{q} = 1 - \hat{p} = 0,4$, tem-se como resultado $[0,54; 0,67]$.

Para n grande, os termos $\frac{1,96^2}{n}$ e $\frac{1}{\sqrt{n}}$ serão muito pequenos, logo pode-se considerar a seguinte expressão:

$$P \left[\hat{p} \pm 1,96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

Considerando os dados informado anteriormente, o I.C. seria então: $[0,53; 0,66]$.

Exemplo 1.15.3. *Seja uma proporção $B(1;p)$ da qual se extrai uma amostra aleatória simples de tamanho $n = 200$, resultando ser a proporção amostral 0,7. Construir um I.C. de 90% para o parâmetro p .*

Exercício 1.15.1. *Seja uma proporção $B(1;p)$ da qual se extrai uma amostra aleatória simples de tamanho $n = 255$, resultando ser a proporção amostral 0,7. Construir um I.C. de 90% para o parâmetro p .*

Exercício 1.15.2. *Seja uma proporção $B(1;p)$ da qual se extrai uma amostra aleatória simples de tamanho $n = 300$, resultando ser a proporção amostral 0,8. Construir um I.C. de 95% para o parâmetro p .*

1.15.1 Intervalos de confiança para a proporção sem utilizar o dobro da aproximação

O I.C. para a proporção sem utilizar o dobro da aproximação e fazendo $c = Z_{1-\frac{\alpha}{2}}$, tem-se:

$$\begin{aligned}
 1 - \alpha &= P[-c \leq N(0, 1) \leq c] \\
 &= P\left[-c \leq \frac{\hat{p} - p}{\sqrt{p(1-p)}} \leq c\right] \\
 &= P[(\hat{p} - p)^2 n \leq c^2 p(1-p)] \\
 &= P[(\hat{p}^2 + p^2 - 2p\hat{p})n \leq \frac{c^2 p(1-p)}{n}] \\
 &= P\left[\left(1 + \frac{c^2}{n}\right)p^2 - \left(2\hat{p} + \frac{c^2}{n}\right)p + \hat{p}^2\right] \\
 p &= \frac{\left(2\hat{p} + \frac{c^2}{n}\right) \pm \sqrt{(2\hat{p} + \frac{c^2}{n})^2 - 4(1 + \frac{c^2}{n})\hat{p}^2}}{2(1 + \frac{c^2}{n})}
 \end{aligned}$$

Portanto, o I.C. será:

$$I.C_{1-\alpha}(p) = \left[\frac{\left(2\hat{p} + \frac{c^2}{n}\right) \pm \sqrt{(2\hat{p} + \frac{c^2}{n})^2 - 4(1 + \frac{c^2}{n})\hat{p}^2}}{2(1 + \frac{c^2}{n})} \right]$$

Ou equivalentemente,

$$I.C_{1-\alpha}(p) = \left[\frac{\hat{p}}{1 + \frac{c^2}{n}} + \frac{1}{2(1 + \frac{n}{c^2})} \pm \frac{c}{\sqrt{n}} \sqrt{4\hat{p}(1 - \hat{p} + \frac{c}{n^2})2(1 + \frac{c^2}{n})} \right]$$

1.15.2 Intervalos de confiança para a proporção utilizando o dobro da aproximação

O I.C. para a proporção utilizando o dobro da aproximação e fazendo $c = Z_{1-\frac{\alpha}{2}}$, tem-se:

$$\begin{aligned} 1 - \alpha &= P[-c \leq N(0, 1) \leq c] \\ &= P\left[-c \leq \frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})}} \sqrt{n} \leq c\right] \\ &= P\left[\hat{p} - \frac{c}{\sqrt{n}} \sqrt{\hat{p}(1 - \hat{p})} \leq p \leq \hat{p} + \frac{c}{n} \sqrt{\hat{p}(1 - \hat{p})}\right] \end{aligned}$$

Onde se deduz o I.C.:

$$I.C_{1-\alpha}(p) = \left[\hat{p} \pm \frac{c}{\sqrt{n}} \sqrt{\hat{p}(1 - \hat{p})} \right]$$

Exemplo 1.15.4. Pretende-se estimar a quantidade de pessoas com crise de ansiedade com 95% de confiança. Toma uma amostra aleatória de tamanho 100, observou-se que 55% tem essa crise de ansiedade. Encontre o I.C. de pessoas que tem crise.

$$\begin{aligned}
 I.C_{1-5\%}(p) &= \left[\hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] \\
 &= \left[0,55 \pm Z_{5\%/2} \sqrt{\frac{0,55(1-0,55)}{100}} \right] \\
 &= \left[0,55 \pm Z_{0,975} \sqrt{\frac{0,55(1-0,55)}{100}} \right] \\
 &= \left[0,55 \pm 1,96 \sqrt{\frac{0,55(1-0,55)}{100}} \right] \\
 &= [0,55 \pm 0,1] \\
 &= [0,45; 0,65]
 \end{aligned}$$

Exercício 1.15.3. Pretende-se estimar a quantidade de pessoas com transtorno de alimentação com 90% de confiança. Toma uma amostra aleatória de tamanho 200, observou-se que 60% tem esse transtorno. Encontre o I.C. de pessoas que tem transtorno de alimentação

Exercício 1.15.4. Pretende-se estimar a quantidade de pessoas com transtornos psicóticos incluindo esquizofrenia com 95% de confiança. Para isso seleciona-se uma amostra aleatória de tamanho 178, observou-se que 65% tem esse transtorno. Encontre o I.C. de pessoas que tem transtornos psicóticos incluindo esquizofrenia.

1.15.3 Determinação do tamanho amostral para proporção

Podemos considerar que p (o valor real) $\in \hat{p}$ (valor estimado da amostra) $\pm 3\%$ (precisão d).

Isso significa que a proporção $\hat{p}$ (a estimada) difere da população p (a real) em 3% (d) ou menos. Ou seja,

$$|\hat{p} - p| \leq 3\% = d$$

Para garantir essa precisão d é preciso que o tamanho da amostra n seja fixado antes de considerar os dados, cuja finalidade tem que igualar a d o valor de $Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ da expressão 1.15 com valores reais de p e colocando n no primeiro membro da equação tem-se:

$$n = \frac{Z_{1-\frac{\alpha}{2}}^2 p(1-p)}{d^2}$$

Uma vez obtido o valor de n , a fórmula garante que se $\hat{p} = 25\%$, por exemplo, então $p \in 25\%(\hat{p}) \pm 3\%(d)$ com uma confiança de $1 - \frac{\alpha}{2}$, ou seja, que a proporção p de jovens com problema estará entre 22% e 27%.

Exercício 1.15.5. *Seja $\hat{p} = 30\%$, $d = 2\%$ e $\alpha = 5\%$. Encontre o tamanho amostral n .*

Exercício 1.15.6. *Seja $\hat{p} = 38\%$, $d = 1\%$ e $\alpha = 10\%$. Encontre o tamanho amostral n .*

1.16 I.C. para diferença de duas proporções dicotômicas

Sejam duas proporções $B(1; p_1)$ e $B(1; p_2)$ das quais se extraem duas amostras aleatórias simples de tamanho n_1 e n_2 , respectivamente, suficientemente grande e independentes entre si., sendo os estimadores de máxima verossimilhança dado por:

$$\hat{p}_1 = \frac{\sum_{i=1}^{n_1} x_i}{n_1} \quad \hat{p}_1 = \frac{\sum_{i=1}^{n_2} y_i}{n_2}$$

Com as seguintes características: $E(\hat{p}_1) = p_1$ e $V(\hat{p}_1) = \frac{p_1(1-p_1)}{n_1}$. Assim,

$$\hat{p}_1 \sim \left(p_1; \sqrt{V(\hat{p}_1)} \right)$$

Com as seguintes características: $E(\hat{p}_2) = p_2$ e $V(\hat{p}_2) = \frac{p_2(1-p_2)}{n_2}$. Assim,

$$\hat{p}_2 \sim \left(p_2; \sqrt{V(\hat{p}_2)} \right)$$

Sendo que p_1 e p_2 se distribuem assintoticamente e $n \rightarrow \infty$. A diferença fica então:

$$E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$$

$$V(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$$

Partindo da distribuição assintótica, tem-se:

$$(\hat{p}_1 - \hat{p}_2) \sim N \left(p_1 - p_2; \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \right)$$

Pode-se determinar a quantidade pivotal da seguinte forma:

$$T(X, Y; \hat{p}_1 - \hat{p}_2) = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \sim N(0; 1)$$

Pode-se então derivar o I.C. para a diferença $p_1 - p_2$ ao nível de $1 - \alpha$ da seguinte forma:

$$\left[\hat{p}_1 - \hat{p}_2 \pm C \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \right]$$

Sendo que o valor C é determinado, como habitual, tendo em conta que:

$$P[-C \leq N(0; 1) \leq C] = 1 - \alpha$$

Exemplo 1.16.1. *Um psicólogo deseja estimar o incremento de preocupação numa escola. Para isso, aplicou-se um teste antes e depois do período provas escolares. A quantidade de jovens que participaram do teste foram $n_1 = 600$ e $n_2 = 1000$, respectivamente. O pesquisador afirma um incremento de 26% antes das provas e, 42% depois. Construir um I.C. de 95%.*

Dado que $\alpha = 0,05$, buscando na tabela normal obteve-se $Z_{\alpha/2} = 1,96$, $\hat{p}_1 = 0,42$, $\hat{p}_2 = 0,26$, $n_1 = 600$ e $n_2 = 1000$. Portanto, o I.C. será:

$$\begin{aligned} & \left[0,42 - 0,26 \pm 1,96 \sqrt{\frac{0,42(1-0,42)}{1000} + \frac{0,26(1-0,26)}{600}} \right] \\ & = [0,11; 0,20] \end{aligned}$$

Exercício 1.16.1. *Um psicólogo deseja estimar o incremento de preocupação numa escola após a Covid 19. Para isso, aplicou-se*

um teste antes e depois do período provas escolares. A quantidade de jovens que participaram do teste foram $n_1 = 700$ e $n_2 = 1200$, respectivamente. O pesquisador afirma um incremento de 30% antes das provas e, 46% depois. Construir um I.C. de 90%.

Exercício 1.16.2. *Encontrar os seguintes I.C. para proporção.*

1. Deseja-se estimar a proporção de jovens estressados que estejam com alto nível de estresse. Tomando-se uma amostra de 40 estressados, identificou-se que 12 estavam com alto nível de estresse. O nosso problema aqui é determinar um intervalo de confiança para a proporção populacional p de estressados que possam estar com alto nível de estresse. Na terminologia utilizada neste texto, o sucesso, neste caso, é considerado com alto nível de estresse.

A partir da amostra, obtemos $\hat{p} = 12/40 = 0,30$

2. Uma amostra aleatória de 300 idosos mostrou que 180 desejavam ser menos ansioso. Encontrar os limites de confiança de 90% e 95% para a proporção da população que ser menos ansioso.
3. Deseja-se estimar a proporção de idosos que usa o mecanismo de defesa super compensação. Tomando-se uma amostra de 40 idosos, identificou-se que 14 usam esse mecanismo. Determine o intervalo de confiança de 90% para a proporção populacional p de idosos que usam nesse mecanismo. Na terminologia utilizada neste texto, o sucesso, neste caso, é considerado idosos que usam esse mecanismo defesa super compensação.

4. Uma amostra aleatória de 300 jovens de uma cidade mostrou que 180 tinha necessidade de autorealização. Encontrar os limites de confiança de 90% e 95% para a proporção de jovens que tem necessidade de autorealização.
5. Uma amostra aleatória de 400 jovens de uma cidade mostrou que 100 tinha necessidade de proteção (segurança, estabilidade). Encontrar os limites de confiança de 90% e 95% para a proporção de jovens que tem essa necessidade.
6. Examinando uma amostra aleatória de 600 pessoas uma avaliação psicológica na adolescência (cenário nacional), encontraram-se 150 pessoas que não fizeram essa avaliação. Ao nível de 95,5% de confiança, construa um intervalo de confiança para as pessoas que não fizeram essa avaliação.
7. Seja uma proporção $B(1; p)$ da qual se extrai uma amostra aleatória simples de tamanho $n = 200$, resultando ser a proporção amostral 0,6. Construir um I.C. de 95% para o parâmetro p .
8. Seja uma proporção $B(1; p)$ da qual se extrai uma amostra aleatória simples de tamanho $n = 280$, resultando ser a proporção amostral 0,6. Construir um I.C. de 90% para o parâmetro p .
9. Em um estudo sobre o consumo de álcool entre os alunos se realiza um questionário e se pretende obter um intervalo de 90% de confiança para a proporção de jovens que consomem álcool regularmente nos finais de semana. Construir um I.C. nos casos que:
 - i. De 1000 alunos, 460 bebem álcool

ii. De 50 alunos, 23 bebem álcool

10. A porcentagem de ansiosos em duas faculdades considerando algumas turmas é a seguinte:

<i>Faculdade(A)</i>	0,81	0,92	0,67	0,95	0,85	0,91	0,98	0,43	0,41
<i>Faculdade(B)</i>	0,75	0,89	0,72	0,85	0,94	0,82			

Pede-se determinar o I.C. a 95% para a diferença da porcentagem do número de estresse nessas duas escolas.

1.17 Intervalo de média para diferença de médias de duas populações normais de variâncias desconhecidas não iguais

Utiliza-se um procedimento conhecido como aproximação de Welch, que consiste em considerar:

$$\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}}$$

É uma variável t de Student com um grau de liberdade fracionário g, que vem dado pela expressão:

$$g = \frac{(\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1})^2}{\frac{s_1^4}{(n_1 + 1)^2(n_1 - 1)} + \frac{s_2^4}{(n_2 + 1)^2(n_2 - 1)}} - 2$$

O I.C. para $\mu_1 = \mu_2$ vem dado por:

$$\bar{x}_1 - \bar{x}_2 \pm t_{(g; 1 - \alpha/2)} \sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}$$

1.17 Intervalo de média para diferença de médias de duas populações normais de variâncias desconhecidas não iguais

100

Exemplo 1.17.1. *As porcentagens de estressados em duas escolas secundárias foram:*

<i>Escola(A)</i>	0,7	0,8	0,6	0,95	0,65	1	0,9	0,2	0,55
<i>Escola(B)</i>	0,4	0,7	0,7	0,85	0,9	1	0,95		

Determinar o I.C. a 95% para a diferença da porcentagem do número de estresse nessas duas escolas. Considerando a escola A temos: $n_A = 9$, $\Sigma x_i = 6,35$, $\Sigma x_i^2 = 4,9675$ de modo que $\bar{x} = 6,35/9 = 0,7055$ e $S_A^2 = \frac{4,9675}{9} - 0,7055^2 = 0,0541$. Considerando a escola B temos: $n_B = 7$, $\Sigma x_i = 5,50$, $\Sigma x_i^2 = 4,575$ de modo que $\bar{x} = 5,50/7 = 0,7857$ e $S_B^2 = \frac{4,575}{7} - 0,7857^2 = 0,0360$.

Suponha igualdade de variâncias populacionais, calculamos:

$$S_T^2 = \frac{9 \cdot 0,0541 + 7 \cdot 0,0362}{14} = 0,0529$$

$$S_T^2 \left(\frac{1}{9} + \frac{1}{7} \right) = 0,0134$$

E assim,

$$S_T \left(\frac{1}{9} + \frac{1}{7} \right) = 0,1159$$

Para 14 graus de liberdade, $t_{14;0,975} = 2,145$, de modo que o I.C. é: $0,7055 - 0,7857 \pm 2,145 \cdot 0,1159 = -0,0802 \pm 0,2486$ para uma confiança de 95%. Não considerando a igualdade da variância, o número fracionário do grau de liberdade será:

$$g = \frac{\left(\frac{0,0541}{8} + \frac{0,0362}{6} \right)^2}{\frac{0,0541^2}{8^2(10)} + \frac{0,0362^2}{6^2 \cdot 8}} - 2 = 15,9468$$

Para 15 graus de liberdade, $t_{15;0,975} = 2,1314$, enquanto que, para 16 graus de liberdade, o valor correspondente é $t_{16;0,975} = 2,1199$, por interpolação linear entre 15 e 16 graus de liberdades, obtemos $t_{g;0,975} = 2,1205$ graus de liberdade. Assim, o I.C. será então:

$$-0,0802 \pm 2,1205 \sqrt{\frac{0,0541}{8} + \frac{0,0362}{6}} = -0,0802 \pm 0,2399.$$

Exemplo 1.17.2. *A porcentagem de jovens que tem inteligência lógico-matemática em duas escolas foram:*

<i>Escola(A)</i>	0,8	0,7	0,7	0,90	0,75	1	0,9	0,2	0,55
<i>Escola(B)</i>	0,5	0,7	0,7	0,85	0,92	1	0,91		

Determinar o I.C. a 90% para a diferença da porcentagem do número jovens que tem inteligência lógico-matemática (soluções de problemas e cálculos) nessas duas escolas.

Exemplo 1.17.3. *As porcentagens de jovens que tem inteligência musical (habilidade musical, habilidade vogal e habilidade instrumental) em duas escolas secundárias A e B foram:*

<i>A</i>	0,83	0,67	0,72	0,85	0,75	0,87	0,92	0,2	0,64
<i>B</i>	0,68	0,75	0,72	0,85	0,92	0,97	0,91	0,63	

Determinar o I.C. a 80% para a diferença da porcentagem do número de jovens tem inteligência musical.

1.18 I.C. para diferença de duas proporções populacionais

às vezes interessa comparar duas proporções p_1 e p_2 de parâmetros desconhecidos de duas populações desconhecidas de duas populações com distribuições de Bernoulli independentes. Sejam n_1 e n_2 as respectivos testes de Bernoulli independentes e sejam X_1 e X_2 as variáveis aleatórias que contabilizam o número de sucessos em cada teste. As estatísticas a utilizar para obter informações sobre p_1 e p_2 são: $p_1 = \frac{X_1}{n_1}$ e $p_2 = \frac{X_2}{n_2}$, em que:

$$\hat{p}_1 - \hat{p}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}$$

Proposição 1.18.1. *Verifica-se que:*

1. $E[\hat{p}_1 - \hat{p}_2] = p_1 - p_2$
2. $Var[\hat{p}_1 - \hat{p}_2] = \frac{p_1(1-p_1)}{n_1} - \frac{p_2(1-p_2)}{n_2}$

Demonstração: Tendo em conta que as variáveis aleatórias X_1 e X_2 seguem distribuições $B(n_1, p_1)$ e $B(n_2, p_2)$, respectivamente tem-se:

1. $E[\hat{p}_1 - \hat{p}_2] = E\left[\frac{X_1}{n_1} - \frac{X_2}{n_2}\right] = \frac{E(X_1)}{n_1} - \frac{E(X_2)}{n_2} = p_1 - p_2$
2. $Var[\hat{p}_1 - \hat{p}_2] = Var\left[\frac{X_1}{n_1} - \frac{X_2}{n_2}\right] = \frac{1}{n_1^2} Var[X_1] + \frac{1}{n_2^2} Var[X_2] = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$

Assim,

$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2; \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}\right)$$

Do teorema central do limite de Lévy segue-se o seguinte resultado que permite aproximar a distribuição na amostra da proporção amostral.

$$\lim_{n_1, n_2 \rightarrow \infty} P \left[\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \leq z \right] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

Pode-se determinar a quantidade pivotal da seguinte forma:

$$T(X_1, X_2, p_1 - p_2) = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} \sim N(0; 1)$$

Da que podemos derivar o I.C. para a diferença $p_1 - p_2$ ao nível de $1 - \alpha$:

$$P \left[-Z_{1-\frac{\alpha}{2}} \leq \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \leq Z_{1-\frac{\alpha}{2}} \right] = 1 - \alpha$$

Sendo o I.C. igual a:

$$I.C_{p_1-p_2} = \left[\hat{p}_1 - \hat{p}_2 Z_{1-\frac{\alpha}{2}} \pm \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \right]$$

A Figura 1.12 mostra a distribuição de $\hat{p}_1 - \hat{p}_2$. O desvio padrão de $\hat{p}_1 - \hat{p}_2$ envolve os parâmetros desconhecidos p_1 e p_2 . Assim, devemos substituí-los por estatimativas para fazer inferências; vamos fazê-lo de maneira um pouco diferente para I.C. e testes.

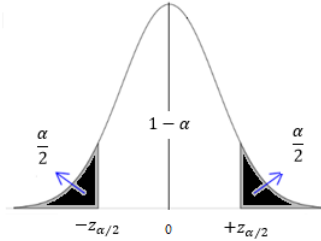


Figura 1.12: Área que tem de manter-se constante quando se deslocam (z_1, z_2) para esquerda ou para direita.

Para um nível de confiança de $(1 - \alpha)100\%$ o I.C. é: $P(\hat{p} - C \leq p \leq \hat{p} + C) = 1 - \alpha$. Vamos encontrar o valor de C de forma que $[LI; LS] = [\hat{p} - C; \hat{p} + C]$. Outro caminho para encontrar o I.C. seria então:

$$P\left(-\frac{C}{\sqrt{\frac{p(1-p)}{n}}} \leq -\frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \leq +\frac{C}{\sqrt{\frac{p(1-p)}{n}}}\right) 1 - \alpha$$

$$P\left(-\frac{C}{\sqrt{\frac{p(1-p)}{n}}} \leq Z \leq +\frac{C}{\sqrt{\frac{p(1-p)}{n}}}\right) 1 - \alpha$$

$$P(-z_{\alpha/2} \leq Z \leq +z_{\alpha/2}) 1 - \alpha$$

$$C = z_{\alpha/2} \frac{p(1-p)}{n}$$

Sendo $z_{\alpha/2}$ o quantil da distribuição $N \sim (0, 1)$, ou seja, $P(Z > z_{\alpha/2}) = \alpha/2 \Leftrightarrow P(Z \leq z_{\alpha/2}) = 1 - \alpha/2$. Assim, o I.C. será:

$$\left[\hat{p} - z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}}; \hat{p} + z_{\alpha/2} \sqrt{\frac{p(1-p)}{n}} \right]$$

Cujo cálculo fica impossibilitado pela presença do parâmetro desconhecido p no intervalo. Para resolver esse problema substitui p por um estimador consistente, ou seja, $\bar{x} = \hat{p}$. Assim, o intervalo fica da seguinte forma:

$$\left[\hat{P} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{P} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

Resolvendo em ordem a p a dupla desigualdade, obtém-se:

$$-z_{\alpha/2} \leq \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq +z_{\alpha/2}$$

Resolvendo, resulta no seguinte I.C.:

$$\frac{\hat{p} - \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z_{\alpha/2}^2}{4n^2}}}{1 + \frac{z_{\alpha/2}^2}{n}}$$

A amplitude do intervalo para diferença de duas proporções será:

$$A = LS - LI = 2z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Exemplo 1.18.1. *Deseja-se testar a igualdade entre as proporções de alunos e alunas que lêem revistas de psicologia. São os seguintes os resultados de amostras aleatórias independentes de alunos e alunas:*

Alunos	Alunas
$x_1 = 70$	$x_2 = 50$
$n_1 = 200$	$n_2 = 200$

Sendo x_1 o número de alunos que tem hábitos de leitura, e x_2 é o correspondente número de alunas. Admita $\alpha = 10\%$. Passos:

- Enunciar as hipóteses: $H_o : p_1 = p_2$ versus $H_a : p_1 \neq p_2$.
- Escolha um α . Vamos considerar $\alpha = 10\%$. Escolher a variável normal padronizada.
- Determinação de RA (região de aceitação) e RR (região de rejeição ou região crítica). Com auxílio da Tabela da distribuição normal padronizada determina-se:

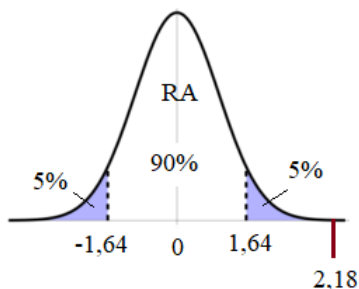


Figura 1.13: Distribuição Z padronizado com $Z_{cal} = 2,18$

- Cálculo da variável:

$$p_1 = \frac{x_1}{n_1} = \frac{70}{200} = 0,35$$

e

$$p_2 = \frac{x_2}{n_2} = \frac{50}{200} = 0,25$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{70 + 50}{200 + 200} = 0,30$$

$$Z_{cal} = \frac{p_1 - p_2}{\sqrt{0,0(0,7) \left(\frac{2}{200}\right)}} = 2,18$$

- Conclusão:

Como o valor $Z_{cal} > 1,64$, rejeita-se a hipótese nula da igualdade entre as proporções, concluindo-se com risco de 10% que as proporções são diferentes.

Exemplo 1.18.2. *Considere uma amostra aleatória simples de 400 adultos e 600 jovens, observando que apenas 100 adultos e 300 jovens tiveram melhoras de um problema psicológico. Para comparar as proporções de adultos e jovens que tiveram melhoras no problema psicológico, a um nível de confiança de 99%. Aqui, usa-se um I.C. de confiança para diferenças de duas populações.*

- p_1 = proporção de jovens que tiveram melhoras psicológicas
- p_2 = proporção de adultos que tiveram melhoras psicológicas

$$\hat{p}_1 = 300/600 = 0,5 \quad \hat{p}_2 = 100/400 = 0,25$$

Substituindo esses valores no intervalo tem-se:

$$\left[0,5 - 0,25 \pm 2,58 \sqrt{\frac{0,5 \cdot 0,5}{600} + \frac{0,25 \cdot 0,75}{400}} \right] = [0,19; 0,31]$$

Assim, o I.C. será:

$$I_{(99\%)}(p_1 - p_2) = [0,19; 0,31]$$

Exercício 1.18.1. *Considere uma amostra aleatória simples de 500 adultos e 700 jovens que tem hábitos de vida contraproducentes, observando que apenas 200 adultos e 400 jovens tiveram melhoras*

no hábitos de vida contraproducentes. Para comparar as proporções de adultos e jovens que tiveram melhoras, a um nível de confiança de 95%. Aqui, usa-se um I.C. de confiança para diferenças de duas populações.

1.19 I.C. para populações não normais

Quando da população de onde se extrai a informação amostral não tem um comportamento probabilístico normal, para a construção de I.C. com seus respectivos parâmetros tem os seguintes critérios:

1.19.1 Caso geral

Se uma população tem uma distribuição não normal definida por $f(x; \theta)$, para a obtenção do I.C. do parâmetro θ pode-se utilizar qualquer dos seguintes métodos:

- Método do pivot

Seja X uma amostra aleatória simples extraída de uma população com distribuição $f(x; \theta)$ em que $\theta \in \Theta$, sendo Θ qualquer intervalo da reta real. Uma função $G(\mathbf{X}; \theta)$ que dependa do parâmetro θ , mas que sua função de probabilidade não dependa do dito parâmetro θ recebe o nome de quantidade pivotal ou, simplesmente, pivot. Este pivot $G(\mathbf{X}; \theta)$ para cada valor fixo de θ é uma estatística.

Exemplo 1.19.1. *Uma amostra aleatória simples extraída de uma população $N(\mu; \sigma)$ com σ conhecido, a média $\bar{x}$ segue a seguinte distribuição:*

$$\bar{x} \sim N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

Então, a função:

$$G(\mathbf{X}; \theta) = \frac{\bar{x} - \mu}{\sigma \sqrt{n}} \sim N(0; 1)$$

$G(\mathbf{X}; \theta)$ é uma quantidade pivotal para o parâmetro μ dado que depende desse parâmetro, mas sua distribuição de probabilidade, $N(0; 1)$, não depende do parâmetro.

Exercício 1.19.1. *Determine o I.C. para o parâmetro θ se a quantidade pivotal $G(\mathbf{X}; \theta)$ é uma função monótona de θ , Considere $C_1(\alpha)$ e $C_2(\alpha)$ que pertençam ao campo de variação de $G(\mathbf{X}; \theta)$, isso é possível pelo fato de que se pode escolher um nível de confiança $1 - \alpha$. Então:*

$$P[C_1(\alpha) \leq G(\mathbf{X}) \leq C_2(\alpha)] = 1 - \alpha, \quad \forall \theta \in \Theta$$

Se $G(\mathbf{X}; \theta)$ é monótona em θ , podem-se resolver as seguintes equações:

$$G(\mathbf{X}; C_1(\alpha)) \quad G(\mathbf{X}; C_2(\alpha))$$

Obtendo dessa forma os limites $\theta_{inf}(\mathbf{X})$ e $\theta_{sup}(\mathbf{X})$ de forma que o I.C. é:

$$P[\theta_{inf}(\mathbf{X}) \leq \theta \leq \theta_{sup}(\mathbf{X})] = 1 - \alpha$$

Exemplo 1.19.2. *Suponha que uma população de jovens com ansiedade tenha a seguinte função densidade:*

$$f(x; \theta) = \frac{3x^2}{\theta^3}, \quad 0 \leq x \leq \theta$$

O estimador de máxima verossimilhança do parâmetro θ é o maior valor da amostra a_n , sendo a função de distribuição dada por:

$$F(a_n) \left(\frac{a_n}{\theta} \right)^{3n} \quad 0 \leq a_n \leq \theta$$

Sua função de densidade será então:

$$f(a_n) \frac{a_n}{\theta^{3n}} a_n^{3n-1} \quad 0 \leq a_n \leq \theta$$

Determinar o I.C. de 80% para o parâmetro θ .

Solução: Podemos considerar como quantidade pivotal: $G(\mathbf{X}; \theta) = \left(\frac{a_n}{\theta} \right)^{3n}$, que seja uma transformação integral de a_n terá uma distribuição uniforme no intervalo $[0; 1]$, pelo que suas funções de distribuição e densidade, respectivamente, são:

$$F_G(t) = t, \quad 0 \leq t \leq f_G(t) = 1, \quad 0 \leq t \leq 1$$

que não dependem do parâmetro θ . Por tanto,

$$P[C_1(\alpha) \leq G(\mathbf{X}; \theta) \leq C_2(\alpha)] = 1 - \alpha = 1 - 30$$

Para a construção de I.C. para um parâmetro θ pode-se obter, para cada amostra $\mathbf{X}$, um número infinito, pois os valores das caudas α_1 e α_2 cumpre a condição $\alpha_1 + \alpha_2 = \alpha$ são infinitos. Assim, o conjunto das probabilidades das caudas (esquerda e direita) pode ser dado por:

$$\int_{C_2(\alpha)}^1 f_G(t) dt = \int_{C_2(\alpha)}^1 dt = 0,15$$

$$\int_0^{C_1(\alpha)} F_G(t) dt = \int_0^{F_G(\alpha)} dt = 0,15$$

Em que $1 - C_2(\alpha) = 0,15$ e $C_1(\alpha) = 0,15$, ou seja, $C_2(\alpha) = 0,85$ e $C_1(\alpha) = 0,15$.

Assim, podem-se formar as equações:

$$G(\mathbf{X}; \theta) = \left(\frac{a_n}{\theta} \right)^{3a} = C_1(\alpha) = 0,15$$

$$G(\mathbf{X}; \theta) = \left(\frac{a_n}{\theta} \right)^{3a} = C_2(\alpha) = 0,85$$

Obtem-se que $\theta_{inf}(\mathbf{X}) = \frac{a_n}{\sqrt[3a]{0,85}}$ e, $\theta_{sup}(\mathbf{X}) = \frac{a_n}{\sqrt[3a]{0,15}}$. Assim, o intervalo de 70% para o parâmetro θ será:

$$I.C. = \left[\frac{a_n}{\sqrt[3a]{0,85}}; \frac{a_n}{\sqrt[3a]{0,15}} \right]$$

- Método geral

No método anterior implica em bucar e localizar um quantidade pivotal $G(\mathbf{X}; \theta)$, cuja distribuição na amostra seja independente do parâmetro, θ . Esse procedimento nem sempre conduz a uma simples determinação dos extremos do I.C., pelo que, em muitas ocasiões, é interessante utilizar um método alternativo.

No método geral de construção de I.C. não se exige a propriedade do pivot como no método anterior, podendo ter a função amostral do dito parâmetro, estimador que se denomina por θ^- , cuja função de probabilidade na amostra é $h(\theta^-, \theta)$.

Dado um nível de confiança $1 - \alpha$, determinam-se os extremos de um intervalo $C_1(\alpha, \theta)$ e $C_2(\alpha, \theta)$ tais que:

$$P[C_1(\alpha, \theta) \leq \theta^- \leq C_2(\alpha, \theta)] = 1 - \alpha$$

Para uma população contínua teremos:

$$\int_{C_1(\alpha, \theta)}^{C_2(\alpha, \theta)} h(\theta^-, \theta) d\theta^- = 1 - \alpha$$

Ou ainda,

$$\int_{-\infty}^{C_1(\alpha_1, \theta)} h(\theta^-, \theta) d\theta^- = \alpha_1 \quad \int_{C_2(\alpha_2, \theta)}^{\infty} h(\theta^-, \theta) d\theta^- = \alpha_2$$

Sendo que $\alpha_1 + \alpha_2 = \alpha$. Como se observa na Figura 1.14.

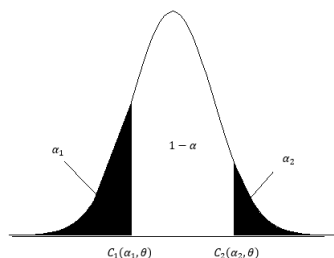


Figura 1.14: Distribuição Normal com as caudas C_1 e C_2

Observa-se que a indicação dos valores α_1 e α_2 (a divisão das áreas da propabilidade) α entre as duas caudas da distribuição, influência na longitude do intervalo obtido.

Ao resolver as integrais anteriores obtem-se os limites: $C_1(\alpha_2, \theta)$ e $C_2(\alpha_2, \theta)$ tal que:

$$P[C_1(\alpha_1, \theta) \leq \theta^- \leq C_2(\alpha_2, \theta)] = 1 - \alpha$$

De onde se deduz as equações integrais:

$$C_1(\alpha_1, \theta) = \theta^- \quad C_2(\alpha_2, \theta) = \theta^-$$

de forma que, se $C_1(\alpha_1, \theta)$ e $C_2(\alpha_2, \theta)$ são funções monótonas em θ , permite encontrar os θ_{inf} e θ_{sup} do I.C. que se procura:

$$I.C. = [\theta_{inf}(\mathbf{X}) \leq \theta \leq \theta_{sup}(\mathbf{X})] \geq 1 - \alpha$$

Exemplo 1.19.3. *Suponha que uma população de jovens com ansiedade tenha a seguinte função densidade:*

$$f(x; \theta) = \frac{5x^2}{\theta^5}, \quad 0 \leq x \leq \theta$$

O esimador de máxima verossimilhança do parâmetro θ é o maior valor da amostra $\theta^- = a_n$, sendo a função de densidade:

$$f(a_n) \frac{5_n}{\theta^{5_n}} a_n^{5_n-1} \quad 0 \leq a_n \leq \theta$$

Determinar o I.C. de 70% para o parâmetro θ pelo método geral.

Para encontrar as equações integrais adotamos o seguinte critério: $\alpha_1 = \alpha_2 = \frac{\alpha}{2} = \frac{1-0,70}{2} = 0,15$. Pelo que, temos:

$$P[C_1(\alpha, \theta) \leq a_n \leq C_2(\alpha, \theta)] = 0,70$$

Que conduz as seguintes integrais:

$$\int_0^{C_1(\alpha, \theta)} \frac{5n}{\theta^{5n}} a_n^{5n-1} dA_n = 0,15$$

$$\int_{C_1(\alpha, \theta)}^0 \frac{5n}{\theta^{5n}} a_n^{5n-1} dA_n = 0,15$$

Ou seja,

$$\left(\frac{C_1(\alpha, \theta)}{\theta} \right)^{5n} = 0,15 \quad 1 - \left(\frac{C_2(\alpha, \theta)}{\theta} \right)^{5n} = 0,15$$

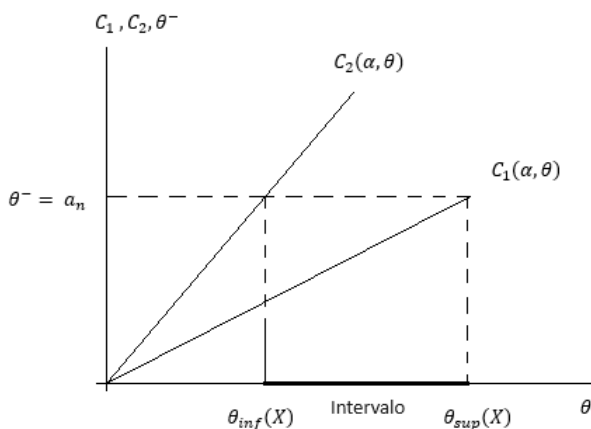


Figura 1.15: Intervalo da coordenada $\theta^- = a_n$

Das equações integrais obtém-se:

$$C_1(\alpha, \theta) = \sqrt[3n]{0,15\theta} = \theta^- = a_n \quad C_2(\alpha, \theta) = \sqrt[3n]{0,85\theta} = \theta^- = a_n$$

Portanto, o I.C. será:

$$I.C. = \left[\frac{a_n}{\sqrt[5n]{0,85}}; \frac{a_n}{\sqrt[5n]{0,15}} \right]$$

Observa-se que o I.C. coincide com o método do pivot para esse parâmetro θ da população considerada.

Exemplo 1.19.4. *Seja uma população definida por: $f(x; \theta) = \frac{2x}{\theta^2}$; $0 \leq x \leq \theta$. Construa mediante o método geral, o I.C. de 90% para o parâmetro θ a partir do estimador de máxima verossimilhança $\hat{\theta} = a_n$, maior valor da amostra, de uma amostra de tamanho 20, cujo valor máximo foi 10.*

Para determinar o I.C. de 90% vamos nos basear na distribuição de probabilidade do estimador de máxima verossimilhança $\hat{\theta} = a_n$ fazendo:

$$P[C_1(\alpha_1; \theta) \leq a_n \leq C_1(\alpha_2; \theta)] = 1 - \alpha = 1 - 0,10 = 0,90$$

A função densidade de probabilidade de a_n é:

$$g(a_n; \theta) = n[F(a_n)]^{n-1} f(a_n)$$

Como

$$F(x) = \int_0^x \frac{2x}{\theta^2} dx = \frac{2}{\theta^2} \left[\frac{x^2}{2} \right]_0^x = \frac{x^2}{\theta^2}, \quad 0 \leq x \leq \theta$$

$$g(a_n; \theta) = 40 \frac{a_n^{39}}{\theta^{40}}; \quad 0 \leq a_n \leq \theta$$

Cuja representação gráfica é:

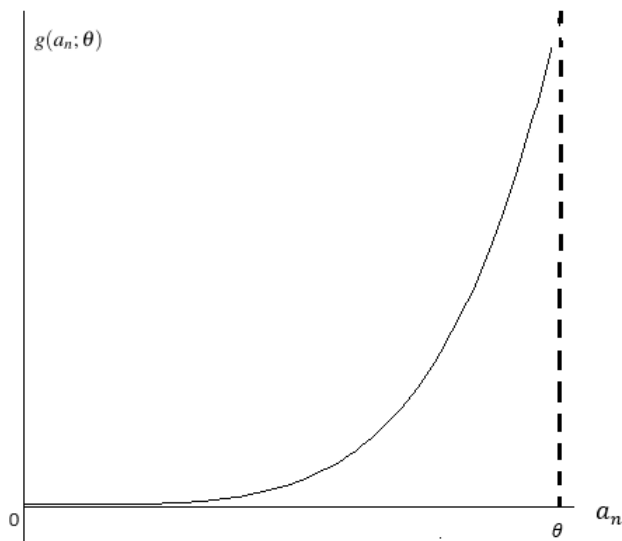


Figura 1.16: Intervalo de confiança para θ

No caso em particular, como $n = 20$, tem-se:

$$g(a_n; \theta) = n \cdot \frac{2n}{\theta^2} \cdot \frac{a_n^{2(n-1)}}{\theta^{2(n-1)}} = \frac{2na_n^{2n-1}}{\theta^{2n}}, \quad 0 \leq a_n \leq \theta$$

Observando a Figura 1.17 da função densidade, o máximo valor da amostra se deduz que o intervalo mais curto se obtém fazendo $\alpha_1 = \alpha$ e $\alpha_2 = 0$. Em consequência, tem-se:

$$P[a_n \leq C_1(\alpha_1; \theta)] = G_a[C_1(\alpha_1; \theta)] = 10\% = 0,10$$

$$P[a_n \leq C_2(\alpha_2; \theta)] = G_a[C_2(\alpha_2; \theta)] = 100\% = 1$$

e como

$$G(a_n) = \int_0^{a_n} 40 \frac{a_n^{39}}{\theta^{40}} dn = \frac{40}{\theta^{40}} \left[\frac{a^{40}}{40} \right]_0^{a_n} = \frac{a^{40}}{\theta^{40}} \quad 0 \leq a_n \leq \theta$$

Assim, temos que:

$$\left[\frac{C_1(\alpha_1; \theta)}{\theta} \right]^{40} = 0,10$$

$$\left[\frac{C_2(\alpha_2; \theta)}{\theta} \right]^{40} = 1$$

Ou seja,

$$C_1(\alpha_1; \theta) = \theta \sqrt[40]{0,10}$$

$$C_2(\alpha_2; \theta) = \theta$$

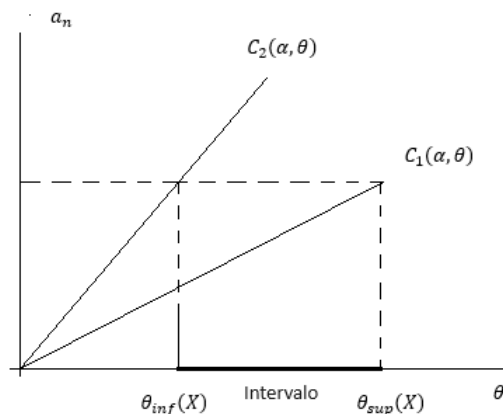


Figura 1.17: Intervalo de confiança para θ

Substituindo $C_1(\alpha_1; \theta)$ e $C_2(\alpha_2; \theta)$ pelo máximo valor obtido na amostra $a_n = 10$, o valor que toma θ em $C_2(\theta)$ constitui o limite inferior do I.C. e $C_1(\theta)$ o limite superior. Logo, o I.C. de θ com 90% será:

$$\left[10; \frac{10}{\sqrt[40]{0,10}} \right] \approx [10; 10,6]$$

- Método da Longitude mínima

Com os mesmos valores amostrais pode-se obter diferentes intervalos com extremos distintos considerando o mesmo nível de confiança $1 - \alpha$. Dentre todos esses intervalos é desejável eleger apenas um, o que apresente menor longitude, de forma que a diferença entre $\theta_{sup} - \theta_{inf}$ seja o menor possível, sempre que se tenha:

$$P[\theta_{inf}(\mathbf{X}) \leq w \leq \theta_{sup}(\mathbf{X})]$$

Para determinar esse intervalo utiliza o método de optimização condicionada dos multiplicadores de Lagrange. Nem sempre é possível determinar o I.C. de nível de confiança $1 - \alpha$ de longitude mínima.

Exemplo 1.19.5. *Suponha que uma população de jovens com ansiedade tenha a seguinte função densidade:*

$$f(x; \theta) = \frac{5x^2}{\theta^5}, \quad 0 \leq x \leq \theta$$

O estimador de máxima verossimilhança do parâmetro θ é o maior valor da amostra a_n , sendo a função de distribuição dada por:

$$f(a_n) \frac{3n}{\theta^{5n}} a_n^{5n-1} \quad 0 \leq a_n \leq \theta$$

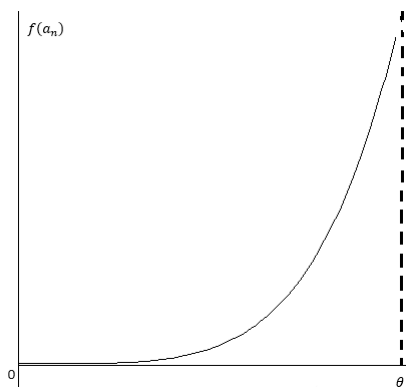


Figura 1.18: Intervalo com limite superior limitado

O intervalo de longitude mínima estaria limitado superiormente pelo limite da variável:

$$P[C_1(\alpha, \theta) \leq a_n \leq \theta] = 1 - \alpha$$

1.20 I.C. assintótico para qualquer parâmetro

Distribuição de Poisson Vamos construir I.C. para qualquer parâmetros do qual se tenha um estimador de máxima verossimilhança. Para isso considera-se o seguinte resultado:

Teorema 1.20.1. *Verifica-se as condições de Fisher-Wolfowitz o estimador de máxima verossimilhança de θ , $\hat{\theta}_{MV}$ é assintoticamente normal.*

$$\frac{\hat{\theta}_{MV}(X_{inf}) - \theta}{\sqrt{I(\theta)^{-1/2}}} \sim N(0; 1)$$

$$\text{Sendo } I(\theta) = -E \left[\frac{\delta^2 \log f(x, \theta)}{\delta \theta^2} \right]$$

Assim, dada uma amostra aleatória simples procedente de uma população cuja distribuição depende de um parâmetro θ desconhecido e supondo conhecido seu estimador de máxima verossimilhança, o I.C. para o dito parâmetro vem dado pela expressão:

$$1 - \alpha = P \left[-Z_{1-\alpha/2} \leq \frac{\hat{\theta}(X_{inf}) - \theta}{\sqrt{I(\theta)}} \leq Z_{1-\alpha/2} \right]$$

Deixando θ no centro da desigualdade temos o seguinte I.C.
:

$$I_{1-\alpha}(\theta) = \left[\hat{\theta}_{MV}(X_{inf}) \pm Z_{1-\alpha/2} \sqrt{I(\theta)} \right]$$

Exemplo 1.20.1. *Seja X_{inf} uma amostra aleatória simples extraída de uma população Poisson de parâmetro desconhecido λ . Sabe-se que o estimador de máxima verossimilhança para λ é $\hat{\lambda}_{MV}(X_{inf}) = \bar{X}$.*

Sabendo que:

Distribuição	Parâmetro	Casos	I.C.
$P(\lambda)$	λ	Grandes amostras	$\bar{X} \pm Z_{1-\alpha/2} \sqrt{\frac{\bar{X}}{n}}$

Tabela 1.11: I.C. para distribuição de Poisson

A quantidade de informação de Fisher para esta distribuição é: $I(\lambda) = \frac{n}{\lambda}$, do qual se deduz:

$$\frac{\bar{X} - \lambda}{\sqrt{\lambda}} \sqrt{n} \sim N(0; 1)$$

Pode-se substituir λ no denominador por um estimador consistente $\bar{X}$ ou resolver diretamente. Assim:

$$\frac{\bar{X} - \lambda}{\sqrt{\bar{X}}} \sqrt{n} \sim N(0; 1)$$

Logo, o I.C. para λ será:

$$I_{1-\alpha}(\lambda) = \left[\bar{X} \pm Z_{1-\alpha/2} \sqrt{\frac{\bar{X}}{n}} \right]$$

Para $\bar{X} = 7$, $Z_{1-\alpha/2} = 1,96$ e $n = 15$. Temos:

$$\begin{aligned}
 I_{1-\alpha}(\lambda) &= \left[\bar{X} \pm Z_{1-\alpha/2} \sqrt{\frac{\bar{X}}{n}} \right] \\
 &= \left[7 \pm 1,96 \sqrt{\frac{7}{15}} \right] \\
 &= [5,66; 8,33]
 \end{aligned}$$

Distribuição	Parâmetro	Casos	I.C.
Bernoulli $B(p)$	p	Grandes amostras	$\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
	p		$\hat{p} = \bar{X}$ e $\hat{q} = 1 - \bar{X}$

Tabela 1.12: I.C. para distribuição de Bernoulli

Distribuição de Bernoulli

Exemplo 1.20.2. Suponha que ($p=1$) indica que a pessoa tem estresse e ($p=0$) que não tem. Sabe-se que a proporção de sucesso (ter estresse) é $\hat{p} = 0,70$, $Z_{1-\frac{\alpha}{2}} = 1,96$ e $n = 20$. Encontre o I.C. para o parâmetro p da distribuição Bernoulli.

$$\begin{aligned}
 I_{1-\alpha}(p) &= \left[\hat{p} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right] \\
 &= \left[0,70 \pm 1,96 \sqrt{\frac{0,70(1-0,70)}{20}} \right] \\
 &= [0,49; 0,90]
 \end{aligned}$$

Distribuição exponencial

Exemplo 1.20.3. *Encontre intervalos de confiança para exponencial com parâmetro λ usando os métodos da desigualdade de Tchebyshev e o métodos assintótico teorema central do limite (T.C.L.) e o método assintótico da máxima verossimilhança (M.V.). Compare os resultado obtidos para um nível de confiança de 95% quando o tamanho amostral seja 120 e a média seja $\bar{X} = 0,72$.*

Método	I.C.	n = 120
Desigualdade de Tchebyshev	$\left[\frac{\bar{X}}{1 + \frac{1}{\sqrt{\alpha n}}}; \frac{\bar{X}}{1 - \frac{1}{\sqrt{\alpha n}}} \right]$	[0,49; 1,30]
T.C.L.	$\left[\frac{\bar{X}}{1 + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}}; \frac{\bar{X}}{1 - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}} \right]$	[0,61; 0,87]
M.V.	$\bar{X} \left(1 - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right); \bar{X} \left(1 + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right)$	[0,59; 0,84]

Tabela 1.13: I.C. para os métodos Tchebyshev, T.C.L. e M.V.

Exercício 1.20.1. *Encontre intervalos de confiança para exponencial com parâmetro λ usando os métodos da desigualdade de Tchebyshev e o métodos assintótico teorema central do limite (T.C.L.) e o método assintótico da máxima verossimilhança (M.V.). Compare os resultado obtidos para um nível de confiança de 95% quando o tamanho amostral seja 160 e a média seja $\bar{X} = 0,81$.*

Método	I.C.	n = 160
Desigualdade de Tchebyshev	$\left[\frac{\bar{X}}{1 + \frac{1}{\sqrt{an}}}; \frac{\bar{X}}{1 - \frac{1}{\sqrt{an}}} \right]$	
T.C.L.	$\left[\frac{\frac{\bar{X}}{Z_{1-\frac{\alpha}{2}}}}{1 + \frac{1}{\sqrt{n}}}; \frac{\frac{\bar{X}}{Z_{1-\frac{\alpha}{2}}}}{1 - \frac{1}{\sqrt{n}}} \right]$	
M.V.	$\bar{X} \left(1 - \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right); \bar{X} \left(1 + \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right)$	

Tabela 1.14: I.C. para os métodos Tchebyshev, T.C.L. e M.V.

1.21 Intervalo de confiança para o correlação entre duas populações normais

Pretende encontrar o I.C. para o coeficiente de correlação populacional *rho* de duas populações com distribuição conjuntamente normal.

1.21.1 Pequenas amostras para correlação

Estimação de ρ (correlação populacional) a partir de:

$$\frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho} \right)$$

Por ln entende-se o logaritmo na base e. O I.C. será então:

$$P \left[\frac{1}{2} \ln \left(\frac{1 + \rho}{1 - \rho} \right) \in \frac{1}{2} \ln \left(\frac{1 + r}{1 - r} \right) - \frac{r}{2(n-1)} \pm Z_{\alpha/2} \frac{1}{\sqrt{n-3}} \right] = 1 - \alpha$$

1.21.2 Grandes amostras para correlação

A estimação de ρ a partir de $\frac{1}{2} \ln \left(\frac{1+r}{1-r} \right)$ terá como I.C.:

$$P \left[\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) \in \frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) \pm Z_{\alpha/2} \frac{1}{\sqrt{n-3}} \right] = 1 - \alpha$$

Até aqui nesta seção foi considerado que as amostras procederam sem reposição o que, em seu efeito, a população em que extraia os dados era ilimitada ou suficientemente grande para buscar uma situação equivalente. Em realidade, a maioria das amostras são aleatórias e os coletivos de onde se extrai a informação constitui populações finitas. Pode-se utilizar fatores de correção do desvio padrão dos estimadores, alterando a margem do erro que estavam considerando nos intervalos. Como norma geral basta multiplicar o erro padrão por $\sqrt{\frac{N-n}{N-1}}$, sendo N o tamanho da população.

Exemplo 1.21.1. *Dado a correlação amostral $r = 0,40$; o tamanho amostral $n = 20$ e $Z_{\alpha/2} = 1,645$. Encontre o I.C. para ρ . Com 90% de confiança, ou seja, com um risco de $\alpha = 10\%$.*

$$P \left[\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) \in \frac{1}{2} \ln \left(\frac{1+0,40}{1-0,40} \right) \pm 1,645 \frac{1}{\sqrt{20-3}} \right] = 1 - 10\%.$$

$$0,0271 \leq \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) \leq 0,8249$$

Assim, o I.C. para correlação populacional, ρ , será: $0,225 \leq \rho \leq 0,678$.

Exemplo 1.21.2. *Seja uma amostra da idade de 14 indivíduos trans-torno do pânico (TP) que varia (0,1 a 10, por exemplo). O TP é caracterizado por crises de ansiedade repentina e intensa com forte sensação de medo ou mal-estar, acompanhadas também de sintomas físicos. Encontrou-se uma correlação igual a 0,83 entre quem*

assiste filme de terror e quem não assiste. Calcule, com 95% de confiança, um I.C. para o coeficiente de correlação entre as duas variáveis.

Com $n = 14$, $r = 0,83$ e $Z_{(\alpha/2)} = 1,96$, temos:

$$\frac{1}{2} \ln \left(\frac{1+r}{1-r} \right) = 1,175 \quad \frac{r}{2(n-1)} = 0,031 \quad \frac{1}{\sqrt{n-3}} = 0,301$$

$$\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) \in 1,175 - 0,031 \pm 1,96 \cdot 0,301$$

$$\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) \in 1,143 \pm 0,590$$

$$0,552 \leq \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) \leq 1,734$$

Sendo as cotas para ρ observando as tabelas:

$$\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) \rightarrow \rho = 0,50$$

$$\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right) \rightarrow \rho = 0,94$$

Exercício 1.21.1. *Seja uma amostra da idade de 16 indivíduos. O transtorno do pânico (TP) que varia (0,1 a 10, por exemplo). Esse transtorno é caracterizado pela crises de ansiedade repentina e intensa com forte sensação de medo ou mal-estar, acompanhadas também de sintomas físicos. Encontrou-se a uma correlação igual a 0,83 entre quem assiste filme de terror e quem não assiste. Calcule, com 95% de confiança, um I.C. para o coeficiente de correlação entre as duas variáveis.*

Exercício 1.21.2. *Seja uma amostra da idade de 20 indivíduos. Suponha que a inteligência musical (varia de 1 a 100) e a inteligência intra pessoal também varia de 1 a 100. Encontrou-se a uma correlação de 0,80 entre essas duas inteligências. Calcule, com 90% de confiança, um I.C. para o coeficiente de correlação entre as duas variáveis (inteligência musical e inteligência intrapessoal).*

Exercício 1.21.3. *Encontrar os seguintes I.C.*

1. Considerando pequenas amostras, encontre o intervalo de confiança de 90% de confiança para $\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right)$. Dado $n = 25$ e $r = 0,87$.
2. Considerando grandes amostras, encontre o intervalo de confiança 90% de confiança para $\frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho} \right)$. Dado $n = 120$ e $r = 0,89$.
3. Seja uma amostra da idade de 14 indivíduos transtorno do pânico (TP) que varia (0,1 a 10, por exemplo). A TP é caracterizado por crises de ansiedade repentina e intensa com forte sensação de medo ou mal-estar, acompanhadas também de sintomas físicos, com os quais se calcula uma correlação igual a 0,83 entre está na residência e fora da residência. Calcule, com 95% de confiança, um I.C. para o coeficiente de correlação entre as duas variáveis.
4. Para obter um I.C. para o coeficiente de correlação entre a renda (re) e consumo de bebidas (be) de homens estressados. Para isso, utilizou-se uma amostra aleatória simples de tamanho 150, resultando em $r_{re,be} = 0,65$. Qual o I.C. ao nível de 95% de confiança. Resposta [0,54; 0,73].

Capítulo 2

Teste de hipóteses

2.1 Relação entre testes de hipóteses e intervalos de confiança

Relação entre testes de hipóteses e intervalos de confiança existe uma correspondência biunívoca entre testes de hipóteses e intervalos de confiança. Esta correspondência permite utilizar um intervalo de confiança para realizar um teste de hipóteses ou utilizar a região crítica de um teste para construir um intervalo de confiança e processa-se da forma seguinte: suponha-se um teste cuja hipótese nula é $H_o : \theta = \theta_o$ a um nível de significância de $(1 - \alpha)100\%$.

Pode-se também construir um intervalo de confiança para θ a um nível de confiança de $100(1 - \alpha)\%$. Então, testa-se H_o , da forma seguinte: se θ_o estiver contido no intervalo calculado, não se rejeita H_o e rejeita-se no caso contrário. Este teste terá um nível de significância de α .

Por outro lado, suponha-se que se construiu uma região crítica do teste de hipóteses. Então, o conjunto de todos os valores de α

que conduzem à não rejeição de H_o formam um intervalo de confiança a um nível de confiança de $100(1 - \alpha)\%$ para θ .

Para que a correspondência entre testes de hipóteses e intervalos de confiança exista, é necessário que o intervalo de confiança tenha a mesma natureza da hipótese alternativa. Assim, se o teste de hipóteses for bilateral, o intervalo de confiança também tem que ser bilateral. Se o teste for unilateral, o intervalo de confiança também tem que o ser e no mesmo sentido.

2.2 Tipos de erros

Nenhum teste de hipótese é 100% correto sem nenhum tipo de erro. Como o teste é baseado em probabilidades, sempre há uma possibilidade de chegar a uma conclusão errada. Ao realizar um teste de hipóteses, dois tipos de erros são possíveis: tipo I e tipo II. Os riscos desses dois erros estão inversamente relacionados e são determinados pelo nível de significância e o poder do teste. Portanto, você deve determinar qual erro tem consequências mais severas para situação antes que você defina seus riscos. Vamos estudar cada um deles. Em outras palavras, quando se utiliza uma estatística de amostra para tomar decisões sobre um parâmetro da população, existe sempre um risco de se chegar a uma conclusão incorreta. Esses dois tipos diferentes de erro podem ocorrer quando aplicamos a metodologia do teste de hipóteses.

Erros possíveis num teste de hipóteses quando se testa H_o contra H_1 , podem ser cometidos os erros indicados como se observa na Tabela 2.1, em que os erros tipo I e II são também designados por erros de primeira e de segunda espécie respectivamente.

2.2.1 Erro tipo I e erro tipo II

- Quando a hipótese nula é verdadeira e você a rejeita, comete um erro do tipo I. A probabilidade de cometer um erro do tipo I é α , que é o nível de significância que você definiu para seu teste de hipóteses. Um α de 0,05 indica que você quer aceitar uma chance de 5% de que está errado ao rejeitar a hipótese nula. Para reduzir este risco, você deve usar um valor inferior para α . Entretanto, usar um valor inferior para alfa significa que você terá menos probabilidade de detectar uma diferença verdadeira, se existir uma realmente.
- Quando a hipótese nula é falsa e você não a rejeita, comete um erro de tipo II. A probabilidade de cometer um erro de tipo II é β , que depende do poder do teste. Você pode diminuir o risco de cometer um erro do tipo II, assegurando que o seu teste tenha potência suficiente. Você pode fazer isso garantindo que o tamanho amostral seja grande o suficiente para detectar uma diferença prática, quando realmente existir uma.

O complemento ($1 - \alpha$) da probabilidade de um erro do tipo I é chamado de coeficiente de confiança, que, ao ser multiplicado por 100% produz o nível de confiança. A probabilidade de rejeitar a hipótese nula sendo falsa é igual a $1 - \beta$. Esse valor é a potência do teste.

População Amostra		Realidade (desconhecida)	
		H_0 é verdadeira	H_0 é falsa
D e c i s ã o	Rejeitar H_0	Erro do tipo I $1 - \alpha$	Decisão correta β
	Não rejeitar H_0	Decisão correta α	Erro do tipo II $1 - \beta$

Figura 2.1: Possíveis erros de acertos de uma decisão

Probabilidades de erro e potência de um teste - supondo primeiro hipóteses simples, $H_0 : \theta = \theta_0$ e $H_1 : \theta = \theta_1$, as probabilidades dos erros indicados são:

$$\begin{aligned}
 P(\text{Erro tipo I}) &= P(\text{Rejeitar } H_0 | H_0 \text{ verdadeiro}) \\
 &= P[(X_1, \dots, X_n) \in RC | \theta = \theta_0] = \alpha
 \end{aligned}$$

O erro tipo I só poderá ser cometido quando se rejeitar H_0 , enquanto que o Erro tipo II poderá ocorrer quando se aceitar H_0 .

$$\begin{aligned}
 P(\text{Erro tipo II}) &= P(\text{Não Rejeitar } H_0 | H_0 \text{ verdadeiro}) \\
 &= P[(X_1, \dots, X_n) \notin RC | \theta = \theta_0] = \beta
 \end{aligned}$$

2.3 Teste de hipóteses através do valor p

O valor p é a probabilidade de ocorrência de um valor da estatística de teste igual ao extremo de que o respectivo valor observado para a amostra particular $(x_1, \dots, x_n)$, supondo que a hipótese nula

é verdadeira. O valor p é também o nível de significância associado ao valor observado da estatística de teste. Outra forma para realizar um teste de hipóteses é através do cálculo de probabilidade de se obter um valor igual ou mais extremos do que o valor obtido pelo pesquisador, dado que a hipótese nula é verdadeira. Essa probabilidade é denominada valor p (p-valor) e um valor p pequeno significa uma das situações:

1. Um psicólogo observou um resultado pouco provável de ocorrer, supondo a hipótese nula verdadeira ou,
2. A hipótese nula é falsa. Na prática rejeitamos a hipótese nula quando o valor p é menor do que o nível de significância (α) fixado pelo pesquisador, o qual representa a probabilidade de rejeição indevida da hipótese nula.

Bicaudal - Temos as seguintes conclusões:

- Se $-t_{\alpha/2} \leq t_{cal} \leq t_{\alpha/2}$, aceita H_o .
- Se $t_{cal} > t_{\alpha/2}$ ou $t_{cal} < -t_{\alpha/2}$, rejeita-se H_o .

Unicaudal à esquerda - Temos as seguintes conclusões:

- Se $t_{cal} > -t_{\alpha}$, aceita H_o .
- Se $t_{cal} < -t_{\alpha}$, rejeita-se H_o .

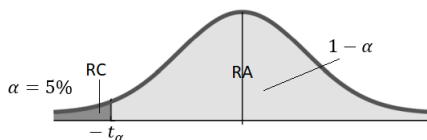


Figura 2.2: Distribuição t de Student unicaudal à esquerda com nível de significância 5%

Unicaudal à direita : Temos as seguintes conclusões:

- Se $t_{cal} < t_{\alpha}$, aceita H_0 .
- Se $t_{cal} > t_{\alpha}$, rejeita-se H_0 .

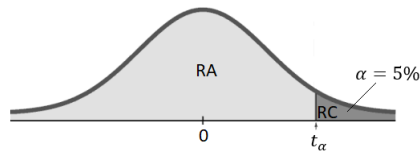


Figura 2.3: Distribuição t de Student unicaudal à direita com nível de significância 5%

Inicialmente, vamos estudar três testes de hipóteses são eles:

- Teste t para uma média
- Teste t para duas amostras independentes
- Teste t para duas amostras relacionadas

2.3.1 O teste de hipótese para uma média

Um dos testes mais importantes dentro do ramo da estatística inferencial é o teste t de Student. O teste t de Student (ou T de Student) para duas amostras é usado para testar se dois grupos (duas populações) são diferentes em termos de uma variável quantitativa, com base na comparação de duas amostras retiradas destes dois grupos. Em outras palavras, esse teste para duas amostras permite determinar se as duas populações das quais suas duas amostras são extraídas são diferentes. As duas amostras sendo medidas em uma variável quantitativa contínua.

O objetivo é desenvolver métodos de testes de hipóteses como uma técnica para alcançar diferenças e tomar decisões em psicologia. Considere uma população da qual retiramos uma amostra $x_1, x_2, \dots, x_n$. Estamos interessados em realizar inferência sobre a média populacional μ , para isso utilizaremos tamanho amostral (n), média amostral $\bar{x}$, desvio padrão amostral s e variância amostral s^2 . Conhecendo os parâmetros populacionais obteremos estimar os parâmetros amostrais usando uma amostra aleatória simples.

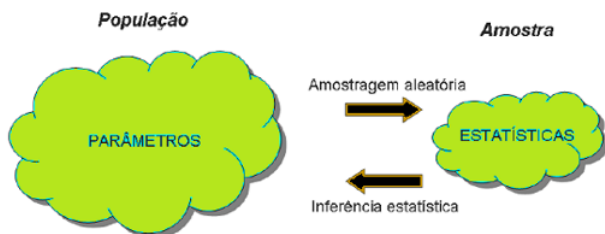


Figura 2.4: População (N) e amostra (n)

Se não conhecemos o valor do desvio padrão populacional σ e a amostra é pequena, $n < 30$, nesse caso utiliza-se a distribuição t de Student.

$$t = \frac{\bar{x} - \mu_o}{\frac{s}{\sqrt{n}}}$$

Ou seja, uma distribuição t de Student com $n - 1$ graus de liberdade. O teste de hipótese passo a passo é o seguinte:

1. Enunciar as hipóteses H_o e H_a

$$\begin{cases} H_o : \text{hipótese nula} \\ H_a : \text{hipótese alternativa} \end{cases}$$

A hipótese nula: é a hipótese que sugere um valor para o parâmetro populacional ou a igualdade dos parâmetros em teste. E a hipótese alternativa é a hipótese que sugere que a afirmação que estamos fazendo na hipótese nula é falsa. No exemplo a seguir se gue alguns exemplos:

Exemplo 2.3.1. *Considere as seguintes hipóteses:*

- *A média de níveis de estresse de 10 idosos (amostra aleatória) após pandemia:*

$$\begin{cases} H_o : \mu = 3, \\ H_a : \mu \neq 3, \end{cases}$$

- *Média da altura desses idosos:*

$$\begin{cases} H_o : \mu = 1,67, \\ H_a : \mu \neq 1,67, \end{cases}$$

2. Fixar α . Escolher a variável t de Student com gl grau de liberdade, que será $gl = n - 1$.
3. Com auxílio da tabela de distribuição t de Student, determinam-se RA (região de aceitação) e RC (região de crítica, ou região de rejeição).

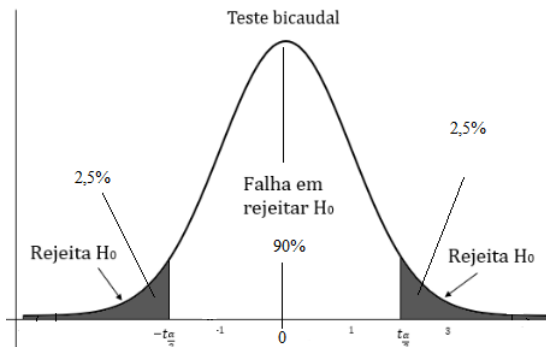


Figura 2.5: Distribuição t de Student bicaudal informando as áreas de aceitação e rejeição com nível de significância 5%

4. Cálculo do valor da variável: t_{cal} e o cálculo do grau de liberdade (gl).
5. Conclusão:
 - Se $t_{cal} > t_{\alpha/2}$ ou $t_{cal} < t_{\alpha/2}$, rejeita-se H_o , com risco α .
 - Se $t_{\alpha/2} \leq t_{cal} \leq t_{\alpha/2}$, não se pode rejeitar H_o , com risco α .

Esse teste de hipótese procura calcular as médias de subgrupos e estatísticas univariadas relacionadas para variáveis dependentes dentro das categorias de uma ou mais variáveis independentes. Podemos obter também uma análise de variância, o teste eta η^2 , que mensura a porção da variância total da variável dependente que, pode ser atribuída a variável de grupo (variável independente). O valor eta é usado como medida do efeito da análise de variância. E também o teste de linearidade.

Exemplo 2.3.2. *Para medir o nível de estresse em três escolas num*

determinado bairro violento eram diferentes. Realizou-se uma análise de variância de um fator para ver se as três médias diferem.

No menu (Analisar - Comparações de Médias) fornece a soma, o número de casos, a média, a mediana, a mediana combinada, o erro padrão da média, valor mínimo, valor máximo, o intervalo, o valor da variável para a primeira categoria da variável de agrupamento, o valor da variável para a última categoria da variável de agrupamento, o desvio padrão, a variância, a curtose e seu erro padrão, a assimetria e seu erro padrão, o porcentagem da soma total, a porcentagem do total (N), a média geométrica e a média harmônica.

As variáveis dependentes são variáveis independentes categóricas. Os valores das variáveis categóricas podem ser numéricos ou string curta (poucas letras juntas). Algumas das estatísticas de subgrupo opcionais como a média e o desvio padrão típicos que são baseados na teoria da distribuição normal e que são adequados para variáveis quantitativas com distribuições simétricas.

Neste capítulo estudaremos quatro procedimentos SPSS:

2.3.2 Teste T para uma média no SPSS

O procedimento One-Sample T-Test testa se a média de uma única variável difere de uma média constante especificada.

Exemplo 2.3.3. *Um pesquisador deseja testar se a pontuação média de QI de um grupo de alunos difere de 100 com um nível de confiança de 95%. Para cada variável contrastante, fornece também a média, o desvio padrão, o erro padrão da média, a diferença média entre cada valor de dados e o valor do contraste hipotético. Também testa se a média é diferença é 0 e faz um intervalo de confiança para a diferença média (para o qual você pode especificar qualquer nível de confiança).*

- No menu: Analisar - Comparação de média - teste T para uma amostra.
- Selecione uma ou mais variáveis para testá-la com o mesmo valor hipotético para a média.

Exemplo 2.3.4. *Testar a hipótese que a média vale 46*

As hipóteses seriam:

$$\begin{cases} H_o : \mu = 46 \\ H_a : \mu \neq 46 \end{cases}$$

- Insira um valor numérico de contraste (46, por exemplo) para comparar com a média de cada amostra.
 - Podemos clicar em Opções para controlar o tratamento de dados perdidos e o nível do intervalo de confiança.
- (a) Intervalo de confiança - Por padrão, ele mostra um intervalo de confiança. 95% para a diferença entre a média e o valor de contraste hipotético. Inserimos um valor entre 1 e 99 para calcular outro nível de confiança diferente do contraste da média igual a 46.
- (b) Valores ausentes - Se quisermos contrastar várias variáveis e se houver dados ausentes em uma ou mais variáveis, indica-se o procedimento de quais casos devem ser incluídos (ou excluídos).

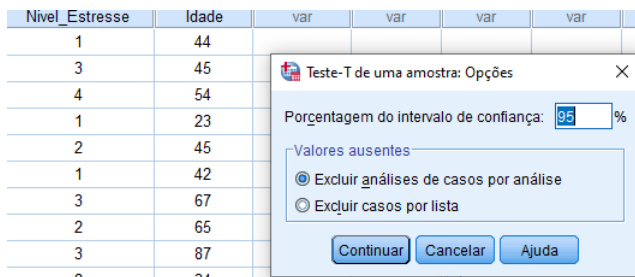


Figura 2.6: Opções para o teste T para uma amostra

- Excluir casos de acordo com a análise. Cada teste T usa todos os casos que tem dados válidos para a variável contrastada. Os tamanhos das amostras podem variar de um teste para outro.
- Excluir casos de acordo com uma lista. Cada teste T usa apenas os casos que tem dados válidos para todas as variáveis usadas em qualquer um dos testes T de Student solicitados. O tamanho da amostra é constante em todos os testes.

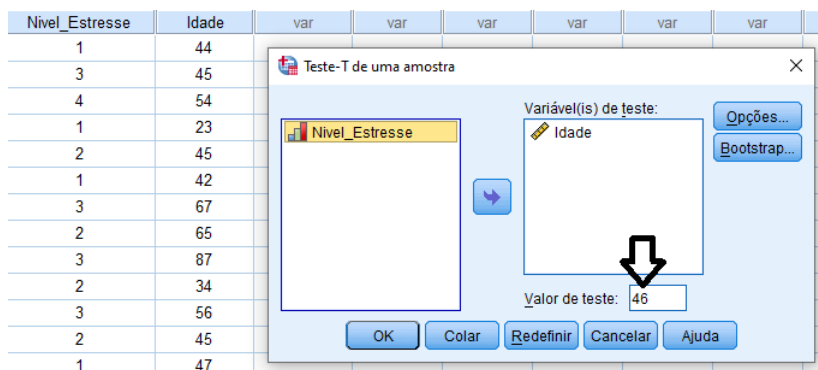


Figura 2.7: Menu: teste T para uma amostra

O resultado será o seguinte:

Teste-T

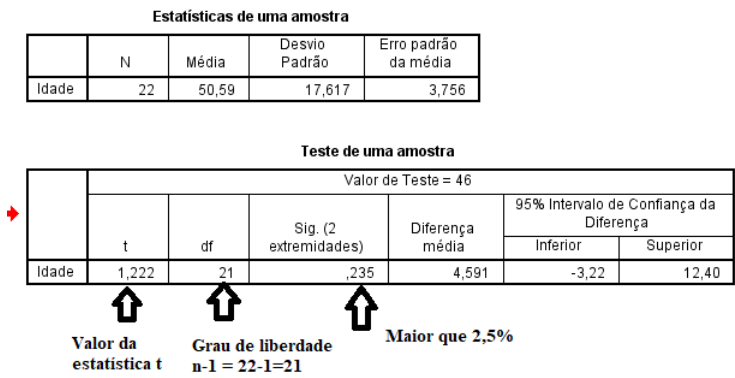


Figura 2.8: Output do teste T para uma amostra

Observar que para contrastar os valores de uma **variável quantitativa** com um valor de contraste de hipótese, temos que escolher uma variável quantitativa e introduzir um valor de contraste de hipóteses. Este teste assume que os dados **são normalmente distribuídos**; no entanto, este teste é bastante robusto contra desvios da normalidade.

Exemplo 2.3.5. Foi selecionada uma amostra aleatória de tamanho 8. Suponha que foi retirada de uma população com distribuição normal. Os valores são: 20,3; 19,8; 20,3; 19,7; 19,8; 19,7; 19,8;...; 19,8. Teste a hipótese que a média é inferior a 20. Use nível de significância de 5%.

$$\begin{cases} H_o : \mu = 20 \\ H_1 : \mu < 20 \end{cases}$$

Sendo $s = 4$, $n = 16$. Então, temos $1,64 = \frac{\bar{x}_c - 20}{4/\sqrt{(16)}}$. Isolando o valor de $\bar{x}_c$ na equação, temos: $\bar{x}_c = 21,64$.

Assim, a regra de decisão para H_o será:

- Rejeitar H_o , quando $\bar{x}_c > 21,64$
- Aceitar H_o , quando $\bar{x}_c \leq 21,64$

Como $H_1 : \mu > 20$, devemos ter um limite crítico à direita para valores de $\bar{x}$. Como se observa na Figura 2.9.

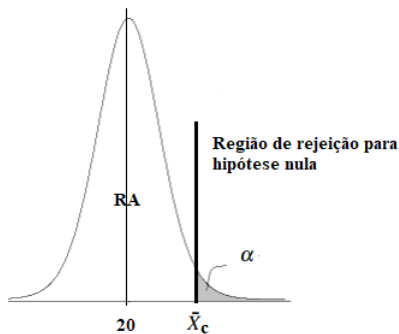


Figura 2.9: Limite crítico para aceitar ou rejeitar a hipótese nula

Na Figura 2.10 temos a distribuição da média amostral para testar $H_o : \mu = 20$. Observa-se que para valores de $\bar{x}$ próximos de 20 a hipótese nula poderá ser aceita.

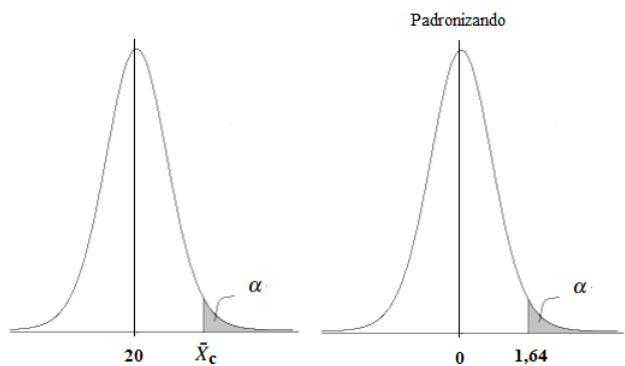


Figura 2.10: Limite para aceitar ou rejeitar a hipótese nula

2.3.3 Resumo de teste de hipótese para uma média

Temos duas condições para esse teste de hipótese:

- 1. Variância populacional desconhecida

A estatística de teste será:

$$t_{cal} = \frac{\bar{x} - \mu_o}{\frac{s}{\sqrt{n}}}$$

H_o	H_1	Rejeita-se H_o	Ao nível α	p valor
$\mu \leq \mu_o$	$\mu > \mu_o$	$t_o > t_{(n-1;\alpha)}$	$p_o \leq \alpha$	$p_o = P[t \geq t_o]$
$\mu \geq \mu_o$	$\mu < \mu_o$	$t_o < -t_{(n-1;\alpha)}$	$p_o \leq \alpha$	$p_o = P[t \leq t_o]$
$\mu = \mu_o$	$\mu \neq \mu_o$	$ t_o > t_{(n-1;\alpha/2)}$	$p_o \leq \alpha$	$p_o = 2P[Uma\ cauda]$

t denota uma variável aleatória t_{n-1} e $t_{(n-1;\alpha)}$ a abscisa de-
baixo dessa densidade deixa a sua direita uma área α .

Exemplo 2.3.6. Sabendo que $\bar{x} = 20$, $s = 4$, $n = 8$ e a esta-
tística de teste:

$$t_{cal} = \frac{\bar{x} - \mu_o}{\frac{s}{\sqrt{n}}} = \frac{\bar{x} - \mu_o}{\frac{s}{\sqrt{n}}}$$

Contrastar as hipóteses:

$$\begin{cases} H_o : \mu = 19 \\ H_1 : \mu < 19 \end{cases}$$

$$t_{cal} = \frac{\bar{x} - \mu_o}{\frac{s}{\sqrt{n}}} = \frac{20 - 19}{\frac{4}{\sqrt{8}}} = 0,7071$$

O valor $t_{n-1,5\%} = 1,89$, logo aceita-se a hipótese nula, pois $1,89 > 0,7071$ ($t_{tabelado} > t_{calculado}$), ou seja, o valor calculado está na área de aceitação, como se verifica na Figura abaixo.



Figura 2.11: Limite para aceitar ou rejeitar a hipótese nula

2. Variância populacional conhecida

A estatística de teste será:

$$z_{cal} = \frac{\bar{x} - \mu_o}{\frac{\sigma}{\sqrt{n}}}$$

H_o	H_1	Rejeita-se H_o ao nível α	p valor
$\mu \leq \mu_o$	$\mu > \mu_o$ $z_o > z_\alpha$	$p_o \leq \alpha$	$p_o = P[Z \geq z_o]$
$\mu \geq \mu_o$	$\mu < \mu_o$ $z_o < -z_\alpha$	$p_o \leq \alpha$	$p_o = P[Z \leq z_o]$
$\mu = \mu_o$	$\mu \neq \mu_o$ $ z_o > z_{\alpha/2}$	$p_o \leq \alpha$	$p_o = 2P[\text{uma cauda}]$

Z denota uma variável aleatória $N(0,1)$ e a abscisa debaixo dessa densidade deixa a sua direita uma área α .

Exemplo 2.3.7. *Seja uma amostra $n = 25$ indivíduos que apresentam transtorno de ansiedade. A média de tempo (em dias) que se apresenta essa transtorno de ansiedade é $\bar{x} = 97$. Pretende-se contrastar se a média populacional do tempo de transtorno de ansiedade é 96,5 a um nível de significância de $\alpha = 5\%$. Sabendo que $\sigma = 2$. Pretende-se testar:*

$$\begin{cases} H_o : \mu = 96,5 \\ H_1 : \mu \neq 96,5 \end{cases}$$

A estatística de teste é $z_{cal} = \frac{\bar{x} - \mu_o}{\frac{\sigma}{\sqrt{n}}}$ que, sob H_o , segue uma distribuição normal padronizada. A região crítica do teste é: $|z_o| > z_{\alpha/2}$. Sendo, neste caso, $z_{\alpha/2} = z_{0,05/2} = z_{0,025} = 1,96$ (Observar na tabela Z padronizado). Assim, o valor de z_o será:

$$z_{cal} = \frac{\bar{x} - \mu_o}{\frac{\sigma}{\sqrt{n}}} = \frac{97 - 96,5}{\frac{2}{\sqrt{25}}} = 1,25$$

Assim, aceita-se H_o com um risco de $\alpha = 5\%$, pois o valor $z_{cal} = 1,25$ está na área de aceitação. Ou seja, $z_{cal} < z_{tabelado}$. Esboce um gráfico para justificar essa evidência estatística.

2.4 Exploração dos dados: teste de normalidade e igualdade de variâncias

O procedimento Explore no SPSS gera estatísticas e gráficos de resumo, para todos os casos ou separadamente para grupos de casos. Existem inúmeras razões para usar este procedimento, como para inspecionar os dados, identificar outliers, obter descrições, testar suposições e caracterizar diferenças entre subpopulações (grupos de casos). A inspeção dos dados pode mostrar que existem valores incomuns, valores extremos, descontinuidades nos dados ou outras peculiaridades. Explorar os dados pode nos ajudar a determinar se eles são as técnicas estatísticas que você está considerando para a análise apropriada dos dados em psicologia. A varredura pode indicar que você precisa transformar os dados se a técnica precisa de uma distribuição normal. Ou, o usuário pode decidir que ele precisa usar testes não paramétricos.

Por exemplo: Observamos a distribuição dos tempos de aprendizagem para labirintos de uma série de ratos submetidos a quatro esquemas diferentes de reforço. Para cada um dos quatro grupos, podemos ver se a distribuição dos tempos é aproximadamente normal e se as quatro variâncias são iguais. Também podemos identificar os casos com os cinco maiores e os cinco menores valores de tempo. Diagramas de caixa (box plot) e diagramas de caule e folha resumem graficamente a distribuição de tempo aprendizagem para cada grupo.

Este menu nos fornece uma série de estatísticas como a média, a mediana, 5% de média aparada (trimedia), erro padrão, variância, desvio padrão, mínimo, máximo, amplitude, amplitude interquartil, assimetria e curtose e seus erros valores, o intervalo de confiança para a média (e o nível de confiança especificado), o percentis, estimador M de Huber, estimador de onda de Andrews, estimador M re-descendente de Hampel, estimador de dois pesos de Tukey, os cinco valores maior e os cinco menores, a estatística Kolmogorov-Smirnov com o nível de significância de Lilliefors para testar a normalidade e os gráficos estatísticos e boxplot de Shapiro Wilks, gráficos de caule e folhas, histograma, gráficos de normalidade e gráficos de dispersão por nível com testes de Levene e suas transformações. Este menu nos fornece uma série de estatísticas como a média, a mediana, 5% de média aparada, erro padrão, variância, desvio padrão, mínimo, máximo, amplitude, amplitude interquartil, assimetria e curtose e seus erros valores, o intervalo de confiança para a média (e o nível de confiança especificado), o percentis, estimador M de Huber, estimador de onda de Andrews, estimador M re-descendente de Hampel, estimador de dois pesos de Tukey, os cinco valores maiores e os cinco menores, a estatística Kolmogorov-Smirnov com o nível de significância de Lilliefors para testar a normalidade e os gráficos estatísticos e boxplot de Shapiro Wilks, gráficos de caule e folhas, histograma, gráficos de normalidade e gráficos de dispersão por nível com testes de Levene e suas transformações. No SPSS: **Analisar - Estatística descritiva - Explorar.**

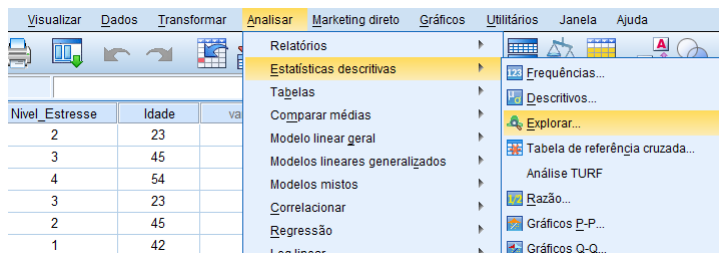


Figura 2.12: Analisar - descritiva e explorar

Como já vimos as estatísticas descritivas, vamos explorar os gráficos e os testes de normalidade e o teste de igualdade de variância. O gráfico de testes de normalidade mostra os gráficos de probabilidade normal e sem tendência. A estatística é exibida Kolmogorov-Smirnov com um nível de significância de Lilliefors para Normalidade de contraste. Se especificarmos pesos não inteiros, irá calcular a estatística Shapiro-Wilks.

2.5 Normalidade univariada

O teste de Shapiro-Wilks é definido da seguinte forma:

$$V = \frac{\left(\sum_{i=1}^G a_{i,n} (X_{n-i+1} - X_{(i)}) \right)^2}{(n-1)S^2}$$

Em que:

- (a) $k = \frac{n}{2}$ se n é par, e $k = \frac{n-1}{2}$ se n é ímpar.
- (b) X_i é a estatística de ordem i -ésimo.

(c) $a_{i,n}$ coeficientes tabulados segundo o tamanho amostral.

A estatística V de Shapiro-Wilks contrasta-se os seguintes testes de hipóteses:

$$\begin{cases} H_0 : X \text{ é normal} \\ H_1 : X \text{ não é normal} \end{cases}$$

Propriedades:

- Verifica-se que dando a igualdade a 1 apenas no caso de que a amostra seja a replica de uma normal.
- Os valores do estatística de Shapiro-Wilks encontram-se tabulados. Na tabela aparecem os pontos críticos $(V_{n,\alpha})$, tal que $P[V \leq V_{n,\alpha}] = \alpha$.

Usaremos um exemplo para verificar os passos do teste de normalidade de Shapiro-Wilks.

Considere os valores de transtorno psicóticos (suponha que o nível varia de 1 a 20) de 11 pessoas. Pergunta-se X segue uma distribuição normal?

$$X = [19.8, 20.5, 19.7, 17.6, 19.2, 18.4, 18.1, 19.1, 17.9, 17.3, 20]$$

Vamos contrastar a hipótese de normalidade da população da qual se extraiu a amostra X usando o Teste de Shapiro-Wilks. Para isso se observa os seguintes passos.

1. Se ordena os 11 valores amostrais em sentido decrescente:
(20.5, 20, 19.8, 19.7, 19.2, 19.1, 18.4, 18.1, 17.9, 17.6, 17.3).

2. Obtém-se as diferenças entre os valores equidistante do centro:
- [20.5 – 17.3 20 – 17.6 19.8 – 17.9 19.7 – 18.1 19.2 – 18.4].
3. Multiplica-se cada uma destas diferenças pelo correspondente coeficiente apresentados por colunas como se observa abaixo.

n										
i	2	3	4	5	6	7	8	9	10	11
1	0,7071	0,7071	0,6872	0,6646	0,6431	0,6233	0,6052	0,5888	0,5739	0,5601
2		0	0,1677	0,2413	0,2806	0,3031	0,3031	0,3244	0,3291	0,3315
3				0	0,0875	0,1401	0,1743	0,1976	0,2141	0,226
4						0	0,0561	0,0947	0,1224	0,1429
5								0	0,0399	0,0695
6										0

Como $n = 11$ temos,

$x_{(n-i+1)}-x_{(i)}$	$a_{i,n}$	$a_{i,n}((n-i+1)-x_{(i)})$
20,5-17,3=3,2	0,5601	$0,5601 \cdot (3,2) = 1,7923$
20,0-17,6=2,4	0,3315	0,7956
19,8-17,9=1,9	0,2260	0,4294
19,7-18,1=1,6	0,1429	0,2286
19,2-18,4= 0,8	0,0695	0,0556

4. Observa-se o valor calculado da estatística V . Antes a variância amostral é:

$$S^2 = \frac{1}{n-1} \left[\sum_{i=1}^{11} x_i^2 - \frac{\left(\sum_{i=1}^{11} x_i \right)^2}{n} \right]$$

$$S^2 = \frac{1}{10} \left[19,8^2 + \dots + 20^2 - \frac{(19,8 + \dots + 20)^2}{11} \right] = 1,14$$

Logo, V será:

$$V = \frac{\left(\sum_{i=1}^k a_{i,n} (X_{n-i+1} - X_{(i)}) \right)^2}{(n-1)S^2}$$

$$V = \frac{(0,5601 \cdot (3,2) + \dots + 0,0695 \cdot (0,8)^2)}{10 \cdot (1,14)} = 0,9493$$

Usando a Tabela Shapiro-Wilks vemos que $V_{11,5\%} = 0,85$.

5. Logo, $v_o = 0,9493 > V_{11,5\%} = 0,85$, ao nível de 5%, assim não há evidências estatística para rejeitar a hipótese de que a amostra tenha sido extraída de uma população normal. Assim, X foi extraído de uma população normal, com 5% de incerteza.

Os passos do teste Shapiro Wilks no SPSS seriam:

- Análise - Estatística descritiva - Explorar
- Inserir a variável IDADE

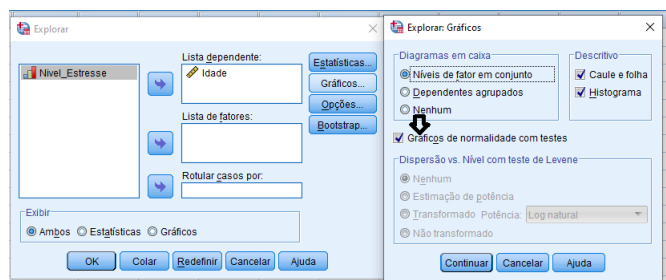


Figura 2.13: Verificando a normalidade da variável Idade

Conferir o resultado abaixo com o valor calculado acima ($V = 0,949$).

Testes de Normalidade ^a						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estatística	df	Sig.	Estatística	df	Sig.
Dados	,144	11	,200 [*]	,950	11	,640

*. Este é um limite inferior da significância verdadeira.

a. Correlação de Significância de Lilliefors

Figura 2.14: Output do teste de Shapiro Wilks

Outro teste para verificar a normalidade é o Teste de Kolmogorov-Smirnov. O procedimento de Kolmogorov-Smirnov para uma amostra compara a função de distribuição cumulativa observada de uma variável com uma distribuição normal teórica (pode ser feito com outras distribuições) O Kolmogorov-Smirnov Z é calculado da maior diferença (em valor absoluto) entre as funções de distribuição acumulada teórica e observada. Este teste de ajuste testa se as observações podem razoavelmente vir da distribuição normal.

Para testar a hipótese de que a função de distribuição da variável é $F(x)$ calcula-se a estatística:

$$D_n = \max |F_n(x) - F_x|$$

Cuja distribuição, quando a hipótese declarada é verdadeira, é tabulada. Se a distância calculada D_n for maior que a encontrada nas tabelas, para um valor de α , é rejeita que a distribuição da variável seja $F(x)$. O teste de Lielliefors é usado na teste de normalidade quando a média e a variância não são conhecidas e são substituídas para suas estimativas. O contraste que é feito é:

$$\begin{cases} H_o : F(x) \text{ é normal} \\ H_1 : F(x) \text{ não é normal} \end{cases}$$

Se o valor p for menor que o nível de significância α , ou seja, $valor p < 0,05$ no qual o teste é realizado chegaríamos à conclusão de rejeitar a hipótese nula.

2.5.1 Dispersão por nível com o teste de Levene

Controlar a transformação dos dados para os gráficos de dispersão por nível. Para todos os diagramas de dispersão por nível mostram a inclinação da linha de regressão e testes robustos de Levene sobre a homogeneidade de variação. Se selecionarmos uma transformação, os testes de Levene são com base nos dados transformados. Se não selecionarmos nenhuma variável de fator, nenhum gráfico de dispersão de nível será gerado.

A estimativa de potência produz um gráfico dos logaritmos naturais das amplitudes interquartis em relação aos logaritmos naturais das medianas de todas as células, bem como uma estimativa da transformação de energia necessária para alcançar variações iguais

nas caixas. Um gráfico de dispersão de nível ajuda a determinar a potência necessária para uma transformação estabilizar (equalizar) a variações do grupo. A opção de transformações permite selecionar uma das alternativas de energia, seguindo as recomendações do estimativa de potência, e gera gráficos dos dados transformados. Eu sei plote o intervalo interquartil e a mediana dos dados transformados. Não transformado gera gráficos dos dados brutos. É equivalente a uma transformação com uma potência de 1. No SPSS seria:

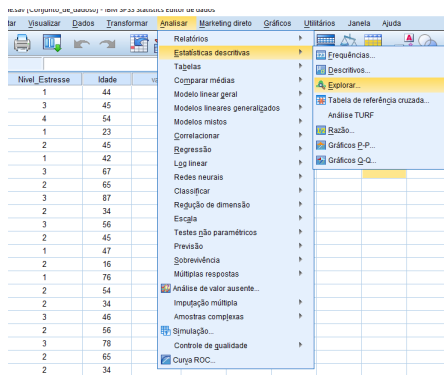


Figura 2.15: Analisar - descritiva e explorar

As vezes é necessário que as variáveis consideradas tenham variâncias iguais, ou seja, hipótese de homocedasticidade. Os testes F de Snedecor são baseados nas variâncias amostrais e o teste de Bartlett assume que as variâncias consideradas tem distribuições normais e é sensível a desvios de normalidade.

Já o teste de Levene é bastante robusto a desvios da normalidade e a estatística de teste é calculada a partir da análise de variância entre grupos e dentro do grupo. Vamos realizar o seguinte contraste, o teste de hipótese:

$$\left\{ \begin{array}{l} H_o : \text{há igualdade de variâncias (homocedasticidade)} \\ H_1 : \text{não há igualdade de variâncias (heterocedasticidade)} \end{array} \right.$$

Se o p valor for **menor que o nível de significância** (α) rejeitamos o hipótese de igualdade de variâncias (homocedasticidade).

Outras opções nesse menu são: as estatísticas descritivas, o intervalo de confiança para média, estimadores M, valores discrepantes e percentis. Como se observa na Figura abaixo.



Figura 2.16: Explorar - Estatísticas

Em **Gráfico** temos: diagrama em caixa, descritivo: ramo e folha, histograma; dispersão versus nível com teste de Levene, Gráfico de normalidade com teste.

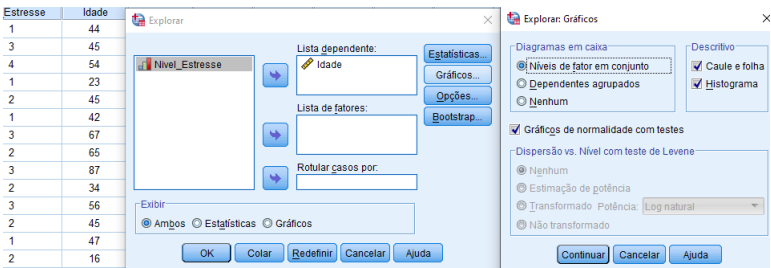


Figura 2.17: Explorar - Gráficos

Click em Opções para manipular valores ausentes:

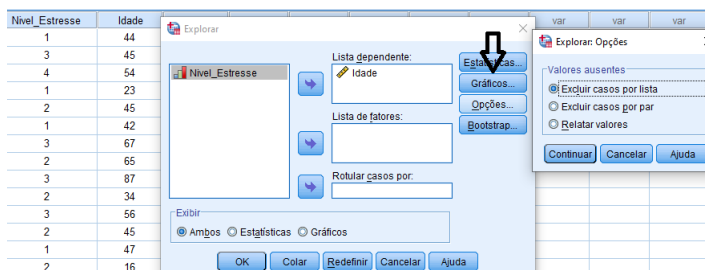


Figura 2.18: Valores ausentes

Temos os seguintes casos usando o SPSS:

- Excluir casos de acordo com uma lista. Nesse caso valores ausentes para qualquer variável fator ou variável dependente são excluídas de todas as análises. Este é o valor padrão.
- Excluir casos de acordo com o pareamento.
- Mostrar valores. Os valores ausentes das variáveis do fator são tratados como categoria separada. Os valores ausentes para uma variável fator são incluídos, mas são rotulados como ausentes.

2.6 Teste de hipóteses para duas amostras

Existem vários tipos do teste para duas amostras, dependendo se as amostras são independentes ou se são dependentes, ou seja, amostras pareadas, e se as variâncias das populações são (des)iguais e/ou (des)conhecidas:

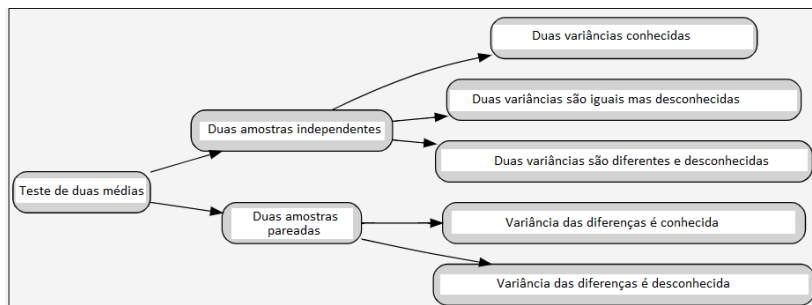


Figura 2.19: Principais testes de hipóteses para duas amostras

2.7 Teste de hipóteses para duas amostras independentes

Exemplo 2.7.1. Dadas duas amostras (A e B) extraídas das populações P_1 e P_2 , o valor obtido para T indica a diferença entre as amostras. As duas populações são homocedástica, ou seja, tem variâncias iguais. Considere as quantidades de alunos ansiosos e com depressão em 5 escolas A e em 5 escolas B . As amostras foram selecionadas aleatoriamente. As das duas populações são independentes e suponha que as variâncias em ambas as populações sejam conhecidas. Testar se as duas médias populacionais são diferentes.

Escolas	A	A	A	A	A	B	B	B	B	B
Valores	13	14	11	14	15	18	16	19	19	18

Exemplo 2.7.2.

Assim temos:

- 5 elementos em cada grupo, ou seja, $n_A = n_B = 5$

- A média do grupo A é $\bar{x} = 13,4$
- A média do grupo B é $\bar{x} = 18$
- O variância populacional de cada grupo vale $\sigma_A^2 = 2,3012$ e $\sigma_B^2 = 1,5006$, respectivamente.

Segue os passos para realizar esse teste de hipóteses:

1. $H_o : \mu_1 - \mu_2 = 0$ e $H_1 : \mu_1 \neq \mu_2$. Queremos testar se as duas médias são diferentes, observe que não impomos uma direção de maior ou menor no teste, assim o teste será bicaudal.
2. A estatística de teste é:

$$S_{conj} = \sqrt{\frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2} \left[\frac{1}{n_A} + \frac{1}{n_B} \right]}$$

$$t_{cal} = \frac{(\bar{x}_A - \bar{x}_B)}{S_{conj}}$$

Aplicando aos dados, temos:

$$S_{conj} = \sqrt{\frac{(5-1)(1,517)^2 + (5-1)(1,225)^2}{5+5-2} \left[\frac{1}{5} + \frac{1}{5} \right]} = 0,8719$$

$$t_{cal} = \frac{(13,4 - 18)}{0,8719} = -5,275$$

A diferença da média é $(13,4 - 18) = -4,6$. O grau de liberdade assumindo que as variâncias são iguais vale $(n_A - n_B - 2 = 5 - 5 - 2 = 8$. O grau de liberdade considerando que as variâncias são diferentes vale: 13,98. O erro padrão da média é $\sigma_A/\sqrt{n_A} = 0,517/\sqrt{5} = 0,225$ e $\sigma_B/\sqrt{n_B} = 1,225/\sqrt{5} = 0,548$.

3. Como $\pm t_{\alpha/2} = \pm t_{5\%/2} = \pm t_{0,025} = \pm 2,7515$. Como se observa na Figura 2.20.

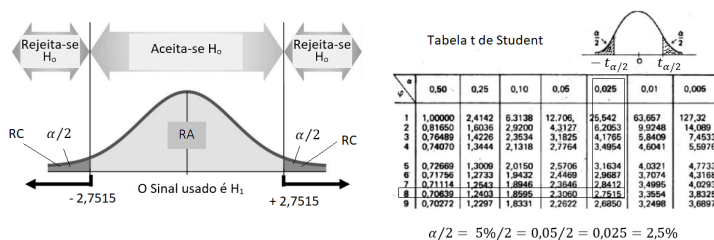


Figura 2.20: Parte da Tabela t de Student

4. Ao nível de significância de 5%, não rejeitamos a hipótese de que as duas médias populacionais são iguais, ou não há evidência suficiente nos dados para concluir que as duas populações consideradas são diferentes, pois o valor -0,063 está na região de rejeição da hipótese nula e aceitação da hipótese alternativa. Logo, A e B tem médias de quantidades de alunos ansiosos e com depressão diferentes.

Observa-se que o valor encontra-se fora da região de aceitação, ou seja, $-5,275 < -2,7515$. Assim, rejeitamos a hipótese H_0 de que existe evidências estatísticas de que tem diferença de médias.

No SPSS seria:

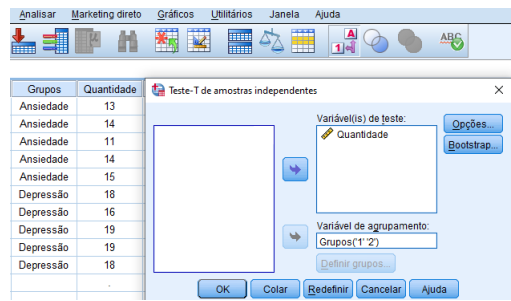


Figura 2.21: Teste para duas amostras independentes com variâncias conhecidas

Conferindo os resultados anterior.

➔ Teste-T

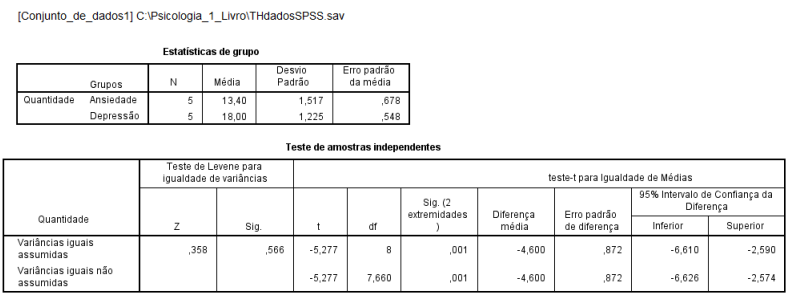


Figura 2.22: Saída do SPSS para o teste para duas amostras independentes com variâncias conhecidas

Exemplo 2.7.3. Dadas duas amostras (C e D) extraídas das populações P_1 e P_2 , o valor obtido para T indica a diferença entre as amostras. As duas populações são homocedástica, ou seja, tem variâncias iguais. Considere as quantidades de alunos com estresse em 5 escolas C e em 5 escolas D . As amostras foram selecionadas

aleatoriamente. As das duas populações são independentes e suponha que as variâncias em ambas as populações sejam conhecidas. Testar se as duas médias populacionais são diferentes.

Escolas	C	C	C	C	C	D	D	D	D	D
Valores	12	145	10	14	15	20	17	18	20	16

Exemplo 2.7.4.

2.8 Teste de hipóteses para diferenças de médias populacionais com variâncias populacionais desconhecidas

Suponha que as duas amostras sejam oriundas de duas populações com distribuição normal independente e variâncias desconhecidas e desiguais. Neste caso, a estatística de teste da desigualdade entre as duas médias será dada por:

$$t_{cal} = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

a qual tem uma distribuição aproximada t de Student com graus de liberdade dado por:

$$gl = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{(\frac{S_1^2}{n_1})^2/(n_1 - 1) + (\frac{S_2^2}{n_2})^2/(n_2 - 1)}$$

No caso do valor do gl ser fracionário, utilizamos um número inteiro resultado do arredondamento para menos.

Exemplo 2.8.1. *As seguintes medidas correspondem ao tempo em hora de tratamento de seções psicológicas. O grupo de controle, ou seja, sem apoio psicológico. Na tabela abaixo temos o tamanho da amostra (n), a média, o desvio padrão e a variância do tempo em hora.*

Tratamento	n	$\bar{x}$	s	s^2
Seis meses	5	20,8h	0,87h	$(0,87h)^2 = 0,756h^2$
Controle	3	18,7h	3,1h	$(3,1h)^2 = 9,61h^2$

Verificar se existe efeito significativo do tratamento psicológico quanto as médias de hora do tratamento.

Para a comparação entre as duas médias relativas ao tratamento e ao controle, temos as seguintes hipóteses: $H_o : \mu_{Trat} = \mu_{Cont}$ versus $H_o : \mu_{Trat} \neq \mu_{Cont}$. Adote um nível de significância de 5%, ou seja, $\alpha = 10\%$.

O passos para realizar esse teste de hipóteses serão:

1. $H_o : \mu_1 - \mu_2 = 0$ e $H_1 : \mu_1 \neq \mu_2$. Queremos testar se as duas médias são diferentes, não impomos uma direção de maior ou menor no teste, assim o teste será bicaudal.
2. A estatística de teste é:

$$t_{cal} = \frac{(20,8 - 18,7)}{\sqrt{\frac{0,87}{5} + \frac{9,61}{3}}} = 1,1427$$

a qual tem uma distribuição aproximada t de Student com

graus de liberdade dado por:

gl = ((0,87/5 + 9,81/3)^2) / ((0,87/5)^2/(5-1) + (9,61/3)^2/(3-1)) = 2,21

3. O valor crítico seria $t_{\alpha;gl} = t_{\alpha/2;gl} = t_{0,10/2;2,21} = 4,303 = t_{tab}$.
Observe a tabela abaixo:

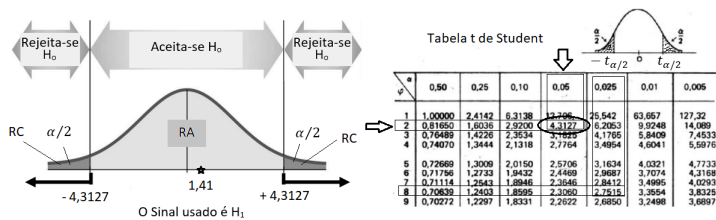


Figura 2.23: Teste para duas amostras independentes com variâncias desiguais

4. Como $4,303 = t_{tab}$, o valor $t_{cal} = 1,1427$ está dentro intervalo que pertence a região de aceitação, ou seja, $[-4,303; +4,303]$.
5. Conclusão: Aceitamos a hipótese nula ao nível de significância de 10%, concluímos que o tratamento psicológico não forneceu um resultado diferente do controle, ao nível de significância de 10%. (CUIDADO: O tamanho amostral é muito pequeno).

Exercício 2.8.1. Dado os grupos 1 e 2 testar as seguintes hipóteses: $H_o : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$. Adotando nível de significância de 5%.

Grupo 1	25	24	25	20	24
Grupo 2	20	24,5	19		

Exercício 2.8.2. Dado os grupos 1 e 2 testar as seguintes hipóteses: $H_o : \mu_1 = \mu_2$ versus $H_1 : \mu_1 > \mu_2$. Adotando nível de significância de 5%.

Grupo 1	25	24	25	20	24
Grupo 2	20	24,5	19		

Exercício 2.8.3. Dado os grupos 1 e 2 testar as seguintes hipóteses: $H_o : \mu_1 = \mu_2$ versus $H_1 : \mu_1 > \mu_2$. Adotando nível de significância de 5%.

Grupo 1	31	20	26	19	24	25	28	24	29
Grupo 2	24	21,5	19,6	29	27	28			

2.9 Teste de hipótese para variância populacional

Seja $X_1, X_2, \dots, X_n$ uma amostra aleatória de tamanho n retirada de uma população normal com média populacional μ e desvio padrão populacional σ . Suponha que desejamos testar uma hipótese sobre a variância σ^2 desta população.

Para isso calcular-se a variável:

$$\chi_{cal}^2 = \frac{(n-1)S^2}{\sigma_o^2}$$

Sendo n o tamanho da amostra, S^2 a variância amostral e σ_o^2 o valor da hipótese nula. Pretende-se testar:

$$\begin{cases} H_0 : \sigma^2 = \sigma_o^2 \\ H_1 : \sigma^2 \neq \sigma_o^2, \sigma^2 > \sigma_o^2, \sigma^2 < \sigma_o^2 \end{cases}$$

Fixar α . Escolher a variável Qui-Quadrada com $gl = (n - 1)$.

Com auxílio da tabela da distribuição Qui-Quadrado, determinam-se RA e RC (ou RR região de rejeição).

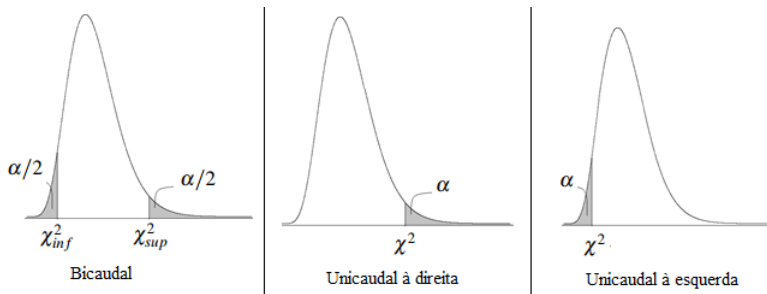


Figura 2.24: Qui-Quadrado unicaudal e bicaudal

Essas regiões críticas são válidas quando as hipóteses nula é: $H_0 : \sigma^2 \leq \sigma_o^2$ ou $H_0 : \sigma^2 \geq \sigma_o^2$. Segue o teste de hipóteses para σ^2 quando μ é conhecido e quando μ é desconhecido.

1. Teste de hipóteses para σ^2 quando μ é conhecido

A estatística de teste será:

$$\chi_{cal} = \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma_o^2}$$

H_o	H_1	Rejeita-se H_o ao nível α	p valor
$\sigma^2 \leq \sigma_o^2$	$\sigma^2 > \sigma_o^2$	$\chi_{cal}^2 > \chi_{n;\alpha}^2$	$p_o = P[C \geq \chi_{cal}^2]$
$\sigma^2 \geq \sigma_o^2$	$\sigma^2 < \sigma_o^2$	$\chi_{cal}^2 < \chi_{n;1-\alpha}^2$	$p_o = P[C \leq \chi_{cal}^2]$
$\sigma^2 = \sigma_o^2$	$\sigma^2 \neq \sigma_o^2$	$\chi_{cal}^2 > \chi_{n;\alpha/2}^2$	$p_o = 2P[\text{uma cola}]$
		$\chi_{cal}^2 < \chi_{n;1-\alpha}^2$	

C denota uma variável aleatória com distribuição χ_n^2 e $\chi_{n;\alpha}^2$ a abscissa debaixo dessa densidade que deixa a sua direita uma área α .

2. Teste de hipóteses para σ^2 quando μ é desconhecida

A estatística de teste será:

$$\chi_{cal} = \frac{(n-1)s^2}{\sigma_o^2}$$

H_o	H_1	Rejeita-se H_o ao nível α	p valor
$\sigma^2 \leq \sigma_o^2$	$\sigma^2 > \sigma_o^2$	$\chi_{cal}^2 > \chi_{n-1;\alpha}^2$	$p_o = P[C \geq \chi_{cal}^2]$
$\sigma^2 \geq \sigma_o^2$	$\sigma^2 < \sigma_o^2$	$\chi_{cal}^2 < \chi_{n-1;1-\alpha}^2$	$p_o = P[C \leq \chi_{cal}^2]$
$\sigma^2 = \sigma_o^2$	$\sigma^2 \neq \sigma_o^2$	$\chi_{cal}^2 > \chi_{n-1;\alpha/2}^2$	$p_o = 2P[\text{uma cola}]$
		$\chi_{cal}^2 < \chi_{n-1;1-\alpha}^2$	

C denota uma variável aleatória com distribuição χ_{n-1}^2 e $\chi_{n-1;\alpha}^2$ a abscissa debaixo dessa densidade que deixa a sua direita uma área α .

Exemplo 2.9.1. Considere uma amostra aleatória de tamanho $n = 25$. A variância amostral dessa amostra é igual a 0.2209. Sabendo que a média populacional μ é desconhecida. Testar as seguintes hipóteses:

$$\begin{cases} H_o : \sigma^2 = 0,25 \\ H_1 : \sigma^2 > 0,25 \end{cases}$$

A região crítica do teste é $\chi_{cal}^2 = \chi_{n-1;\alpha}^2$. A estatística de teste é:

$$\chi_{cal} = \frac{(n-1)s^2}{\sigma_o^2} = \frac{(25-1)0,2209}{0,25} = 21,20$$

O valor tabela será: $\chi_{n-1;\alpha}^2 = \chi_{25-1;0,10}^2 = 33,19$. O valor p será: $p_o = P[\chi_{cal}^2 \leq 21,20] = 0,6265 > 0,10$. Assim, aceita-se H_o com bastante evidência estatística.

Exercício 2.9.1. Considere uma amostra aleatória de tamanho $n = 31$. Sabendo que a média populacional μ é desconhecida, o desvio padrão amostral é igual a 4,56. Fazendo $\alpha = 5\%$ testar as seguintes hipóteses:

$$\begin{cases} H_o : \sigma^2 = 0,30 \\ H_1 : \sigma^2 > 0,30 \end{cases}$$

Exercício 2.9.2. Considere uma amostra aleatória de tamanho $n = 36$. A variância populacional dessa amostra é igual a 0,40 e a média populacional μ vale 4. Dado também que $\sum_1^{36} (x_i - 4)^2 = 8$. Pede-se

a) A região crítica do teste

b) Testar as seguintes hipóteses:

$$\begin{cases} H_o : \sigma^2 \leq 0,60 \\ H_1 : \sigma^2 > 0,60 \end{cases}$$

2.10 Teste de hipótese para T pareado

Consideremos duas amostras dependentes $x_1, \dots, x_n$ e $y_1, \dots, y_n$. Neste caso consideraremos observações pareadas, isto é, podemos considerar que temos na realidade uma amostra de pares $(x_1, y_1), \dots, (x_n, y_n)$. Define-se que $D_i = x_i - y_i$, para $i = 1, 2, \dots, n$. Assim obteremos a amostra $D_1, \dots, D_n$, resultante das diferenças entre os valores de cada par. Como as amostras são dependentes, vamos considerar que $D_i \sim N(\mu_D, \sigma_D^2)$. Para realizar o teste T pareado devemos primeiramente enunciar as seguintes hipóteses:

$$\begin{cases} H_0 : \mu_D = 0 \\ H_1 : \mu_D \neq 0, \mu_D > 0, \mu_D < 0 \end{cases}$$

O parâmetro μ_D será estimado pela média amostral das diferenças, ou seja $\bar{D}$ o parâmetro, σ_D^2 será estimado pela variância amostral das diferenças, ou seja,

$$s_D^2 = \frac{\sum_{i=1}^n (D_i - \bar{D})^2}{n-1}$$

O teste será realizado pela expressão:

$$t = \frac{\bar{D} - \mu_D}{\frac{s_D}{\sqrt{n}}}$$

Segue uma distribuição T de Student com $n-1$ graus de liberdade. Tudo o que foi dito para o teste T comum serve para o teste T pareado, basta substituir a média por μ_D e o desvio padrão amostral por s_D . Com isto, temos que a um nível de significância α .

Os pontos críticos são determinados por $t_{\alpha/2}$ e $-t_{\alpha/2}$ para o caso

bilateral, $t\alpha$ para o caso unilateral à direita e $-t\alpha$ para o unilateral à esquerda.

Observação: A medição se realiza sobre as mesmas pessoas, portanto, não temos duas amostras aleatórias independentes, mas somente uma.

Exemplo 2.10.1. *Encontrar evidência a favor de que o tratamento contra ansiedade surta o efeito desejado (baixa do colesterol). Os valores de antes e depois do tratamento se encontram abaixo (dados fictícios).*

Antes	200	210	330	240	260	300	245	210	190	225
Depois	150	200	275	250	200	250	200	180	190	205
Diferença	50	10	55	-10	60	50	45	30	0	20

Enuciando as hipóteses:

$$\begin{cases} H_o : \mu_D = 0 \\ H_1 : \mu_D > 0 \end{cases}$$

Se $\mu_D = 0$ significa que não houve nenhuma melhora no nível de colesterol em virtude da ansiedade. Se $\mu_D > 0$ houve algum bom resultado.

Elevando os valores das diferenças ao quadrado, temos:

Diferenças											Soma
D_i	50	10	55	-10	60	50	45	30	0	20	310
D_i^2	2500	100	3025	100	3600	2500	2025	900	0	400	15150

$$\bar{D} = \frac{310}{10} = 31$$

$$s_D = \sqrt{\frac{10.15150 - (310)^2}{10(10-1)}} = \sqrt{615,55} = 24,81$$

O erro padrão da média é calculado por:

$$s_{\bar{D}} = \frac{s_D}{\sqrt{n}} = \frac{24,81}{\sqrt{10}} = 7,84$$

O grau de liberdade é: $t_{n-1} = t_{10-1} = t_9$ (distribuição t de Student com 9 grau de liberdade).

O valor de t_{cal} será:

$$t_{cal} = \frac{31 - 0}{7,84} = 3,951$$

O valor calculado encontra-se claramente na região crítica (região de rejeição) do teste ($t_{cal} > t_{ab}$), portanto, concluímos que existe evidência estatisticamente significativa contrária à hipótese nula e a favor da hipótese alternativa (ao menos com um nível de significância de 5%). O valor $t_{(9,5\%)} = 1,8331$ (veja na tabela t-Student para o caso unilateral com 9 grau de liberdade).

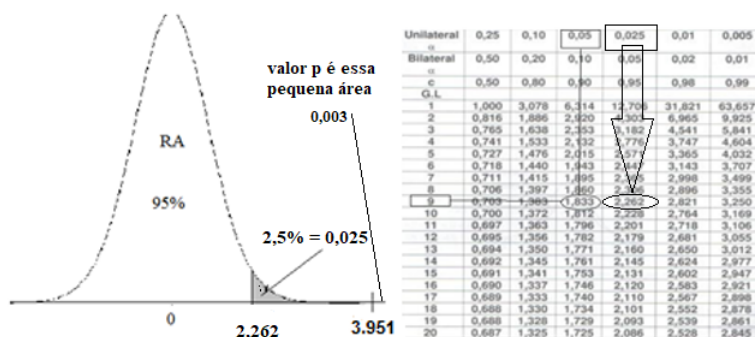


Figura 2.25: Distribuição t de Student unilateral

Rejeitar H_o , pois $0,003 < 0,05$. Também observa-se que $t_{cal} = 3,951 > t_{tab} = 2,262$. Observe também que a área (valor p) é 0,003 está dentro da área 0,025, ou seja, $0,025 > 0,003$. Assim está na região de rejeição (RR ou RC) como se observa na figura acima.

Exemplo 2.10.2. O presidente de uma grande empresa pretende implementar um programa de atividade desportiva para os empregados com intuito de diminuir a ansiedade nos empregados. Assim, foram registrados os níveis de estresse no ano anterior e no ano seguinte ao do início do programa de 15 pessoas que aderiram ao mesmo selecionados ao acaso. Suponha distribuição normal os níveis de estresse referidos, pronuncie-se sobre a eficácia do programa através de um intervalo de confiança a 95%.

Empregados	1	2	3	4	5	6	7	8
Antes	104	190	101	133	153	176	128	130
Depois	85	130	75	110	100	89	102	98
Empregados	9	10	11	12	13	14	15	
Antes	80	53	123	101	173	102	72	
Depois	45	46	92	84	165	81	55	

$$P\left(-t_{0,025} \leq \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} \leq +t_{0,025}\right) = 1 - 5\%$$

A média amostral é $\bar{D} = 30,8$ e o desvio padrão é $S_D = 21,27$. A estimativa pontual de $\mu_D = \bar{D} = 30,8$ e, sendo o desvio padrão de D desconhecido, terá que ser estimado a partir da amostra, ou seja, $\hat{\sigma}_D = 21,273$. O intervalo de confiança dado por:

$$\left[\bar{D} - t_{\alpha/2} \frac{S_D}{\sqrt{n}}; \bar{D} + t_{\alpha/2} \frac{S_D}{\sqrt{n}} \right]$$

Em que $t_{\alpha/2}$ é o quantil apropriado da distribuição $t_{(n-1)}$.

$$P\left(-t_{0,025} \leq \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} \leq +t_{0,025}\right) = 1 - 5\%$$

Sabe-se que $t_{0,025;14} = 2,14$. Assim, o I.C. será:

$$\left[30,8 - 2,14 \frac{21,273}{\sqrt{15}}; 30,8 + 2,14 \frac{21,273}{\sqrt{15}} \right] = [19,05; 42,55]$$

Estima-se com 95% de confiança que a diferença média dos níveis de estresse referidos antes e depois do programa de atividade desportiva esteja compreendido entre 19,05 e 42,55.

Testar também as seguintes hipóteses:

$$\begin{cases} H_o : \mu_D = 0 \\ H_1 : \mu_D > 0 \end{cases}$$

No SPSS seria:

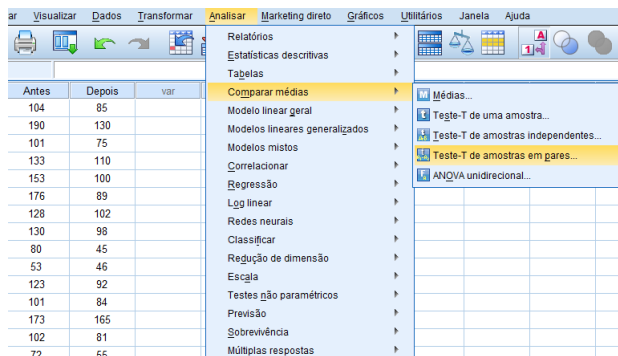


Figura 2.26: 1Passos no SPSS

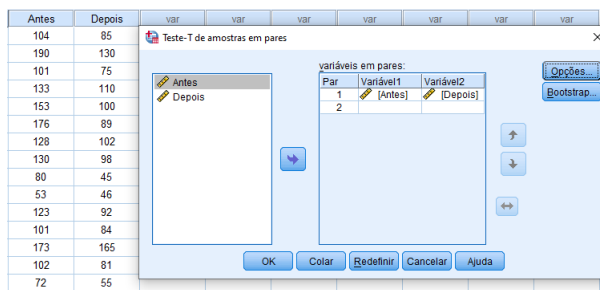


Figura 2.27: Passos no SPSS continuação

O resultado seria:

➔ Teste-T

Estatísticas de amostras emparelhadas

	Média	N	Desvio Padrão	Erro padrão da média	
Par 1	Antes	121,27	15	39,593	10,223
	Depois	90,47	15	31,039	8,014

Correlações de amostras emparelhadas

	N	Correlação	Sig.
Par 1 Antes & Depois	15	,846	,000

Teste de amostras emparelhadas

Diferenças emparelhadas						t	df	Sig. (2 extremidades)
	Média	Desvio Padrão	Erro padrão da média	95% Intervalo de Confiança da Diferença				
				Inferior	Superior			
Par 1 Antes - Depois	30,800	21,268	5,491	19,022	42,578	5,609	14	,000

Figura 2.28: Output do SPSS

$$T = \frac{\bar{D}}{\frac{S_D}{\sqrt{n}}} \sim t_{(n-1)}$$

Exemplo 2.10.3. Considere as medições sanguíneas sistólicas pré-tretamento e pós-tratamento de um problema de ansiedade.

Tabela 2.1:			
Pré-tratamento	Pós-tratamento	diferenças	d ²
217	209	8	8 ² = 64
252	241	11	11 ² = 121
229	230	1	1 ² = 1
200	208	8	8 ² = 64
209	206	3	3 ² = 9
213	211	2	2 ² = 4

No R seria:

```
A<-c(217,252,229,200,209,213)
D<-c(209,241,230,208,206,211)
d<-A-D; d
```

```

n<-length(A)
md<-mean(d); md
s<-sd(d); s
d^2
t<-md/(s/sqrt(6)); t
Li<- md-qt(0.95,n-1)*s/sqrt(n); Li
Ls<- md+qt(0.95,n-1)*s/sqrt(n); Ls
To<-md/(s/sqrt{n}); To

```

$$T_o < -\frac{\bar{D}}{s_d/\sqrt{n}} = 0,9118 \quad (2.1)$$

Não se verifica a condição de rejeição e se conclui que não se pode rejeitar H_o a nível de $\alpha = 5\%$, de maneira que os dados são indicativos de que o tratamento não tem propriedade redutora de ansiedade ao nível fixado. A mesma conclusão chega-se mediante o p-valor dado por:

$$p_o = P[T \geq 0,9118] = 0,2018 > 0,05 \quad (2.2)$$

2.10.1 Resumo de teste de hipóteses para a diferenças de médias populacionais relacionadas

Como vimos que a estatística de teste é:

$$t_{cal} = \frac{\bar{D}}{\frac{s_D}{\sqrt{n}}}$$

H_o	H_1	Rejeita-se H_o ao nível α	p valor
$\mu_1 - \mu_2 \leq 0$	$\mu_1 - \mu_2 > 0$	$t_o > t_{n-1;\alpha}$ $p_o \leq \alpha$	$p_o = P[T \geq t_o]$
$\mu_1 - \mu_2 \geq 0$	$\mu_1 - \mu_2 < 0$	$t_o > -t_{n-1;\alpha}$ $p_o \leq \alpha$	$p_o = P[T \leq t_o]$
$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq 0$	$ t_o > t_{n-1;\alpha/2}$ $p_o \leq \alpha$	$p_o = 2P[\text{uma cauda}]$

T denota uma variável aleatória com distribuição t_{n-1} e $t_{n-1;\alpha}$ a abscissa debaixo dessa densidade que deixa a sua direita uma área α .

Exemplo 2.10.4. *Uma empresa farmaceutica está interresada na investigação preliminar de um novo medicamento que podia ter propriedades redutoras de colesterol. Com essa finalidade selecionou-se uma amostra aleatória de 6 indivíduos comparáveis, e se determinou nível de colesterol antes e depois do tratamento. Os resultados foram os seguintes:*

Antes	220	252	229	200	209	213
Depois	209	241	230	208	206	211
Diferença	8	11	-1	-8	3	2

Assumindo que a diferença entre os níveis de colesterol é uma variável aleatória normal. As hipóteses a serem testadas serão: $H_o : \mu_1 = \mu_2$ contra $H_1 : \mu_1 > \mu_2$.

A média da diferença é: $\bar{D} = 3$. O desvio padrão da diferença é: $s_D = 7,29$ e $n = 6$. Assim, tem-se:

$$T_o < -\frac{\bar{D}}{s_d/\sqrt{n}} = 1,0074 \tag{2.3}$$

```
D<-c(11,11,-1,-8,3,2)
mD<-mean(D)
```

```

Sd<-sd(D)
n<-6
t<- mD / (Sd/sqrt(n)) ; t
1-pt(1.007491,n-1) % valor p
[1] 0.179969

```

T_o sob H_o , segue uma distribuição t_5 . Fixando $\alpha = 5\%$ a região crítica do teste é $t_o > t_{5,5\%} = 2,015$. Assim, $t_{cal} < t_{tab}$, logo aceita-se a hipótese nula. Observe também que $p_o = P[T \geq 1,0074] = 0,179969$, uma área maior que 0,05, logo aceita-se H_o com um risco de 5%.

2.11 Teste para uma proporção populacional

Feita uma afirmação sobre uma proporção (p), deseja-se estabelecer um critério para aceitar ou não a hipótese, com base nos dados de uma amostra aleatória simples. Os passos são os seguintes:

1. Enunciar as hipóteses H_o e H_a

$$\begin{cases} H_o : p = p_o \\ H_1 : p \neq p_o, \quad p > p_o, \quad p < p_o \end{cases}$$

2. Fixa-se o nível α
3. Retira-se uma amostra de tamanho n e define-se x : número de sucesso, calculando $\hat{p}_o = \frac{x}{n}$.
4. Determina-se a proporção dado no teste de hipótese H_o , $\sigma_{\hat{p}} =$

$$\sqrt{\frac{p_{H_o} \cdot q_{H_o}}{n}}$$

5. Define-se como variável critério: $Z = \frac{\hat{p}_o - p_{H_0}}{\sigma_{\hat{p}}}$
6. Com auxílio da tabela de distribuição Z normal, determinam-se RA e RC.
7. Conclusão:
 - Se $-Z_{\alpha/2} \leq Z_{cal} \leq +Z_{\alpha/2}$, aceita-se H_0
 - Se $Z_{cal} > Z_{\alpha/2}$ ou $Z_{cal} < -Z_{\alpha/2}$, rejeita-se H_0 .

Exemplo 2.11.1. *Deseja-se testar a igualdade entre as proporções de homens e mulheres que têm diversidade de motivo (fisiológico: fome, sede, esquivar à dor). Os resultados estão na Tabela abaixo. As amostras foram aleatórias independentes de homens e mulheres:*

1. Enunciar as hipóteses

Tabela 2.2: Número de homens e mulheres

Homens	Mulheres
$x_1 = 70$	$x_2 = 50$
$n_1 = 200$	$x_2 = 200$

Sendo x_1 o número de homens e x_2 o número de mulheres. Admita $\alpha = 10\%$.

$$\begin{cases} H_0 : p_1 = p_2 \\ H_1 : p_1 \neq p_2 \end{cases}$$

2. Para $\alpha = 10\%$. Escolher a variável normal padronizada.
3. Com auxílio da tabela de distribuição Z normal, determinam-se RA e RC.

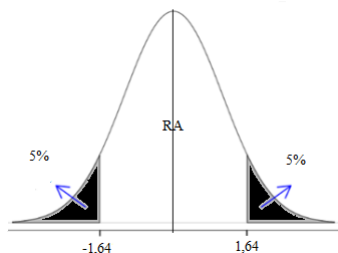


Figura 2.29: Distribuição normal padronizada bicaudal

4. Cálculo dos valores das variáveis:

$$p_1 = \frac{x_1}{n_1} = \frac{70}{200} = 0,35$$

$$p_2 = \frac{x_2}{n_2} = \frac{50}{200} = 0,25$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{70 + 50}{200 + 200} = 0,30$$

$$Z_{cal} = \frac{0,35 - 0,25}{\sqrt{0,3(0,7)(\frac{2}{200})}} = 2,18$$

5. Conclusão: Como $Z_{cal} > 1,64$, rejeita-se a hipótese de igualdade entre as proporções, concluindo-se com risco de 10% que as proporções são diferentes.

2.12 Teste T para duas amostras independentes

O teste T (ou t de Student) para duas amostras independentes permite testar hipóteses referentes à diferença entre duas médias independentes. Considere duas populações normais, com médias μ_1 e μ_2 , de cada uma das quais selecionamos uma amostra aleatória de tamanhos n_1 e n_2 , respectivamente.

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

Considere também a variância amostral combinada:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

1. Quando σ_1^2 e σ_2^2 são conhecidos. A estatística de teste será:

$$z_o = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

H_o	H_1	Rejeita-se H_o ao nível α	p valor
$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$	$z_o > z_\alpha$ $p_o \leq \alpha$	$p_o = P[Z \geq z_o]$
$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$	$z_o < -z_\alpha$ $p_o \leq \alpha$	$p_o = P[Z \leq z_o]$
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$ z_o > z_{\alpha/2}$ $p_o \leq \alpha$	$p_o = 2P[\text{uma cauda}]$

Z denota uma variável aleatória com distribuição $N(0, 1)$ e z_α a abscissa debaixo dessa densidade que deixa a sua direita uma área α .

2. Quando σ_1^2 e σ_2^2 são desconhecidas e $\sigma_1^2 = \sigma_2^2$. A estatística de teste será:

$$t_o = \frac{\bar{X}_1 - \bar{X}_2}{s_p^2 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

H_o	H_1	Rejeita-se H_o ao nível α	p valor
$\mu_1 \leq \mu_2$	$\mu_1 > \mu_2$	$t_o > t_{n_1+n_2-2;\alpha}$ $p_o \leq \alpha$	$p_o = P[T \geq t_o]$
$\mu_1 \geq \mu_2$	$\mu_1 < \mu_2$	$t_o < -t_{n_1+n_2-2;\alpha}$ $p_o \leq \alpha$	$p_o = P[T \leq t_o]$
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	$ t_o > t_{n_1+n_2-2;\alpha/2}$ $p_o \leq \alpha$	$p_o = 2P[\text{uma cauda}]$

T denota uma variável aleatória com distribuição $t_{n_1+n_2-2}$ e $t_{n_1+n_2-2;\alpha/2}$ a abscissa debaixo dessa densidade que deixa a sua direita uma área α .

Exemplo 2.12.1. *Foram selecionadas de forma aleatória duas amostras de tamanho $n_1 = 28$ e $n_2 = 25$. As médias amostrais dessas amostras foram: $\bar{x}_1 = 10$ e $\bar{x}_2 = 9$ e as variâncias amostrais foram $s_1^2 = 0,43$ e $s_2^2 = 0,55$. Assumindo que os dados foram extraídos de uma distribuições normais com $\sigma_1^2 = \sigma_2^2$. Pede-se testar as hipóteses:*

$$\begin{cases} H_o : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{cases}$$

No R seria:

```
x1b<-10
x2b<-9
n1<- 28; n2<-25
s12<-0.43
s22<-0.55
sp2<- ((n1-1)*s12 + (n2-1)*s22)/(n1+n2-2); sp2
sp<-sqrt(sp2); sp
to<- (x1b - x2b)/(sp*(sqrt(1/n1 + 1/n2))); to
valorp<- 2*(1-pt(to,50)); valorp
[1] 3.578536e-06 =0,0000
```

A estatística de contraste T, sob a H_o , segue uma distribuição $t_{28+25-2} = t_{51}$. Assim,

$$t_o = \frac{\bar{X}_1 - \bar{X}_2}{s_p^2 \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 5,21054$$

A um nível $\alpha = 5\%$ verifica-se a condição de rejeição. O p valor será: $p_o = 2P[T \geq 5,2104] = 0,0000$, como $0,05 > 0,000$, o valor 0,0000 na área de rejeição, logo rejeita-se a hipótese nula, aceitando que $\mu_1 \neq \mu_2$.

2.12.1 Uso do SPSS para um teste T para duas amostras independentes

Para pequenas amostra $n < 30$, usa-se a estatística t de Student. Para amostra maior que 30 usa-se as fórmulas vista no item acima.

Exemplo 2.12.2. *Considere dois grupos de alunos de psicologia que tiveram metodologias pedagógicas diferentes. Queremos saber*

com um nível de significância de 0,05 se existe diferença entre a média dos dois grupos. As hipóteses são:

$$\begin{cases} H_o : \mu_A = \mu_B \\ H_a : \mu_A \neq \mu_B \end{cases}$$

$Grupo_A = [15, 12, 11, 18, 15, 15, 9, 19, 14, 13, 11, 12, 18, 15, 16, 14, 16, 17, 15, 17, 13, 14, 13, 15, 17, 19, 17, 18, 16, 14]$

$Grupo_B = [11, 16, 14, 18, 6, 8, 9, 14, 12, 12, 10, 15, 12, 9, 13, 16, 17, 12, 8, 7, 15, 5, 14, 13, 13, 12, 11, 13, 11, 7]$

A média e o desvio padrão do Grupo A e B são respectivamente, $\bar{X}_A = 14,93$, $s_A = 2,50$ e $\bar{X}_B = 11,77$, $s_B = 3,30$. O tamanho amostra é $n=30$. Assim, o erro padrão da diferença entre as médias estimadas são:

$$Epd = \sqrt{s_A^2/N_1 + s_B^2/N_B}$$

Sendo Epd o erro padrão da diferença. E s_A^2 é a variância do grupo A, s_B^2 a variância do grupo B. N_A e N_B é o número de observações em cada grupo.

Observação: para amostras relativamente pequenas ($N < 30$) a distribuição amostral tende para a distribuição T de Student. Vamos calcular a probabilidade que o Epd esteja nessa faixa necessária, aplicando a fórmula na definição. Temos:

$$Epd = \sqrt{2,50^2/30 + 3,30^2/30} = 0,756$$

$$t = \frac{\bar{x}_A - \bar{x}_B}{Epd} = \frac{14,83 - 11,77}{0,756} = 4,188$$

Se procurarmos esse valor na Tabela da distribuição T de Student para 29 graus de liberdade $(N - 1) = 30 - 1 = 29$, veremos que devemos rejeitar H_0 e concluir que existe uma diferença estatisticamente significativa entre as duas amostras A e B. Temos razão em acreditar que o método pedagógico no curso de psicologia influencia os resultados finais dos alunos.

2.13 Teste com amostras independentes e variâncias iguais

Considere duas amostras independentes, oriundas de duas populações com distribuição normal, com médias μ_1 e μ_2 , e com variâncias σ_1^2 e σ_2^2 , desconhecidas e iguais. A estatística de teste para verificar se as duas médias populacionais são iguais ou não é a seguinte:

$$t = \frac{\bar{x}_A - \bar{x}_B}{S_d}$$

Sendo $d = \bar{x}_1 - \bar{x}_2$ e o estimador do desvio padrão da diferença entre as médias amostrais é:

$$S_d = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}$$

Sob a hipótese nula, tem distribuição t de Student com $n_1 + n_2 - 2$ graus de liberdade. Senão n_1 e n_2 o tamanho da amostra da população 1 e 2 respectivamente, e S_1^2 e S_2^2 são as respectivas variâncias amostrais.

Exemplo 2.13.1. *Suponha que temos as amostras de duas populações normais: $A=(1,3; 1,4; 1,1; 1,4; 1,5)$ e $B=(1,8; 1,6; 1,9; 1,9; 1,8)$. Assumindo que as duas populações tenham variâncias populacionais iguais. Dado $\bar{x}_A = 1,34$; $\bar{x}_B = 1,80$; $\sigma_A^2 = 0,023$ e $\sigma_B^2 = 0,0150$.*

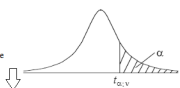
o valor de t_{cal} será:

$$t_{cal} = \frac{1,34 - 1,80}{\sqrt{\frac{(5-1)0,023 + (5-1)0,015}{5+5-2} \left[\frac{1}{5} + \frac{1}{5} \right]}} = -5,26$$

Observando a Tabela da distribuição T de Student observamos que: $t_{8,5\%/2} = 2,306$.

Tabela t de Student com v grau de liberdade

$\Pr(T_v > t_{\alpha;v}) = \alpha$



α	0,1	0,05	0,025	0,01	0,005	0,001	0,0005
1	3,078	6,314	12,076	31,821	63,657	318,310	636,620
2	1,886	2,920	4,303	6,965	9,925	22,326	31,598
3	1,638	2,353	3,182	4,541	5,841	10,213	12,924
4	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	1,383	1,833	2,262	2,821	3,260	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,930	4,318

Figura 2.30: Tabela t de Student com v grau de liberdade

Na Figura, rejeitamos a hipótese de igualdade das populações A e B, pois o valor -5,26 está fora do intervalo -2,306 e +2,306 (área de aceitação).

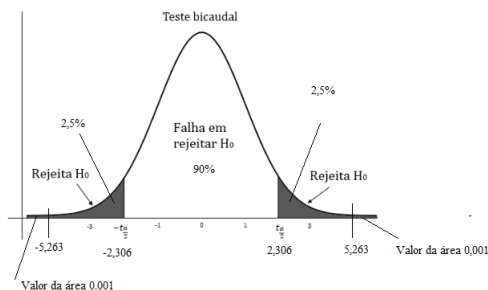


Figura 2.31: Tabela T de Student bicaudal com 8 graus de liberdade

A estimação do quanto mais A em relação a B pode ser feita através do intervalo de confiança para a diferença entre as médias, $\mu_B - \mu_A$. O intervalo de confiança é dado pela expressão:

$$IC(\mu_B - \mu_A; 95\%) = (\bar{x}_B - \bar{x}_A) \pm t_{(n_A + n_B - 2; 5\%)} \cdot S_d$$

Concluimos que:

$$IC(\mu_B - \mu_A; 95\%) = 0,46 \pm 2,306(0,0873)$$

$$IC(\mu_B - \mu_A; 95\%) = 0,46 \pm 0,20131$$

$$IC(\mu_B - \mu_A; 95\%) = [0,2587; 0,6613]$$

Exemplo 2.13.2. *Suponha que temos as amostras de duas populações normais: $A=(1,8; 1,5; 1,3; 1,8; 1,3)$ e $B=(1,1; 2,6; 1,8; 1,7; 1,6)$. Assumindo que as duas populações tenham variâncias populacionais iguais. Dado $\bar{x}_A = 1,64$; $\bar{x}_B = 1,91$; $\sigma_A^2 = 0,028$ e $\sigma_B^2 = 0,061$.*

2.14 Teste com amostras independentes com variâncias desconhecidas e diferentes

Considere duas amostras independentes, oriundas de duas populações com distribuição normal, com variâncias σ_1^2 e σ_2^2 , desconhecidas e diferentes. Esse teste é válido para $n_1 + n_2 \leq 30$. A estatística de teste para o teste de hipótese será:

$$t_{cal} = \frac{\bar{x}_d - \mu}{S_{\bar{x}_d}}$$

$$S_{\bar{x}_d} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Então, o valor de de t_{cal} será então:

$$t_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Com grau de liberdade igual a:

$$gl = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1+1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2+1}} - 2$$

Exemplo 2.14.1. *O QI de 16 estudantes de uma zona pobre de certa cidade apresenta a média de 107 pontos com desvio padrão de 10 pontos, enquanto os 14 estudantes de outra região rica da cidade apresentam média de 112 pontos com desvio padrão de 8 pontos. O QI em ambas as regiões tem distribuição normal. Há uma diferença entre os QIs médios dos dois grupos a 5%?*

Procedimento para efetuar o teste de hipótese:

1. Enunciar as hipóteses H_o e H_a

$$H_o: \mu_1 - \mu_2 = 0 \text{ ou } \mu_1 = \mu_2$$

$$H_a: \mu_1 - \mu_2 \neq 0 \text{ ou } \mu_1 \neq \mu_2$$

2. Fixar α . Escolher a variável t de Student com gl grau de liberdade
3. Com auxílio da tabela de distribuição t de Student, determinam-se RA e RC

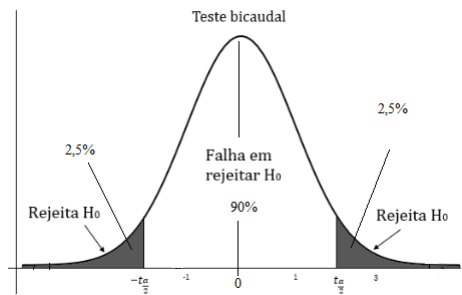


Figura 2.32: Distribuição t de Student bicaudal informando as áreas de aceitação e rejeição com nível de significância 5%

4. Cálculo do valor da variável: t_{cal} e o cálculo do grau de liberdade.
5. Conclusão:
- Se $t_{cal} < t_{tab}$, aceita H_o .
 - Caso contrário, rejeita H_o .

Aplicando aos dados do exemplo, temos:

Tabela 2.3: Valores calculados			
Tamanho	Médias	Variâncias	Diferenças das médias
$n_1 = 16$	$\bar{x}_1 = 107$	$\sigma_1^2 = 100$	$\bar{x}_d = \bar{x}_1 - \bar{x}_2$
$n_2 = 14$	$\bar{x}_2 = 112$	$\sigma_2^2 = 64$	$\bar{x}_d = 107 - 112 = -5$

$$gl = \frac{\left(\frac{100}{16} + \frac{64}{14}\right)^2}{\frac{\left(\frac{100}{16}\right)^2}{16+1} + \frac{\left(\frac{64}{14}\right)^2}{14+1}} - 2 = 29,74$$

$$S_{\bar{x}_d} = \sqrt{\frac{100}{16} + \frac{64}{14}} = \sqrt{10,82} = 3,28$$

$$t_{cal} = \frac{(-5 - 0)}{3,28} = -1,52$$

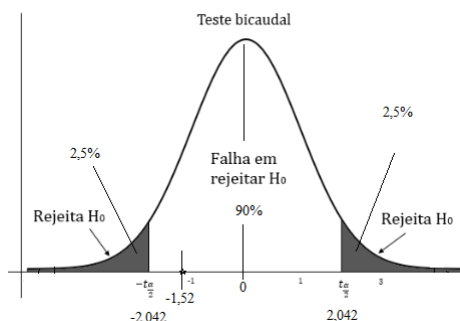


Figura 2.33: Distribuição t de Student com 29,74 graus de liberdade, informando as áreas de aceitação e rejeição

Como $t_{cal} \in RA$, não se rejeita H_o , isto é, ao nível de 5% não é significativa a diferença entre os QIs das duas regiões da cidade.

2.15 Teste de hipóteses para a comparação de variâncias com médias populacionais conhecidas e desconhecidas

Sejam X_1 e X_2 duas variáveis aleatórias independentes tais que $X_1 \sim N(\mu_1; \sigma_1^2)$ e $X_2 \sim N(\mu_2; \sigma_2^2)$. Sendo n_1 e n_2 os tamanhos da

amostra de X_1 e X_2 , respectivamente.

1. Quando μ_1 e μ_2 são conhecidos. A estatística de teste será:

$$f_o = \frac{n_1 \sum_{i=1}^{n_1} (x_i - \mu_1)^2}{n_2 \sum_{j=1}^{n_2} (x_j - \mu_2)^2}$$

H_o	H_1	Rejeita-se H_o ao nível α	p valor
$\sigma_1^2 \leq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$f_o > F_{n_1, n_2; \alpha}$	$p_o = P[F \geq f_o]$
$\sigma_1^2 \geq \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$f_o < F_{n_1, n_2; 1-\alpha}$	$p_o = P[F \leq f_o]$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$f_o > F_{n_1, n_2; \alpha/2}$ $f_o < F_{n_1, n_2; 1-\alpha/2}$	$p_o = 2P[\text{uma cauda}]$

F denota uma variável aleatória com distribuição F_{n_1, n_2} e $F_{n_1, n_2; \alpha}$ a abscissa debaixo dessa densidade que deixa a sua direita uma área α .

2. Quando μ_1 e μ_2 são desconhecidas. A estatística de teste será:

$$f_o = \frac{s_1^2}{s_2^2}$$

H_o	H_1	Rejeita-se H_o ao nível α	p valor
$\sigma_1^2 \leq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$f_o > F_{n_1-1, n_2-1; \alpha}$	$p_o = P[F \geq f_o]$
$\sigma_1^2 \geq \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$f_o < F_{n_1-1, n_2-1; 1-\alpha}$	$p_o = P[F \leq f_o]$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$f_o > F_{n_1-1, n_2-1; \alpha/2}$ $f_o < F_{n_1-1, n_2-1; 1-\alpha/2}$	$p_o = 2P[\text{uma cauda}]$

F denota uma variável aleatória com distribuição F_{n_1-1, n_2-1} e $F_{n_1-1, n_2-1; \alpha}$ a abscissa debaixo dessa densidade que deixa a sua direita uma área α .

Exemplo 2.15.1. Usando as tabelas abaixo. Encontre os seguintes valores.

Limites unilaterais da distribuição F de Fisher-Snedecor ao nível de 5% de probabilidade.																				
GL	V1																			
V2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	40	60	120	240
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.0	243.9	244.7	245.4	245.9	246.0	251.1	252.2	253.3	253.8
2	18.513	19.000	19.164	19.247	19.299	19.329	19.353	19.371	19.385	19.398	19.405	19.412	19.419	19.424	19.429	19.440	19.471	19.479	19.487	19.492
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785	8.763	8.745	8.729	8.715	8.703	8.680	8.694	8.672	8.649	8.638
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.935	5.912	5.891	5.873	5.858	5.803	5.717	5.688	5.658	5.643
5	6.608	5.786	5.409	5.192	5.050	4.960	4.876	4.818	4.772	4.735	4.704	4.678	4.655	4.636	4.619	4.658	4.464	4.431	4.398	4.382
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.027	4.000	3.976	3.956	3.938	3.874	3.774	3.740	3.705	3.687
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.603	3.575	3.550	3.529	3.511	3.445	3.340	3.304	3.267	3.249
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347	3.313	3.284	3.259	3.237	3.218	3.150	3.043	3.005	2.967	2.947
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.102	3.073	3.048	3.025	3.006	2.936	2.826	2.787	2.748	2.727
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978	2.943	2.913	2.887	2.865	2.845	2.774	2.661	2.621	2.580	2.559

Limites unilaterais da distribuição F de Fisher-Snedecor ao nível de 2,5% de probabilidade.																				
GL	V1																			
V2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	20	40	60	120	240
1	647.8	799.5	864.2	889.6	921.6	937.1	948.2	956.6	963.3	968.6	973.0	976.7	979.8	982.5	984.9	993.1	1005.6	1009.8	1014.0	1016.1
2	38.506	39.000	39.166	39.248	39.298	39.331	39.356	39.373	39.387	39.398	39.407	39.415	39.421	39.427	39.431	39.448	39.473	39.481	39.489	39.494
3	17.443	16.044	15.439	15.101	14.885	14.735	14.624	14.540	14.473	14.419	14.374	14.337	14.305	14.277	14.253	14.167	14.036	13.992	13.947	13.925
4	12.218	10.649	9.979	9.604	9.364	9.197	9.074	8.980	8.905	8.844	8.794	8.751	8.715	8.684	8.657	8.560	8.411	8.360	8.309	8.283
5	10.307	8.834	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.568	6.525	6.488	6.455	6.428	6.329	6.175	6.123	6.069	6.042
6	8.813	7.260	6.595	6.227	5.980	5.820	5.695	5.600	5.523	5.461	5.410	5.366	5.329	5.297	5.269	5.168	5.012	4.959	4.904	4.877
7	8.073	6.542	5.880	5.523	5.285	5.119	4.995	4.899	4.823	4.761	4.709	4.666	4.628	4.595	4.568	4.467	4.309	4.254	4.199	4.171
8	7.571	6.059	5.416	5.063	4.817	4.652	4.529	4.433	4.357	4.295	4.243	4.200	4.162	4.130	4.101	3.999	3.840	3.784	3.728	3.698
9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964	3.912	3.868	3.831	3.798	3.769	3.667	3.505	3.449	3.392	3.363
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.665	3.621	3.583	3.550	3.522	3.419	3.255	3.198	3.140	3.110
11	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526	3.474	3.430	3.392	3.359	3.330	3.226	3.061	3.004	2.944	2.914

Figura 2.34: Limites unilaterais da distribuição F Fisher-Snedecor aos níveis 2,5% e 5%

- a) $F_{6,7;2,5\%}$
- b) $F_{9,8;5\%}$
- c) $F_{7,8;5\%}$
- d) $F_{8,9;2,5\%}$
- e) $F_{4,5;2,5\%}$

Exemplo 2.15.2. Testar a hipóteses para a comparação de variâncias sabendo que as médias populacionais de duas populações sejam desconhecidas. Sabendo que $s_1^2 = 0,4$, $s_2^2 = 0,2$, $n_1 = 10$ e

$n_2 = 10$. *Assumindo que os dados foram extraídos de distribuições normais. Testar as seguintes hipóteses:*

$$\begin{cases} H_o : \sigma_1^2 = \sigma_2^2 \\ H_a : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

Observando a tabela acima, a região crítica do teste será: $f_o > F_{n_1-1, n_2-1; \alpha/2}$ ou $f_o < F_{n_1-1, n_2-1; 1-\alpha/2}$. Nessa situação, temos:

$$F_{n_1-1, n_2-1; \alpha/2} = F_{10-1, 10-1; 0,05/2} = F_{9,9; 0,0025} = 4,026$$

$$F_{n_1-1, n_2-1; 1-\alpha/2} = F_{10-1, 10-1; 1-0,05/2} = F_{9,9; 0,975} = 0,248$$

A estatística de teste será:

$$f_o = \frac{s_1^2}{s_2^2} = \frac{0,4}{0,2} = 2$$

Rejeita-se a hipótese nula se $f_o > F_{n_1-1, n_2-1; \alpha/2}$ ou $f_o < F_{n_1-1, n_2-1; 1-\alpha/2}$. Como essas condições não são satisfeitas, aceita-se H_o , pois $4,026 > 2$.

O valor p será: $p_o = 2P[F \geq 2] = 0,3164$, ou seja, $0,3164 > 0,05$ (está na área de aceitação da hipótese nula).

No R seria:

```
n1=n2=10  
g11<-n1-1; g12<-n2-1  
a2<-0.05/2
```

```
qf(a2, gl1, gl2)
qf(1-a2, gl1, gl2)
s12<-0.4
s22<-0.2
fo<-s12/s22; fo
fo<-s12/s22; fo
valorp<- 2*(1-pf(fo, gl1, gl2)); valorp
```

Exemplo 2.15.3. *A quantidade de jovens estressados em 22 cidades tiveram as seguintes estatísticas: $S_1 = 2$, $n_1 = 22$, $S_2 = 3$ e $n_2 = 22$. Suponha que as amostras foram extraídas de populações normais. Adote $\alpha = 5\%$. Pede-se:*

a) Testar as seguintes hipóteses:

$$\begin{cases} H_o: \frac{\sigma_2^2}{\sigma_1^2} = 1 \\ H_a: \frac{\sigma_2^2}{\sigma_1^2} \neq 1 \end{cases}$$

b) Encontrar a probabilidade de cometer o erro tipo II (β) se a verdadeira relação seja $\frac{\sigma_1^2}{\sigma_2^2} = 3$.

A estatística de teste será: $F = \frac{S_1^2}{S_2^2} = \frac{2^2}{3^2} = 2,25$. No R será:

```
alfa<-0.05
alfa/2
n1<-22
n2<-22
A<-qf(0.025, n1-1, n2-1); A
B<-qf(1-0.025, n1-1, n2-1); B
s1<-2
```

```
s2<-3
F<-s2^2/s1^2; F
F1<-1.23
F2<-7.23
pf(F1,n1-1,n2-2)
pv<-1-pf(F1,n1-1,n2-2)
pf(F2,n1-1,n2-2)
beta<- 1-pv; beta
```

Conclui-se que aceita-se a hipótese nula H_o , pois 0,25 está entre [0,41;2,41, ou seja, não existe evidência estatística para rejeitar H_o .

$$f_o = \frac{s_2^2}{s_1^2} \frac{\sigma_1^2}{\sigma_2^2}$$

Para 0,41, tem-se:

$$\frac{s_2^2}{s_1^2} \frac{\sigma_1^2}{\sigma_2^2} = 0,41.3 = 1,23$$

Para 2,41, tem-se:

$$\frac{s_2^2}{s_1^2} \frac{\sigma_1^2}{\sigma_2^2} = 2,41.3 = 7.23$$

Observe esses valores na figura abaixo.

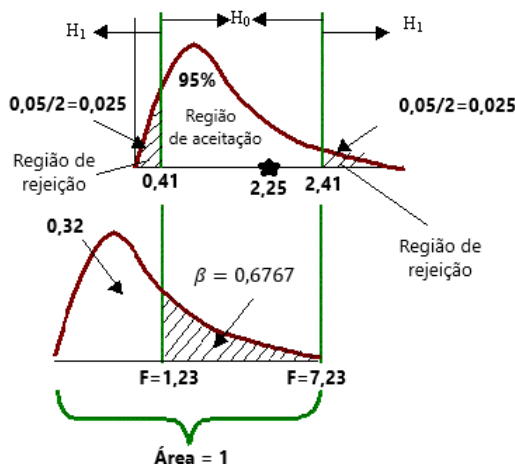


Figura 2.35: Erros tipo I e erro tipo II

Exercício 2.15.1. A quantidade de jovens estressados em 22 cidades tiveram as seguintes estatísticas: $S_1 = 3$, $n_1 = 25$, $S_2 = 4$ e $n_2 = 25$. Suponha que as amostras foram extraídas de populações normais. Adote $\alpha = 5\%$. Pede-se:

a) Testar as seguintes hipóteses:

$$\begin{cases} H_o : \frac{\sigma_2^2}{\sigma_1^2} = 1 \\ H_a : \frac{\sigma_2^2}{\sigma_1^2} \neq 1 \end{cases}$$

b) Encontrar a probabilidade de cometer o erro tipo II (β) se a verdadeira relação seja $\frac{\sigma_1^2}{\sigma_2^2} = 6$.

Exercício 2.15.2. A quantidade de jovens estressados em 22 cidades tiveram as seguintes estatísticas: $S_1 = 7$, $n_1 = 30$, $S_2 = 8$ e $n_2 = 35$. Suponha que as amostras foram extraídas de populações normais. Adote $\alpha = 5\%$. Pede-se:

a) Testar as seguintes hipóteses:

$$\begin{cases} H_o : \frac{\sigma_2^2}{\sigma_1^2} = 1 \\ H_a : \frac{\sigma_2^2}{\sigma_1^2} \neq 1 \end{cases}$$

b) Encontrar a probabilidade de cometer o erro tipo II (β) se a verdadeira relação seja $\frac{\sigma_1^2}{\sigma_2^2} = 9$.

Exercício 2.15.3. *Testar a hipóteses para a comparação de variâncias sabendo que as médias populacionais de duas populações sejam desconhecidas. Sabendo que $s_1^2 = 0,57$, $s_2^2 = 0,31$, $n_1 = 9$ e $n_2 = 8$. Assumindo que os dados foram extraídos de distribuições normais. Seja $\alpha = 5\%$ testar as seguintes hipóteses:*

a)

$$\begin{cases} H_o : \sigma_1^2 \leq \sigma_2^2 \\ H_a : \sigma_1^2 > \sigma_2^2 \end{cases}$$

b)

$$\begin{cases} H_o : \sigma_1^2 \geq \sigma_2^2 \\ H_a : \sigma_1^2 < \sigma_2^2 \end{cases}$$

Índice Remissivo

- afetivo, 43
 alimentação, 93
 amostra, 26, 27, 29, 36–39, 42–45, 54, 55, 76, 97, 98, 105
 ansiedade, 18, 36, 51, 57, 63, 109, 119, 125, 126
 Bernoulli, 85, 102
 bipolar, 43
 conduta, 59
 contraproducentes, 107
 crise, 92
 defesa, 44
 desenvolvimento, 52
 distribuição, 27, 29, 37, 41, 52–54, 73, 84, 106
 distúrbios, 52
 déficit, 84
 espacial, 84
 esquizofrenia, 93
 hiperatividade, 84
 hipertensão, 62, 70
 hipótese, 19, 106, 107, 128–136, 138, 140, 141, 148, 150–154, 157, 158, 161–163, 167–169, 171, 176, 178, 182, 184–187
 hábitos, 107
 intelectualização, 44
 inteligência, 60, 84, 101, 127
 inteligência , 101
 interpessoal, 53
 intervalo, 19, 25–30, 32, 37, 42, 48, 53, 54, 56, 76, 128, 129, 137, 138, 146, 154, 162, 170, 171, 185
 intuitiva, 60

- liberdade, 37, 47, 54, 71–73, 76, variável, 99, 106, 187, 188
99–101, 134–136, 158, 160,
162, 167, 169, 183, 184,
187, 188
- normal, 44, 45, 56–58, 84, 106
- nível, 26, 52, 76, 128–130, 132,
136–138, 140, 146–148,
150, 152–154, 158, 161–
163, 167–169, 176, 182,
189
- pacientes, 45
- parâmetro, 26
- população, 25, 27, 29, 30, 32,
41, 42, 48, 55, 56, 129,
134, 140, 148, 150, 163,
184
- probabilidade, 26, 27, 54, 182
- proporção, 97, 98, 176
- psicóticos, 93
- recuperação, 43
- região, 128, 135, 158, 162, 164,
169, 170, 187
- social, 59
- tamanho, 27, 29
- Tchebyshev, 123
- transtorno, 126
- transtornos, 93

Referências Bibliográficas

Martins, G. de Andrade. *Estatística Geral e Aplicada*. 3.ed.- 2.reimp.- São Paulo: Atlas, 2006.

Arango, Héctor Gustavo. *Bioestatística. Teórica e Computacional. Com bancos de dados reais*. 3.ed.-[Reimpr.]-Rio de Janeiro: Guanabara Koogan,2012. 2012.

Ávila, M. J. del M., García, J. M. T. *Técnicas Estadísticas Aplicadas*. Grupo Editorial Universitario, 2006.

Bento, Murteira e Antunes, Marília. *Probabilidade e Estatística*. Volume II. ed. Lisboa: Escolar Editora, 2012.

Cristóbal, J. A.C. *Inferencia Estadística*. Prensas Universitarias de Zaragoza, 2.ed., 1995.

Fernández-Abascal, H., Guijarro, M. M., Rojo, J. L., Sanz, J. A. *Cálculo de probabilidades y estadística*. Barcelona: Editorial Ariel, S. A., 1994.

Figueiredo, F; Figueirede, A; Ramos, A; Teles, P. *Inferência*

Estatística. Problemas resolvidos e propostos com aplicação em R. Lisboa: Escolar Editora, 2017.

García, J.E.; Nebot, J.M.B.; Escortell, A.I.; Rodríguez, Ma.I.L., Olivas, C.R.; Ponce, F.R. *Inferencia Estadística*. 2ed revisada, Madrid: Garceta grupo editorail, 2018.

F.J. Martín-Pliego, J.Maria, Montero Lorenzo, L. Ruíz-Maya Pérez *Problemas de Inferencia Estadística*. Paraninfo, 3.ed., Madrid, 2000.

López F.J. Martín-Pliego; Lorenzo, J.Ma, Monteiro; Pérez L.Ruíz-Maya. *Problemas de Inferencia Estadística*. 3ed., Madrid: Thomson, 2000.

Luis Ruiz-Maya e F. Javier Martin Pliego. *Estadística II: Inferencia*. Madrid: Colección Plan Nuevo. ed. Editorial AC, 1995.

Maindonald, J. H., Braun, W. J. *DAAG: Data Analysis and Graphics Data and Functions*. note = R package version 1.22, 2015. <http://CRAN.R-project.org/package=DAAG>.

Kamper, S. J. *Confidence intervals: linking evidence to practice. journal of orthopaedic and sports physical therapy*. 49(10), 763-764, 2006.

Moore, David. *A Estatística Básica e Sua Prática*. W.H.Freeman and Company. New York: LTC, 1995.

Paradis, E. *R for Beginners*. Institut des Sciences de l'Évolution, Université Montpellier II, France, 2005. https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf

Monteiro; Pérez L. Ruíz-Maya; López F.J. Martín-Pliego.
Fundamentos de Inferencia Estadística. 3ed., Madrid: Thomson,
2002.

R Core Team. *R: A Language and Environment for Statistical
Computing*. R Foundation for Statistical Computing, Vienna:
Austria, 2015. <http://www.R-project.org/>.

Rmetrics Core Team., Wuertz, D., Setz, T., Chalabi, Y. *fBasics:
Rmetrics - Markets and Basic Statistics*. R package version
3011.87, 2014, <http://CRAN.R-project.org/package=fBasics>.

Sarkar, D., Andrews, F. *latticeExtra: Extra Graphical
Utilities Based on Lattice*. R package version 0.6-28, 2016.
<http://CRAN.R-project.org/package=latticeExtra>

Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T. *ROCR:
visualizing classifier performance in R*. Bioinformatics, 21(20),
pp. 7881. 2005. <http://rocr.bioinf.mpi-sb.mpg.de>

Organizador e Autores

Dr Edwirde Luiz Silva Camêlo (Brasil) - (Organizador)

Professor Associado da Universidade Estadual da Paraíba (UEPB). Pós-doutorado em *Estadística Aplicada* (2016) e Doutor em *Estadística e Investigación Operativa* (2007) pela *Universidad de Granada*. Mestrado em Biometria e Estatística Aplicada (2001) pela Universidade Federal Rural de Pernambuco (UFRPE). Atualmente, é professor associado da Universidade Estadual da Paraíba, com atuação nos Departamentos de Estatística e Psicologia e no Programa de Pós-Graduação em Psicologia da Saúde (PPGPS), sendo membro do grupo de pesquisa de Psicologia da Saúde (CNPq/UEPB). E-mail: edwirde@servidor.uepb.edu.br

Dra Dalila Camêlo Aguiar (Brasil) - Doutora em *Estadística*

Matemática y Aplicada pela *Universidad de Granada* (UGR), Mestre em *Estadística Aplicada* (2016) pela UGR, Especialista em Estatística Aplicada (2011) pela Fundação de Apoio, Pesquisa e Extensão (FURNE) e Bacharela em Estatística (2010) pela Universidade Estadual da Paraíba (UEPB). Pesquisadora na área de estatís-

tica multivariada aplicada. E-mail: dalilacamel@correo.ugr.es

Dr Ramón Gutiérrez Sánchez (España) - Professor Titular da *Universidad de Granada* (UGR), Diretor do *Departamento de Estadística e Investigación Operativa*, Coordenador do Mestrado em *Estadística Aplicada*, *CATEDRÁTICO DE UNIVERSIDAD Departamento de Estadística e Investigación Operativa*.

E-mail: ramongs@ugr.es

Dr Ivan Olier (England) His is a *Reader in Artificial Intelligence and Data Science*. My research interests lie in algorithms for Artificial Intelligence, with a specific focus on Causal AI, Digital Twins, and the modelling of large-scale, highly structured, and/or relational data (including multivariate time series, graphs, networks, etc.). My research finds its applications in various domains such as cardiovascular science, drug development, bioinformatics, healthcare, astrophysics, engineering, etc. Additionally, I hold a senior membership position at the Liverpool Centre for Cardiovascular Science (LCCS). E-mail:

I.A.OlierCaparroso@ljmu.ac.uk.

A estatística básica e sua prática constitui uma introdução à estatística para estudantes de estatística e psicologia I e II e também os tópicos especiais em estatística para o mestrado de Pós-Graduação em Psicologia da Saúde da Universidade Estadual da Paraíba. Em linhas gerais esse livro dar ênfase ao pensamento estatístico aplicado a psicologia; mais dados e conceitos, menos teorias e menos receitas; promover o ensino ativo que é caracterizado como métodos de ensino que fortalecem a formação de vínculos democráticos na relação entre professor, alunos e conteúdo.