

10 – INFERÊNCIA ESTATÍSTICA – Testes de Hipóteses

10.1 - Introdução

Viu-se anteriormente que uma determinada população pode ser descrita através de um modelo probabilístico, que apresenta características e parâmetros. Muitas vezes estes parâmetros são desconhecidos e há interesse em estimá-los para obter um melhor conhecimento sobre a população: retira-se então uma amostra aleatória da população e através das técnicas de **Estimação de Parâmetros**¹ procura-se obter uma estimativa de algum parâmetro de interesse, e associamos uma probabilidade de que a estimativa esteja correta. A Estimação de Parâmetros é uma subdivisão da Inferência Estatística (que consiste em fazer afirmações probabilísticas sobre o modelo probabilístico da população a partir de uma amostra aleatória desta população), a outra grande subdivisão constitui os **Testes de Hipóteses**.

Contrariamente à Estimação de Parâmetros os Testes de Hipóteses permitem fazer inferências sobre outras características do modelo probabilístico da população além dos parâmetros (como, por exemplo, a *forma* do modelo probabilístico da população). Quando os Testes são feitos sobre os parâmetros da população são chamados de Testes Paramétricos, e quando são feitos sobre outras características são chamados de Testes Não Paramétricos². Não obstante vamos nos restringir aos Testes Paramétricos³.

10.2 - Lógica dos Testes de Hipóteses

Imagine-se que um determinado pesquisador está interessado em alguma característica de uma população. Devido a estudos prévios, ou simplesmente por bom senso (melhor ponto de partida para o estudo) ele estabelece que a característica terá um determinado comportamento. Formula então uma **hipótese estatística** sobre a característica da população, e esta hipótese é aceita como válida até **prova estatística** em contrário.

Para testar a hipótese é coletada uma amostra aleatória representativa da população, sendo calculadas as estatísticas necessárias para o teste. Naturalmente (devido ao fato de ser utilizada uma amostra aleatória) haverá diferenças entre o que se esperava (sob a condição da hipótese verdadeira) e o que realmente foi obtido na amostra. A questão a ser respondida é: as diferenças são **significativas** o bastante para que a hipótese estatística estabelecida seja rejeitada? Esta não é uma pergunta simples de responder: dependerá do que está sob teste (que parâmetro, por exemplo), da confiabilidade desejada para o resultado, etc. Basicamente, porém, será necessário comparar as diferenças com uma **referência** (a distribuição amostral de um parâmetro, por exemplo), que supõe que a hipótese sob teste é verdadeira: a comparação costuma ser feita através de uma estatística de teste que envolve os valores da amostra e os valores sob teste.

¹ Ver o Capítulo 9.

² Na realidade a denominação correta deveria ser “Testes dependentes de distribuição de referência” (porque para fazer inferências sobre os parâmetros devemos supor que o modelo probabilístico populacional é normal, por exemplo, ou que a distribuição amostral do parâmetro pode ser aproximada por uma normal), e “Testes livres de distribuição” (porque os Testes Não Paramétricos não exigem que os dados tenham uma aderência a um certo modelo).

³ Ao leitor interessado em Testes Não Paramétricos recomendam-se as referências: SIEGEL, S., CASTELLAN, N. J. Estatística Não Paramétrica (para as Ciências do Comportamento); 2ª ed. – Porto Alegre: Artmed, 2006, e o Capítulo 10 de BARBETTA, P.A., REIS, M.M., BORNIA, A.C. Estatística para Cursos de Engenharia e Informática, 3ª ed., São Paulo: Atlas, 2010.

A **tomada de decisão** é feita da seguinte forma:

- se a diferença entre o que foi observado na amostra e o que era esperado (sob a condição da hipótese verdadeira) não for **SIGNIFICATIVA** a hipótese será **aceita**.
- se a diferença entre o que foi observado na amostra e o que era esperado (sob a condição da hipótese verdadeira) for **SIGNIFICATIVA** a hipótese será **rejeitada**.

O valor a partir do qual a diferença será considerada significativa será determinado pelo **Nível de Significância** do teste. O Nível de Significância geralmente é fixado pelo pesquisador, muitas vezes de forma arbitrária, e também será a probabilidade de erro do Teste de Hipóteses: a probabilidade de cometer um erro no teste, rejeitando uma hipótese válida. Como a decisão do teste é tomada a partir dos dados de uma amostra aleatória da população há **SEMPRE** a probabilidade de estar cometendo um erro, mas com a utilização de métodos estatísticos é possível **calcular** o valor desta probabilidade⁴. O Nível de Significância é uma probabilidade, portanto é um número real que varia de 0 a 1 (0 a 100%), e como é a probabilidade de se cometer um erro no teste é interessante que seja o mais próximo possível de zero: valores típicos são 5%, 10%, 1% e até menores dependendo do problema sob análise. Contudo, *não é possível* usar um Nível de Significância igual a zero porque devido ao uso de uma amostra aleatória sempre haverá chance de erro, a não ser que a amostra fosse do tamanho da população. O complementar do Nível de Significância é chamado de **Nível de Confiança**, pois ele indica a confiabilidade do resultado obtido, a probabilidade de que a decisão tomada esteja correta⁵.

10.3 - Tipos de Hipóteses

Para realizar um Teste de Hipóteses é necessário definir (enunciar) duas **Hipóteses Estatísticas complementares** (que abrangem todos os resultados possíveis): a chamada **Hipótese Nula** (denotada por H_0) e a **Hipótese Alternativa** (denotada por H_1 ou H_a). Enunciar as hipóteses é o primeiro e possivelmente mais importante passo de um Teste de Hipóteses, pois todo o procedimento dependerá dele.

A Hipótese Nula (H_0) é a hipótese estatística aceita como verdadeira até prova estatística em contrário: pode ser o ponto de partida mais adequado para o estudo, ou exatamente o *contrário* do que o pesquisador quer provar (ou o contrário daquilo que o preocupa).

A Hipótese Alternativa (H_1), que será uma hipótese complementar de H_0 , fornecerá uma alternativa à hipótese nula: muitas vezes é justamente o que o pesquisador quer provar (ou o que o preocupa).

Quando as hipóteses são formuladas sobre os parâmetros do modelo probabilístico da população o Teste de Hipóteses é chamado de Paramétrico. Quando as hipóteses são formuladas sobre outras características do modelo o Teste é chamado de Não Paramétrico.

A decisão do teste consiste em **ACEITAR** ou **REJEITAR** a Hipótese Nula (H_0): vai-se aceitar ou não a hipótese até então considerada verdadeira.

⁴ Usando outros métodos (empíricos) não há como ter ideia da chance de erro (pode ser um erro de 0% ou de 5000%...).

⁵ O leitor deve estar lembrado destes dois conceitos de Estimação de Parâmetros: Nível de Confiança era a probabilidade de que o Intervalo de Confiança contivesse o valor real do parâmetro, e Nível de Significância, complementar daquele, era a probabilidade de que o Intervalo NÃO contivesse o parâmetro, em suma a probabilidade da Estimação estar correta ou não, respectivamente.

É importante ter a noção exata do que significa ACEITAR ou REJEITAR a Hipótese Nula (H_0). A decisão é tomada sobre esta hipótese e não sobre a Hipótese Alternativa porque é a Hipótese Nula que é considerada verdadeira (até prova em contrário). Quando se aceita a Hipótese Nula significa que **não há provas suficientes para rejeitá-la**. Já quando a decisão é por rejeitar a Hipótese Nula **há evidências suficientes** de que as diferenças obtidas (entre o que era esperado e o que foi observado na amostra) não ocorreram por acaso. Usando uma analogia com o direito dos EUA, aceitar H_0 seria comparável a um veredito de **não culpado** (“not guilty”), ou seja, não há provas suficientes para condenar o réu... Por outro lado rejeitar H_0 seria comparável a um veredito de **culpado** (“guilty”), ou seja, as provas reunidas são suficientes para condenar o réu. O Nível de Significância será a probabilidade assumida de **Rejeitar H_0 sendo H_0 verdadeira**.

10.4 - Tipos de Testes Paramétricos

A formulação das hipóteses é o ponto inicial do problema, e deve depender ÚNICA E EXCLUSIVAMENTE das conclusões que se pretende obter com o teste. A formulação da hipótese alternativa determinará o tipo de teste: se **Unilateral** ou **Bilateral**.

Se a formulação da hipótese alternativa indicar que o parâmetro é **maior** ou **menor** do que o valor de teste (valor considerado verdadeiro até prova em contrário) o teste será **Unilateral**: somente há interesse se as diferenças entre os dados da amostra e o valor de teste forem em uma determinada direção. Se a formulação da hipótese alternativa indicar que o parâmetro é **diferente** do valor de teste o teste será **Bilateral**: há interesse nas diferenças em qualquer direção. As hipóteses então seriam:

Testes Unilaterais

H_0 : parâmetro = valor de teste

H_1 : parâmetro < valor de teste

H_0 : parâmetro = valor de teste

H_1 : parâmetro > valor de teste

Testes Bilaterais

H_0 : parâmetro = valor de teste

H_1 : parâmetro \neq valor de teste

A escolha do tipo de teste dependerá das condições do problema sob estudo, sejam as três situações abaixo:

- Um novo protocolo de atendimento foi implementado no Banco RMG, visando reduzir o tempo que as pessoas passam na fila do caixa. O protocolo será considerado satisfatório se a média do tempo de fila for **menor** do que 30 minutos. Um teste **Unilateral** seria o adequado.
- Cerca de 2000 formulários de pedidos de compra estão sendo analisados. Os clientes podem ficar insatisfeitos se houver erros nos formulários. Neste caso admite-se que a proporção máxima de formulários com erros seja de 5%. Ou seja, um valor **maior** do que 5% causaria problemas. Um teste **Unilateral** seria o adequado.
- Uma peça automotiva precisa ter 100 mm de diâmetro, exatamente. Neste caso, a dimensão não pode ser **maior** ou **menor** do que 100 mm (em outras palavras não pode ser **diferente** de 100 mm), pois isso indicará que a peça não está de acordo com as especificações. Um teste **Bilateral** seria o adequado.

Após definir as hipóteses é coletada uma amostra aleatória da população para testá-las. É importante ressaltar novamente:

“A MONTAGEM DAS HIPÓTESES DEVE DEPENDER APENAS DAS CONCLUSÕES QUE SE PRETENDE OBTER E JAMAIS DE UMA EVENTUAL EVIDÊNCIA AMOSTRAL DISPONÍVEL”.

A decisão de aceitar ou rejeitar H_0 dependerá das **regiões de aceitação e rejeição de H_0** , que por sua vez dependem dos seguintes fatores:

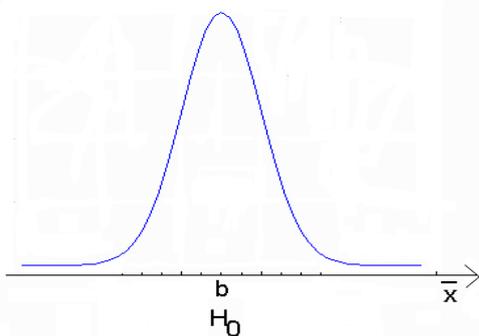
- do parâmetro sob teste (e da estatística ou variável de teste usada para testá-lo).
- do tipo de teste, se Unilateral ou Bilateral.
- do valor de teste (valor do parâmetro considerado verdadeiro até prova em contrário).
- do Nível de Significância (α) ou Nível de Confiança ($1 - \alpha$) adotado.
- de um valor crítico da estatística ou variável de teste a partir do qual a hipótese será rejeitada, e este valor dependerá por sua vez do Nível de Significância, do tipo de teste e da **Distribuição Amostral** do parâmetro.

A **Região de Aceitação de H_0** será a faixa de valores da estatística (ou da variável de teste) associada ao parâmetro em que as diferenças entre o que foi observado na amostra e o que era esperado **não são significativas**. A **Região de Rejeição de H_0** será a faixa de valores da estatística (ou da variável de teste) associada ao parâmetro em que as diferenças entre o que foi observado na amostra e o que era esperado **são significativas**.

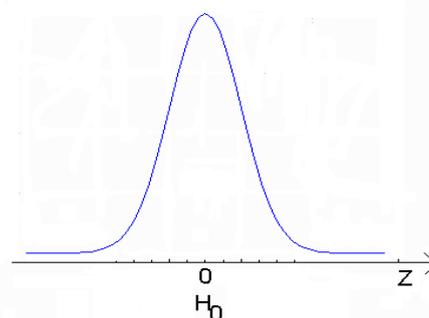
Esta abordagem é chamada de abordagem **clássica** dos testes de hipóteses. Há também a do **valor-p**, muito usada por programas computacionais: “a probabilidade de significância, ou valor p, é definida como a probabilidade da estatística do teste acusar um resultado tão ou mais distante do esperado como o resultado ocorrido na particular amostra observada, supondo H_0 como a hipótese verdadeira” (BARBETTA, REIS, BORNIA, 2010). O valor-p obtido é comparado com o nível de significância: se for MENOR do que o nível de significância, rejeita-se H_0 , se for maior ou igual aceita-se H_0 .

Para entender melhor os conceitos acima observe o exemplo a seguir:

Há interesse em realizar um teste de hipóteses sobre o comprimento médio de uma das dimensões de uma peça mecânica. O valor nominal da média (aceito como verdadeiro até prova em contrário) é igual a “**b**” (valor genérico), $H_0: \mu = b$. Supondo que a distribuição amostral do parâmetro (distribuição de \bar{X}) seja normal⁶, e será centrada em **b**: é possível fazer a conversão para a distribuição normal padrão (média zero e desvio padrão 1, variável **Z**).



$$H_0: \mu = b$$



$$H_0: \mu = 0$$

O valor de **b** (média da dimensão e média de \bar{X}) corresponde a zero, média da variável **Z**. Dependendo da formulação da Hipótese Alternativa haveria diferentes Regiões de Rejeição de H_0 .

⁶ Ver o Capítulo 9 para maiores detalhes.

Se a Hipótese Alternativa fosse $H_1: \mu < b$ ($H_1: \mu < 0$), ou seja, se o teste fosse Unilateral à esquerda a Região de Rejeição de H_0 seria:

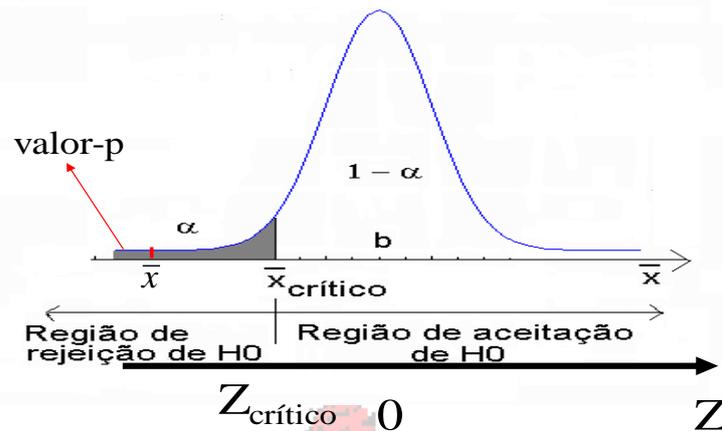


Figura 1 - $H_1: \mu < b$ $\mu < 0$

Observe que há um valor crítico de \bar{x} : se a média amostral estiver abaixo dele a Hipótese Nula será rejeitada, acima dele será aceita. A determinação do valor é feita com base no Nível de Significância, a área abaixo da curva normal até o valor crítico de \bar{x} . Geralmente, obtém-se o valor crítico da variável de teste (Z neste caso) através de uma tabela ou pacote computacional, que corresponde ao valor crítico de \bar{x} , faz-se a transformação de variáveis $Z = (\bar{x} - \mu_0)/\sigma'$ e obtém-se o valor crítico de \bar{x} . μ_0 é o valor sob teste (b no exemplo) e σ' é um desvio padrão (cujo valor será explicitado posteriormente).

Pela abordagem **clássica** a decisão será tomada comparando o valor da média amostral \bar{x} com o seu valor crítico: se for menor do que $\bar{x}_{critico}$, (ou seja, está na região de **Rejeição de H_0**), então se rejeita a Hipótese Nula. É possível tomar a decisão comparando o valor da variável de teste (Z neste caso), obtido com base nos dados da amostra, com o $Z_{critico}$ desta mesma variável (obtido de uma tabela ou programa computacional): se for menor do que o valor crítico rejeita-se a Hipótese Nula. Observe que o valor do Nível de Significância α é colocado na curva referente à Hipótese Nula, porque é esta que é aceita como válida até prova em contrário. Observe também que a faixa de valores da região de Rejeição pertence à curva da Hipótese Nula, assim o valor α é a probabilidade de Rejeitar H_0 sendo H_0 verdadeira.⁷

Se a decisão for tomada com base na variável de teste (Z , por exemplo), como o teste é Unilateral à esquerda o valor $Z_{critico}$ será NEGATIVO, uma vez que a região de Rejeição de H_0 está à ESQUERDA de 0 (menor do que zero). No teste Unilateral à direita, que veremos a seguir, o valor de $Z_{critico}$ será positivo, pois a região de Rejeição de H_0 estará à DIREITA de 0 (maior do que zero). Se, por exemplo, o Nível de Significância fosse de 5% (0,05) o valor de $Z_{critico}$ para o teste Unilateral à esquerda seria -1,645. Se houvesse interesse em obter o valor de $\bar{x}_{critico}$ correspondente bastaria usar a expressão $Z = (\bar{x} - \mu_0)/\sigma'$ substituindo Z por -1,645. O sinal correto é importante para que o valor de $\bar{x}_{critico}$ seja coerente com a posição da região de Rejeição de H_0 .

Pela abordagem do **valor-p** calcula-se a probabilidade de que \bar{x} assumia valores *menores* que aquele obtido na amostra. Veja na Figura 1 uma posição hipotética de \bar{x} : o valor-p é a probabilidade de que ela seja menor (porque o teste é unilateral à esquerda) do que aquele valor. Se o valor-p for MENOR do que 0,05 (nível de significância α) rejeita-se H_0 .

⁷ Probabilidade de tomar uma decisão errada FIXADA pelo pesquisador.

Se a Hipótese Alternativa fosse $H_1: \mu > b$ ($H_1: \mu > 0$), ou seja, se o teste fosse Unilateral à direita a Região de Rejeição de H_0 seria (Figura 2)

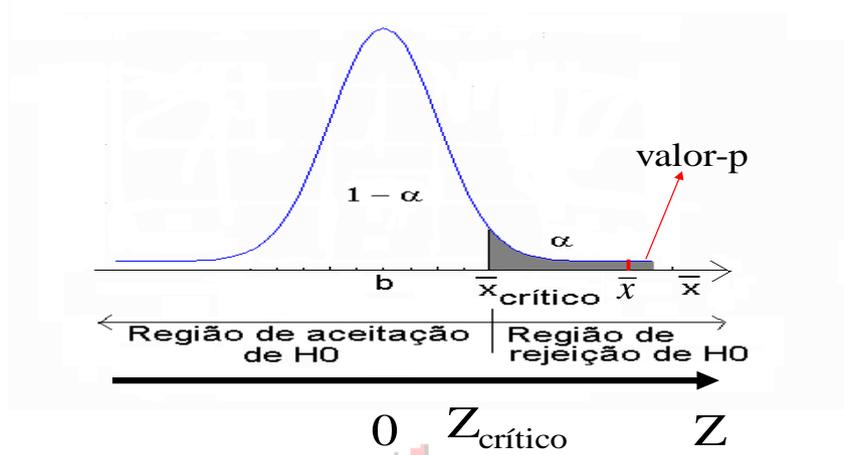


Figura 2 - $H_1: \mu > b$ $\mu > 0$

Neste caso o valor crítico está à direita: se a média amostral \bar{x} ou a variável de teste Z tiverem valores superiores aos respectivos valores críticos a Hipótese Nula será rejeitada, pois os valores “caíram” na região de Rejeição de H_0 . O valor de $Z_{crítico}$ será positivo, pois é maior do que zero: usando o mesmo Nível de Significância de 5%, $Z_{crítico}$ seria 1,645, igual em módulo ao anterior, porque a distribuição normal padrão é simétrica em relação à sua média, que é igual a zero.

Pela abordagem do **valor-p** é preciso calcular a probabilidade de que \bar{x} assuma valores *maiores* que aquele obtido na amostra. Veja na Figura 2 uma posição hipotética de \bar{x} : o valor-p é a probabilidade de que ela seja maior (porque o teste é unilateral à direita) do que aquele valor. Se o valor-p for MENOR do que 0,05 (nível de significância α) rejeita-se H_0 .

Se a Hipótese Alternativa fosse $H_1: \mu \neq b$ ($H_1: \mu \neq 0$), ou seja, o teste fosse Bilateral a Região de Rejeição de H_0 seria (Figura 3)

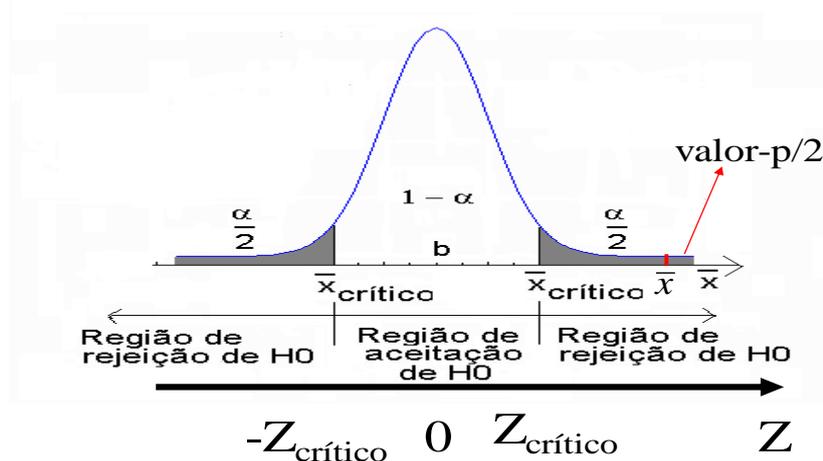


Figura 3 - $H_1: \mu \neq b$ $\mu \neq 0$

Neste caso a região de Rejeição se divide em duas iguais (probabilidades iguais à metade do Nível de Significância α), semelhante ao que acontece na Estimação por Intervalo. Existirão dois valores críticos, um abaixo do valor de teste e outro acima: se a média amostral \bar{x} ou a variável de teste Z tiverem valores acima do valor crítico “superior” ou abaixo do valor crítico “inferior” a Hipótese Nula será rejeitada, pois os valores “caíram” em uma das 2 regiões de Rejeição. Se for usada a variável de teste Z os valores críticos serão iguais em módulo, pois estão à mesma distância do valor sob teste (zero).

Pela abordagem do **valor-p** é preciso calcular a probabilidade de que \bar{x} assuma valores *maiores* que aquele obtido na amostra e depois multiplicar esta probabilidade por 2, se \bar{x} for **maior** do que o valor testado μ_0 . Se \bar{x} for **menor** do que o valor testado μ_0 , o procedimento exigiria calcular a probabilidade de que \bar{x} assuma valores *menores* do que aquele obtido na amostra, e também multiplicar esta probabilidade por 2. Veja na Figura 3 uma posição hipotética de \bar{x} : \bar{x} é *maior* do que b , portanto o valor-p é a probabilidade de que \bar{x} seja *maior* do que aquele valor *multiplicada por 2* (porque o teste é bilateral). Se o valor-p for MENOR do que 0,05 (nível de significância α), rejeita-se H_0 .

Recordando as três situações que foram abordadas anteriormente, seria interessante definir completamente as Hipóteses Estatísticas. Nos dois primeiros casos optou-se por um Teste Unilateral e no terceiro por um Teste Bilateral.

a) Um novo protocolo de atendimento foi implementado no Banco RMG, visando reduzir o tempo que as pessoas passam na fila do caixa. O protocolo será considerado satisfatório se a média do tempo de fila for **menor** do que 30 minutos. Um teste **Unilateral** seria o adequado. Mas Unilateral à Esquerda ou à Direita? Como está grifado na frase anterior haverá problema se a média do tempo fosse menor do que 30, resultando:

Teste UNILATERAL À ESQUERDA

$$H_0 : \mu = 30 \quad \text{onde } \mu_0 = 30 \text{ (valor de teste)}$$

$$H_1 : \mu < 30 \quad \text{Teste Unilateral à Esquerda.}$$

b) Cerca de 2000 formulários de pedidos de compra estão sendo analisados. Os clientes podem ficar insatisfeitos se houver erros nos formulários. Neste caso admite-se que a proporção máxima de formulários com erros seja de 5%. Ou seja, um valor **maior** do que 5% causaria problemas. Um teste **Unilateral** seria o adequado. Neste caso, um teste de *proporção*, o problema será um valor maior do que 5%, resultando:

Teste UNILATERAL À DIREITA

$$H_0 : \pi = 5\% \quad \text{onde } \pi_0 = 5\% \text{ (valor de teste)}$$

$$H_1 : \pi > 5\%$$

c) Uma peça automotiva precisa ter 100 mm de diâmetro, exatamente. Neste caso, a dimensão não pode ser **maior** ou **menor** do que 100 mm (em outras palavras não pode ser **diferente** de 100 mm) pois isso indicará que a peça não está de acordo com as especificações. Um teste **Bilateral** seria o adequado, resultando:

Teste BILATERAL

$$H_0 : \mu = 100 \text{ mm} \quad \text{onde } \mu_0 = 100 \text{ mm (valor de teste)}$$

$$H_1 : \mu \neq 100 \text{ mm}$$

Para a definição correta das hipóteses é imprescindível a correta identificação do valor de teste, pois se trata de um dos aspectos mais importantes do teste: o resultado da amostra será comparado ao valor de teste.

Lembrando novamente que a tomada de decisão depende da correta determinação da região de Rejeição (e, por conseguinte, de Aceitação) da Hipótese Nula, e isso, por sua vez, depende diretamente da formulação das Hipóteses Estatísticas.

10.4.1 - Testes de Hipóteses sobre a Média de uma Variável em uma População

Neste caso há interesse em testar a hipótese de que o parâmetro média populacional (μ) de uma certa variável QUANTITATIVA seja **maior**, **menor** ou **diferente** de um certo valor. Para a realização deste teste é necessário que uma das duas condições seja satisfeita:

- sabe-se, ou é razoável supor, que a variável de interesse segue uma distribuição normal na população: isso significa que a distribuição amostral da média também será normal, permitindo realizar a inferência estatística paramétrica.

- a distribuição da variável na população é desconhecida, mas a amostra retirada desta população é considerada “suficientemente grande”⁸ o que, de acordo com o Teorema Central do Limite, permite concluir que a distribuição amostral da média é normal.

Supõe-se também que a amostra é representativa da população e foi retirada de forma aleatória.

Tal como na Estimação de Parâmetros por Intervalo existirão diferenças nos testes dependendo do conhecimento ou não da variância populacional da variável.

a) Se a variância populacional (σ^2) da variável (cuja média populacional queremos testar) for conhecida.

Neste caso a variância amostral da média poderá ser calculada através da expressão:

$$V(\bar{x}) = \sigma^2/\sqrt{n}, \text{ e, por conseguinte, o “desvio padrão”}^9 \text{ será } \sigma' = \sigma/\sqrt{n}$$

A variável de teste será a variável **Z** da distribuição normal padrão, lembrando que:

$$Z = \frac{\text{valor} - \text{“média”}}{\text{“desvio padrão”}}$$

A “**média**” será o valor de teste (suposto verdadeiro até prova em contrário), denotado por μ_0 . O **valor** (obtido pela amostra) será a média amostral (que é o melhor estimador da média populacional¹⁰) denotada por \bar{x} , e o “desvio padrão” será o valor obtido anteriormente. Sendo assim a expressão do valor calculado da variável de teste **Z**:

$$Z_{\text{calculado}} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Pela abordagem **clássica** compara-se o valor da variável de teste com o valor crítico ($Z_{\text{crítico}}$ que depende do Nível de Significância adotado) de acordo com o tipo de teste:

Se $H_1 \mu: > \mu_0$ Rejeitar H_0 se $Z_{\text{calculado}} > Z_{\text{crítico}}$ ($\bar{x} > \bar{x}_{\text{crítico}}$)

Se $H_1 \mu: < \mu_0$ Rejeitar H_0 se $Z_{\text{calculado}} < Z_{\text{crítico}}$ ¹¹ ($\bar{x} < \bar{x}_{\text{crítico}}$)

Se $H_1 \mu: \neq \mu_0$ Rejeitar H_0 se $|Z_{\text{calculado}}| > |Z_{\text{crítico}}|$

Pela abordagem do **valor-p** obtém-se a probabilidade associada ao valor da variável de teste:

Se $H_1 \mu: > \mu_0$ Rejeitar H_0 se $P(Z > Z_{\text{calculado}}) < \alpha$

Se $H_1 \mu: < \mu_0$ Rejeitar H_0 se $P(Z < Z_{\text{calculado}})^{12} < \alpha$

Se $H_1 \mu: \neq \mu_0$ Rejeitar H_0 se $2 \times P(Z > |Z_{\text{calculado}}|)^{13} < \alpha$

⁸ Há muita controvérsia a respeito do que seria uma amostra “suficientemente grande”, vejam a discussão a respeito no Capítulo 9, na seção sobre distribuição amostral.

⁹ O desvio padrão é a raiz quadrada positiva da variância.

¹⁰ Ver o Capítulo 9 – Estimação de Parâmetros.

¹¹ Neste caso $Z_{\text{crítico}}$ será NEGATIVO, já que a região de Rejeição de H_0 está à esquerda de zero.

¹² Neste caso calcula-se a probabilidade de Z ser MENOR do que o $Z_{\text{calculado}}$, pois o teste é unilateral à esquerda.

¹³ Neste caso multiplica-se por 2 a probabilidade de Z ser MAIOR do que o valor em módulo de $Z_{\text{calculado}}$ pois o teste é bilateral.

b) Se a variância populacional σ^2 da variável for *desconhecida*.

Naturalmente este é o caso mais encontrado na prática. Procede-se como no item anterior, apenas fazendo com que $\sigma = s$, ou seja, considerando que o desvio padrão da variável na população é uma boa estimativa do desvio padrão da variável na amostra (suposição razoável para grandes amostras). Contudo, não será possível usar a distribuição normal padrão (Z), mas a distribuição **t de Student**. Esta distribuição já é conhecida (ver Capítulo 9). Trata-se de uma distribuição de probabilidades que possui média zero (como a distribuição normal padrão, variável Z), mas sua variância é igual a $n/(n-2)$, ou seja, a variância depende do tamanho da amostra. Quanto maior for o tamanho da amostra mais o quociente acima se aproxima de 1 (a variância da distribuição normal padrão), e mais a distribuição t de Student aproxima-se da distribuição normal padrão. A partir de $n = 50$, já é possível considerar a variância da distribuição t de Student aproximadamente igual a 1.

A variável de teste será então t_{n-1} (t com $n - 1$ graus de liberdade), e o seu valor calculado:

$$t_{n-1, \text{calculado}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

onde s é o desvio padrão amostral e os outros valores têm o mesmo significado anterior.

Pela abordagem **clássica** compara-se o valor da variável de teste com o valor crítico ($t_{n-1, \text{crítico}}$ que depende do Nível de Significância adotado) de acordo com o tipo de teste:

Se $H_1 \mu: > \mu_0$ Rejeitar H_0 se $t_{n-1, \text{calculado}} > t_{n-1, \text{crítico}}$ ($\bar{x} > \bar{x}_{\text{crítico}}$)

Se $H_1 \mu: < \mu_0$ Rejeitar H_0 se $t_{n-1, \text{calculado}} < t_{n-1, \text{crítico}}$ ($\bar{x} < \bar{x}_{\text{crítico}}$)¹⁴

Se $H_1 \mu: \neq \mu_0$ Rejeitar H_0 se $|t_{n-1, \text{calculado}}| > |t_{n-1, \text{crítico}}|$

Pela abordagem do **valor-p** obtém-se a probabilidade associada ao valor da variável de teste:

Se $H_1 \mu: > \mu_0$ Rejeitar H_0 se $P(t_{n-1} > t_{n-1, \text{calculado}}) < \alpha$

Se $H_1 \mu: < \mu_0$ Rejeitar H_0 se $P(t_{n-1} < t_{n-1, \text{calculado}})$ ¹⁵ $< \alpha$

Se $H_1 \mu: \neq \mu_0$ Rejeitar H_0 se $2 \times P(|t_{n-1}| > |t_{n-1, \text{calculado}}|)$ ¹⁶ $< \alpha$

10.4.2 - Testes de Hipóteses sobre a Proporção de um dos Valores de uma Variável em uma População

Neste caso há interesse em testar a hipótese de que o parâmetro proporção populacional (π) de um dos valores de uma certa variável seja **maior**, **menor** ou **diferente** de um certo valor. Para a realização deste teste, tal como será descrito é necessário que DUAS condições sejam satisfeitas:

- que o produto $n \times \pi_0$ seja maior ou igual a 5;
- que o produto $n \times (1 - \pi_0)$ seja maior ou igual a 5.

Onde n é o tamanho da amostra e π_0 é a proporção sob teste (de um dos valores da variável). Se AMBAS as condições forem satisfeitas a distribuição amostral da proporção que é binomial (uma Bernoulli repetida n vezes) pode ser aproximada por uma normal. Obviamente supõe-se que a amostra é representativa da população e foi retirada de forma aleatória, e que a variável pode assumir apenas dois valores, aquele no qual há interesse e o seu complementar¹⁷.

Se as condições acima forem satisfeitas a distribuição amostral da proporção poderá ser aproximada por uma normal com:

¹⁴ Neste caso $t_{n-1, \text{crítico}}$ será negativo, já que a região de Rejeição de H_0 está à esquerda de zero.

¹⁵ Neste caso calcula-se a probabilidade de t_{n-1} ser MENOR do que o $t_{n-1, \text{calculado}}$ pois o teste é unilateral à esquerda.

¹⁶ Neste caso multiplica-se por 2 a probabilidade de t_{n-1} ser MAIOR do que o valor em módulo de $t_{n-1, \text{calculado}}$ pois o teste é bilateral.

¹⁷ Ou seja, que se trata de um “experimento de Bernoulli”. Praticamente qualquer experimento pode ser reduzido a um experimento de Bernoulli, simplesmente isolando o valor da variável no qual há interesse e agrupando todos os outros como seu valor complementar.

$$\text{Média} = \pi_0$$

$$\text{Desvio Padrão} = \sqrt{\frac{\pi_0 \times (1 - \pi_0)}{n}}$$

Lembrando-se da expressão da variável Z :

$$Z = \frac{\text{valor} - \text{"média"}}{\text{"desvio padrão"}}$$

O **valor** será a proporção amostral (melhor estimador da proporção populacional) do valor da variável denotada por p . A "**média**" e o "**desvio padrão**" são os valores definidos acima, então a expressão do valor calculado de Z será:

$$Z_{\text{calculado}} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 \times (1 - \pi_0)}{n}}}$$

Pela abordagem **clássica** compara-se o valor da variável de teste com o valor crítico ($Z_{\text{crítico}}$ que depende do Nível de Significância adotado) de acordo com o tipo de teste:

Se $H_1 \pi: > \pi_0$ Rejeitar H_0 se $Z_{\text{calculado}} > Z_{\text{crítico}}$ ($p > p_{\text{crítico}}$)

Se $H_1 \pi: < \pi_0$ Rejeitar H_0 se $Z_{\text{calculado}} < Z_{\text{crítico}}$ ($p < p_{\text{crítico}}$)¹⁸

Se $H_1 \pi: \neq \pi_0$ Rejeitar H_0 se $|Z_{\text{calculado}}| > |Z_{\text{crítico}}|$

Pela abordagem do **valor-p** obtém-se a probabilidade associada ao valor da variável de teste:

Se $H_1 \pi: > \pi_0$ Rejeitar H_0 se $P(Z > Z_{\text{calculado}}) < \alpha$

Se $H_1 \pi: < \pi_0$ Rejeitar H_0 se $P(Z < Z_{\text{calculado}})^{19} < \alpha$

Se $H_1 \pi: \neq \pi_0$ Rejeitar H_0 se $2 \times P(Z > |Z_{\text{calculado}}|)^{20} < \alpha$

Exemplo 10.1- Uma peça automotiva precisa ter 100 mm de diâmetro, exatamente. Foram medidas 15 peças, aleatoriamente escolhidas. Obteve-se média de 100,7 mm e variância de 0,01 mm². Supõe-se que a dimensão segue distribuição normal na população. A peça está dentro das especificações? Usar 1% de significância.

1) Enunciar as hipóteses.

Conforme visto no item 10.4 o teste mais adequado para este caso é um Teste Bilateral:

$H_0: \mu = 100$ mm onde $\mu_0 = 100$ mm (valor de teste)

$H_1: \mu \neq 100$ mm

2) Nível de significância. O problema declara que é necessário usar 1% de significância.

3) Variável de teste.

Uma vez que a variância populacional da variável é DESCONHECIDA (o valor fornecido é a variância AMOSTRAL), a variável de teste será t_{n-1} da distribuição t de Student.

4) Definir a região de aceitação de H_0 .

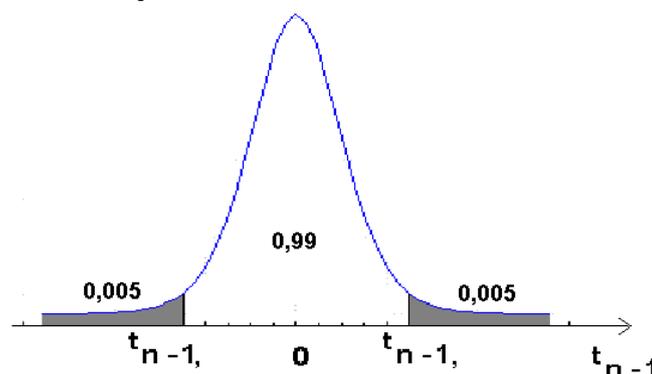


Figura 4 - Regiões de aceitação e rejeição - Teste de 1 média - bilateral

¹⁸ Neste caso $Z_{\text{crítico}}$ será NEGATIVO, já que a região de Rejeição de H_0 está à esquerda de zero.

¹⁹ Neste caso calcula-se a probabilidade de Z ser MENOR do que o $Z_{\text{calculado}}$, pois o teste é unilateral à esquerda.

²⁰ Neste caso multiplica-se por 2 a probabilidade de Z ser MAIOR do que o valor em módulo de $Z_{\text{calculado}}$ pois o teste é bilateral.

Observe que por ser um teste Bilateral o Nível de Significância α foi dividido em dois, metade para cada região de rejeição de H_0 .

Pela abordagem clássica:

Procurando o valor crítico na tabela da distribuição de Student, na linha correspondente a $n-1$ graus de liberdade, ou seja, em $15 - 1 = 14$ graus de liberdade, para uma área na cauda superior igual a $0,005$ ($0,01/2$): os valores críticos serão $t_{14;0,005}$ e $t_{14;0,995}$ os quais serão iguais em módulo. E o valor de $t_{14,\text{crítico}}$ será igual a $2,977$ (em módulo).

Pela abordagem do valor-p: ao calcular a probabilidade de t_{14} ser maior/menor do que o valor calculado é preciso multiplicar por 2, porque o teste é bilateral.

5) Através dos valores da amostra avaliar o valor da variável.

Neste ponto é preciso encontrar o valor da variável de teste:

$$t_{n-1,\text{calculado}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

O valor de teste μ_0 é igual a 100 mm, a média amostral \bar{X} vale $100,7$ mm, o tamanho de amostra n é igual a 15 e o desvio padrão amostral s é a raiz quadrada de $0,01$ mm². Substituindo na equação acima:

$$t_{n-1,\text{calculado}} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = t_{15-1,\text{calculado}} = t_{14,\text{calculado}} = \frac{100,7 - 100}{\sqrt{0,01}/\sqrt{15}} = 27,11$$

$$\text{então } |t_{14,\text{calculado}}| = 27,11$$

6) Decidir pela aceitação ou rejeição de H_0 .

Como se trata de um teste bilateral:

- Pela abordagem clássica

Rejeitar H_0 se $|t_{n-1,\text{calculado}}| > |t_{n-1,\text{crítico}}|$

Como $|t_{14,\text{calculado}}| = 27,11 > |t_{n-1,\text{crítico}}| = |t_{14;0,995}| = 2,977$

REJEITAR H_0 a 1% de Significância (há 1% de chance de erro)

- Pela abordagem do valor-p

Rejeitar H_0 se $2 \times P(|t_{n-1}| > |t_{n-1,\text{calculado}}|) < \alpha$

Como $27,11$ é maior do que zero, calcula-se a probabilidade de que t_{14} seja maior do que $27,11$, que é praticamente igual a zero²¹. Mesmo multiplicando-a por 2, $\text{valor-p} \cong 0 \ll \alpha = 0,01$

REJEITAR H_0 a 1% de Significância (há 1% de chance de erro)

7) Interpretar a decisão no contexto do problema.

Há provas estatísticas suficientes de que a dimensão da peça não está dentro das especificações²².

Exemplo 10.2 - Um novo protocolo de atendimento foi implementado no Banco RMG, visando reduzir o tempo que as pessoas passam na fila do caixa. O protocolo será considerado satisfatório se a média do tempo de fila for **menor** do que 30 minutos. Suponha que o tempo que 35 clientes (selecionados aleatoriamente) passaram na fila foi monitorado, resultando em uma média de 29 minutos e desvio padrão de 5 minutos. O protocolo pode ser considerado satisfatório a 5% de significância?

1) Enunciar as hipóteses.

Conforme o item 10.4 o teste para este caso é um Teste Unilateral à Esquerda:

$H_0 : \mu = 30$ onde $\mu_0 = 30$ (valor de teste)

$H_1 : \mu < 30$

²¹ Usando a função DIST.T.CD(27,11;14) do Microsoft Excel®, $\text{valor-p} = 0,0000000000000842165$.

²² Cuidado com os casos de FRONTEIRA, em que o valor da variável de teste é muito próximo do valor crítico, ou o valor-p muito próximo do nível de significância α . Nesses casos a rejeição ou aceitação de H_0 pode ocorrer *por acaso*. Sempre que apresentar o resultado de fronteira recomende que uma nova amostra seja retirada para avaliar novamente o problema. Mas neste caso rejeita-se H_0 com folga.

2) Nível de significância. O problema declara que é necessário usar 5% ($\alpha = 0,05$).

3) Variável de teste.

Uma vez que a variância populacional da variável é DESCONHECIDA (o valor fornecido é o desvio padrão AMOSTRAL), a variável de teste será t_{n-1} da distribuição t de Student.

4) Definir a região de aceitação de H_0 .

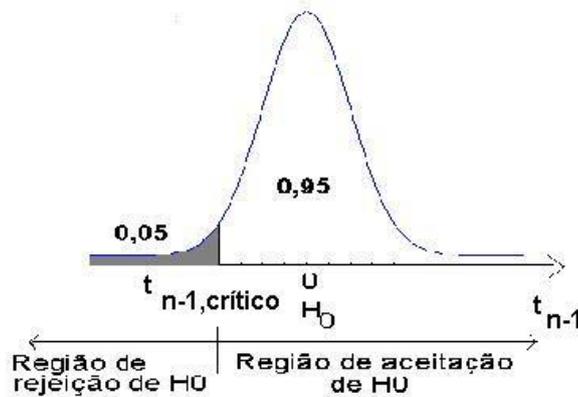


Figura 5 - Regiões de aceitação e rejeição - Teste de 1 média – unilateral à esquerda

Observe que por ser um teste Unilateral à Esquerda o Nível de Significância α está todo concentrado em um dos lados da distribuição, definindo a região de rejeição de H_0 .

Pela abordagem clássica:

Para encontrar o valor crítico devemos procurar na tabela da distribuição de Student, na linha correspondente a $n-1$ graus de liberdade, ou seja, em $35 - 1 = 34$ graus de liberdade.

$P(t > t_{n-1, crítico}) = 0,95$. Deve-se procurar a probabilidade complementar 0,05 e mudar o sinal do valor encontrado, pois o $t_{n-1, crítico}$ aqui é menor do que zero: o valor crítico será $t_{34;0,05}$. E o valor de $t_{34, crítico}$ será igual a $-1,691$ ²³.

Pela abordagem do valor-p: calcular a probabilidade de t_{n-1} ser menor do que o valor calculado, porque o teste é unilateral à esquerda.

5) Através dos valores da amostra avaliar o valor da variável.

$$t_{n-1, calculado} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

O valor de teste μ_0 é igual a 30, a média amostral \bar{X} vale 29, o tamanho de amostra n é igual a 35 e o desvio padrão amostral s é 5. Substituindo na equação acima:

$$t_{34, calculado} = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{29 - 30}{5/\sqrt{35}} = -1,183$$

6) Decidir pela aceitação ou rejeição de H_0 .

- Pela abordagem clássica

Rejeitar H_0 se $t_{n-1, calculado} < t_{n-1, crítico}$

Como $t_{34, calculado} = -1,183 > t_{n-1, crítico} = t_{34,0,05} = -1,691$

ACEITAR H_0 a 5% de Significância (há 5% de chance de erro)

- Pela abordagem do valor-p

Rejeitar H_0 se $P(t_{n-1} < t_{n-1, calculado}) < \alpha$

Calcula-se a probabilidade de que t_{34} seja menor do que $-1,183$ (porque o teste é unilateral à esquerda): usando uma função do Microsoft Excel ®²⁴ obtém-se 0,123. Valor-p = 0,123 $>$ $\alpha = 0,05$

ACEITAR H_0 a 5% de Significância (há 5% de chance de erro)

²³ Obtido através do Microsoft Excel ® pela função INV.T(0,05;34).

²⁴ 1-DIST.T.CD(-1,183;34), é preciso usar o complementar da função, porque a função DIST.T.CD calcula valores acumulados ACIMA do $t_{34, calculado}$ (-1,183) no caso, e como o teste é unilateral à esquerda busca-se obter a probabilidade ABAIXO de $t_{34, calculado}$.

7) Interpretar a decisão no contexto do problema.

Não há provas estatísticas suficientes para concluir que o protocolo tem um desempenho satisfatório.

EX.10.3 - Cerca de 2000 formulários de pedidos de compra estão sendo analisados. Os clientes podem ficar insatisfeitos se houver erros nos formulários. Neste caso admite-se que a proporção máxima de formulários com erros seja de 5%. Suponha que dentre os 2000 formulários 7% apresentavam erros. A proporção máxima foi ultrapassada a 1% de significância?

1) Enunciar as hipóteses.

Conforme o item 10.4 o teste o para este caso é um Teste Unilateral à Direita:

$$H_0 : \pi = 5\% \quad \text{onde } \pi_0 = 5\% \text{ (valor de teste)}$$

$$H_1 : \pi > 5\%$$

2) Nível de significância. O problema declara que é necessário usar 1% de significância.

3) Variável de teste.

Como se trata de um teste de proporção é necessário verificar o valor dos produtos:

$n \times \pi_0 = 2000 \times 0,05 = 100$ e $n \times (1 - \pi_0) = 2000 \times 0,95 = 1900$. Como ambos são maiores do que 5 é possível fazer uma aproximação pela normal, e a variável de teste será Z .

4) Definir a região de aceitação de H_0 .

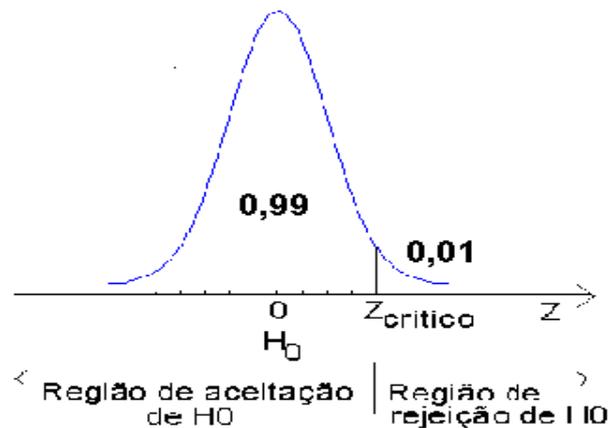


Figura 6 - Regiões de aceitação e rejeição - Teste de 1 proporção – unilateral à direita

Observe que por ser um teste Unilateral à Direita o Nível de Significância α está todo concentrado em um dos lados da distribuição, definindo a região de rejeição de H_0 .

Pela abordagem clássica:

Para encontrar o valor crítico devemos procurar na tabela da distribuição normal padrão, pela probabilidade acumulada 0,01 (o Z crítico aqui é maior do que zero). O valor crítico será igual a 2,326.

Pela abordagem do valor-p: calcular a probabilidade de Z ser maior do que o valor calculado, porque o teste é unilateral à direita.

5) Através dos valores da amostra avaliar o valor da variável.

Neste ponto é preciso encontrar o valor da variável de teste:

$$Z_{\text{calculado}} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 \times (1 - \pi_0)}{n}}}$$

O valor de teste π_0 é igual a 0,05 (5%), a proporção amostral p vale 0,07 (7%), e o tamanho de amostra n é igual a 2000. Substituindo na equação acima:

$$Z_{\text{calculado}} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 \times (1 - \pi_0)}{n}}} = \frac{0,07 - 0,05}{\sqrt{\frac{0,05 \times (0,95)}{2000}}} = 4,104$$

6) Decidir pela aceitação ou rejeição de H_0 .

- Pela abordagem clássica

Rejeitar H_0 se $Z_{\text{calculado}} > Z_{\text{crítico}}$

Como $Z_{\text{calculado}} = 4,104 > Z_{\text{crítico}} = 2,326$

REJEITAR H_0 a 1% de Significância (há 1% de chance de erro)

- Pela abordagem do valor-p

Rejeitar H_0 se $P(t_{n-1} > t_{n-1,\text{calculado}}) < \alpha$

Calcula-se a probabilidade de que Z seja maior do que 4,104 (porque o teste é unilateral à direita): usando uma função do Microsoft Excel ®^{25} obtém-se 0,0000203. Valor-p = 0,0000203 < $\alpha = 0,01$

REJEITAR H_0 a 1% de Significância (há 1% de chance de erro)

7) Interpretar a decisão no contexto do problema.

Há provas estatísticas suficientes de que a proporção está acima do máximo admitido²⁶.

Provavelmente os vendedores/compradores precisarão passar por novo treinamento.

10.4.3 - Teste de diferenças entre as Proporções de um dos Valores de uma Variável em duas Populações independentes

Frequentemente precisamos avaliar se a proporção de um dos valores de certa variável em uma população apresenta uma diferença significativa da mesma proporção em outra população (ou seja, comparar duas populações **independentes**). Por exemplo, se a proporção de eleitores favoráveis a um candidato em uma região de SC é diferente da proporção em outra região. Temos uma situação então em que queremos realizar um teste de diferença entre duas proporções populacionais, que chamaremos π_1 e π_2 .

Neste caso há interesse em testar a hipótese de que a diferença entre o parâmetro proporção populacional de um dos valores de uma certa variável, em duas populações **independentes** ($\pi_1 - \pi_2$) seja **maior**, **menor** ou **diferente** de um certo valor (que chamaremos π_0), a partir dos dados coletados em duas amostras aleatórias das respectivas populações. Vamos chamar as proporções **amostrais** do valor de interesse de p_1 e p_2 , e os respectivos tamanhos de amostra de n_1 e n_2 . Para a realização deste teste, tal como será descrito é necessário que QUATRO condições sejam satisfeitas:

- que o produto $n_1 \times p_1$ seja maior ou igual a 5; que o produto $n_1 \times (1 - p_1)$ seja maior ou igual a 5.

- que o produto $n_2 \times p_2$ seja maior ou igual a 5; que o produto $n_2 \times (1 - p_2)$ seja maior ou igual a 5.

Se TODAS as condições forem satisfeitas as distribuições amostrais das proporções, que são binomiais, podem ser aproximadas por uma normal. Obviamente supõe-se que as amostras são representativas das populações e foram retiradas de forma aleatória, e que a variável pode assumir apenas dois valores, aquele no qual há interesse e o seu complementar.

Se as condições acima forem satisfeitas a distribuição amostral da diferença entre as proporções poderá ser aproximada por uma normal com:

$$\text{Média} = \pi_0 \quad \text{Desvio Padrão} = \sqrt{\left(\frac{\pi_1 \times (1 - \pi_1)}{n_1}\right) + \left(\frac{\pi_2 \times (1 - \pi_2)}{n_2}\right)}$$

²⁵ 1-DIST.NORMP.N(4,104;VERDADEIRO), é preciso usar o complementar da função, porque a função DIST.NORMP.N calcula valores acumulados ABAIXO do $Z_{\text{calculado}}$ (4,104) no caso, e como o teste é unilateral à direita busca-se obter a probabilidade ACIMA de $Z_{\text{calculado}}$.

²⁶ Este NÃO é um caso de fronteira.

Onde π_0 é o valor da diferença entre as proporções populacionais π_1 e π_2 que queremos testar²⁷. Como não conhecemos π_1 e π_2 vamos estimá-las através das proporções amostrais, assim o "desvio padrão" passa a ser:

$$\text{Desvio Padrão} = \sqrt{\left(\frac{p_1 \times (1-p_1)}{n_1}\right) + \left(\frac{p_2 \times (1-p_2)}{n_2}\right)}$$

Lembrando-se da expressão da variável Z : $Z = \frac{\text{valor} - \text{"média"}}{\text{"desvio padrão"}}$

O **valor** será a diferença entre as proporções amostrais do valor da variável, denotada por $p_1 - p_2$. A "**média**" e o "**desvio padrão**" são os valores definidos acima, então a expressão de $Z_{\text{calculado}}$ será:

$$Z_{\text{calculado}} = \frac{p_1 - p_2 - \pi_0}{\sqrt{\left(\frac{p_1 \times (1-p_1)}{n_1}\right) + \left(\frac{p_2 \times (1-p_2)}{n_2}\right)}}$$

Pela abordagem **clássica** compara-se o valor da variável de teste com o valor crítico ($Z_{\text{crítico}}$ que depende do Nível de Significância adotado) de acordo com o tipo de teste (as expressões abaixo também estão no apêndice):

Se $H_1 \pi_1 - \pi_2 > \pi_0$ **Rejeitar H_0 se $Z_{\text{calculado}} > Z_{\text{crítico}}$**
 Se $H_1 \pi_1 - \pi_2 < \pi_0$ **Rejeitar H_0 se $Z_{\text{calculado}} < Z_{\text{crítico}}$** ²⁸
 Se $H_1 \pi_1 - \pi_2 \neq \pi_0$ **Rejeitar H_0 se $|Z_{\text{calculado}}| > |Z_{\text{crítico}}|$**

Pela abordagem do **valor-p** calcula-se a probabilidade associada ao valor da variável de teste:

Se $H_1 \pi_1 - \pi_2 > \pi_0$ **Rejeitar H_0 se $P(Z > Z_{\text{calculado}}) < \alpha$**
 Se $H_1 \pi_1 - \pi_2 < \pi_0$ **Rejeitar H_0 se $P(Z < Z_{\text{calculado}})^{29} < \alpha$**
 Se $H_1 \pi_1 - \pi_2 \neq \pi_0$ **Rejeitar H_0 se $2 \times P(Z > |Z_{\text{calculado}}|)^{30} < \alpha$**

EX. 10.4 - Uma lei extremamente polêmica está em tramitação na Assembleia Legislativa de Santa Catarina. Parece que homens e mulheres apresentam opiniões divergentes. Para avaliar se os dois grupos apresentam proporções de favoráveis diferentes realizou-se uma pesquisa, em que foi entrevistada uma amostra de 200 homens e uma amostra de 300 mulheres de todo o estado. Na amostra de homens 50% declaram-se favoráveis à lei, enquanto que na amostra das mulheres houve 47% de favoráveis. Usando 5% de significância, há diferença entre as proporções populacionais de favoráveis nos dois grupos?

Chamamos a população masculina de grupo 1 e a feminina de 2. Há interesse em verificar apenas se as proporções populacionais de favoráveis são diferentes, ou seja, se a diferença entre elas é igual a zero: então $\pi_0 = 0$. O teste será bilateral, pois não interessa qual das proporções será maior ou menor. Seguindo o roteiro do Apêndice:

1) *Enunciar as hipóteses.* *Rearranjando as hipóteses:*

$H_0 : \pi_1 - \pi_2 = 0$ onde $\pi_0 = 0$ $H_0 : \pi_1 = \pi_2$
 $H_1 : \pi_1 - \pi_2 \neq 0$ $H_1 : \pi_1 \neq \pi_2$

2) *Nível de significância. O problema declara que é necessário usar 5% de significância.*

3) *Variável de teste.*

Como se trata de um teste de diferença entre proporções é necessário verificar o valor dos produtos:

$n_1 \times p_1 = 200 \times 0,5 = 100$ $n_1 \times (1 - p_1) = 200 \times 0,50 = 100.$
 $n_2 \times p_2 = 300 \times 0,47 = 147$ $n_2 \times (1 - p_2) = 300 \times 0,53 = 153.$

²⁷ Usualmente π_0 é igual a zero, queremos saber se há qualquer diferença significativa entre as proporções.

²⁸ Neste caso $Z_{\text{crítico}}$ será NEGATIVO, já que a região de Rejeição de H_0 está à esquerda de zero.

²⁹ Neste caso calcula-se a probabilidade de Z ser MENOR do que o $Z_{\text{calculado}}$, pois o teste é unilateral à esquerda.

³⁰ Neste caso multiplica-se por 2 a probabilidade de Z ser MAIOR do que o valor em módulo de $Z_{\text{calculado}}$ pois o teste é bilateral.

Como todos são maiores do que 5 é possível fazer uma aproximação pela normal, e a variável de teste será Z .

4) Definir a região de aceitação de H_0 .

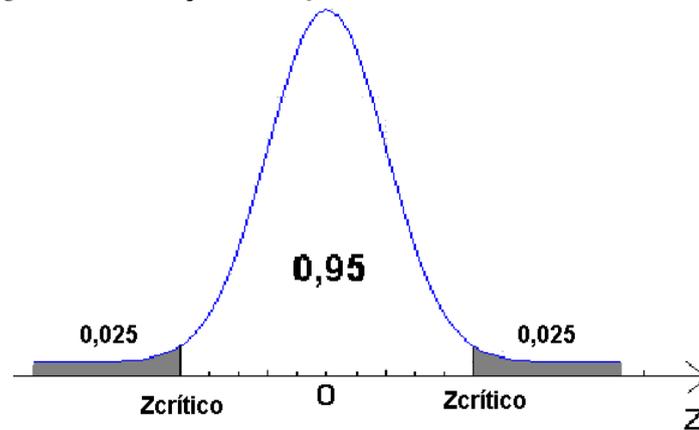


Figura 7 - Regiões de aceitação e rejeição - Teste de diferença entre 2 proporções - bilateral

Observe que por ser um teste Bilateral o Nível de Significância α foi dividido em dois, metade para cada região de rejeição de H_0 .

Pela abordagem clássica:

Procurando o valor crítico na tabela da distribuição normal padrão pela probabilidade 0,025 e 0,975 (0,95+ 0,025) para uma área na cauda superior igual a 0,025 (0,05/2): os valores críticos serão $Z_{0,025}$ e $Z_{0,975}$ os quais serão iguais em módulo. E o valor de $Z_{crítico}$ será igual a 1,96 (em módulo).

Pela abordagem do valor-p: ao calcular a probabilidade de Z ser maior/menor do que o valor calculado é preciso multiplicar por 2, porque o teste é bilateral.

5) Através dos valores da amostra avaliar o valor da variável.

Neste ponto é preciso encontrar o valor da variável de teste:

$$Z_{calculado} = \frac{p_1 - p_2 - \pi_0}{\sqrt{\left(\frac{p_1 \times (1 - p_1)}{n_1}\right) + \left(\frac{p_2 \times (1 - p_2)}{n_2}\right)}}$$

O valor de teste π_0 é igual a 0, a proporção amostral p_1 vale 0,5 (50%), a proporção amostral p_2 vale 0,47 (47%), o tamanho de amostra n_1 é igual a 200, e o tamanho de amostra n_2 é igual a 300. Substituindo na equação acima:

$$Z_{calculado} = \frac{p_1 - p_2 - \pi_0}{\sqrt{\left(\frac{p_1 \times (1 - p_1)}{n_1}\right) + \left(\frac{p_2 \times (1 - p_2)}{n_2}\right)}} = \frac{0,5 - 0,47 - 0}{\sqrt{\left(\frac{0,5 \times (0,5)}{200}\right) + \left(\frac{0,47 \times 0,53}{300}\right)}} = 0,658$$

6) Decidir pela aceitação ou rejeição de H_0 .

Como se trata de um teste bilateral:

- Pela abordagem clássica

Rejeitar H_0 se $|Z_{calculado}| > |Z_{crítico}|$

Como $|Z_{calculado}| = 0,658 > |Z_{crítico}| = 1,96$

ACEITAR H_0 a 5% de Significância (há 5% de chance de erro)

- Pela abordagem do valor-p

Rejeitar H_0 se $2 \times P(|Z| > |Z_{calculado}|) < \alpha$

Como 0,6577 é maior do que zero, obtém-se a probabilidade de que Z seja maior do que 0,658, aproximadamente 0,2546 através da

tabela da normal padrão. Multiplicando-a por 2, valor-p $\cong 0,5092 > \alpha = 0,05$

ACEITAR H_0 a 5% de Significância (há 5% de chance de erro)

7) Interpretar a decisão no contexto do problema.

Não há provas estatísticas suficientes que indiquem diferenças nas proporções de favoráveis nas populações de homens e mulheres.

10.5 - Teste do Quiquadrado de independência

O teste do Quiquadrado de independência está vinculado à análise bidimensional de variáveis qualitativas. No Capítulo 3 de INE7001 estudamos as tabelas de contingências e o coeficiente de contingência modificado, que permitia mensurar a força do relacionamento entre duas variáveis qualitativas. Vamos rever alguns daqueles conceitos antes de apresentar o teste do Quiquadrado de independência.

10.5.1 - Variáveis qualitativas e tabelas de contingência

É comum haver interesse em saber se duas variáveis quaisquer estão relacionadas, e o quanto estão relacionadas, seja na vida prática, seja em trabalhos de pesquisa, por exemplo:

- se a satisfação com um produto está relacionada à faixa etária do consumidor;
- o quanto o nível de pluviosidade em uma certa região influencia a produtividade de uma cultura agrícola;
- se a função exercida por uma pessoa em uma organização está associada a seu sexo.

A Análise Bidimensional (vista no Capítulo 3, em INE7001) propõe-se a tentar responder as perguntas do parágrafo anterior. As duas variáveis abordadas podem ser qualitativas ou quantitativas, e para cada tipo haverá técnicas apropriadas.

Variáveis Qualitativas são as variáveis cujas realizações são atributos, categorias. Como exemplo de variáveis qualitativas tem-se: sexo de uma pessoa (duas categorias, masculino e feminino), grau de instrução (analfabeto, primeiro grau incompleto, etc.), opinião sobre um assunto (favorável, desfavorável, indiferente).

Em estudos sobre variáveis qualitativas é extremamente comum registrar as frequências de ocorrência de cada valor que as variáveis podem assumir, e quando há duas variáveis envolvidas é comum registrar-se a frequência de ocorrência dos cruzamentos entre valores: por exemplo, quantas pessoas do sexo masculino são favoráveis a uma certa proposta de lei, quantas são desfavoráveis, quantas pessoas do sexo feminino são favoráveis. E, para facilitar a análise dos resultados estes resultados costumam ser dispostos em uma Tabela de Contingências. A Tabela de Contingências relaciona os possíveis valores de uma variável qualitativa com os possíveis valores da outra, registrando quantas ocorrências foram verificadas de cada cruzamento.

Exemplo 10.5 - Seja a tabela de contingências abaixo, que relaciona as funções exercidas e o sexo de 474 funcionários de uma organização.

Sexo	Função			Total
	Escritório	Serviços gerais	Gerência	
Masculino	157	27	74	258
Feminino	206	0	10	216
Total	363	27	84	474

Fonte: hipotética

Podemos apresentar os percentuais calculados em relação aos totais das colunas:

Sexo	Função			Total
	Escritório	Serviços gerais	Gerência	
Masculino	43,25%	100%	88,10%	54%
Feminino	56,75%	0%	11,90%	46%
Total	100%	100%	100%	100%

Fonte: hipotética

Seria interessante saber se as duas variáveis são estatisticamente dependentes, e o quão forte é esta associação. Repare que os percentuais de homens e mulheres em cada função são diferentes dos percentuais marginais (de homens e mulheres no total de funcionários), sendo que em duas funções as diferenças são bem grandes.

O Teste do Quiquadrado é uma das ferramentas estatísticas mais utilizadas quando se deseja estudar o relacionamento entre duas variáveis QUALITATIVAS. Permite verificar se duas variáveis qualitativas são **independentes**, se as proporções de ocorrência dos valores das variáveis observadas estão de acordo com o que era esperado, etc. Neste texto haverá interesse em usar o teste para avaliar se duas variáveis qualitativas são independentes.

Como todo teste de hipóteses o Teste do Quiquadrado consiste em comparar os valores observados em uma amostra com os valores de uma referência (referência esta que supõe que a hipótese nula seja válida).

As frequências observadas da variável são representadas em uma tabela de contingências, e a Hipótese Nula (H_0) do teste será que as duas variáveis não diferem em relação às frequências com que ocorre uma característica particular, ou seja, as variáveis são independentes, que será testada contra a Hipótese Alternativa (H_1) de que as variáveis NÃO SÃO independentes.

O teste pode ser realizado porque o grau de dependência pode ser quantificado descritivamente através de uma estatística, que se chama justamente Quiquadrado (χ^2), cuja expressão é:

$$\chi^2_{gl, calculado} = \sum_{i=1}^L \sum_{j=1}^C \left[\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right]$$

$$E_{ij} = \frac{\text{total da linha } i \times \text{total da coluna } j}{\text{total geral}}$$

Onde

- E_{ij} é a frequência esperada, sob a condição de independência entre as variáveis, em uma célula qualquer da tabela de contingências. TODAS as frequências esperadas precisam ser MAIORES OU IGUAIS A 5 para que o resultado do teste seja válido³¹.
- O_{ij} é a frequência observada em uma célula qualquer da tabela de contingências.
- L é o número total de linhas da tabela de contingências (número de valores que uma das variáveis pode assumir)
- C é o número total de colunas da tabela (número de valores que a outra variável pode assumir).

Então, para cada célula da tabela de contingências calcula-se a diferença entre a frequência observada e a esperada. Para evitar que as diferenças positivas anulem as negativas as diferenças são elevadas ao quadrado. E para evitar que uma diferença grande em termos absolutos, mas pequena em termos relativos, "inflacione" a estatística, ou que uma diferença pequena em termos absolutos, mas grande em termos relativos, tenha sua influência reduzida, divide-se o quadrado da

³¹ Se isso não ocorrer recomenda-se agrupar as categorias (de uma ou outra variável, ou de ambas) até obter todas as frequências pelo menos iguais a 5.

diferença pela frequência esperada. Somam-se os valores de todas as células e obtêm-se o valor da estatística: quanto maior χ^2 , mais o Observado afasta-se do Esperado, portanto maior a dependência.

Sob a hipótese de independência a estatística χ^2 terá distribuição Quiquadrado, uma distribuição assimétrica, cujo comportamento depende do seu número de **graus de liberdade**.

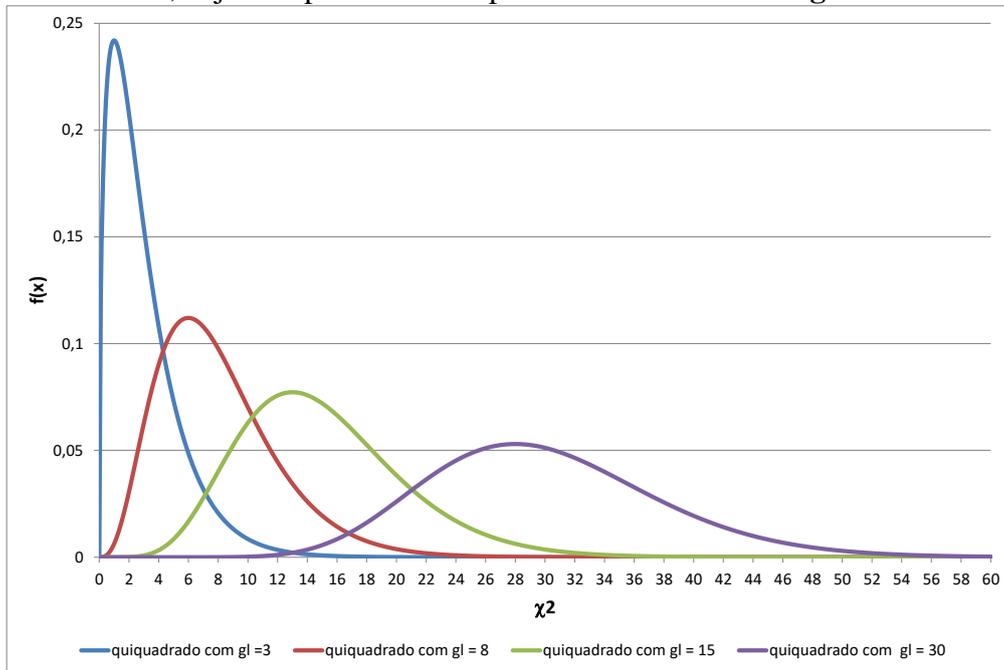


Figura 8 - Distribuição Quiquadrado com 3, 8, 15 e 30 graus de liberdade

A Figura 8 mostra as curvas da distribuição Quiquadrado para 3, 8, 15 e 30 graus de liberdade. Observe que as figuras são assimétricas, e variam de forma dependendo do número de graus de liberdade da estatística, mas que para os maiores valores elas começam a ficar simétricas.

O número de graus de liberdade (gl) da estatística é calculado da seguinte forma:

$$\text{graus de liberdade} = (L - 1) \times (C - 1)$$

Sendo o número de linhas (L) e o número de colunas (C) referentes à tabela de contingências (o número de valores que cada variável pode assumir). O número de graus de liberdade assume este valor porque para calcular as frequências esperadas não é necessário calcular os valores de TODAS as células, as últimas podem ser calculadas por diferença já que os totais são fixos. Por exemplo, para duas variáveis que somente podem assumir 2 valores cada, o número de graus de liberdade seria igual a 1 [(2-1)×(2-1)]: bastaria calcular a frequência esperada de uma célula e obter as outras por diferença em relação ao total.

O teste do Quiquadrado para avaliar se duas variáveis são independentes será sempre UNILATERAL:

Pela abordagem **clássica** compara-se o valor da variável de teste com o valor crítico ($\chi^2_{\text{crítico}}$ que depende do Nível de Significância adotado e do número de graus de liberdade) de acordo com o tipo de teste (as expressões abaixo também estão no apêndice):

$$\text{Rejeitar } H_0 \text{ se } \chi^2_{\text{gl,calculado}} > \chi^2_{\text{crítico}}$$

Pela abordagem do **valor-p** calcula-se a probabilidade associada ao valor da variável de teste:

$$\text{Rejeitar } H_0 \text{ se } P(\chi^2 > \chi^2_{\text{gl,calculado}}) < \alpha$$

Exemplo 10.6 - Para o conjunto do Exemplo 10.5, supondo que os resultados são uma amostra aleatória, verificar se as variáveis são independentes a 1% de significância.

1) Enunciar as Hipóteses:

H_0 : as variáveis sexo e função são independentes

H_1 : as variáveis sexo e função não são independentes

2) Nível de significância: determinado pelo problema, $\alpha = 0,01$; $1 - \alpha = 0,99$

3) Retirar as amostras aleatórias e montar a tabela de contingências (isso já foi feito):

Sexo	Função			Total
	Escritório	Serviços gerais	Gerência	
Masculino	157	27	74	258
Feminino	206	0	10	216
Total	363	27	84	474

Fonte: hipotética

Na tabela acima encontram-se os totais marginais e o total geral:

$L1 = \text{total Masculino} = 258$ $L2 = \text{total Feminino} = 216$ $C1 = \text{total Escritório} = 157$

$C2 = \text{total S.Gerais} = 27$ $C3 = \text{total gerência} = 84$ $N = \text{total geral} = 474$

Repare que somando os totais das linhas o resultado é o total geral, e que somando os totais das colunas o resultado é o total geral também.

4) Calcular as frequências esperadas:

Calculando as frequências esperadas de acordo com a fórmula vista anteriormente:

Masculino - Escritório $E = (258 \times 363) / 474 = 197,58$

Masculino - Serviços Gerais $E = (258 \times 27) / 474 = 14,70$

Masculino - Gerência $E = (258 \times 84) / 474 = 45,72$

Feminino - Escritório $E = (216 \times 363) / 474 = 165,42$

Feminino - Serviços Gerais $E = (216 \times 27) / 474 = 12,30$

Feminino - Gerência $E = (216 \times 84) / 474 = 38,28$

Observe que os resultados são os mesmos obtidos no Exemplo 3.2.

5) Calculando a estatística χ^2 para cada célula:

Agora podemos calcular as diferenças entre as frequências e as demais operações, que serão mostradas nas tabelas a seguir.

O - E	Função		
Sexo	Escritório	Serviços gerais	Gerência
Masculino	157 - 197,58	27 - 14,70	74 - 45,72
Feminino	206 - 165,42	0 - 12,30	10 - 38,28

(O-E) ²	Função		
Sexo	Escritório	Serviços gerais	Gerência
Masculino	1646,921	151,383	799,672
Feminino	1646,921	151,383	799,672

Finalmente:

(O-E) ² /E	Função		
Sexo	Escritório	Serviços gerais	Gerência
Masculino	8,336	10,301	17,490
Feminino	9,956	12,304	20,891

Agora podemos somar os valores:

$$\chi^2 = 8,336 + 10,301 + 17,490 + 9,956 + 12,304 + 20,891 = 79,227$$

Os graus de liberdade: (número de linhas - 1) x (número de colunas - 1) = (2 - 1) x (3 - 1) = 2

Então $\chi^2_2 = 79,227$

6) Pela abordagem clássica:

O χ^2 crítico será: procurando na tabela do Apêndice, ou em um programa, para 2 graus de liberdade e 99% de confiança (1% de significância): $\chi^2_{2,\text{crítico}} = 9,21$

Pela abordagem do valor-p

É preciso calcular a probabilidade de χ^2_2 ser maior do que 79,227.

7) *Pela abordagem clássica:*

$$\chi^2_2 = 79,227 > \chi^2_{2, \text{crítico}} = 9,21$$

REJEITAR H_0 a 1% de Significância (há 1% de chance de erro)

Pela abordagem do valor-p

$P(\chi^2_2 > 79,227) < 0,0005$ (esta é a menor probabilidade listada na tabela para 2 graus de liberdade, correspondente a $\chi^2_2 = 15,20$)³².

Como valor-p < $\alpha = 0,01$

REJEITAR H_0 a 1% de Significância (há 1% de chance de erro)

8) *Há evidência estatística de que as variáveis Sexo e Função estão associadas.*

É importante ressaltar que ao calcularmos o Coeficiente de Contingência Modificado para os mesmos dados (Exemplo 3.4, Capítulo 3 de INE7001), obtivemos um valor igual a 0,54 (indicando uma associação moderada). Ele pode ser usado em conjunto com o teste do Quiquadrado. O teste do Quiquadrado provou que há associação, e o coeficiente a quantificou.

10.6 - Testes de diferenças entre as médias de uma variável em duas populações (testes t)

É extremamente comum comparar as médias de uma variável QUANTITATIVA em duas populações³³, através das médias de duas amostras aleatórias de valores da variável retiradas destas populações, geralmente com um dos dois objetivos abaixo:

- comparar as médias dos valores da variável provenientes de grupos PAREADOS (usualmente o MESMO grupo), procurando verificar se houve diferença (comumente, estudos do tipo antes-depois);
- comparar as médias provenientes de grupos DISTINTOS (INDEPENDENTES), para verificar se a variável possui a mesma média nas duas populações.

Estes testes são TESTES PARAMÉTRICOS. Exigem que uma das duas condições abaixo seja satisfeita:

- sabe-se, ou é razoável supor, que a variável de interesse segue uma distribuição normal em ambas as populações: isso significa que a distribuição amostral das médias também será normal, permitindo realizar a inferência estatística paramétrica.
- as distribuições da variável nas populações são desconhecidas, mas as amostras retiradas destas populações são consideradas “suficientemente grandes”³⁴ o que, de acordo com o Teorema Central do Limite, permite concluir que a distribuição amostral das médias é normal.

Supõe-se também que as amostras são representativas das populações e foram retiradas de forma ALEATÓRIA (sorteio não viciado).

³² Usando a função DIST.QUIQUA.CD(79,227;2) no Microsoft Excel ® o valor-p é igual a 0,000000000000000006253

³³ Aliás, é muito comum comparar as médias de vários grupos, o que consiste um capítulo especial da Estatística a ANÁLISE DE VARIÂNCIA (ANOVA) que não será vista nesta disciplina.

³⁴ Há muita controvérsia a respeito do que seria uma amostra “suficientemente grande”, mas geralmente uma amostra com pelo menos 30 elementos costuma ser considerada grande o bastante para que a distribuição amostral da média possa ser aproximada por uma normal. Mas há autores que mencionam 50 elementos, outros ainda mencionam que depende do comportamento da variável na população: se a distribuição for unimodal (tiver apenas um pico) 30 elementos é considerada amostra grande, mas se for multimodal seriam necessários mais dados, o problema é que geralmente não temos este conhecimento sobre a variável na população.

O teste t para 2 amostras pareadas será visto no item 10.6.1. Se os grupos forem independentes e as variâncias populacionais são conhecidas deve ser utilizado o teste Z para 2 amostras independentes (item 10.6.2). Se não forem conhecidas, através do teste F de diferença entre 2 variâncias (no item 10.6.3) é possível inferir se são semelhantes ou não. No item 10.6.4 será apresentado o teste t para 2 amostras independentes: se o teste F (item 10.6.3) indicar que as variâncias podem ser consideradas iguais a variável de teste t terá n_1+n_2-2 graus de liberdade (onde n_1 e n_2 são os tamanhos das amostras); caso contrário a variável de teste t terá ν graus de liberdade (ν é calculado através de uma expressão que envolve os valores dos desvios padrões amostrais e os próprios tamanhos de amostra). A Figura 9 mostra os possíveis cursos de ação.

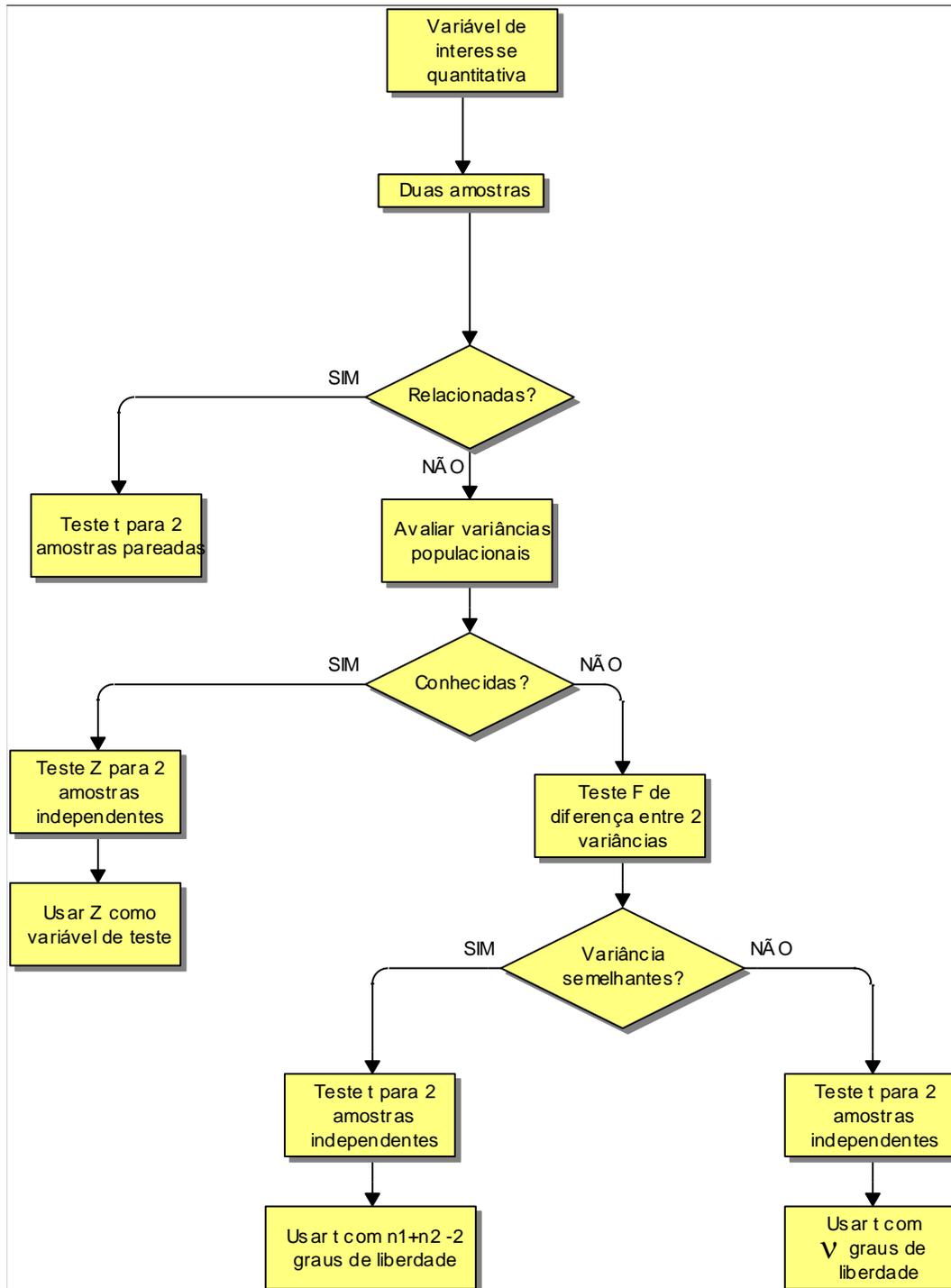


Figura 9 - Teste de diferença entre 2 médias - resumo

10.6.1 - Teste t para 2 amostras pareadas

Os resultados das 2 amostras são relacionados 2 a 2 segundo algum critério, exigindo que as amostras precisem ter o mesmo tamanho. Calculam-se as diferenças entre cada par de valores, obtendo uma única amostra com n diferenças.

EX 10.7 - Dez cobaias foram submetidas ao tratamento de engorda com certa ração. Os pesos em gramas, antes e após o teste são dados a seguir (supõe-se que provenham de distribuições normais). A 1% de significância, podemos concluir que o uso da ração contribuiu para o aumento do peso médio dos animais?

Cobaia	1	2	3	4	5	6	7	8	9	10
Antes	635	704	662	560	603	745	698	575	633	669
Depois	640	712	681	558	610	740	707	585	635	682

Trata-se de uma situação em que queremos comparar as MÉDIAS DE UMA VARIÁVEL EM DUAS distribuições normais, supondo que se trata da MESMA população, mas em dois momentos diferentes: antes e após um tratamento de engorda. Há interesse em verificar se a dieta contribuiu para o aumento do peso médio dos animais: ou seja, queremos verificar se a média de peso antes do tratamento é MENOR do que a média de peso após o tratamento (se a dieta fez efeito os animais estarão em média mais pesados ao final do tratamento). Reparem que é exigido que se tome uma decisão, o que configura um problema de TESTE DE HIPÓTESES.

Iremos então aplicar um TESTE DE DIFERENÇAS ENTRE AS MÉDIAS DE UMA VARIÁVEL EM 2 POPULAÇÕES. E como as amostras são relacionadas (MESMA POPULAÇÃO: ANTES E DEPOIS), usaremos o teste t para 2 amostras pareadas.. Supõe-se que se ambas as distribuições populacionais são normais a distribuição da diferença entre os valores também será.

1) Enunciar as hipóteses

De acordo com o que foi dito acima queremos verificar se a média antes é menor do que a média depois; o melhor ponto de partida, que servirá para a definição da hipótese H_0 , é que a dieta NÃO FAZ EFEITO, ou seja, as médias antes e após o tratamento são iguais (costumamos colocar em H_0 o CONTRÁRIO do que queremos provar), ou seja, a DIFERENÇA ENTRE AS MÉDIAS DEVE SER SUPOSTA IGUAL A ZERO, teremos então:

$$\begin{aligned} H_0: \mu_d &= 0 \\ H_1: \mu_d &< 0 \quad \text{onde } \mu_d = \mu_{\text{antes}} - \mu_{\text{depois}} \end{aligned}$$

2) Estabelecer o nível de significância ou nível de confiança.

Conforme foi estabelecido no enunciado do problema:

$$\alpha = 0,01 \quad 1 - \alpha = 0,99$$

3) Identificar a variável de teste.

No presente problema temos uma amostra de apenas 10 elementos. Como a variância populacional das diferenças é desconhecida a variável de teste que será utilizada será a variável t_{n-1} da distribuição t de Student.

4) Definir a região de aceitação de H_0 , de acordo com o tipo de teste e variável.

Trata-se de um teste unilateral à esquerda (com 1% de significância), e a variável de teste é t_{n-1} (a amostra tem 10 elementos).

Pela abordagem clássica: então o valor crítico (obtido da tabela da distribuição t de Student do Apêndice) será: $t_{n-1, \text{critico}} = t_{10-1; 0,01} = t_{9; 0,99} = -t_{9; 0,01} = -2,821$

Rejeita-se H_0 se o $t_{n-1, \text{calculado}}$ for menor do que $-2,821$

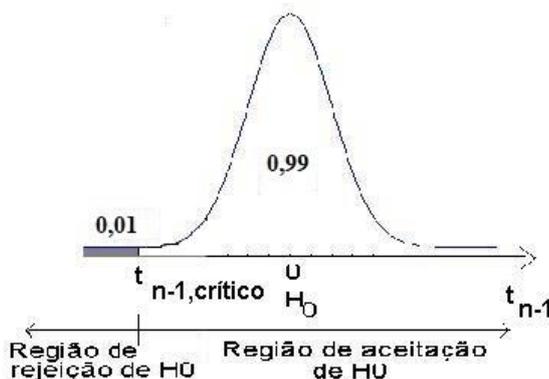


Figura 10 - Região de rejeição - teste unilateral à esquerda - 1% de significância

Pela abordagem do valor-p calcula-se a probabilidade de que t_{n-1} seja MENOR do que $t_{n-1, calculado}$, porque o teste é unilateral à ESQUERDA. Se o valor-p for menor do que 0,01, rejeita-se H_0 .

5) Através dos valores das amostras antes e depois, calcular a diferença d_i entre cada par de valores, onde $d_i = X_{antes} - X_{depois}$.

Para o conjunto sob análise teremos:

Cobaia	1	2	3	4	5	6	7	8	9	10
Antes	635	704	662	560	603	745	698	575	633	669
Depois	640	712	681	558	610	740	707	585	635	682
d_i	-5	-8	-19	2	-7	5	-9	-10	-2	-13
d_i^2	25	64	361	4	49	25	81	100	4	169

6) e 7) Calcular a diferença média e o desvio padrão da diferença média.

Para o presente problema: $\bar{d} = \frac{\sum d_i}{n} = \frac{-66}{10} = -6,6$ gramas

$$s_d = \sqrt{\frac{\sum d_i^2 - [(\sum d_i)^2/n]}{n-1}} = \sqrt{\frac{882 - [(-66)^2/10]}{10-1}} = 7,04 \text{ gramas}$$

8) Calcular o valor da variável de teste. Neste problema é a variável t_{n-1} :

$$t_{n-1, calculado} = \frac{\bar{d}}{(s_d/\sqrt{n})} \quad t_{10-1, calculado} = t_{9, calculado} = \frac{-6,6}{(7,04/\sqrt{10})} = -2,96$$

9) Decidir pela aceitação ou rejeição de H_0 .

- Pela abordagem clássica

Rejeitar H_0 se $t_{n-1, calculado} < t_{n-1, crítico}$

Como $t_{9, calculado} = -2,96 < t_{n-1, crítico} = t_{9, 0,01} = -2,821$

REJEITAR H_0 a 1% de Significância (há 1% de chance de erro)

- Pela abordagem do valor-p

Rejeitar H_0 se $P(t_{n-1} < t_{n-1, calculado}) < \alpha$

Calcula-se a probabilidade de que t_9 seja menor do que -2,96 (porque o teste é unilateral à esquerda): usando uma função do Microsoft Excel³⁵ obtém-se 0,00798³⁶.

Valor-p = 0,00798 < $\alpha = 0,01$

REJEITAR H_0 a 1% de Significância (há 1% de chance de erro)

³⁵ 1-DIST.T.CD(-2,96;9), é preciso usar o complementar da função, porque a função DIST.T.CD calcula valores acumulados ACIMA do $t_{9, calculado}$ (-2,96) no caso, e como o teste é unilateral à esquerda busca-se obter a probabilidade ABAIXO de $t_{9, calculado}$.

³⁶ Na tabela da distribuição t, para 9 graus de liberdade, observa-se que quando a área na cauda é igual a 0,005 t vale 3,250, então para t igual a 2,96 o valor-p deve estar entre 0,01 (que corresponde a t igual a 2,821) e 0,005, portanto MENOR do que 0,01.

10) Interpretar a decisão dentro do contexto do problema.

Assim, concluímos com 99% de confiança (ou uma chance de erro de 1%) que a dieta contribuiu para o aumento do peso médio dos animais.

10.6.2 – Teste Z para 2 amostras independentes – variâncias populacionais CONHECIDAS.

Neste caso há interesse em comparar as médias de uma variável quantitativa em dois grupos (populações) distintas, independentes, através de duas amostras aleatórias retiradas de cada grupo, respectivamente. Além disso, as variâncias da variável nas duas populações (σ_1^2 e σ_2^2) são conhecidas³⁷.

EX.10.8 – A JabyL Circuits está avaliando o tempo de montagem de um novo modelo de circuito em 2 de suas unidades. Suspeita-se que o desempenho da fábrica 1 seja pior do que o da fábrica 2 (aquela seria mais lenta). Sabe-se que a variância populacional do tempo na fábrica 1 é de 1,5 minutos² e na fábrica 2 de 2,5 minutos², além disso, supõe-se que as distribuições dos tempos podem ser consideradas normais. Foram coletadas duas amostras de tempos de montagem: 8 na fábrica 1, resultando em média de 2,35 minutos, e 10 na fábrica 2, resultando em média de 1,85 minutos. A suspeita é procedente a 1% de significância?

Trata-se de uma situação em que queremos comparar as MÉDIAS DE DUAS distribuições normais, supondo que se tratam de duas populações distintas, podemos supor que as amostras são independentes: está sendo avaliado o tempo de montagem dos circuitos em 2 fábricas DIFERENTES. Há interesse em verificar se a média da população da fábrica 1 é maior do que a de fábrica 2 (ou seja, que leve mais tempo para montar os circuitos na fábrica 1, significando um desempenho pior). Reparem que é exigido que se tome uma decisão, o que configura um problema de TESTE DE HIPÓTESES.

Iremos então aplicar um TESTE DE DIFERENÇA ENTRE AS MÉDIAS DE UMA VARIÁVEL EM 2 POPULAÇÕES. E como as amostras são independentes, e as variâncias populacionais são CONHECIDAS, deve-se usar o teste Z para 2 amostras independentes.

1) Enunciar as hipóteses

De acordo com o que foi dito acima queremos verificar se a média da fábrica 1 é maior do que a da fábrica 2; o melhor ponto de partida, que servirá para a definição da hipótese H_0 , será considerar que NÃO HÁ DIFERENÇA entre as médias, ou seja, A MÉDIA DO TEMPO DE MONTAGEM NA FÁBRICA 1 É IGUAL AO TEMPO DE MONTAGEM NA FÁBRICA 2 (costumamos colocar em H_0 o CONTRÁRIO do que queremos provar), teremos então:

$$\begin{aligned} H_0: \mu_1 &= \mu_2 \\ H_1: \mu_1 &> \mu_2 \end{aligned} \quad \text{onde } \mu_1 = \mu_{FÁBRICA1} \text{ e } \mu_2 = \mu_{FÁBRICA2}$$

2) Estabelecer o nível de significância ou nível de confiança.

Conforme foi estabelecido no enunciado do problema:

$$\alpha = 0,01 \quad 1 - \alpha = 0,99$$

3) Identificar a variável de teste.

Há 3 variáveis de teste possíveis, dependendo das condições do problema, mais especificamente das variâncias das duas populações.

Como as variâncias de ambas as populações são conhecidas deverá ser usada a variável Z da distribuição normal padrão.

4) Definir a região de aceitação de H_0 , de acordo com o tipo de teste e variável.

³⁷ Este caso é muito raro na prática, mas a compreensão do seu procedimento é útil.

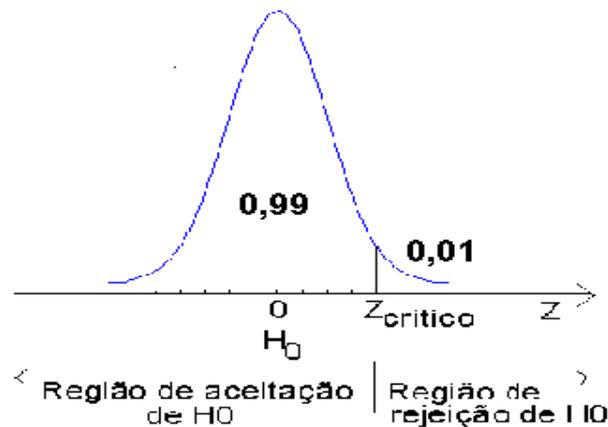


Figura 11 - Regiões de aceitação e rejeição - Teste de diferença entre 2 médias – variâncias populacionais conhecidas – unilateral à direita

Observe que por ser um teste Unilateral à Direita o Nível de Significância α está todo concentrado em um dos lados da distribuição, definindo a região de rejeição de H_0 .

Pela abordagem clássica:

Para encontrar o valor crítico devemos procurar na tabela da distribuição normal padrão, pela probabilidade acumulada 0,01 (o $Z_{\text{crítico}}$ aqui é maior do que zero). O valor crítico será igual a 2,326.

Pela abordagem do valor-p: calcular a probabilidade de Z ser MAIOR do que o valor calculado, porque o teste é unilateral à direita.

5) Calcular o desvio padrão das diferenças.

Como as duas variâncias são conhecidas, iremos utilizar a primeira expressão para calcular o desvio padrão das diferenças que está no roteiro:

$$\sigma_d = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{1,5}{8} + \frac{2,5}{10}} = 0,6614 \text{ min}$$

6) Calcular a variável de teste.

Novamente, como as duas variâncias são conhecidas, iremos utilizar a primeira expressão para calcular o valor da variável de teste que está no roteiro:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma_d} \Rightarrow Z = \frac{2,35 - 1,85}{0,6614} = 0,7559$$

7) Decidir pela aceitação ou rejeição de H_0 .

- Pela abordagem clássica

Rejeitar H_0 se $Z_{\text{calculado}} > Z_{\text{crítico}}$

Como $Z_{\text{calculado}} = 0,7559 < Z_{\text{crítico}} = 2,326$

ACEITAR H_0 a 1% de Significância (há 1% de chance de erro)

- Pela abordagem do valor-p

Rejeitar H_0 se $P(Z > Z_{\text{calculado}}) < \alpha$

Calcula-se a probabilidade de que Z seja maior do que 0,7559 (porque o teste é unilateral à direita): usando uma função do Microsoft Excel ³⁸ obtém-se 0,2249³⁹.

Valor-p = 0,2249 > $\alpha = 0,01$

ACEITAR H_0 a 1% de Significância (há 1% de chance de erro)

³⁸ 1-DIST.NORMP.N(0,7559;VERDADEIRO), é preciso usar o complementar da função, porque a função DIST.NORMP.N calcula valores acumulados ABAIXO do $Z_{\text{calculado}}$ (0,7559) no caso, e como o teste é unilateral à direita busca-se obter a probabilidade ACIMA de $Z_{\text{calculado}}$.

³⁹ Na tabela da distribuição Z, para $Z = 0,7559 \cong 0,76$, a probabilidade de $Z > 0,76$ vale 0,2236.

8) Interpretar a decisão dentro do contexto do problema.

Assim, concluímos com 99% de confiança (ou uma chance de erro de 1%) que NÃO há evidências estatísticas suficientes para declarar que a média do tempo de montagem dos circuitos na Fábrica 1 é maior do que a média na Fábrica 2. A suspeita NÃO é procedente.

10.6.3 – Teste F de diferença entre 2 variâncias.

Para realizar um teste de diferença entre as médias de uma variável em 2 populações, sendo as amostras independentes, se as variâncias populacionais da variável DESCONHECIDAS, é preciso avaliar se tais variâncias podem ser consideradas iguais ou não. Tal avaliação é necessária, pois dependendo da sua conclusão a variável de teste t terá diferentes graus de liberdade, o que implicará em valores críticos (e valores-p) diferentes, o que influenciará diretamente a decisão de aceitação ou rejeição da hipótese nula. Portanto, é preciso realizar um teste para avaliar se há diferenças entre as variâncias populacionais, a partir das variâncias amostrais.

As hipóteses do teste são sempre as mesmas: na hipótese nula supõe-se que as duas variâncias populacionais (estimadas a partir das respectivas variâncias amostrais) são iguais, e na hipótese alternativa declara-se que elas são diferentes (trata-se então de um teste BILATERAL).

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Para realizar o teste é preciso calcular o quociente entre as variâncias amostrais. Então a variável de teste do teste F (F calculado) será:

$$F_{n_1-1, n_2-1, calculado} = \frac{S_1^2}{S_2^2}$$

Esta variável segue uma distribuição amostral chamada F de Fisher (ou de Snedecor), trata-se de uma distribuição assimétrica, que somente pode assumir valores positivos, e que possui graus de liberdade associados ao numerador e ao denominador de um quociente. Veja a figura abaixo, de uma distribuição F com 14 graus de liberdade no numerador e 14 graus de liberdade no denominador, considerando uma confiança de 95% (significância de 5%):

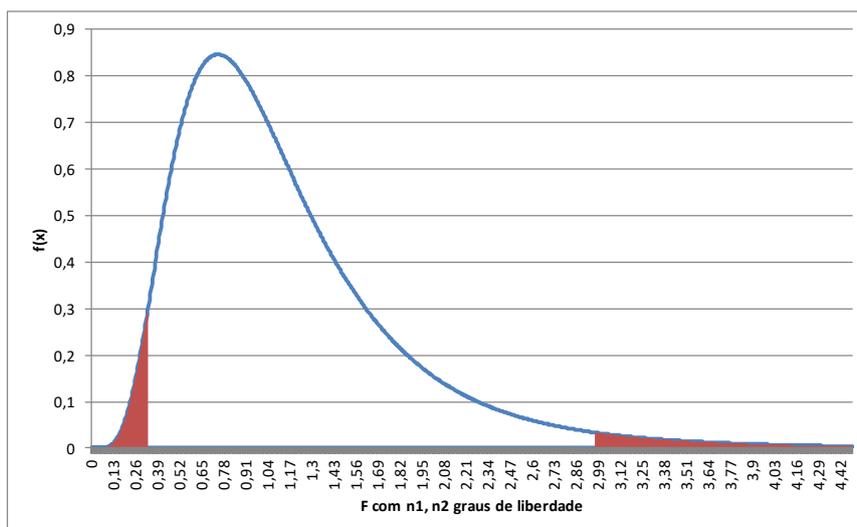


Figura 12 - Distribuição F - 14 graus de liberdade no numerador e denominador – 5% de significância

Na Figura 12 é possível perceber duas áreas hachuradas: estas áreas correspondem a 0,05 (valor comum para o nível de significância α de testes de hipóteses) dividido por 2, resultando 0,025: isso foi feito porque o teste F implementado neste capítulo será sempre bilateral.

Para o caso da Figura 12, pela abordagem *clássica*⁴⁰ é preciso obter os valores de F correspondentes a 0,025 e 0,975 (0,025 + 0,95): os F críticos. Isso é possível através da tabela da distribuição F existente no Apêndice, ou através da função INV.F⁴¹ na planilha eletrônica Microsoft Excel®:

- se o F calculado estiver ENTRE os F críticos, ACEITA-SE H_0 para o nível de significância α ;
- se o F calculado estiver ABAIXO do menor F crítico, ou ACIMA do maior F crítico, REJEITA-SE H_0 para o nível de significância α .

Exemplos de aplicação do teste F serão apresentados na próxima seção.

10.6. 4 - Teste t para 2 amostras independentes – variâncias populacionais DESCONHECIDAS.

Neste caso há interesse em comparar as médias de uma variável quantitativa em dois grupos (populações) distintas, independentes, através de duas amostras aleatórias retiradas de cada grupo, respectivamente. Mas, as variâncias da variável nas duas populações (σ_1^2 e σ_2^2) são DESCONHECIDAS: então é preciso verificar se tais variâncias podem ser consideradas iguais ou não, para definir quantos graus de liberdade terá a variável de teste t.

Se as variâncias forem supostas iguais, o número de graus de liberdade será $n_1 + n_2 - 2$ (onde n_1 e n_2 são os tamanhos das amostras). Se as variâncias forem supostas diferentes, o número de graus de liberdade será igual a v , cuja expressão é:

$$v = \left[\frac{(\omega_1 + \omega_2)^2}{\frac{\omega_1^2}{(n_1 + 1)} + \frac{\omega_2^2}{(n_2 + 1)}} \right]$$

$$\text{Onde } \omega_1 = \frac{s_1^2}{n_1} \quad \text{e} \quad \omega_2 = \frac{s_2^2}{n_2}$$

n_1 e n_2 são os tamanhos das amostras e s_1^2 e s_2^2 são as variâncias AMOSTRAIS.

Exemplo 10.9 - A Jabyt Circuits está avaliando o tempo de montagem de um novo modelo de circuito em 2 de suas unidades. Suspeita-se que o desempenho da fábrica 1 seja pior do que o da fábrica 2 (aquela seria mais lenta). Supõe-se que as distribuições dos tempos podem ser consideradas normais. Foram coletadas duas amostras de tempos de montagem: 8 na fábrica 1, resultando em média de 3,52 minutos, variância de 1,5 minutos² e 10 na fábrica 2, resultando em média de 1,85 minutos e variância de 1,7 minutos². A suspeita é procedente a 1% de significância?

Trata-se de uma situação em que queremos comparar as MÉDIAS DE DUAS distribuições normais, supondo que se trata de duas populações distintas, podemos supor que as amostras são independentes: está sendo avaliado o tempo de montagem dos circuitos em 2 fábricas DIFERENTES. Há interesse em verificar se a média da população da fábrica 1 é maior do que a de fábrica 2 (ou seja, que leve mais tempo para montar os circuitos na fábrica 1, significando um desempenho pior). Reparem que é exigido que se tome uma decisão, o que configura um problema de TESTE DE HIPÓTESES.

⁴⁰ Para a resolução manual a abordagem clássica é mais simples, os aplicativos computacionais usam a abordagem do valor-p, as conclusões têm que ser as mesmas.

⁴¹ No presente caso os valores seriam obtidos por INV.F(0,025;14;14), resultando 0,336 e INV.F(0,975;14;14), resultando 2,979.

Iremos então aplicar um **TESTE DE DIFERENÇA ENTRE AS MÉDIAS DE UMA VARIÁVEL EM 2 POPULAÇÕES**. Como as amostras são independentes, e as variâncias populacionais são **DESCONHECIDAS**, vamos usar o teste *t* para 2 amostras independentes. Antes, porém precisamos saber se as variâncias populacionais podem ser consideradas iguais ou não: é preciso aplicar o teste *F* de diferença entre 2 variâncias.

TESTE F

$$H_0: \sigma^2_1 = \sigma^2_2$$

$$H_1: \sigma^2_1 \neq \sigma^2_2$$

Queremos apenas verificar se há **DIFERENÇA** entre as variâncias, por isso o teste será sempre **BILATERAL**.

Nível de significância: podemos usar 1% de significância, tal como especificado para o teste das médias.

Neste problema teremos⁴²:

$$s_1^2 = 1,5 \text{ min}^2 \quad \text{e} \quad n_1 = 8$$

$$s_2^2 = 1,7 \text{ min}^2 \quad \text{e} \quad n_2 = 10$$

Então a variável de teste do teste *F* será:

$$F_{n_1-1, n_2-1} = \frac{s_1^2}{s_2^2} \rightarrow F_{7,9} = \frac{1,5}{1,7} = 0,882$$

Procurando os valores críticos na tabela da distribuição F^{43} para 1% de significância (0,025 e 0,975 de probabilidade acumulada), para 7 graus de liberdade no numerador e 9 no denominador: $F_{7,9;0,005} = 0,117$ e $F_{7,9;0,995} = 6,885$

Como o valor da variável $F_{7,9}$ (0,882) está **ENTRE** os valores críticos podemos **ACEITAR** a hipótese de que as variâncias populacionais, desconhecidas, são iguais, com uma chance de erro de 1%. Sendo assim, a nossa variável *t* de Student terá n_1+n_2-2 graus de liberdade.

Agora podemos partir para o teste *t* propriamente dito.

1) Enunciar as hipóteses

De acordo com o que foi no enunciado do problema queremos verificar se a média da fábrica 1 é maior do que a da fábrica 2; o melhor ponto de partida, que servirá para a definição da hipótese H_0 , será considerar que **NÃO HÁ DIFERENÇA** entre as médias, ou seja, **A MÉDIA DO TEMPO DE MONTAGEM NA FÁBRICA 1 É IGUAL AO TEMPO DE MONTAGEM NA FÁBRICA 2** (costumamos colocar em H_0 o **CONTRÁRIO** do que queremos provar), teremos então:

$$H_0: \mu_1 = \mu_2 \quad \text{onde} \quad \mu_1 = \mu_{FÁBRICA1} \quad \text{e} \quad \mu_2 = \mu_{FÁBRICA2}$$

$$H_1: \mu_1 > \mu_2$$

2) Estabelecer o nível de significância ou nível de confiança.

Conforme foi estabelecido no enunciado do problema:

$$\alpha = 0,01 \quad 1 - \alpha = 0,99$$

3) Identificar a variável de teste.

Como as variâncias de ambas as populações são **desconhecidas**, e o teste *F* de diferença entre 2 variâncias indicou que podem ser supostas iguais, deverá ser usada a variável **t** da distribuição *t* de Student com n_1+n_2-2 graus de liberdade.

4) Definir a região de aceitação de H_0 , de acordo com o tipo de teste e variável.

Pela abordagem clássica: trata-se de um teste Unilateral à Direita (com 1% de significância), e a variável de teste é *t* tem n_1+n_2-2 ($8 + 10 - 2 = 16$) graus de liberdade. Então o valor crítico (obtido da tabela da distribuição normal padrão) será:

$$t_{\text{critico}} = t_{16;0,01} = 2,583$$

Rejeita-se H_0 se o $t_{n_1+n_2-2, \text{calculado}}$ for maior do que 2,583.

⁴² A variância é o quadrado do desvio padrão. Se forem dados os desvios padrões amostrais basta elevá-los ao quadrado.

⁴³ Ou a função INV.F no Microsoft Excel ®: = INV.F(0,005;7;9) e =INV.F(0,995;7;9)

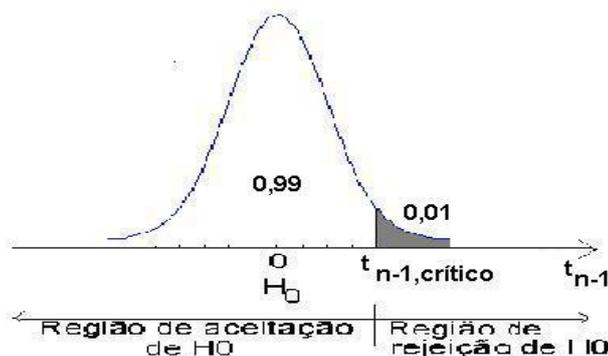


Figura 13 - Região de rejeição - teste unilateral à direita - 1% de significância

Pela abordagem do valor-p calcula-se a probabilidade de que $t_{n_1+n_2-2}$ seja MAIOR do que $t_{n_1+n_2-2, calculado}$, porque o teste é unilateral à DIREITA. Se o valor-p for menor do que 0,01, rejeita-se H_0 .

5) Calcular o desvio padrão das diferenças.

Como as duas variâncias são desconhecidas e supostas iguais, iremos utilizar a segunda expressão para calcular o desvio padrão das diferenças:

$$s_d = \sqrt{\frac{[(n_1 - 1) \times s_1^2 + (n_2 - 1) \times s_2^2] \times \left[\frac{1}{n_1} + \frac{1}{n_2}\right]}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{[(8 - 1) \times 1,5 + (10 - 1) \times 1,7] \times \left[\frac{1}{8} + \frac{1}{10}\right]}{8 + 10 - 2}} = 0,6023$$

6) Calcular a variável de teste.

Novamente, como as duas variâncias são desconhecidas e supostas iguais, iremos utilizar a segunda expressão para calcular o valor da variável de teste:

$$t_{16} = \frac{\bar{x}_1 - \bar{x}_2}{s_d} \Rightarrow t_{16} = \frac{3,52 - 1,85}{0,6023} = 2,773$$

7) Decidir pela aceitação ou rejeição de H_0 .

- Pela abordagem clássica

Rejeitar H_0 se $t_{n_1+n_2-2, calculado} > t_{n_1+n_2-2, crítico}$

Como $t_{16, calculado} = 2,773 > t_{n_1+n_2-2, crítico} = t_{16; 0,01} = 2,583$

REJEITAR H_0 a 1% de Significância (há 1% de chance de erro)

- Pela abordagem do valor-p

Rejeitar H_0 se $P(t_{n_1+n_2-2} > t_{n_1+n_2-2, calculado}) < \alpha$

Calcula-se a probabilidade de que t_{16} seja maior do que 2,773 (porque o teste é unilateral à direita): usando uma função do Microsoft Excel ⁴⁴ obtém-se 0,006788⁴⁵.

Valor-p = 0,006788 < $\alpha = 0,01$

REJEITAR H_0 a 1% de Significância (há 1% de chance de erro)

⁴⁴ DIST.T.CD(2,773;16), a função DIST.T.CD calcula valores acumulados ACIMA do $t_{16, calculado}$ (2,773) no caso, e como o teste é unilateral à direita busca-se exatamente obter a probabilidade ACIMA de $t_{16, calculado}$.

⁴⁵ Na tabela da distribuição t, para 16 graus de liberdade, observa-se que quando a área na cauda é igual a 0,005 t vale 2,921, então para t igual a 2,773 o valor-p deve estar entre 0,01 (que corresponde a t igual a 2,583) e 0,005, portanto MENOR do que 0,01.

8) Interpretar a decisão dentro do contexto do problema.

Assim, concluímos com 99% de confiança (ou uma chance de erro de 1%) que HÁ evidências estatísticas suficientes para declarar que a média do tempo de montagem dos circuitos na Fábrica 1 é maior do que a média na Fábrica 2. A suspeita é procedente. Observe que é uma conclusão diferente daquela obtida no Exemplo 10.8.

Exemplo 10.10 - A Jabyl Circuits está avaliando o tempo de montagem de um novo modelo de circuito em 2 de suas unidades. Suspeita-se que o desempenho da fábrica 1 seja pior do que o da fábrica 2 (aquela seria mais lenta). Supõe-se que as distribuições dos tempos podem ser consideradas normais. Foram coletadas duas amostras de tempos de montagem: 8 na fábrica 1, resultando em média de 3,52 minutos, variância de 4,5 minutos² e 10 na fábrica 2, resultando em média de 1,85 minutos e variância de 0,6 minutos². A suspeita é procedente a 1% de significância?

Situação semelhante a do Exemplo 10.9, apenas os valores das variâncias amostrais sofreram modificação.

Trata-se de uma situação em que queremos comparar as MÉDIAS DE DUAS distribuições normais, supondo que se trata de duas populações distintas, podemos supor que as amostras são independentes: está sendo avaliado o tempo de montagem dos circuitos em 2 fábricas DIFERENTES. Há interesse em verificar se a média da população da fábrica 1 é maior do que a de fábrica 2 (ou seja, que leve mais tempo para montar os circuitos na fábrica 1, significando um desempenho pior). Reparem que é exigido que se tome uma decisão, o que configura um problema de TESTE DE HIPÓTESES.

Iremos então aplicar um TESTE DE DIFERENÇA ENTRE AS MÉDIAS DE UMA VARIÁVEL EM 2 POPULAÇÕES. Como as amostras são independentes, e as variâncias populacionais são DESCONHECIDAS, vamos usar o teste t para 2 amostras independentes. Antes, porém precisamos saber se as variâncias populacionais podem ser consideradas iguais ou não: é preciso aplicar o teste F de diferença entre 2 variâncias.

TESTE F

$$H_0: \sigma^2_1 = \sigma^2_2$$

$$H_1: \sigma^2_1 \neq \sigma^2_2$$

Queremos apenas verificar se há DIFERENÇA entre as variâncias, por isso o teste será sempre BILATERAL.

Nível de significância: como a tabela da distribuição F (que está na página 13 do Apêndice) apresenta valores apenas para 5% de significância (teste bilateral) este será o nível adotado em todos os testes F.

Encontrar a maior variância amostral, que será chamada de s_A^2 (e, por conseguinte, n_A) e a menor que será chamada s_B^2 (e por conseguinte, n_B). Neste problema teremos:

$$s_1^2 = 4,5 \text{ min}^2 \quad \text{e} \quad n_1 = 8$$

$$s_2^2 = 0,6 \text{ min}^2 \quad \text{e} \quad n_2 = 10$$

Então a variável de teste do teste F será:

$$F_{n_1-1, n_2-1} = \frac{s_1^2}{s_2^2} \rightarrow F_{7,9} = \frac{4,5}{0,6} = 7,5$$

Procurando os valores críticos na tabela da distribuição F^{46} para 1% de significância (0,025 e 0,975 de probabilidade acumulada), para 7 graus de liberdade no numerador e 9 no denominador: $F_{7,9;0,005} = 0,117$ e $F_{7,9;0,995} = 6,885$

Como o valor da variável $F_{7,9}$ (7,5) está ACIMA do MAIOR valor crítico (6,885) podemos REJEITAR a hipótese de que as variâncias populacionais, desconhecidas, são iguais, com uma chance de erro de 1%. Sendo assim, a nossa variável t de Student terá ν graus de liberdade.

⁴⁶ Ou a função INV.F no Microsoft Excel ®: =INV.F(0,005;7;9) e =INV.F(0,995;7;9)

Agora podemos partir para o teste *t* propriamente dito.

1) Enunciar as hipóteses

De acordo com o que foi no enunciado do problema queremos verificar se a média da fábrica 1 é maior do que a da fábrica 2; o melhor ponto de partida, que servirá para a definição da hipótese H_0 , será considerar que **NÃO HÁ DIFERENÇA** entre as médias, ou seja, **A MÉDIA DO TEMPO DE MONTAGEM NA FÁBRICA 1 É IGUAL AO TEMPO DE MONTAGEM NA FÁBRICA 2** (costumamos colocar em H_0 o **CONTRÁRIO** do que queremos provar), teremos então:

$$H_0: \mu_1 = \mu_2 \quad \text{onde } \mu_1 = \mu_{FÁBRICA1} \text{ e } \mu_2 = \mu_{FÁBRICA2}$$

$$H_1: \mu_1 > \mu_2$$

2) Estabelecer o nível de significância ou nível de confiança.

Conforme foi estabelecido no enunciado do problema:

$$\alpha = 0,01 \quad 1 - \alpha = 0,99$$

3) Identificar a variável de teste.

Como as variâncias populacionais são desconhecidas e supostas diferentes (de acordo com o teste *F*) a variável *t* terá **v** graus de liberdade, que será calculado a seguir:

$$v = \left[\frac{(\omega_1 + \omega_2)^2}{\frac{\omega_1^2}{n_1+1} + \frac{\omega_2^2}{n_2+1}} \right] - 2 \quad \text{onde } \omega_1 = \frac{s_1^2}{n_1} = \frac{4,5}{8} = 0,5625 \quad \omega_2 = \frac{s_2^2}{n_2} = \frac{0,6}{10} = 0,06$$

$$v = \left[\frac{(0,5625 + 0,06)^2}{\frac{(0,5625)^2}{8+1} + \frac{(0,06)^2}{10+1}} \right] - 2 \cong 8,92 \quad \text{assim, } v \cong 9$$

Então a variável de teste *t* de Student terá 9 graus de liberdade⁴⁷.

4) Definir a região de aceitação de H_0 , de acordo com o tipo de teste e variável.

Pela abordagem clássica: trata-se de um teste Unilateral à Direita (com 1% de significância), e a variável de teste é *t* tem *v* (9) graus de liberdade. Então o valor crítico (obtido da tabela da distribuição normal padrão) será:

$$t_{\text{crítico}} = t_{9,0,01} = 2,821$$

Rejeita-se H_0 se o $t_{v,\text{calculado}}$ for maior do que 2,821.

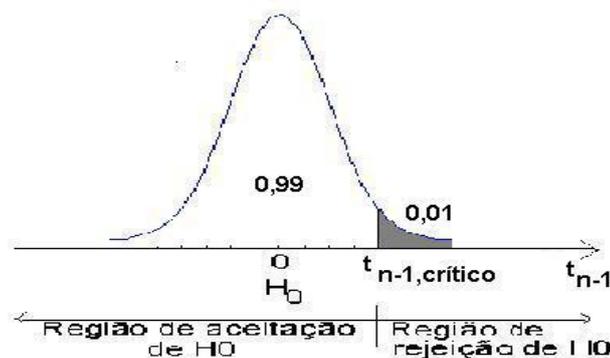


Figura 14 - Região de rejeição - teste unilaterial à direita - 1% de significância

Pela abordagem do valor-p calcula-se a probabilidade de que t_v seja **MAIOR** do que $t_{v,\text{calculado}}$, porque o teste é unilaterial à **DIREITA**. Se o valor-p for menor do que 0,01, rejeita-se H_0 .

5) Calcular o desvio padrão das diferenças.

Como as duas variâncias são desconhecidas e supostas diferentes, iremos utilizar a terceira expressão para calcular o desvio padrão das diferenças que está no roteiro.

⁴⁷ Observe que é um processo razoavelmente trabalhoso.

$$s_d = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = s_d = \sqrt{\frac{4,5}{8} + \frac{0,6}{10}} = 0,789$$

6) Calcular a variável de teste.

Novamente, como as duas variâncias são desconhecidas e supostas diferentes, iremos utilizar a terceira expressão para calcular o valor da variável de teste que está no roteiro.

$$t_9 = \frac{\bar{x}_1 - \bar{x}_2}{s_d} \Rightarrow t_9 = \frac{3,52 - 1,85}{0,789} = 2,117$$

7) Decidir pela aceitação ou rejeição de H_0 .

- Pela abordagem clássica

Rejeitar H_0 se $t_v, \text{calculado} > t_{v, \text{crítico}}$

Como $t_9, \text{calculado} = 2,117 < t_{v, \text{crítico}} = t_{9; 0,01} = 2,821$

ACEITAR H_0 a 1% de Significância (há 1% de chance de erro)

- Pela abordagem do valor-p

Rejeitar H_0 se $P(t_v > t_{v, \text{calculado}}) < \alpha$

Calcula-se a probabilidade de que t_9 seja maior do que 2,117 (porque o teste é unilateral à direita): usando uma função do Microsoft Excel ⁴⁸ obtém-se 0,031673⁴⁹.

Valor-p = 0,031673 > $\alpha = 0,01$

ACEITAR H_0 a 1% de Significância (há 1% de chance de erro)

8) Interpretar a decisão dentro do contexto do problema.

Assim, concluímos com 99% de confiança (ou uma chance de erro de 1%) que **NÃO** há evidências estatísticas suficientes para declarar que a média do tempo de montagem dos circuitos na Fábrica 1 é maior do que a média na Fábrica 2. A suspeita **NÃO** é procedente.

MUITO IMPORTANTE

Testes de hipóteses (todos os vistos aqui e também os não abordados) são ferramentas estatísticas para “small data”: isso significa que em um cenário em que houver milhares, dezenas de milhares, ou mesmo milhões de observações, ou seja, “big data”, eles não são as ferramentas estatísticas mais apropriadas para detectar diferenças entre grupos, por exemplo. Para tais situações ainda é possível recorrer aos intervalos de confiança, Análise Exploratória de Dados, e Análise Multivariada.

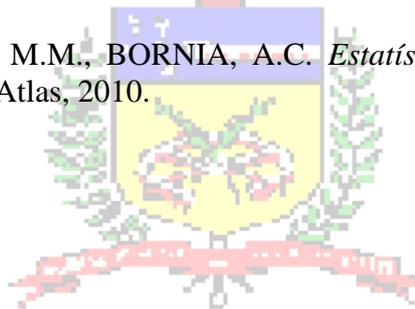
Hoje está na moda “Data Science”, que na essência é Estatística, mas com um nome aparentemente mais palatável.

⁴⁸ DIST.T.CD(2,117;9), a função DIST.T.CD calcula valores acumulados ACIMA do $t_{9, \text{calculado}}$ (2,117) no caso, e como o teste é unilateral à direita busca-se exatamente obter a probabilidade ACIMA de $t_{9, \text{calculado}}$.

⁴⁹ Na tabela da distribuição t, para 9 graus de liberdade, observa-se que quando a área na cauda é igual a 0,025 t vale 2,262, e que para área na cauda igual a 0,05, t vale 1,833, então para t igual a 2,117 o valor-p deve estar entre 0,025 (que corresponde a t igual a 2,262) e 0,05 (que corresponde a t igual a 1,833, portanto MAIOR do que 0,01).

BIBLIOGRAFIA

1. BARBETTA, P. A. *Estatística Aplicada às Ciências Sociais*. Ed. da UFSC, 7 ed. Florianópolis, 2007.
2. STEVENSON, Willian J. *Estatística Aplicada à Administração*. – São Paulo: Harbra, 2001
3. MOORE, D.S., McCABE, G.P., DUCKWORTH, W.M., SCLOVE, S. L., *A prática da estatística empresarial: como usar dados para tomar decisões*. Rio de Janeiro: LTC, 2006.
4. TRIOLA, M. F. – *Introdução à Estatística*, 9ª ed., Rio de Janeiro: LTC, 2005.
5. SOARES, J. F., FARIAS, A. A., CESAR, C. C. – *Introdução à Estatística*, 2ª ed., Rio de Janeiro: LTC, 2003.
6. BRAULE, Ricardo. *Estatística Aplicada com Excel: para cursos de administração e economia*. Rio de Janeiro: Campus, 2001.
7. LEVINE, D. M., STEPHAN, D., KREHBIEL, T. C., BERENSON, M. L. *Estatística: Teoria e Aplicações - Usando Microsoft Excel em Português*. 5ª ed. – Rio de Janeiro: LTC, 2005.
8. DOWNING, D., CLARK, J. *Estatística Aplicada*. São Paulo: Saraiva, 2000.
9. SILVA, Paulo Afonso Lopes da, *Probabilidades & Estatística*. Rio de Janeiro: Reichmann & Affonso Editores, 1999.
10. BARBETTA, P.A., REIS, M.M., BORNIA, A.C. *Estatística para Cursos de Engenharia e Informática*, 3ª ed., São Paulo: Atlas, 2010.



Marcelo
Menezes
Reis