

CEC3214 – ROTEIRO PARA RESOLUÇÃO DA LISTA DE INFERÊNCIA ESTATÍSTICA

1) A variável sob análise (tempo de atendimento) é QUANTITATIVA. Portanto serão feitas inferências sobre a MÉDIA.

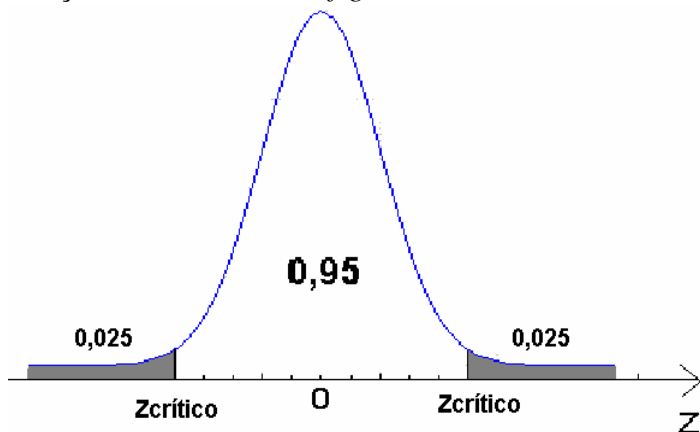
a) O parâmetro de interesse é a média populacional μ do tempo de atendimento.

Adotou-se um nível de confiança de 95%, então $1 - \alpha = 0,95$ $\alpha = 0,05$ $\alpha/2 = 0,025$.

Estatísticas disponíveis são: **média amostral** = 195 segundos $s = 15$ segundos $n = 40$

Definição da variável de teste: uma vez que a variância populacional da variável é DESCONHECIDA (o valor fornecido é o desvio padrão AMOSTRAL), mas a amostra retirada apresenta 40 elementos (portanto mais de 30) a variável de teste será Z da distribuição normal.

Encontrar o valor de $Z_{\text{crítico}}$: como o Intervalo de Confiança para a média é bilateral, teremos uma situação semelhante à da figura abaixo:



Para encontrar o valor crítico devemos procurar na tabela da distribuição normal padrão pela probabilidade 0,025 e 0,975 ($0,95 + 0,025$). O valor da probabilidade pode ser visto na figura ao lado: os valores críticos serão $Z_{0,025}$ e $Z_{0,975}$ os quais serão iguais em módulo. $P(Z > Z_{\text{crítico}}) = 0,025$. Então $Z_{\text{crítico}}$ será igual a 1,96 (em módulo).

Passa-se agora a determinação dos limites do intervalo:

$$L_I = \bar{x} - \frac{Z_{\text{crítico}} \times s}{\sqrt{n}} = 195 - \frac{1,96 \times 15}{\sqrt{40}} = 195 - 4,65 = 190,35 \text{ segundos}$$

$$L_S = \bar{x} + \frac{Z_{\text{crítico}} \times s}{\sqrt{n}} = 195 + \frac{1,96 \times 15}{\sqrt{40}} = 195 + 4,65 = 199,65 \text{ segundos}$$

Então o intervalo de 95% de confiança para a média populacional do tempo de atendimento é $[190,35; 199,65]$ segundos. Interpretação: há 95% de probabilidade de que a média populacional do tempo de atendimento esteja entre 190,35 e 199,65 segundos.

b) Como a variância populacional é DESCONHECIDA, e o tamanho da amostra é maior do que 30 elementos, pode ser usada a variável de teste Z da distribuição normal padrão. Assim será empregada a seguinte expressão para calcular o tamanho mínimo de amostra para a estimação por intervalo da média populacional.

$$n = \left(\frac{Z_{\text{crítico}} \times s}{e_0} \right)^2$$

O nível de significância é o mesmo do item a. Sendo assim, o valor crítico continuará sendo o mesmo: $Z_{\text{crítico}} = 1,96$. O desvio padrão amostral vale 15 segundos, e o valor de e_0 , a precisão, foi fixado em 1 minuto, ou seja 60 segundos. Basta então substituir os valores na expressão:

$$n = \left(\frac{Z_{\text{crítico}} \times s}{e_0} \right)^2 = \left(\frac{1,96 \times 15}{60} \right)^2 = 0,24 \cong 1 \text{ elementos}$$

Observe que o tamanho mínimo de amostra necessário para atender a 95% de confiança e precisão de 60 segundos deveria ser de 1 elemento. Como a amostra coletada possui 40 elementos ela é plenamente SUFICIENTE para a significância e precisão exigidas.

c) Observe que é preciso tomar uma decisão: com base nos dados da amostra a afirmação do dono da agência é verdadeira ou ele deve contratar mais um atendente? Trata-se então de um **teste de hipóteses**. A amostra foi coletada para avaliar se o tempo médio de atendimento de 3 minutos (180 segundos) ainda é válido: não haverá problema algum se o tempo for igual ou menor do que 180 segundos, mas se for **maior**, o dono da agência precisaria contratar um novo atendente. Então faremos um teste **unilateral à direita**.

Conforme visto acima o teste mais adequado para este caso é um Teste Unilateral à Direita:

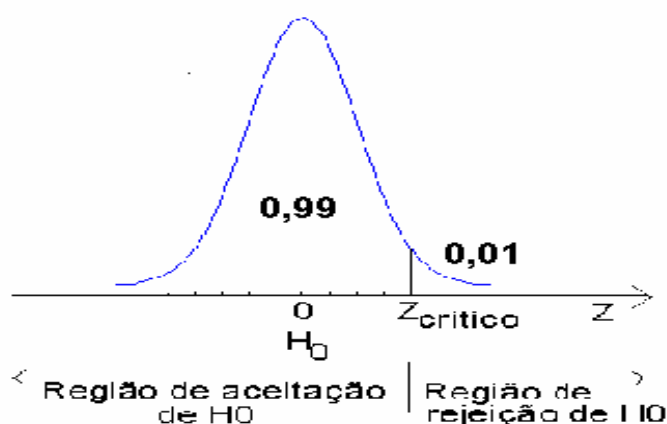
$$H_0 : m = 180 \text{ onde } m_0 = 180 \text{ segundos (valor de teste)}$$

$$H_1 : m > 180$$

Nível de significância. O problema recomenda usar 1%. Então $\alpha = 0,01$ e $1 - \alpha = 0,99$

Variável de teste. Uma vez que a variância populacional da variável é DESCONHECIDA (o valor fornecido é o desvio padrão AMOSTRAL), mas a amostra retirada apresenta 40 elementos (portanto mais de 30) a variável de teste será **Z** da distribuição normal.

Definir a região de aceitação de H_0 .



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 . Para encontrar o valor crítico devemos procurar na tabela da distribuição normal, pela probabilidade acumulada 0,01. Repare que o **Zcrítico** aqui é maior do que zero:

$$P(Z > Z_{\text{crítico}}) = 0,01.$$

Então **Zcrítico @ 2,33**

Através dos valores da amostra avaliar o valor da variável.

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

$$Z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

O valor de teste m_0 é igual a 180, a média amostral \bar{X} vale 195, o tamanho de amostra n é igual a 40 e o desvio padrão amostral s é 15. Substituindo na equação acima:

$$Z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{195 - 180}{15 / \sqrt{40}} = 6,32$$

Decidir pela aceitação ou rejeição de H_0 . Trata-se de um teste Unilateral à Direita, devemos calcular a probabilidade de Z ser maior do que 6,32.

Calculando o valor-p: pela tabela da distribuição observa-se que não há valores de Z superiores a 5: $P(Z > 5) = 0,000000287$. Como o Z encontrado vale 6,32, $P(Z > 6,32)$ deve ser ainda menor (na verdade é $1,31 \times 10^{-10}$), portanto muito menor do que o nível de significância. Então a decisão deve ser:

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

Há provas estatísticas suficientes de que o tempo médio de atendimento é maior do que 180 segundos. A afirmação do dono da agência não é verdadeira, um novo atendente deveria ser contratado.

2) A variável sob análise (tempo de montagem) é QUANTITATIVA. Portanto serão feitas inferências sobre a MÉDIA.

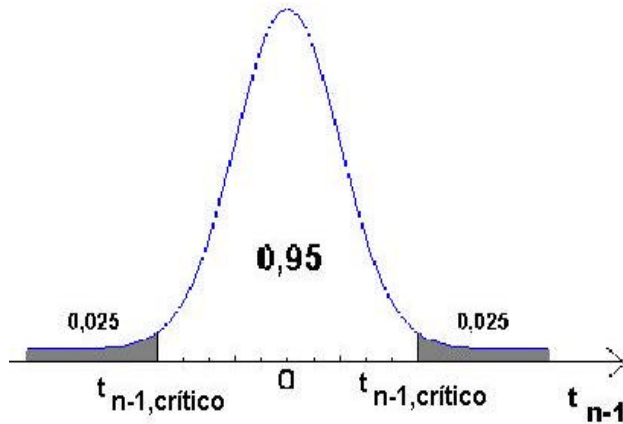
a) O parâmetro de interesse é a média populacional m do tempo de montagem do novo processo.

Adotou-se um nível de confiança de 95%, então $1 - \alpha = 0,95$ $\alpha = 0,05$ $\alpha/2 = 0,025$.

Estatísticas: **média amostral** = 3,005 segundos **s** = 0,5083 segundos **n** = 20

Definição da variável de teste: uma vez que a variância populacional da variável é DESCONHECIDA (o valor fornecido é o desvio padrão AMOSTRAL), e a amostra retirada apresenta 20 elementos (portanto menos de 30) a distribuição amostral da média será **t** de Student, e a variável de teste será **t_{n-1}**.

Encontrar o valor de **t_{n-1,crítico}**: como o Intervalo de Confiança para a média é bilateral, teremos uma situação semelhante à da figura abaixo:



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 20 - 1 = 19 graus de liberdade. O valor da probabilidade pode ser visto na figura ao lado: $P(t > t_{n-1,crítico}) = 0,025$ e $P(t > t_{n-1,crítico}) = 0,975$ (os valores são iguais em módulo).

E o valor de **t_{n-1,crítico}** será igual a **2,093** (em módulo)

Passa-se agora a determinação dos limites do intervalo:

$$L_I = \bar{x} - \frac{t_{n-1,crítico} \times s}{\sqrt{n}} = 3,005 - \frac{2,093 \times 0,5083}{\sqrt{20}} = 3,005 - 0,238 = 2,767 \text{ segundos}$$

$$L_S = \bar{x} + \frac{t_{n-1,crítico} \times s}{\sqrt{n}} = 3,005 + \frac{2,093 \times 0,5083}{\sqrt{20}} = 3,005 + 0,238 = 3,243 \text{ segundos}$$

Então o intervalo de 95% de confiança para a média populacional do tempo de montagem pelo novo processo é [2,767;3,243] segundos. Interpretação: há 95% de probabilidade de que a verdadeira média populacional do tempo de montagem pelo novo processo esteja entre 2,767 e 3,243 segundos.

b) Como a variância populacional é DESCONHECIDA, e o tamanho da amostra é menor do que 30 elementos a distribuição amostral da média será **t** de Student, e a variável de teste será **t_{n-1}**. Assim será usada a seguinte expressão para calcular o tamanho mínimo de amostra para a estimação por intervalo da média populacional.

$$n = \left(\frac{t_{n-1,crítico} \times s}{e_0} \right)^2$$

O nível de significância é o mesmo do item a. Sendo assim, o valor crítico continuará sendo o mesmo: **t_{n-1,crítico}** = 2,093. O desvio padrão amostral vale 0,5083 segundos, e o valor de **e₀**, a precisão, foi fixado em 0,5 segundos. Basta então substituir os valores na expressão:

$$n = \left(\frac{t_{n-1,crítico} \times s}{e_0} \right)^2 = \left(\frac{2,093 \times 0,5083}{0,5} \right)^2 = 4,53 \cong 5 \text{ elementos}$$

Observe que o tamanho mínimo de amostra necessário para atender a 95% de confiança e precisão de 0,5 segundos deveria ser de 5 elementos. Como a amostra coletada possui 20 elementos ela é plenamente SUFICIENTE para a significância e precisão exigidas.

c) Observe que é preciso tomar uma decisão: com base nos dados da amostra deve-se mudar para o novo processo? Trata-se então de um **teste de hipóteses**. A amostra foi coletada para avaliar se o

tempo médio de montagem do novo processo é de 3,5 segundos: se o tempo for igual ou maior não há razão para mudar, mas se for **menor**, a mudança será interessante pois haverá um ganho de produtividade. Então faremos um teste **unilateral à esquerda**.

Enunciar as hipóteses. Conforme visto acima o teste mais adequado para este caso é um Teste Unilateral à esquerda:

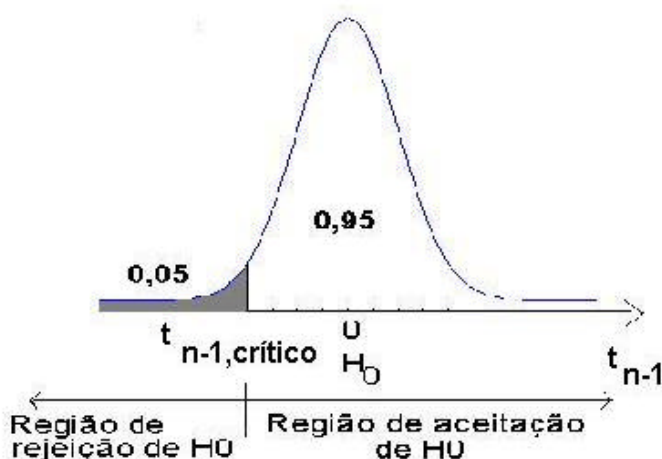
$$H_0 : m = 3,5 \quad \text{onde } m_0 = 3,5 \text{ segundos (valor de teste)}$$

$$H_1 : m < 3,5$$

Nível de significância. O problema diz que é necessário usar 5%, então $\alpha = 0,05$ e $1 - \alpha = 0,95$

Variável de teste. Uma vez que a variância populacional da variável é DESCONHECIDA (o valor fornecido é a variância AMOSTRAL), e a amostra retirada apresenta apenas 20 elementos (portanto menos de 30) a variável de teste será t_{n-1} da distribuição t de Student.

Definir a região de aceitação de H_0 .



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 .

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 $20 - 1 = 19$ graus de liberdade. O valor da probabilidade pode ser visto na figura ao lado: $P(t > t_{n-1, crítico}) = 0,95$. 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 igual a $-1,729$.

Através dos valores da amostra avaliar o valor da variável. Neste ponto é preciso encontrar o valor da variável de teste:

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

O valor de teste m_0 é igual a 3,5, a média amostral \bar{X} vale 3,005, o tamanho de amostra n é igual a 20 e o desvio padrão amostral s é 0,5083. Substituindo na equação acima:

$$t_{n-1} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{3,005 - 3,5}{0,5083 / \sqrt{20}} = -4,36$$

Decidir pela aceitação ou rejeição de H_0 . Trata-se de um teste Unilateral à Esquerda, deve-se calcular a probabilidade de t_{n-1} seja menor do que $-4,36$, ou pela simetria da distribuição t de Student, a probabilidade de t_{n-1} seja maior do que $4,36$.

Calculando o valor-p: pela tabela da distribuição t , na linha de 19 graus de liberdade, observa-se que não há valores de t superiores a 3,883: $P(Z > 3,883) = 0,0005$, portanto $P(Z < -3,883) = 0,0005$. Como o t encontrado vale 4,36, $P(t > 4,36)$ deve ser ainda menor (na verdade é 0,00016), portanto $P(t < -4,36)$ também deve ser menor do que 0,0005. Tal valor-p é muito menor do que o nível de significância (5%, 0,05). Então a decisão deve ser:

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

Há provas estatísticas suficientes de que o tempo médio de montagem dos conectores pelo novo processo é menor do que o atual processo. A empresa deve mudar para o novo processo pois terá ganhos de produtividade.

3) A variável sob análise (tempo de atraso) é QUANTITATIVA. Portanto serão feitas inferências sobre a MÉDIA.

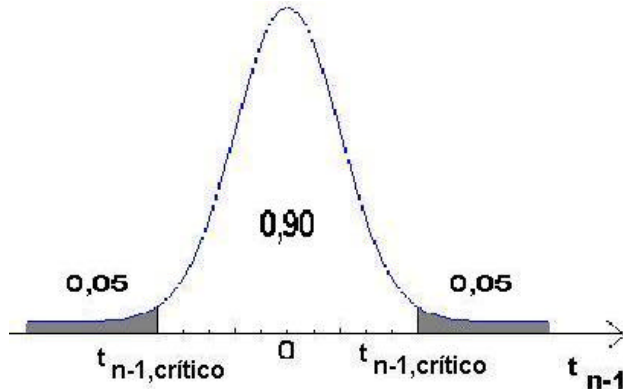
a) O parâmetro de interesse é a média populacional μ do tempo de atraso nas entregas

Adotou-se um nível de confiança de 90%, então $1 - \alpha = 0,90$ $\alpha = 0,10$ $\alpha/2 = 0,05$.

Estatísticas disponíveis são: **média amostral** = 3,3 dias $s = 3,0105$ dias $n = 20$

Definição da variável de teste: uma vez que a variância populacional da variável é DESCONHECIDA (o valor fornecido é o desvio padrão AMOSTRAL), e a amostra retirada apresenta 20 elementos (portanto menos de 30) a distribuição amostral da média será t de Student, e a variável de teste será t_{n-1} .

Encontrar o valor de $t_{n-1,crítico}$: como o Intervalo de Confiança para a média é bilateral, teremos uma situação semelhante à da figura abaixo:



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 $20 - 1 = 19$ graus de liberdade. O valor da probabilidade pode ser visto na figura ao lado: $P(t > t_{n-1,crítico}) = 0,05$ e

$P(t > t_{n-1,crítico}) = 0,95$ (os valores são iguais em módulo).

E o valor de $t_{n-1,crítico}$ será igual a 1,729 (em módulo)

Passa-se agora a determinação dos limites do intervalo:

$$L_I = \bar{x} - \frac{t_{n-1,crítico} \times s}{\sqrt{n}} = 3,3 - \frac{1,729 \times 3,0105}{\sqrt{20}} = 3,3 - 1,164 = 2,136 \text{ dias}$$

$$L_S = \bar{x} + \frac{t_{n-1,crítico} \times s}{\sqrt{n}} = 3,3 + \frac{1,729 \times 3,0105}{\sqrt{20}} = 3,3 + 1,164 = 4,464 \text{ dias}$$

Então o intervalo de 90% de confiança para a média populacional do tempo de atraso é $[2,136; 4,464]$ dias.

Interpretação: há 90% de probabilidade de que a verdadeira média populacional do tempo de atraso na entrega dos pedidos esteja entre 2,136 e 4,464 dias.

b) Neste caso a variância populacional é conhecida (foi expressamente declarado que o desvio padrão populacional, σ , vale 2 dias). Independente do tamanho da amostra é possível utilizar a variável Z , da distribuição normal padrão. Para obter o valor crítico basta obter o valor de Z tal que: $P(Z > Z_{crítico}) = 0,05$. Procurando na tabela da distribuição normal padrão encontra-se $Z_{crítico} = 1,645$.

Passa-se agora a determinação dos limites do intervalo, através da expressão abaixo (cujo resultado será somado e subtraído da média amostral) para determinar os limites do intervalo:

$$e_0 = \frac{Z_{crítico} \times \sigma}{\sqrt{n}} = \frac{1,645 \times 2}{\sqrt{20}} = 0,736 \text{ dias}$$

$$L_I = \bar{x} - e_0 = 3,3 - 0,736 = 2,564 \text{ dias}$$

$$L_S = \bar{x} + e_0 = 3,3 + 0,736 = 4,036 \text{ dias}$$

Então o intervalo de 90% de confiança para a média populacional do tempo de atraso é $[2,564; 4,036]$ dias.

Interpretação: há 90% de probabilidade de que a verdadeira média populacional do tempo de atraso na entrega dos pedidos esteja entre 2,564 e 4,036 dias.

c) Para a mesma situação do item a.

Como a variância populacional é **DESCONHECIDA**, e o tamanho da amostra é menor do que 30 elementos a distribuição amostral da média será **t** de Student, e a variável de teste será **t_{n-1}** . Assim será usada a seguinte expressão para calcular o tamanho mínimo de amostra para a estimação por

intervalo da média populacional.
$$n = \left(\frac{t_{n-1, \text{crítico}} \times s}{e_0} \right)^2$$

O nível de significância é o mesmo do item a. Sendo assim, o valor crítico continuará sendo o mesmo: **$t_{n-1, \text{crítico}} = 1,729$** . O desvio padrão amostral vale 3,0105 dias, e o valor de **e_0** , a precisão, foi fixado em 0,5 dias. Basta então substituir os valores na expressão:

$$n = \left(\frac{t_{n-1, \text{crítico}} \times s}{e_0} \right)^2 = \left(\frac{1,729 \times 3,0105}{0,5} \right)^2 = 108,37 \cong 109 \text{ elementos}$$

Observe que o tamanho mínimo de amostra necessário para atender a 90% de confiança e precisão de 0,5 dias deveria ser de 109 elementos. Como a amostra coletada possui 20 elementos ela é **INSUFICIENTE** para a significância e precisão exigidas, seria necessário obter mais 89 medidas.

d) Observe que é preciso tomar uma decisão: com base nos dados da amostra deve-se confiar na declaração da empresa sobre o tempo de atraso? Trata-se então de um **teste de hipóteses**. A amostra foi coletada para avaliar se o tempo médio de atraso na entrega dos pedidos é maior do que 1 dia: se o tempo for igual ou menor não há razão para o cliente reclamar, mas se for **maior**, a reclamação tem fundamento. Então faremos um teste **unilateral à direita**.

Enunciar as hipóteses. Conforme visto acima o teste mais adequado para este caso é um Teste Unilateral à direita:

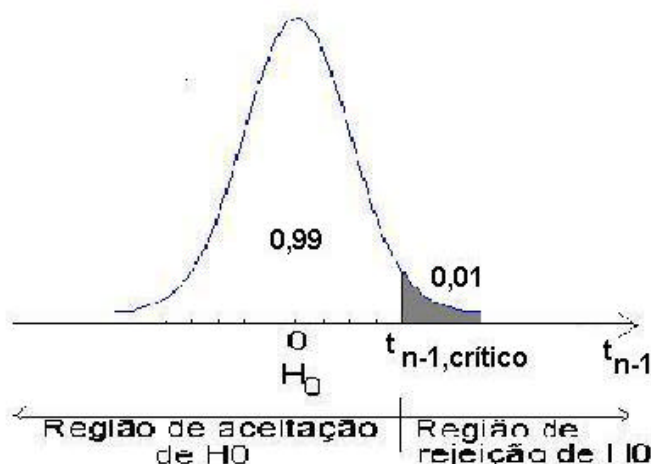
$$H_0 : m = 1 \quad \text{onde } m_0 = 1 \text{ dia (valor de teste)}$$

$$H_1 : m > 1$$

Nível de significância. O problema declara que é necessário usar uma confiança de 99%, então **$1 - \alpha = 0,99$** e **$\alpha = 0,01$**

Variável de teste. Uma vez que a variância populacional da variável é **DESCONHECIDA** (o valor fornecido é a variância **AMOSTRAL**), e a amostra retirada apresenta apenas 20 elementos (portanto menos de 30) a variável de teste será **t_{n-1}** da distribuição **t** de Student.

Definir a região de aceitação de **H_0** .



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** . 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 $20 - 1 = 19$ graus de liberdade. O valor da probabilidade pode ser visto na figura ao lado: $P(t > t_{n-1, \text{crítico}}) = 0,01$. Então **$t_{n-1, \text{crítico}}$** será igual a 2,539.

Através dos valores da amostra avaliar o valor da variável. Neste ponto é preciso encontrar o valor

da variável de teste:

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

O valor de teste m_0 é igual a 1, a média amostral \bar{X} vale 3,3, o tamanho de amostra n é igual a 20 e o desvio padrão amostral s é 3,0105. Substituindo na equação acima:

$$t_{n-1} = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{3,3 - 1}{3,0105 / \sqrt{20}} = 3,42$$

Decidir pela aceitação ou rejeição de H_0 . Trata-se de um teste Unilateral à Direita: deve-se calcular a probabilidade de t_{n-1} seja maior do que 3,42.

Calculando o valor-p: pela tabela da distribuição t, na linha de 19 graus de liberdade, observa-se que o valor 3,42 está entre 3,174 e 3,579, que correspondem às probabilidades 0,0025 e 0,001 respectivamente: ambas menores do que o nível de significância. Na realidade, $P(t > 3,42) = 0,0014$. Tal valor-p é muito menor do que o nível de significância (1%, 0,01). Então a decisão deve ser:

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

Há provas estatísticas suficientes de que o tempo médio de atraso na entrega dos pedidos é maior do que 1 dia. O cliente tem razão na sua reclamação.

4) A variável sob análise (opinião sobre a administração estadual) é QUALITATIVA, e somente admite dois resultados: satisfeita ou insatisfeita. Portanto serão feitas inferências sobre a proporção de pessoas insatisfeitas ou satisfeitas.

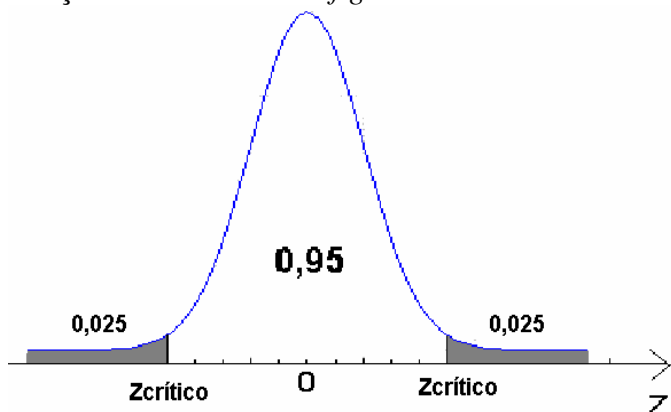
a) O parâmetro de interesse é a proporção populacional p de pessoas insatisfeitas com a administração estadual.

Adotou-se um nível de significância de 5%, então $\alpha = 0,05$ $\alpha/2 = 0,025$ $1 - \alpha = 0,95$

As estatísticas são: **proporção amostral de pessoas insatisfeitas** $p = 585/1000 = 0,585$, o seu complementar $1 - p = 0,415$ e $n = 1000$ elementos.

Definição da variável de teste: precisamos verificar se é possível fazer a aproximação pela normal, então $n \times p = 1000 \times 0,585 = 585 > 5$ e $n \times (1 - p) = 1000 \times 0,415 = 415 > 5$. Como ambos os produtos satisfazem as condições para a aproximação podemos usar a variável Z da distribuição normal padrão

Encontrar o valor de $Z_{\text{crítico}}$: como o Intervalo de Confiança para a média é bilateral, teremos uma situação semelhante à da figura abaixo:



Para encontrar o valor crítico devemos procurar na tabela da distribuição normal padrão pela probabilidade 0,025 e 0,975 ($0,95 + 0,025$). O valor da probabilidade pode ser visto na figura ao lado: os valores críticos serão $Z_{0,025}$ e $Z_{0,975}$ os quais serão iguais em módulo. $P(Z > Z_{\text{crítico}}) = 0,025$. Então $Z_{\text{crítico}}$ será igual a 1,96 (em módulo).

Passa-se agora a determinação dos limites do intervalo, através da expressão abaixo (cujo resultado será somado e subtraído da proporção amostral de pessoas insatisfeitas) para determinar os limites do intervalo:

$$e_0 = Z_{\text{crítico}} \times \sqrt{\frac{p \times (1 - p)}{n}} = 1,96 \times \sqrt{\frac{0,585 \times 0,415}{1000}} = 0,0305$$

$$L_i = p - e_0 = 0,585 - 0,0305 = 0,5545 \quad L_s = p + e_0 = 0,585 + 0,0305 = 0,6155$$

Então, o intervalo de 95% de confiança para a proporção populacional de pessoas insatisfeitas com a administração estadual é [55,45%; 61,55%].

Interpretação: há 95% de probabilidade de que a verdadeira proporção populacional de pessoas insatisfeitas esteja entre 55,45% e 61,55%.

b) De acordo com o item anterior é possível utilizar a aproximação pela distribuição normal. Assim, a expressão para o cálculo do tamanho mínimo de amostra para a proporção populacional será:

$$n = \left(\frac{Z_{\text{crítico}}}{e_0} \right)^2 \times p \times (1 - p)$$

Os valores de p e $1 - p$ já são conhecidos: $p = 0,585$ $1 - p = 0,415$

O nível de confiança exigido é de 95%: para encontrar o valor crítico devemos procurar na tabela da distribuição normal padrão pela probabilidade 0,025 e 0,975 (0,95+0,025); os valores críticos serão $Z_{0,025}$ e $Z_{0,975}$ os quais serão iguais em módulo. $P(Z > Z_{\text{crítico}}) = 0,025$. Então $Z_{\text{crítico}}$ será igual a 1,96 (em módulo). A precisão foi fixada em 2,5% (0,025). Substituindo os valores na expressão acima:

$$n = \left(\frac{Z_{\text{crítico}}}{e_0} \right)^2 \times p \times (1 - p) = \left(\frac{1,96}{0,025} \right)^2 \times 0,585 \times 0,415 = 1492,23 \approx 1493$$

Observe que o tamanho mínimo de amostra necessário para atender a 95% de confiança e precisão de 2,5% deveria ser de 1493 elementos. Como a amostra coletada possui apenas 1000 elementos ela é INSUFICIENTE para a confiança e precisão exigidas. Recomenda-se o retorno à população para a retirada aleatória de mais 493 pessoas.

c) Observe que é preciso tomar uma decisão: com base nos dados da amostra deve ser redirecionado o plano governamental? Trata-se então de um **teste de hipóteses**. A amostra foi coletada para avaliar se a proporção de insatisfeitos com a administração estadual é igual a 50% (0,5): se a proporção for igual ou menor não haveria razão para redirecionar o plano, mas se for **maior** significa que a maioria da população está insatisfeita, e algo precisa ser feito. Então faremos um teste **unilateral à direita**.

Enunciar as hipóteses. Conforme visto acima o teste mais adequado para este caso é um Teste Unilateral à Direita:

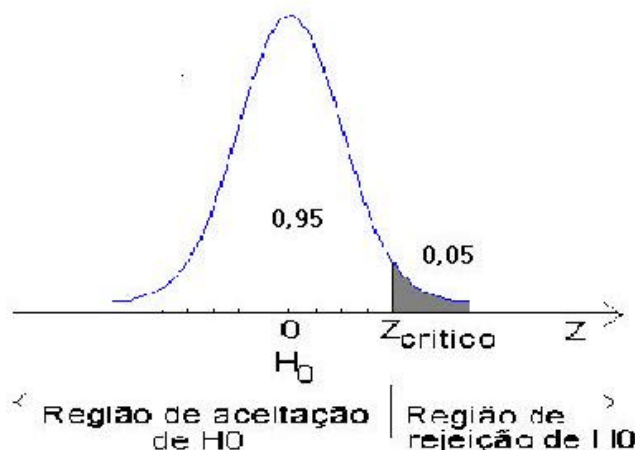
$H_0 : p = 0,5$ (50%) onde $p_0 = 0,5$ (valor de teste)

$H_1 : p > 0,5$ (50%)

Nível de significância. É necessário usar uma significância de 5%, então $\alpha = 0,05$ e $1 - \alpha = 0,95$.

Variável de teste. Como se trata de um teste de proporção é necessário verificar o valor dos produtos: $n \times p_0 = 1000 \times 0,5 = 500$ e $n \times (1 - p_0) = 1000 \times 0,5 = 500$. Como ambos são maiores do que 5 é possível fazer uma aproximação pela normal, e a variável de teste será Z .

Definir a região de aceitação de H_0 .



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 . Para encontrar o valor crítico devemos procurar na tabela da distribuição normal, pela probabilidade acumulada 0,05 (o $Z_{\text{crítico}}$ aqui é maior do que zero). $P(Z > Z_{\text{crítico}}) = 0,05$. Então $Z_{\text{crítico}}$ será igual a 1,645.

Através dos valores da amostra avaliar o valor da variável. Neste ponto é preciso encontrar o valor da variável de teste:

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 \times (1 - \pi_0)}{n}}}$$

O valor de teste p_0 é igual a 0,5 (50%), a proporção amostral p vale 0,585, e o tamanho de amostra n é igual a 1000. Substituindo na equação acima:

$$Z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 \times (1 - \pi_0)}{n}}} = \frac{0,585 - 0,5}{\sqrt{\frac{0,5 \times 0,5}{1000}}} = 5,375$$

Decidir pela aceitação ou rejeição de H_0 . Trata-se de um teste Unilateral à Direita: deve-se calcular a probabilidade de Z seja maior do que 5,375.

Calculando o valor-p Calculando o valor-p: pela tabela da distribuição observa-se que não há valores de Z superiores a 5: $P(Z > 5) = 0,000000287$. Como o Z encontrado vale 5,375, $P(Z > 5,375)$ deve ser ainda menor (na verdade é $4,407 \times 10^{-8}$), portanto muito menor do que o nível de significância. Então a decisão deve ser:

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

Há provas estatísticas suficientes para recomendar o redirecionamento do plano governamental, mais de 50% das pessoas estão insatisfeitas com a administração estadual.

15) A variável sob análise (consumo de gasolina em km/l) é QUANTITATIVA, portanto serão feitas inferências sobre as médias antes e depois da instrução.

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 curso. Há interesse em verificar se o curso contribuiu para o aumento da média de km/l (diminuição do consumo): ou seja, queremos verificar se a média de km/l antes do curso é MENOR do que a média de km/l após o tratamento (se o curso auxiliou os motoristas os valores de km/l estarão em média maiores após a instrução). 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 MÉDIAS POPULACIONAIS, PARA DADOS PAREADOS** (MESMA POPULAÇÃO: ANTES E DEPOIS). Supõe-se que se ambas as distribuições populacionais são normais a distribuição da diferença entre os valores também será. 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 hipóteses H_0 , é que o curso NÃO FAZ EFEITO, ou seja as médias antes e após o curso 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:

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

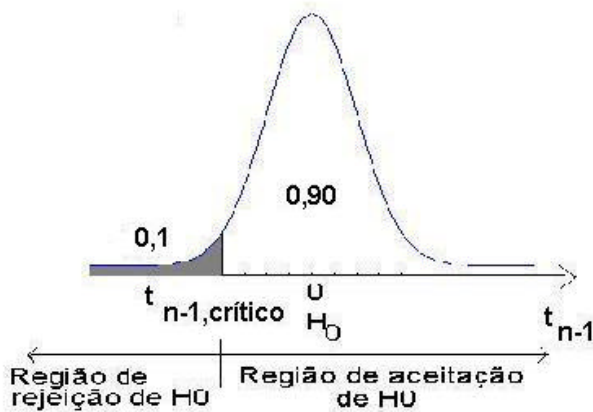
Estabelecer o nível de significância ou nível de confiança. Conforme foi estabelecido no enunciado do problema: $1 - \alpha = 0,9$ $\alpha = 0,1$

Identificar a variável de teste. A variável de teste que será utilizada será a variável t_{n-1} da distribuição t de Student.

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 10% de significância), e a variável de teste é t_{n-1} (a amostra tem 10 elementos), então o valor crítico (obtido da tabela da distribuição t de Student) será:

$$t_{n-1, \text{critico}} = t_{10-1; 0,9} = t_{9; 0,9} = -t_{9; 0,1} = -1,383$$

Observe a região de aceitação de H_0 :



Para valores maiores do que $-1,383$ aceitaremos H_0 (ou seja o curso não faz efeito, a diferença entre as médias é nula). Se t_{n-1} for menor do que $-1,383$ rejeitaremos H_0 (a média DEPOIS aumentou demais em relação à média ANTES do curso para que a diferença seja devida apenas ao acaso). Claro que há uma chance de 10% de que venhamos a rejeitar H_0 sendo ela verdadeira.

Através dos valores das amostras antes e depois, calcular a diferença d_i entre cada par de valores, onde

$$d_i = X_{\text{antes}} - X_{\text{depois}}$$

Para o conjunto sob análise teremos:

| Motorista | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------|-----|------|------|------|------|------|------|------|------|------|
| Antes | 7,6 | 7,9 | 6,5 | 7,5 | 8,9 | 7,5 | 8,2 | 7,8 | 6,7 | 8,0 |
| Depois | 7,6 | 8,2 | 7,2 | 7,2 | 8,5 | 7,3 | 7,8 | 7,9 | 6,4 | 7,3 |
| d_i | 0 | -0,3 | -0,7 | 0,3 | 0,4 | 0,2 | 0,4 | -0,1 | 0,3 | 0,7 |
| d_i^2 | 0 | 0,09 | 0,49 | 0,09 | 0,16 | 0,04 | 0,16 | 0,01 | 0,09 | 0,49 |

Calcular a diferença média e o desvio padrão da diferença média.

Para o presente problema: $\bar{d} = \frac{\sum d_i}{n} = \frac{1,2}{10} = 0,12 \text{ km/l}$

$$s_d = \sqrt{\frac{\sum d_i^2 - [(\sum d_i)^2 / n]}{n - 1}} = \sqrt{\frac{1,62 - [(1,2)^2 / 10]}{10 - 1}} = 0,405 \text{ km/l}$$

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

$$t_{n-1} = \frac{\bar{d} \times \sqrt{n}}{s_d} \quad t_{10-1} = t_9 = \frac{0,12 \times \sqrt{10}}{0,405} = 0,937$$

Decidir pela aceitação ou rejeição de H_0 . Trata-se de um teste Unilateral à Esquerda: deve-se calcular a probabilidade de t_{n-1} seja menor do que 0,937, ou pela simetria da distribuição t de Student, o complementar da probabilidade de t_{n-1} seja maior do que 0,937.

Calculando o valor-p: pela tabela da distribuição t , na linha de 9 graus de liberdade, observa-se que o valor 0,937 está entre 0,703 e 1,383, que correspondem às probabilidades 0,25 e 0,10 respectivamente. Calculando as probabilidades complementares obteremos 0,75 e 0,90 respectivamente (na realidade, $P(t < 0,937) = 1 - P(t > 0,937) = 0,8134$), ambas maiores do que o nível de significância. Tal valor-p é muito maior do que o nível de significância (10%, 0,10). Então a decisão deve ser:

ACEITAR H_0 a 10% de Significância (há 10% de chance de erro)

Interpretar a decisão dentro do contexto do problema. Assim, concluímos com 90% de confiança (ou uma chance de erro de 10%) que não há evidência estatística suficiente que o curso aumenta a média de km/l (diminui o consumo).

18) A variável de interesse, nota dos alunos, é QUANTITATIVA. Então serão feitas inferências sobre as MÉDIAS nos cursos de economia e administração.

a) Os parâmetros de interesse são as médias populacionais μ das notas dos cursos de economia e administração.

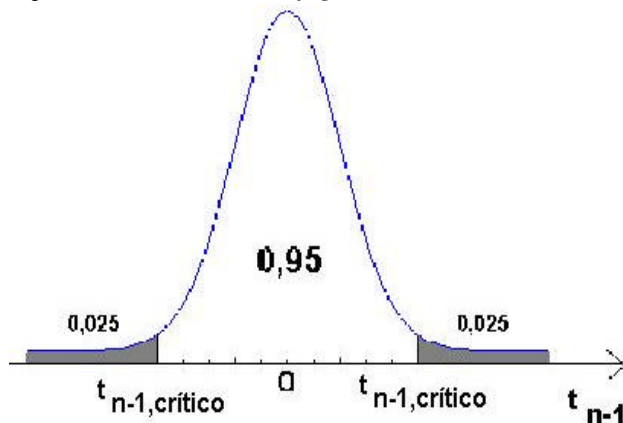
Adotou-se um nível de confiança de 95%, então $1 - \alpha = 0,95$ $\alpha = 0,05$ $\alpha/2 = 0,025$.

Estatísticas: **média amostral de economia** = 7,3 $s = 2,6$ $n = 10$

média amostral de administração = 7,1 $s = 3,1$ $n = 10$

Definição da variável de teste: uma vez que a variância populacional da variável é DESCONHECIDA (o valor fornecido é o desvio padrão AMOSTRAL), e as amostras retiradas apresentam 10 elementos em cada grupo a distribuição amostral da média será t de Student, e a variável de teste será t_{n-1} (tanto para o curso de economia quanto para o de administração)

Encontrar o valor de $t_{n-1,crítico}$: como o Intervalo de Confiança para a média é bilateral, teremos uma situação semelhante à da figura abaixo:



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 $10 - 1 = 9$ graus de liberdade. O valor da probabilidade pode ser visto na figura ao lado: $P(t > t_{n-1,crítico}) = 0,025$ e

$P(t > t_{n-1,crítico}) = 0,975$ (os valores são iguais em módulo).

E o valor de $t_{n-1,crítico}$ será igual a 2,262 (em módulo)

Podemos usar o valor obtido para os intervalos de confiança das médias dos dois cursos.

Passa-se agora a determinação dos limites dos intervalos, através da expressão abaixo (cujo resultado será somado e subtraído das médias amostrais) para determinar os limites dos intervalos:

$$\text{Economia: } e_0 = \frac{t_{n-1,crítico} \times s}{\sqrt{n}} = \frac{2,262 \times 2,6}{\sqrt{10}} = 1,86 \quad L_I = \bar{x} - e_0 = 7,3 - 1,86 = 5,44$$

$$L_S = \bar{x} + e_0 = 7,3 + 1,86 = 9,16$$

$$\text{Administração: } e_0 = \frac{t_{n-1,crítico} \times s}{\sqrt{n}} = \frac{2,262 \times 3,1}{\sqrt{10}} = 2,22 \quad L_I = \bar{x} - e_0 = 7,1 - 2,22 = 4,88$$

$$L_S = \bar{x} + e_0 = 7,1 + 2,22 = 9,32$$

Economia: o intervalo de 95% de confiança para a média populacional das notas em economia é [5,44;9,16]. Interpretação: há 95% de probabilidade de que a verdadeira média populacional das notas em economia esteja entre 5,44 e 9,16.

Administração: o intervalo de 95% de confiança para a média populacional das notas em administração é [4,88;9,32]. Interpretação: há 95% de probabilidade de que a verdadeira média populacional das notas em economia esteja entre 4,88 e 9,32.

b) Como a variância populacional é DESCONHECIDA, e o tamanho da amostra é menor do que 30 elementos, nos dois cursos, a distribuição amostral da média será t de Student, e a variável de teste será t_{n-1} . Assim será usada a seguinte expressão para calcular o tamanho mínimo de amostra para a estimação por intervalo da média populacional.

$$n = \left(\frac{t_{n-1,crítico} \times s}{e_0} \right)^2$$

O nível de significância é o mesmo do item a. Sendo assim, o valor crítico continuará sendo o mesmo: $t_{n-1,crítico} = 2,262$. Vamos calcular o tamanho mínimo de amostra para os dois cursos. O desvio padrão amostral das notas em economia vale 2,6, e em administração vale 3,1, e o valor de

e_0 , a precisão, foi fixado em 1 para ambos os cursos. Basta então substituir os valores nas expressões.

Economia:

$$n = \left(\frac{t_{n-1, \text{crítico}} \times s}{e_0} \right)^2 = \left(\frac{2,262 \times 2,6}{1} \right)^2 = 34,59 \cong 35 \text{ elementos}$$

Observe que o tamanho mínimo de amostra necessário para atender a 95% de confiança e precisão de 1 deveria ser de 35 elementos. Como a amostra coletada possui 10 elementos ela é **INSUFICIENTE** para a significância e precisão exigidas, para o curso de Economia.

Administração:

$$n = \left(\frac{t_{n-1, \text{crítico}} \times s}{e_0} \right)^2 = \left(\frac{2,262 \times 3,1}{1} \right)^2 = 49,17 \cong 50 \text{ elementos}$$

Observe que o tamanho mínimo de amostra necessário para atender a 95% de confiança e precisão de 1 deveria ser de 50 elementos. Como a amostra coletada possui 10 elementos ela é **INSUFICIENTE** para a significância e precisão exigidas, para o curso de Administração.

c) Queremos avaliar se há diferenças entre as médias obtidas pelos alunos de dois cursos distintos. Trata-se de uma situação em que queremos comparar as MÉDIAS DE DUAS distribuições normais, supondo que se tratam de duas populações distintas, podemos considerá-las independentes: está sendo avaliada a média obtida por grupos DIFERENTES de alunos. Não há indícios se este ou aquele curso apresenta maior média: deseja-se saber simplesmente se há diferenças significativas entre as médias. 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 MÉDIAS POPULACIONAIS** (teste t), **PARA DADOS NÃO PAREADOS (POPULAÇÕES DISTINTAS)**.

Um fator crucial para a aplicação do teste é observar o conhecimento que temos acerca das variâncias populacionais das notas nos dois cursos. As informações disponíveis são **AMOSTRAIS**: as variâncias populacionais são **DESCONHECIDAS**. Neste caso utilizaremos a variável t de Student com $2 \times (n-1)$ graus de liberdade, uma vez que as amostras têm o mesmo tamanho.

Enunciar as hipóteses. De acordo com o que foi dito acima queremos verificar se há diferenças (não importa qual dos cursos tem a maior média) entre as médias dos dois cursos; 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 DOS ALUNOS DO CURSO DE ECONOMIA É IGUAL À MÉDIA DOS ALUNOS DO CURSO DE ADMINISTRAÇÃO** (costumamos colocar em H_0 o **CONTRÁRIO** do que queremos provar), teremos então:

$$\begin{aligned} H_0 : \mu_1 &= \mu_2 \\ H_1 : \mu_1 &\neq \mu_2 \end{aligned} \quad \text{onde } \mu_1 = \mu_{\text{Economia}} \text{ e } \mu_2 = \mu_{\text{Administração}}$$

Estabelecer o nível de significância ou nível de confiança. Podemos escolher:

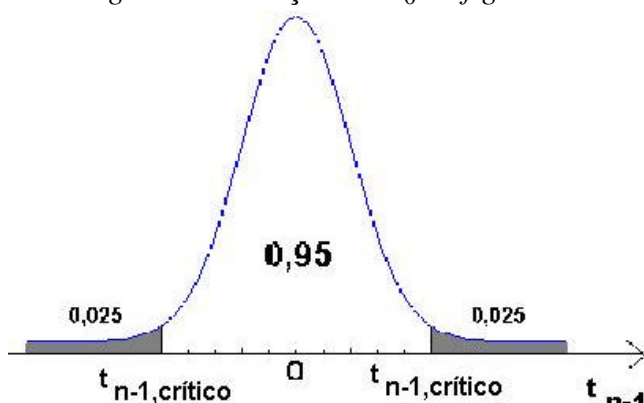
$$\alpha = 0,05 \quad 1 - \alpha = 0,95 \quad \alpha/2 = 0,025$$

Identificar a variável de teste. Deverá ser usada a variável t da distribuição t de Student, com $2 \times (n-1)$ graus de liberdade, onde $n = 10$ são os tamanhos das amostras): $t_{2 \times (n-1)} = t_{2 \times (10-1)} = t_{18}$

Definir a região de aceitação de H_0 , de acordo com o tipo de teste e variável. Trata-se de um teste Bilateral (com 5% de significância), e a variável de teste é t com 18 graus de liberdade então o valor crítico (obtido da tabela da distribuição t de Student) será:

$$t_{18; 0,025} = 2,101$$

Observe a região de aceitação de H_0 na figura abaixo:



Para valores de t_{18} (em módulo) maiores do que 2,101, deve-se REJEITAR H_0 , ou seja há diferenças entre as médias dos alunos dos cursos de economia e administração (claro que há 5% de chance de que venhamos a rejeitar H_0 sendo ela verdadeira): as diferenças entre o que era esperado e o que foi encontrado na amostra serão consideradas grandes demais para serem casuais.

Calcular o desvio padrão das diferenças. Como as duas amostras têm o mesmo tamanho podemos usar a seguinte fórmula:

$$s_a^2 = \frac{s_1^2 + s_2^2}{2} = \frac{2,6^2 + 3,1^2}{2} = 8,185$$

Calcular a variável de teste.

$$t_{18} = (\bar{x}_1 - \bar{x}_2) \times \sqrt{\frac{n}{2 \times s_a^2}} = (7,3 - 7,1) \times \sqrt{\frac{10}{2 \times 8,185}} = 0,1563$$

Decidir pela aceitação ou rejeição de H_0 . Trata-se de um teste Bilateral: deve-se calcular a probabilidade de t_{18} seja maior do que 0,1563, e multiplicá-la por 2.

Calculando o valor-p: pela tabela da distribuição t, na linha de 18 graus de liberdade, observa-se que o valor 0,1563 está abaixo de 0,688 (menor valor disponível), que corresponde à probabilidade 0,25, que já é maior do que o nível de significância. Multiplicando-a por 2 temos 0,50, que é bem maior do que o nível de significância (5%), na realidade $2 \times P(t_{18} > 0,1563)$ vale 0,8776. Então a decisão deve ser: **ACEITAR H_0** a 5% de Significância (há 5% de chance de erro)

Interpretar a decisão dentro do contexto do problema. Assim, concluímos com 95% de confiança (ou uma chance de erro de 5%) que **NÃO** há evidências estatísticas suficientes para declarar que as médias das notas dos alunos dos cursos de economia e administração sejam diferentes.

25) Trata-se de um exercício de teste do Chi-Quadrado de independência: queremos saber se a variável lembrança do consumidor está associada à variável meio de comunicação (e vice-versa).

Enunciar as Hipóteses:

H_0 : as variáveis são independentes

H_1 : as variáveis não são independentes

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

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

| Lembrança | Meio de comunicação | | | |
|-------------|---------------------|-----|-------|-------|
| | Revista | TV | Rádio | Total |
| Lembram | 25 | 93 | 7 | 125 |
| Não lembram | 73 | 10 | 108 | 191 |
| Total | 98 | 103 | 115 | 316 |

Fonte: hipotética

Calcular as frequências esperadas: devemos calculá-las para todas as células da tabela (6 no presente problema). Os resultados estão na tabela abaixo:

| | Meio de comunicação | | | |
|-------------|---------------------|----------|-----------|-------|
| Lembrança | Revista | TV | Rádio | Total |
| Lembram | 38,76582 | 40,74367 | 45,490506 | |
| Não lembram | 59,23418 | 62,25633 | 69,509494 | |

Calculando a estatística χ^2 para cada célula: agora podemos calcular as diferenças entre as frequências e as demais operações. Os valores finais estão na tabela abaixo:

| (O-E) ² /E | Meio de comunicação | | | |
|-----------------------|---------------------|----------|----------|-------|
| Lembrança | Revista | TV | Rádio | Total |
| Lembram | 4,888272 | 67,02204 | 32,56765 | |
| Não lembram | 3,199131 | 43,86259 | 21,31391 | |

Agora podemos somar os valores:

$$\chi^2 = 172,8536$$

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

Então $\chi^2_2 = 172,8536$

Calculando o valor-p. Como o teste do Chi-Quadrado de independência é SEMPRE unilateral à direita, precisamos encontrar $P(\chi^2_2 > 172,8536)$. Procurando na tabela apropriada, para 2 graus de liberdade observamos que o maior valor é 13,8, que corresponde à probabilidade de 0,001, que já é menor do que o nível de significância (na realidade $P(\chi^2_2 > 172,8536) = 2,92 \times 10^{-38}$), o que nos leva a REJEITAR H_0 a 1% de significância.

HÁ evidência estatística suficiente que indica que as variáveis meio de comunicação da propaganda e lembrança do consumidor são dependentes.

b) A força da associação pode ser medida através do coeficiente de contingência modificado:

$$C^* = \sqrt{\frac{\chi^2}{\chi^2 + N}} \times \sqrt{\frac{k}{k-1}} = \sqrt{\frac{172,8536}{172,8536 + 316}} \times \sqrt{\frac{3}{3-1}} = 0,84094$$

Como C^* é próximo de 1 (seu valor máximo possível) a associação pode ser considerada forte.