

Resumo para a Ps - Estatística (PRO 3200) 2015

Por Arthur Salles

R#5, 2015

I - Conceitos Básicos

Ao estudar probabilidade, vimos que existem várias distribuições teóricas de probabilidade às quais podem ser relacionadas certas variáveis.

Vimos por exemplo que a altura da população mundial possui uma distribuição normal com média μ e variância σ^2 . A questão que nos faz estudar estatística é: μ e σ^2 são desconhecidas como a partir de alguns dados coletados (uma amostra), posso estimar os valores de μ e σ^2 , por exemplo?

Algumas medidas usadas pela estatística de grande utilidade são:

- Média amostral ($\bar{X}$): para n valores ($x_1, \dots, x_n$) de uma variável, definimos a média amostral como:

$$\bar{X} = \frac{\sum x_i}{n}$$

- Mediana: colocamos os valores de X em ordem crescente buscamos aquele para o qual
 - Metade ou mais dos valores são maiores ou iguais
 - Metade ou mais dos valores são menores ou iguais

- Moda : valor de X mais frequente na amostra
- Variancia amostral (s^2)

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

- Desvio padrão amostral : $s = \sqrt{s^2}$

- Amplitude (R)

$$R = \text{Max}(X_i) - \text{Min}(X_i)$$

II - Identificando valores atípicos

Algumas vezes pode ser útil determinar se há um valor atípico ou 'outlier' na amostra, que pode ser um erro de medida ou alguma outra falha no método de coleta (ainda que não sempre).

Para identificar um outlier definimos o IQR - Intervalo Interquartil

$$IQR = 3^\circ Q - 1^\circ Q$$

Em que $1^\circ Q$ é o primeiro quantil e é o valor para o qual um quarto dos dados estão abaixo e $3^\circ Q$ é o terceiro quantil, abaixo do qual estão $3/4$ dos dados.

Um dado será um outlier se estiver fora de $[LI; LS]$ com:

$$LI = 1^\circ Q - 1,5 IQR \quad \text{e} \quad LS = 3^\circ Q + 1,5 IQR$$

Estimativa pontual de Parâmetros

Para estimar os parâmetros de cada tipo de distribuição, usaremos seus Estimadores de Máxima Verossimilhança*:

A. Normal

$X \sim \text{Normal}(\mu, \sigma^2)$

Vamos estimar μ por $\hat{\mu}$ e σ^2 por $\hat{\sigma}^2$:

$$\boxed{\hat{\mu} = \bar{X}}$$

$$\hat{\sigma}^2 = \frac{(n-1)}{n} s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$

⚠ não confundir: se enunciado pedir variância amostral é: $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$

B. Binomial / Bernoulli

$$\boxed{\hat{p} = \bar{X}}$$

C. Poisson

$$\boxed{\hat{\lambda} = \bar{X}}$$

* Ver mais no ANEXO da Fim

- Intervalos de Confiança (IC)

Informar apenas a média amostral ($\bar{X}$) de uma amostra pode não ser uma informação satisfatoriamente útil. Por vezes, queremos saber também a precisão dessa estimativa. Para tanto, recorremos aos intervalos de confiança.

Você provavelmente já ouviu que certo político tem $X\%$ das intenções de voto, com margem de erro de $Y\%$. O que isso quer dizer? Que a intenção de voto, com uma probabilidade bem alta, está entre $(X-Y)\%$ e $(X+Y)\%$.

Isso é um intervalo de confiança:

Para determinado nível de confiança (γ), por exemplo 95%, determinamos um intervalo que tem uma probabilidade γ de conter o valor real do que se estima.

Vejam os intervalos de confiança mais comuns

1. IC para média, com σ^2 conhecida

Vamos determinar um IC para a média (μ), a partir da média amostral ($\bar{X}$), com um nível de confiança γ .

$$\boxed{\bar{X} - z_{\frac{\gamma}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\gamma}{2}} \frac{\sigma}{\sqrt{n}}}$$

Onde n = tamanho da amostra

Obs: se $\gamma = 90\%$, $z_{\frac{\gamma}{2}} = z_{45\%}$ será o valor de z da normal padrão para o qual $P(0 \leq z \leq z_{45\%}) = 45\%$.

2. Intervalo de Confiança para σ^2

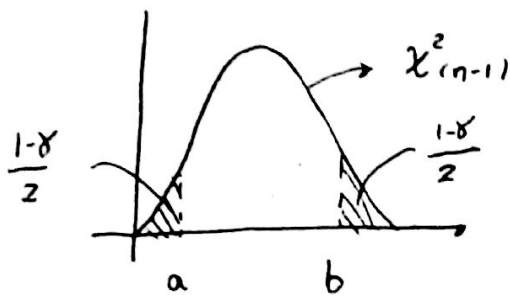
Supondo que X tenha uma distribuição normal, vamos querer determinar um IC para σ^2 a partir da variância amostral (s^2). Lembrando que

$$s^2 = \hat{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

O IC para σ^2 será:

$$\boxed{\frac{s^2(n-1)}{b} < \sigma^2 < \frac{s^2(n-1)}{a}}$$

Onde a e b são determinados usando a distribuição Qui-Quadrado com $(n-1)$ graus de liberdade



Ou seja :

$$\boxed{a = \chi^2\left(\frac{1-\gamma}{2}, n-1\right)}$$

$\uparrow$ $\uparrow$
 probab. - graus de
 liberd. liberd.

$$\boxed{b = \chi^2\left(\frac{1+\gamma}{2}, n-1\right)}$$

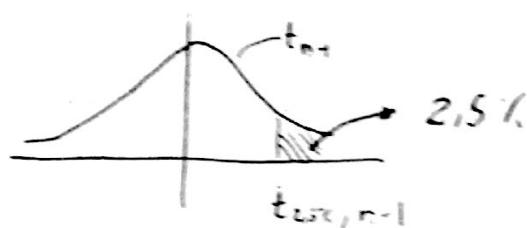
3. IC para μ com σ^2 desconhecido

Se não conhecemos σ^2 , vamos nos virar com S^2 :

$$\bar{x} - t_{(\frac{1-\gamma}{2}, n-1)} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{(\frac{1-\gamma}{2}, n-1)} \frac{s}{\sqrt{n}}$$

Em que $t_{(\frac{1-\gamma}{2}, n-1)}$ é o valor obtido na tabela da distribuição t-student

Por exemplo se $\gamma = 95\%$: $t_{2,5\%, n-1}$



4. IC para proporção da população p

Seja X uma v.a. com distribuição de Bernoulli, p é estimado por:

$$\hat{p} = \frac{\sum x_i}{n}$$

Na teoria, o intervalo de confiança seria

$$p = \hat{p} \pm z_{\frac{\gamma}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

O problema é que não conhecemos p , então não podemos calcular o que está dentro da raiz

Tomamos duas opções:

1ª Solução : estimar p por $\hat{p}$

$$P = \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

2ª Solução : usar o pior caso possível : $p(1-p)$ é máxima em $[0;1]$ para $p = 0,5$

$$P = \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{0,5 \times 0,5}{n}}$$

Observação : p está sempre entre 0 e 1. Assim se seu limite inferior for negativo, use 0 na resposta; se o limite superior for > 1 , use 1 na resposta.

5. IC para λ da distribuição de Poisson

Não é muito cobrado, mas vale a pena mencionar:

$$\hat{\lambda} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}}{n}} < \lambda < \hat{\lambda} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}}{n}}$$

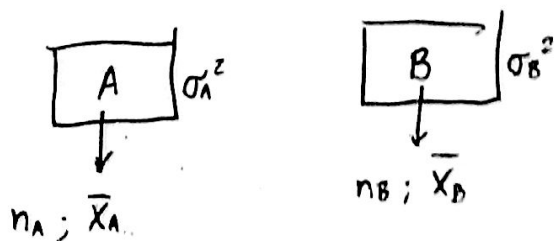
Até agora vimos intervalos de confiança baseados em uma única amostra, por exemplo: IC para a altura da população brasileira.

Muitas vezes, no entanto, queremos comparar duas populações. Por exemplo: qual a diferença de altura média entre homens e mulheres brasileiras?

Nesse caso usaremos Intervalos de Confiança para duas populações:

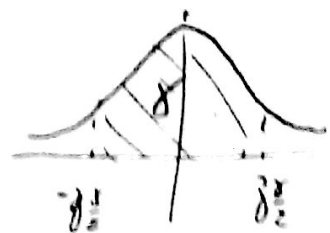
6. IC para $\mu_A - \mu_B$ com σ_A^2 e σ_B^2 conhecidos

Supomos que temos duas populações A e B. Tiro uma amostra de cada uma e quero saber quanto vale a diferença das médias.



O intervalo de confiança será:

$$\left| \bar{X}_A - \bar{X}_B \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} \right|$$



Obs: se perguntarmos se as médias são iguais ($\mu_A = \mu_B$) ou ($\mu_A - \mu_B = 0$), temos que ver se 0 está dentro do IC. Se estiver, consideramos que as médias são iguais.

7. IC para $\mu_A - \mu_B$ com σ_A^2 e σ_B^2 desconhecidos

Se não conhecemos as variâncias como fazemos?

Neste curso trabalharemos apenas com o caso das 2 populações terem a mesma variância

$$\sigma_A^2 = \sigma_B^2 = \sigma^2$$

O intervalo de confiança para $\mu_A - \mu_B$ será:

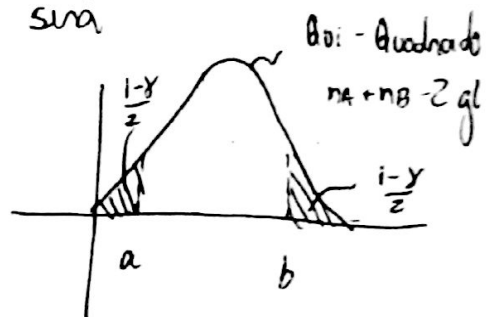
$$\left[\bar{X}_A - \bar{X}_B \pm t_{\left(\frac{1-\alpha}{2}, n_A+n_B-2\right)} \sqrt{sp^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)} \right]$$

Onde:

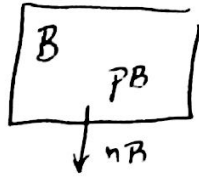
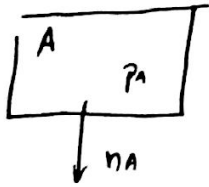
$$\left[sp^2 = \frac{(n_A-1)s_A^2 + (n_B-1)s_B^2}{n_A + n_B - 2} \right]$$

Além disso, o IC para σ^2 será

$$\left[\frac{1}{b} < \frac{\sigma^2}{sp^2 (n_A + n_B - 2)} < \frac{1}{a} \right]$$



o. IC para $p_A - p_B$ (variáveis com distribuição de Bernoulli)



$X_1, X_2, \dots, X_{n_A}$

(Bernoulli)

$$\hookrightarrow \hat{p}_A = \sum X_i / n_A$$

$Y_1, Y_2, \dots, Y_{n_B}$

(Bernoulli)

$$\hookrightarrow \hat{p}_B = \sum Y_i / n_B$$

Queremos estimar $p_A - p_B$: usaremos o intervalo de confiança:

$$\left| \hat{p}_A - \hat{p}_B \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}} \right|$$

V - Teste de Hipóteses

Muitas vezes, ao coletar uma amostra, não queremos saber o valor médio de uma variável ou seu I.C., mas sim decidir entre duas afirmações contraditórias.

Por exemplo

- A média é maior que 500 vs A média é menor que 500
- A média é nula vs a média é $\neq 0$
- A média de A é igual a de B vs $\mu_A \neq \mu_B$

Em casos como esses, vamos usar o método do teste de hipóteses.

Em um teste de hipóteses temos sempre uma hipótese nula (H_0), que é aquilo que queremos ver se é verdade, e a sua negação, chamada de hipótese alternativa (H_a).

Nos exemplos anteriores teríamos

- $H_0 : \mu > 500$ e $H_a : \mu < 500$
- $H_0 : \mu = 0$ e $H_a : \mu \neq 0$
- $H_0 : \mu_A = \mu_B$ e $H_a : \mu_A \neq \mu_B$

Para poder decidir qual das hipóteses é verdadeira com base nos dados coletados em uma amostra, definiremos dois tipos de erro

- Erro do tipo I (α) : rejeitar H_0 quando H_0 é verdadeira
- Erro do tipo II (β) : aceitar H_0 quando H_0 é falsa

Em um exercício típico, temos um valor de α , que a probabilidade máxima aceitável para um erro do tipo I, e vamos querer estabelecer uma regra: para quais valores da minha estimativa aceito que H_0 é verdadeira e para quais valores rejeito H_0 em favor da hipótese alternativa.

Mãos à obra:

1. Teste de Hipótese para a média

Supondo que X seja uma variável normal e que tomarmos uma amostra aleatória com tamanho n , o quadro a seguir resume os testes que podemos fazer e como decidimos qual hipótese aceitar / rejeitar

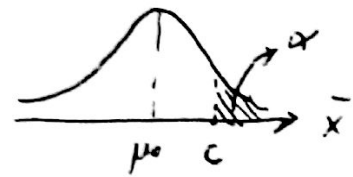
Hipóteses	Regra	σ^2 conhecido	σ^2 desconhecido
$H_0: \mu = \mu_0$ $H_a: \mu > \mu_0$	Se $\bar{x} > C$, rejeito H_0	$C = \mu_0 + z_{\alpha} \frac{\sigma}{\sqrt{n}}$	$C = \mu_0 + t_{\alpha, n-1} \frac{s}{\sqrt{n}}$
$H_0: \mu = \mu_0$ $H_a: \mu < \mu_0$	Se $\bar{x} < C$, rejeito H_0	$C = \mu_0 - z_{\alpha} \frac{\sigma}{\sqrt{n}}$	$C = \mu_0 - t_{\alpha, n-1} \frac{s}{\sqrt{n}}$
$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$	Se $\bar{x} > S$ ou $\bar{x} < I$, rejeito H_0	$I = \mu_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ $S = \mu_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$	$I = \mu_0 - t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}$ $S = \mu_0 + t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}$

Para atender às dúvidas mais curiosas, vejamos como surgem alguns desses resultados

A. $H_0: \mu = \mu_0$ e $H_a: \mu > \mu_0$ para σ^2 conhecido

O erro do tipo I acontecerá para o caso em que a média é μ_0 , mas obtemos na amostra um valor maior e rejeitamos H_0 . Quão maior deve ser o valor obtido? Definimos um limitante C , tal que $P(\bar{X} > C \mid \mu = \mu_0) = \alpha$

Ou seja: a probabilidade de rejeitar H_0 com $\mu = \mu_0$ é α



Calculamos C :

$$P(\bar{X} > C \mid \mu = \mu_0) = \alpha$$

$$\alpha = P\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > \frac{C - \mu_0}{\sigma/\sqrt{n}}\right)$$



$$\frac{C - \mu_0}{\sigma/\sqrt{n}} = z_\alpha$$

$$C = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

Assim a regra será:

Se $\bar{X}$ observado for maior que $C = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$, rejeite H_0 .

2. Teste para Variância

Hipóteses	Regra de decisão	Limites
$H_0: \sigma^2 = \sigma_0^2$ $H_a: \sigma^2 > \sigma_0^2$	Se $S^2_{obs} > C$, rejeita H_0	$C = \chi^2_{(1-\alpha, n-1)} \frac{\sigma_0^2}{n-1}$
$H_0: \sigma^2 = \sigma_0^2$ $H_a: \sigma^2 < \sigma_0^2$	Se $S^2_{obs} < C$, rejeita H_0	$C = \chi^2_{(\alpha, n-1)} \frac{\sigma_0^2}{n-1}$
$H_0: \sigma^2 = \sigma_0^2$ $H_a: \sigma^2 \neq \sigma_0^2$	Se $S^2_{obs} > I$ ou $S^2_{obs} < S$, rejeita H_0	$I = \frac{a\sigma_0^2}{n-1}$ $S = \frac{b\sigma_0^2}{n-1}$ $a = \chi^2_{(\frac{\alpha}{2}, n-1)}$ $b = \chi^2_{(1-\frac{\alpha}{2}, n-1)}$

Obs: assumindo que a variável X segue distribuição normal

3. Teste para proporção (p) com n grande

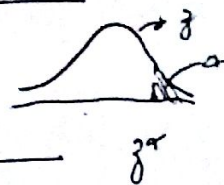
Para X uma variável de Bernoulli, tomamos $X_1, X_2, \dots, X_n$ com n grande ($n > 30$ ou $np > 5$ e $n(1-p) > 5$).
 A proporção é estimada por $\hat{p} = \frac{\sum X_i}{n} = \bar{X}$

Com n grande, podemos aproximar pela normal

$$\hat{p} = \bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$$

Assim, os testes de hipóteses serão:

Hipóteses	Crítéria	Limitantes
$H_0: p = p_0$ $H_a: p > p_0$	Se $\bar{X} > C$, rejeito H_0	$C = p_0 + z_{\alpha} \sqrt{\frac{p_0(1-p_0)}{n}}$
$H_0: p = p_0$ $H_a: p < p_0$	Se $\bar{X} < C$, rejeito H_0	$C = p_0 - z_{\alpha} \sqrt{\frac{p_0(1-p_0)}{n}}$
$H_0: p = p_0$ $H_a: p \neq p_0$	Se $\bar{X} < I$ ou $\bar{X} > S$, rejeito H_0	$I = p_0 - z_{\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}$ $S = p_0 + z_{\frac{\alpha}{2}} \sqrt{\frac{p_0(1-p_0)}{n}}$



4. Teste para proporção (p) : n pequeno

Se n for pequeno ($n < 30$ ou $np < 5$ / $n(1-p) < 5$) não vale aproximar pela normal.

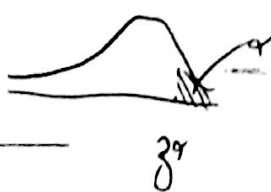
Teremos qui usam a propriedade da distribuição binomial, para a qual a probabilidade de ter x sucessos em n experimentos i :

$$P(X = x) = \binom{n}{x} p_0^x (1-p_0)^{n-x}$$

Vejamos um exemplo: questão 44 da lista

Vamos agora estudar testes de hipóteses envolvendo duas populações:

Teste para $\mu_A - \mu_B$ com σ_A^2 e σ_B^2 conhecidos

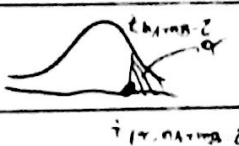
Hipóteses	Critérios	Limitantes	
$H_0: \mu_A - \mu_B = 0$ $H_a: \mu_A - \mu_B > 0$	se $\bar{X}_A - \bar{X}_B > C$, rejeito H_0	$C = z_\alpha \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$	
$H_0: \mu_A - \mu_B = 0$ $H_a: \mu_A - \mu_B < 0$	se $\bar{X}_A - \bar{X}_B < C$, rejeito H_0	$C = -z_\alpha \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$	
$H_0: \mu_A - \mu_B = 0$ $H_a: \mu_A - \mu_B \neq 0$	se $\bar{X}_A - \bar{X}_B > S$ ou $\bar{X}_A - \bar{X}_B < I$, rejeito H_0	$S = z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$	$I = -z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$

6. Teste de hipótese para $\mu_A - \mu_B$ com σ_A^2 e σ_B^2 desconhecidos

Novamente, trabalharemos com a suposição de que $\sigma_A^2 = \sigma_B^2 = \sigma^2$, que é estimada por

$$S_p^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}$$

Teremos:

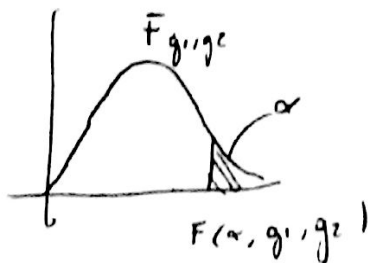
Hipóteses	Critérios	Limitante	
$H_0: \mu_A - \mu_B = 0$ $H_a: \mu_A - \mu_B > 0$	se $\bar{X}_A - \bar{X}_B > C$, rejeito H_0	$C = t(\alpha, n_A + n_B - 2) S_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$	
$H_0: \mu_A - \mu_B = 0$ $H_a: \mu_A - \mu_B < 0$	se $\bar{X}_A - \bar{X}_B < C$, rejeito H_0	$C = -t(\alpha, n_A + n_B - 2) S_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$	
$H_0: \mu_A - \mu_B = 0$ $H_a: \mu_A - \mu_B \neq 0$	se $\bar{X}_A - \bar{X}_B > S$ ou $\bar{X}_A - \bar{X}_B < I$, rejeito H_0	$S = t(\frac{\alpha}{2}, n_A + n_B - 2) S_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$	$I = -S$

7. Teste de Hipótese para Variâncias: σ_A^2 e σ_B^2

Hipóteses	Critério para rejeição de H_0
$H_0: \sigma_A^2 = \sigma_B^2$ $H_a: \sigma_A^2 > \sigma_B^2$	$F = \frac{S_A^2}{S_B^2} \geq F_{(\alpha; n_A-1; n_B-1)}$
$H_0: \sigma_A^2 = \sigma_B^2$ $H_a: \sigma_A^2 < \sigma_B^2$	$F = \frac{S_A^2}{S_B^2} \leq F_{(1-\alpha; n_A-1; n_B-1)}$
$H_0: \sigma_A^2 = \sigma_B^2$ $H_a: \sigma_A^2 \neq \sigma_B^2$	$F = \frac{S_A^2}{S_B^2} \geq F_{(\frac{\alpha}{2}; n_A-1; n_B-1)} \quad \text{ou} \quad F \leq F_{(1-\frac{\alpha}{2}; n_A-1; n_B-1)}$

Em que F é a distribuição de Fischer

$F(\alpha; g_1; g_2)$ indica para uma Fischer com graus de liberdade g_1 e g_2 o valor para o qual $P(F \geq F) = \alpha$



5. Testes Emparelhados - 2 populações

De 5 a 7 vimos testes para duas amostras independentes. Vimos vir agora como proceder para amostras dependentes.

Imagine que medimos um um dia de manhã a produtividade de 5 operários $(X_1, \dots, X_5)$ e, à tarde, repetimos a medida $(Y_1, \dots, Y_5)$.

Facilmente vemos que para certo i , X_i e Y_i são dependentes, pois se trata do mesmo funcionário. Podemos, no entanto, querer saber se a produtividade média de manhã é igual ou não à da tarde. O que fazemos?

1. Calculamos $\Delta_i = X_i - Y_i$.

x	y	Δ
x_1	y_1	$\Delta_1 = x_1 - y_1$
$\vdots$	$\vdots$	$\vdots$
x_n	y_n	$\Delta_n = x_n - y_n$

2. Fazemos teste de hipóteses para a média de Δ (μ_Δ), como já aprendemos anteriormente (teste de hipótese para μ com σ^2 desconhecido).

Os testes podem ser:

$$H_0 : \mu_\Delta = 0$$

$$H_a \left\{ \begin{array}{l} \mu_\Delta > 0 \\ \mu_\Delta < 0 \\ \mu_\Delta \neq 0 \end{array} \right.$$

9. Comparação de K médias

Aprendemos a comparar as médias de 2 populações. Mas e se quisermos comparar as médias de K populações com $K > 2$?

Nossas hipóteses são:

$$\begin{array}{l} H_0 : \mu_1 = \mu_2 = \dots = \mu_K \\ H_a : \text{existe alguma diferente} \end{array}$$

Suposições

- variável de interesse normal
- variâncias das K populações iguais ($\sigma_1^2 = \dots = \sigma_K^2 = \sigma^2$)

Vamos calcular:

$$S_p = \frac{\sum (n_i - 1) s_i^2}{\sum (n_i - 1)}$$

e

$$S_E^2 = \frac{\sum [n_i (\bar{x}_i - \bar{\bar{x}})^2]}{K - 1}$$

$\bar{x}_i$ = média população i

$\bar{\bar{x}}$ = média de todos juntos

Calculamos

$$F_{obs} = \frac{S_E^2}{S_p^2} \text{ e rejeitamos } H_0 \text{ se } F_{obs} \geq F_{(\alpha; K-1, \sum (n_i-1))}$$

Em que

Observação : se quisermos testar a suposição de que as variâncias são iguais, usamos o seguinte teste

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

H_a : existe algum diferente

$$\text{Calcula : } D = \sum_{i=1}^k \left[(n_i - 1) \ln \left(\frac{sp^2}{si^2} \right) \right]$$

$$\text{Rejeito } H_0 \text{ se } D \geq \chi^2_{(\alpha; k-1)}$$

EXERCÍCIO - Estimador de Máxima Verossimilhança (EMV)

Se algum exercício pedir o EMV de uma distribuição
faremos o seguinte:

- Dada uma amostra $x_1, \dots, x_n$ escrevermos a sua
função de verossimilhança ($L(x_1, \dots, x_n)$) que é
dada por

(DISCRETA) $L(x_1, \dots, x_n) = P(X=x_1) \times P(X=x_2) \times \dots \times P(X=x_n)$

(CONTÍNUA) $L(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i)$

↑
produto

- Para descobrir o estimador do parâmetro δ (pode
ser $p, \lambda, \mu, \sigma, \dots$), vamos achar o valor de δ
que maximiza L . Assim, derivamos e igualamos a 0

$$\frac{\partial L}{\partial \delta} = 0$$

Exemplo: qual o EMV para uma distribuição exponencial?

$$f(x) = \lambda e^{-\lambda x}$$

Qual EMV de λ ?

$$L(x_1, \dots, x_n) = \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} \cdot \dots \cdot \lambda e^{-\lambda x_n}$$

$$L(x_1, \dots, x_n) = \lambda^n e^{-\lambda \sum x_i}$$

Para achar o valor que maximiza L , é mais simples
achar λ que maximiza $\ln L$

$$\ln L = n \ln \lambda - \lambda \sum x_i$$

$$\frac{\partial \ln L}{\partial \lambda} = \frac{n}{\lambda} - \sum x_i = 0$$

$$\hat{\lambda} = \frac{n}{\sum x_i}$$

$$\boxed{\hat{\lambda} = \frac{1}{\bar{x}}}$$

Propriedades do EMV

- ① Assintoticamente, todo estimador $\hat{\theta}$ de máxima verossimilhança segue uma distribuição normal

Para amostra grande $\rightarrow \hat{\theta} \sim \text{Normal} (E(\hat{\theta}), \text{Var}(\hat{\theta}))$

Além disso: $\sqrt{n} (\hat{\theta} - \theta) \approx N \left(0, \frac{1}{I_F(\theta)^*} \right)$

$\underbrace{\hspace{1cm}}$
erro do estimador

Assim, para n grande $\hat{\theta} - \theta$ se aproxima de zero e $E(\hat{\theta}) = \theta$, o que torna o EMV um estimador justo.

$$I_F(\theta) = E \left[\left(\frac{\partial \log L}{\partial \theta} \right)^2 \right] = - E \left[\frac{\partial^2 \log L}{\partial \theta^2} \right]$$

↑
média

1.

LISTA

Um paciente fez seis exames de sangue em 6 meses consecutivos para medir seu nível de fósforo por decilitro de sangue. Os resultados obtidos foram:

5,6; 5,2; 4,6; 4,9; 5,7; 6,4

Calcule a média, a mediana e o desvio padrão amostral.

i) Média

$$\bar{x} = \frac{\sum x_i}{n} = \frac{5,6 + 5,2 + 4,6 + 4,9 + 5,7 + 6,4}{6}$$

$$\boxed{\bar{x} = 5,4}$$

ii) Desvio - padrão

$$s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

$$s^2 = \frac{(5,6^2 + 5,2^2 + 4,6^2 + 4,9^2 + 5,7^2 + 6,4^2) - 6 \cdot 5,4^2}{6-1}$$

$$s^2 = 0,412$$

$$s = \sqrt{0,412} \Rightarrow \boxed{s = 0,642}$$

iii) Mediana ordenando

4,6 4,9 5,2 (5,6) 5,7 6,4

$$\boxed{Md = 5,2 \text{ ou } 5,6}$$

2.

LISTA (modificado)

Calcule a média e o desvio padrão para a distribuição de frequências abaixo:

Classes	Frequências
150 - 180	3
180 - 210	8
210 - 240	10
240 - 270	13
270 - 300	25

Quando temos intervalos superiores que as pontas estão no ponto médio

$$\bar{x} = \frac{\sum x_i}{n} = \frac{3 \cdot 165 + 8 \cdot 195 + 10 \cdot 225 + 13 \cdot 255 + 25 \cdot 285}{3 + 8 + 10 + 13 + 25}$$

$$\boxed{\bar{x} = 249,935}$$

$$s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1}$$

$$s^2 = \frac{(3 \cdot 165^2 + 8 \cdot 195^2 + 10 \cdot 225^2 + 13 \cdot 255^2 + 25 \cdot 285^2) - 59\bar{x}^2}{59-1}$$

$$s^2 = 1432,450$$

$$\boxed{s = 37,848}$$

INTERVALOS DE CONFIANÇA PARA UMA POPULAÇÃO

3.

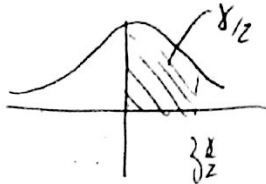
LISTA

De experiências passadas sabe-se que o desvio padrão de altura de crianças de 5a série é 5cm. Pede-se:

- Colhendo uma amostra de 36 destas crianças observou-se média 150cm. Qual o intervalo de confiança para a média da população? (em $\delta = 95\%$)
- Que tamanho deve ter uma amostra para que o intervalo $150 \pm 0,98$ tenha 95% de confiança?

a) IC para μ com σ^2 conhecido

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



No caso $z_{\frac{\delta}{2}} = z_{47,5\%} = 1,96$

$$\left[150 - 1,96 \cdot \frac{5}{\sqrt{36}} ; 150 + 1,96 \cdot \frac{5}{\sqrt{36}} \right]$$

$$\boxed{\mu = 150 \pm 1,633 \text{ com } \delta = 95\%}$$

b) Queremos $z_{\frac{\delta}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 0,98$

$$z_{47,5\%} \cdot \frac{\sigma}{\sqrt{n}} = 0,98$$

$$n = \left(\frac{1,96 \cdot 5}{0,98} \right)^2$$

$$\boxed{n = 100}$$

4.

LISTA (modificado)

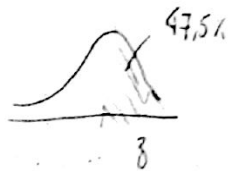
Para os dados do exercício 2 Pede-se:

- Construa um intervalo para a média com confiança de 95%, supondo que a variância seja conhecida e igual a 100.
- Construa um intervalo de confiança para a variância dessa variável.

a) O intervalo para μ com σ^2 conhecida é:

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

No caso $z_{\frac{\delta}{2}} = z_{47,5\%}$
 $z_{\frac{\delta}{2}} = 1,96$



$$\Rightarrow \left[249,915 - 1,96 \frac{10}{\sqrt{98}} ; 249,915 + 1,96 \frac{10}{\sqrt{98}} \right]$$

Assim

$$\boxed{\mu = 249,915 \pm 2,552 \text{ com } 95\% \text{ de confiança}}$$

b)

IC para σ^2

$$\left[\frac{s^2(n-1)}{b} ; \frac{s^2(n-1)}{a} \right]$$

Onde χ^2_{n-1} $a = \chi^2_{0,025}$ $b = \chi^2_{0,975}$

Da tabela de χ^2 : $a = 40,48$ e $b = 63,3$

↓
Usamos χ^2_{60} (mais próximo)

$$\Rightarrow \boxed{\text{IC para } \sigma^2 [997,384 ; 2052,83]}$$

5.

LISTA

A precipitação pluviométrica anual em uma certa região tem desvio padrão igual a 3,1. Nos últimos 9 anos, observou-se 28,3/31,7/29,8/30,5/34,1/27,9/35,0/26,9/30,2. Pede-se:

- Construa um intervalo para a precipitação média com 98% de confiança.
- Utilizando ainda esses dados, construa um intervalo para a precipitação média com 98% de confiança caso o desvio padrão não fosse conhecido.

a) Calculamos:

$$\bar{x} = 30,489$$

$$s^2 = 7,474$$

$$z_{\frac{\alpha}{2}} = z_{0,01} = 2,33$$

$\Rightarrow$ IC para μ com σ^2 conhecido

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

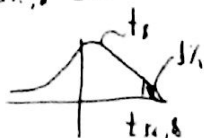
$$\left[30,489 - 2,33 \cdot \frac{3,1}{\sqrt{9}} ; 30,489 + 2,33 \cdot \frac{3,1}{\sqrt{9}} \right]$$

$$\Rightarrow \boxed{\mu = 30,489 \pm 2,408}$$

b) IC para μ com σ^2 desconhecido

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

No caso: $t_{\left(\frac{1-\alpha}{2}; n-1\right)} = t_{0,01; 8} = 2,896$



Assim temos

$$IC: 30,489 \pm 2,896 \cdot \frac{\sqrt{7,474}}{\sqrt{9}}$$

$$\boxed{\mu = 30,489 \pm 2,639}$$

6.

LISTA

A resistência à tração de 20 corpos de prova é:

131 132 134 135 135 138 139 139 140 142
143 144 144 145 146 147 148 149 150 138

Construa um intervalo de confiança de 95% para a variância.

Calculamos

$$\bar{x} = 140,95$$

$$s^2 = 32,787$$

IC para σ^2

$$\left[\frac{s^2(n-1)}{b} ; \frac{s^2(n-1)}{a} \right]$$

Onde $\begin{cases} a = \chi^2_{\left(\frac{1-\alpha}{2}; n-1\right)} \\ b = \chi^2_{\left(\frac{1+\alpha}{2}; n-1\right)} \end{cases}$

$$\begin{cases} a = \chi^2_{0,025; 19} = 8,91 \\ b = \chi^2_{0,975; 19} = 32,9 \end{cases}$$

$$IC: \left[\frac{32,787 \cdot 19}{32,9} ; \frac{32,787 \cdot 19}{8,91} \right]$$

IC para σ^2

$$\boxed{[18,935 ; 69,916]}$$

7.

LISTA

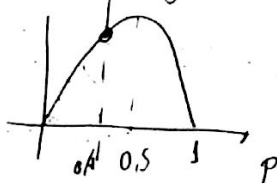
Uma empresa vai lançar um novo produto. O gerente de marketing encomenda uma pesquisa de mercado, entrevistando 500 pessoas na qual apenas 157 manifestaram intenção de comprar o novo produto. Construa um I.C. com 92% de confiança nas seguintes situações:

- Se soubermos que a real proporção de compradores é inferior a 40%.
- Se soubermos que a relação é inferior a 80%.

a) para p , o I.C. é:

$$p = \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

Nesse caso, como desconhecemos p , podemos usar o valor que maximiza $p(1-p)$



Para $0 \leq p \leq 0.4$, $p(1-p)$ é máximo para $p = 0.4$

$$p = \frac{157}{500} + 1.75 \sqrt{\frac{0.4 \cdot 0.6}{500}}$$

$$p = 0.314 \pm 0.038 \quad \text{com } \delta = 92\%$$

b) para $0 \leq p \leq 0.8$, $p(1-p)$ é máximo em $p = 0.5$

$$p = \frac{157}{500} + 1.75 \sqrt{\frac{0.5 \cdot 0.5}{500}}$$

$$p = 0.314 \pm 0.039$$

8.

LISTA

Uma amostra preliminar de 500 famílias verificou que 240 delas possuem forno de micro ondas. Para estimar a proporção de famílias com esse utensílio, qual o tamanho de amostra necessário para que tenhamos 95% de confiança em que o erro da nova estimativa não seja superior a 0.02?

i) A amostra preliminar fornece uma estimativa de p .

$$p^* = \frac{240}{500} = 0.48$$

ii) Para o 2º experimento o

I.C. será:

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

ERRO

$$\text{Se } z_{\frac{\alpha}{2}} \sqrt{\frac{p^*(1-p^*)}{n}} = 0.02$$

$$\text{Temos } n = \left(\frac{z_{\frac{\alpha}{2}}}{0.02} \right)^2 p^*(1-p^*)$$

$$n = \left(\frac{1.96}{0.02} \right)^2 0.48 \cdot 0.52$$

$$n = 2397.16$$

Como n deve ser inteiro

$$n = 2398$$

Obs, para ter um erro menor que 0.02, n deve ser maior ou igual a 2397.16. Por isso não podemos arredondar para baixo

9

LISTA

Uma amostra de 10 mil itens de um lote de produção foi inspecionada e o número de defeitos por item foi registrado:

Defeitos	0	1	2	3	4
Quantidade	6.000	3.000	600	350	50

Pede-se:

- Determine um intervalo para a proporção de itens com defeitos com 98% de confiança.
- Determine um intervalo para o número médio de defeitos nos itens com 98% de confiança.

a) Para λ de Poisson

$$IC: \hat{\lambda} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{\lambda}}{n}}$$

No caso

$$\hat{\lambda} = \bar{X} = \frac{0 \cdot 6000 + 1 \cdot 3000 + 2 \cdot 600 + 3 \cdot 350 + 4 \cdot 50}{10000}$$

$$\hat{\lambda} = 0,545$$

$$E: z_{\frac{\alpha}{2}} = z_{49\%} = 2,33$$

$$IC: 0,545 \pm 2,33 \sqrt{\frac{0,545}{10000}}$$

$$\lambda = 0,545 \pm 0,017$$

com $\alpha = 98\%$

INTERVALOS DE CONFIANÇA PARA DUAS POPULAÇÕES

10

LISTA

Dois processos de conservação de alimentos estão sendo utilizados e a variável de interesse é o tempo de duração. Duas amostras independentes foram colhidas. A com 16 latas apresentou tempo médio de 50 dias e B com 25 latas apresentou 60 dias. Construa um intervalo para a diferença de médias com 95% de confiança e interprete os resultados. Use desvio padrão do processo igual a 10.

IC para $\mu_A - \mu_B$ com σ_A^2 e σ_B^2 conhecidos:

$$\bar{X}_A - \bar{X}_B \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

$$50 - 60 \pm 1,96 \sqrt{\frac{10^2}{16} + \frac{10^2}{25}}$$

$$\mu_A - \mu_B = -10 \pm 6,28$$

$$OU: [-16,28, -3,72]$$

Note que como zero não está contida no IC, não podemos considerar que

$$\mu_A = \mu_B$$

Para se estudar o desempenho de duas corretoras de ações, selecionou-se de cada uma delas amostras aleatórias das ações negociadas. Para cada ação selecionada, computou-se a porcentagem de lucro apresentada durante um período fixado de tempo. Os dados estão a seguir:

Corretora A: 45; 60; 54; 62; 55; 38; 48; 64

Corretora B: 57; 55; 58; 52; 59; 55; 59

Supondo que as variâncias sejam iguais para as duas corretoras, pede-se para verificar se elas possuem desempenhos diferentes.

Calculamos

	$\bar{x}$	s^2	n
A	53,25	81,37	8
B	56,42	6,619	7

Além disso:

$$sp^2 = \frac{(n_A - 1) s_A^2 + (n_B - 1) s_B^2}{n_A + n_B - 2}$$

$$sp^2 = 46,863$$

O IC para $\mu_A - \mu_B$ será

$$\bar{X}_A - \bar{X}_B \pm z_{\frac{\alpha}{2}} \sqrt{sp^2 \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}$$

$$53,25 - 56,42 \pm 2,16 \sqrt{46,863 \left(\frac{1}{8} + \frac{1}{7} \right)}$$

$$\boxed{\mu_A - \mu_B = -3,17 \pm 7,65}$$

$$\text{OU } [-10,14 ; 3,47]$$

Como 0 (zero) está contido no IC, não podemos dizer com o nível de significância adotado que $\mu_A \neq \mu_B$ ($\mu_A - \mu_B \neq 0$)

Em uma pesquisa de intenção de voto para presidente conduzida em dois estados brasileiros revelou os seguintes dados:

	entrevistados	favoráveis ao candidato X
Estado A	2000	850
Estado B	1500	650

$$\hat{p}_A = 0,425$$

$$\hat{p}_B = 0,433$$

Pede-se para determinar um intervalo de confiança ($\gamma = 98\%$) para a diferença $p_A - p_B$ de intenção de voto no candidato X entre os dois estados.

Calculamos

$$\hat{p}_A = \frac{\sum x_{Ai}}{n_A} = \frac{850}{2000} = 0,425$$

$$\hat{p}_B = \frac{\sum x_{Bi}}{n_B} = \frac{650}{1500} = 0,433$$

O IC para $p_A - p_B$ será

$$\hat{p}_A - \hat{p}_B \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_A(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B(1-\hat{p}_B)}{n_B}}$$

$$0,425 - 0,433 \pm 2,33 \sqrt{\frac{0,425 \cdot 0,575}{2000} + \frac{0,433 \cdot 0,567}{1500}}$$

$$\boxed{p_A - p_B = 0,008 \pm 0,039}$$

com $\gamma = 98\%$

TESTES DE HIPÓTESE PARA UMA POPULAÇÃO

13

Na ausência de treinamento os escores de um exame de admissão em um MBA variam normalmente com média 475 e desvio padrão 100. Suponhamos que o treinamento possa melhorar a média, mas não altere o desvio padrão. Uma equipe treina 100 estudantes. Suas notas acusam $\bar{x} = 478$. Pede-se:

LISTA

- Para um nível de significância de 5% é possível afirmar que as notas aumentaram?
- O que ocorreria se a amostra tivesse 1000 e não 100 alunos?

a) $H_0: \mu = 475$
 $H_a: \mu > 475$ ← o que queremos ver

Vamos rejeitar H_0 se $\bar{x}$ for maior que C:

$$C = \mu_0 + z_{\alpha} \frac{\sigma}{\sqrt{n}}$$

$$C = 475 + 1,64 \frac{100}{\sqrt{100}}$$

$$C = 491,4$$

Como $\bar{x} = 478 < C$, não rejeito H_0 e digo que a média não aumentou

b) Neste caso:

$$C = \mu_0 + z_{\alpha} \frac{\sigma}{\sqrt{n}} = 475 + 1,64 \frac{100}{\sqrt{1000}}$$

$$C = 476,64$$

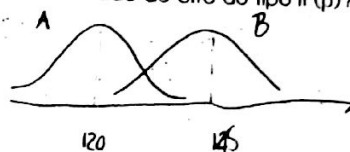
Assim como $\bar{x} > C$, digo que a média dos que passaram pelo treinamento aumentou
 $(\mu > \mu_0) \leftrightarrow \text{Rejeito } H_0$

14

LISTA (modificado)

Um estudo pretende identificar se um grupo de indígenas pertence a uma determinada tribo. Sabe-se que os índios da tribo A possuem altura média de 120 cm, enquanto que os da tribo B possuem altura média 145 cm. O desvio padrão nos dois casos é 40. O critério de decisão é o seguinte: se para uma amostra de 100 pessoas for observada média amostral superior a 130, considera-se que o grupo é da tribo B, caso contrário, é da tribo A. Você desconfia que os indivíduos são do grupo B. Pede-se:

- Qual o erro do tipo I?
- Determine a probabilidade do erro do tipo I (α).
- Qual deve ser o critério de decisão para que $\alpha = 5\%$?
- Para o critério definido no enunciado, qual a probabilidade de erro do tipo II (β)?



$$H_0: \mu = 145$$

$$H_a: \mu = 120$$

a) O erro do tipo I é rejeitar H_0 e H_0 ser verdadeira $\Rightarrow$ Dizer que é da A quando é da B

$$b) \alpha = P(\bar{x} < 130 \mid \mu = 145)$$

↑ ↑
parar em A seguir para B

$$\alpha = P\left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < \frac{C - \mu}{\sigma/\sqrt{n}}\right)$$

Z

$$\alpha = P\left(Z < \frac{130 - 145}{40/\sqrt{100}}\right)$$

$$\alpha = P(Z < -3,75)$$

$$\alpha \approx 0$$

(VERSO)

12

LISTA

Desconfiando-se de que uma moeda fosse viciada, realizou-se um experimento que consiste em lançar essa moeda cem vezes. Obtivemos 59 caras. Pode-se afirmar a existência de vício? Considere o nível de significância igual a 5%.

$$H_0: p = 0,5$$

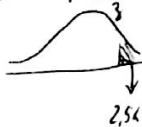
$$H_a: p \neq 0,5$$

$$\hat{p} = \frac{59}{100} = 0,59$$

O critério é rejeitar H_0 se $\hat{p} > S$ ou $\hat{p} < I$ com:

$$\rightarrow I = p_0 - z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} = p_0 - z_{2,5\%} \cdot \sqrt{\frac{0,5 \cdot 0,5}{100}}$$

$$I = 0,5 - 1,96 \cdot \sqrt{\frac{0,5 \cdot 0,5}{100}}$$



$$I = 0,402$$

$$+ S = p_0 + z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}$$

$$S = 0,598$$

Como $\hat{p} = 0,58 < S$ e $\hat{p} > I$,

Aceitamos H_0 e dizemos que a moeda não é viciada.

18

LISTA

Pretende-se testar a honestidade de uma moeda com base no número de caras obtidas em 15 lançamentos. (Atenção: não é possível utilizar a distribuição normal). Pede-se

- Qual o teste de hipóteses mais adequado?
- Com um nível de significância de 5%, quantas caras deveriam ocorrer para rejeitarmos H_0 ?
- Qual o real nível de significância no item (b)?

$$a) H_0: p = 0,5$$

$$H_a: p \neq 0,5$$

b) Vamos ter que calcular as probabilidades da binomial tal que

$$P(X \leq I) = \frac{\alpha}{2}$$

$$P(X \geq S) = \frac{\alpha}{2}$$

Para cada $0 \leq x \leq 15$

$$\begin{aligned} P(X=x) &= \binom{15}{x} \cdot p_0^x \cdot (1-p_0)^{15-x} \\ &= \binom{15}{x} \cdot 0,5^x \cdot (1-0,5)^{15-x} \\ &= \frac{15!}{x!(15-x)!} \cdot 0,5^{15} \end{aligned}$$

Vamos que (com ajuda do Excel)

$$P(X \leq 3) = 1,76\% \quad \text{e} \quad P(X \leq 4) = 5,92\%$$

Podemos adotar como I qualquer um deles. Vou usar 3 que tem probabilidade mais perto de 2,5%.

$$P(X \geq 12) = 1,76\% \quad \text{e} \quad P(X \geq 11) = 5,92\%$$

Podemos adotar 10 ou 11 como S , já que 2,5% está entre eles. usarei $S = 12$

(VERSO)

TESTE DE HIPÓTESES PARA DUAS POPULAÇÕES

19.

LISTA (modificado)

Os dados abaixo foram coletados em duas empresas diferentes no dia 18/05 e referem-se ao tempo (minutos) que profissionais gastaram com pesquisas na Internet naquele dia.

Empresa A	12,2	13	11,5	12,6	10,5
	8,4	9,7	8,1	12,3	
Empresa B	11,2	23,1	12,4	10,4	12,1
	19,3	17,5	11,1	12,4	16,6

Suponha que $\sigma_A^2 = 2,3$ e $\sigma_B^2 = 3$, teste a hipótese de que o tempo médio é diferente entre as duas empresas. $\gamma_{0,05}$, $\alpha = 5\%$

Calculamos

	n	$\bar{X}$	S^2
A	9	10,92	3,37
B	10	11,61	18,24

O teste será:

$$H_0: \mu_A = \mu_B$$

$$H_1: \mu_A \neq 0$$

Vamos rejeitar H_0 se $\bar{X}_A - \bar{X}_B > S$

ou $\bar{X}_A - \bar{X}_B < I$ com:

$$I = -z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = -1,96 \sqrt{\frac{2,3^2}{9} + \frac{3^2}{10}}$$


$$I = -2,39$$

$$S = +z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = -I$$

$$\Rightarrow S = 2,39$$

$$\text{No caso: } \bar{X}_A - \bar{X}_B = 10,92 - 11,61$$

$$| \bar{X}_A - \bar{X}_B | = 0,69$$

Como $\bar{X}_A - \bar{X}_B < I$, rejeitamos H_0 e dizemos que as médias são diferentes.

20.

LISTA (modificado)

Para os dados do exercício anterior, supondo que as variâncias das duas populações fossem iguais e desconhecidas, repita o teste de hipóteses.

i) Calculamos

$$s_p^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}$$

$$s_p^2 = \frac{8 \cdot 3,37 + 9 \cdot 18,24}{17}$$

$$s_p^2 = 11,242$$

ii) Para $\mu_A - \mu_B$ com σ^2 desconhecido

$$I = -t\left(\frac{\alpha}{2}; n_A + n_B - 2\right) s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

$$I = -t(2,5\%; 17) \cdot \sqrt{11,242} \sqrt{\frac{1}{9} + \frac{1}{10}}$$

$\downarrow$
2,11

$$I = -3,251$$

$$S = t\left(\frac{\alpha}{2}; n_A + n_B - 2\right) s_p \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} = -I$$

$$S = 3,251$$

$$\text{Como } \bar{X}_A - \bar{X}_B = -0,69 < I$$

Rejeitamos H_0 e dizemos que há diferença de médias

21

LISTA (modificado)

Pretende-se comparar a variância do preço de duas ações (com distribuição de preços normal). Amostras de tamanhos 21 e 25 forneceram variâncias $s_A^2 = (67,233)^2$ e $s_B^2 = (37,128)^2$. Teste com $\alpha = 1\%$ se as duas variâncias são iguais nos casos:

- a) Com hipótese alternativa $\sigma_A^2 > \sigma_B^2$
 b) Com hipótese alternativa $\sigma_A^2 < \sigma_B^2$

$$F_{obs} = \frac{s_A^2}{s_B^2} = \frac{(67,233)^2}{(37,128)^2} = 3,279$$

a) $H_0: \sigma_A^2 = \sigma_B^2$

$H_a: \sigma_A^2 > \sigma_B^2$

O critério é: rejeito H_0 se $F_{obs} > C$, com

$$C = F(\alpha; n_A - 1; n_B - 1)$$

$$C = F(1\%; 20; 24)$$

$$C = 1,73$$

Como $F_{obs} > C \Rightarrow$ rejeito H_0
 e digo que $\sigma_A^2 > \sigma_B^2$

b) $H_0: \sigma_A^2 = \sigma_B^2$

$H_a: \sigma_A < \sigma_B^2$

O critério é: rejeito H_0 se

$F_{obs} < C$ com

$$C = F(1-\alpha; n_A - 1; n_B - 1)$$

$$C = F(99\%; 20; 24)$$

$$C = \frac{1}{F(1\%; 24, 20)} = \frac{1}{4,12} = 0,24$$

Como $F_{obs} > C$, não rejeito H_0 .

TESTES EMPARELHADOS

22

LISTA

No mês de dezembro, 9 pessoas vão para um spa. Elas são pesadas no dia de chegada e no dia de saída, obtendo-se o seguinte:

Pessoa	X Antes	Y Depois	$\Delta = X - Y$
1	76	98	-22
2	92	57	35
3	99	85	14
4	87	89	-2
5	65	60	5
6	87	53	34
7	86	52	34
8	85	93	-8
9	88	95	-7

Você pode concluir com 5% de significância que passar o mês no spa reduz o peso dos pacientes?

i) $H_0: \mu_{ANTES} = \mu_{DEPOIS}$ *imagina*

$H_a: \mu_{ANTES} > \mu_{DEPOIS}$

OU

$H_0: \mu_{\Delta} = 0$

$H_a: \mu_{\Delta} > 0$

ii) Teste de hipótese para μ com s^2 desconhecido

Rejeito H_0 se $\bar{\Delta} > C$ com

$$C = 0 + t(\alpha; n-1) \frac{s}{\sqrt{n}}$$

iii) No caso, para Δ :

$$\bar{\Delta} = 9,222 \quad s^2 = 449,194$$

$$s = 21,194$$

iv) Assim, $C = t_{5\%; 8} \frac{21,194}{\sqrt{9}}$

$$C = 1,86 \cdot \frac{21,194}{3} = 13,4$$

Como $\bar{\Delta}$ não é maior que C , não rejeito H_0 . Concluindo que o spa NÃO imagina.

Os salários horários médios de mecânicos de automóveis foram estudados por um grupo de consumidores. A finalidade era determinar a eventual existência de diferenças entre 4 localidades. Obteve-se:

Observação	Loc A	Loc B	Loc C	Loc D
1	6	12	11	9
2	9	11	8	7
3	9	10	12	10
4	6	8	9	10
5	5	9	10	9

Supondo que as variâncias sejam iguais, no nível de significância 5%, podemos afirmar que existe diferença?

Calculamos

	$\bar{x}$	s^2
A	7	3,5
B	10	2,5
C	10	2,5
D	9	1,5

Além disso:

$$\bar{x} = \frac{7 + 10 + 10 + 9}{4} = 9$$

$$s_p^2 = \frac{\sum (n_i - 1) s_i^2}{\sum (n_i - 1)} = \frac{4 \cdot 3,5 + 4 \cdot 2,5 + 4 \cdot 2,5 + 4 \cdot 1,5}{4 + 4 + 4 + 4}$$

$$s_p^2 = 2,5$$

$$s_E^2 = \frac{\sum [n_i (\bar{x}_i - \bar{x})^2]}{k - 1}$$

$$s_E^2 = \frac{5[(7-9)^2 + (10-9)^2 + (10-9)^2 + (9-9)^2]}{3}$$

$$s_E^2 = 10$$

$$F_{obs} = \frac{s_E^2}{s_p^2} = \frac{10}{2,5}$$

$$F_{obs} = 4$$

O teste é:

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D$$

H_a : existe uma diferente

Rejeitamos H_0 se $F_{obs} > C$
com

$$C = F(\alpha; k-1; \sum (n_i-1))$$

$$C = F(5\%; 3; 16)$$

$$C = 3,24$$

Como $F_{obs} = 4 > C$

Rejeitamos H_0 e dizemos

que as médias não são todas iguais $\Rightarrow$ há diferenças de remuneração média entre as cidades.