IMPACTA

Inferência Estatística

AC2: Comparações Envolvendo Médias

Prof. Gilberto Perez

AC2: Banco de Dados => **Salario Alpha IT Company 1 (Amostras de Salários Homem x Mulher)**

Como o N. de cada Amostra é >= 30 – Assumir Normalidade

1) Efetuar Estatística Descritiva dos Salários de Homen e Mulheres

2) Testar Hipótese de Normalidade de Salário de Homen e Mulheres (Sig = 0.05, ou 95% Confiabilidade).

3) Fazer a Análise de Intervalos de Média para a amostra completa. O que se pode dizer da Média de salários Homem x Mulher

4) Fazer a Análise de Intervalos de Média para amostras com N1, e N2 = 50 e com N1, e N2 = 100. O que se pode inferir?

AC2: Banco de Dados => **Vestibular 2022**

1) Efetuar Estatística Descritiva das Notas: Matemática, Linguas, Humanidades, Naturais, Português.

2) Separar notas por cursos, sendo:

Curso
3 = "Administracao"
5 = "Psicologia"
7 = "Ciencia da computacao"
8 = "Sistemas de Informacao"
9 = "Direito"

3) Testar a Normalidade dessas Notas.

Inferência Estatística em Comparações de Médias

Estimação da diferença entre médias de duas Populações: Amostras Independentes

Vamos começar a abordagem mostrando como a estimação da diferença entre as médias de duas populações pode ser desenvolvida, a partir de um estudos de amostras simples dessas populações.

Isso pode ser feito relativamente fácil, utilizando-se o Excel.

Vamos tomar como exemplo duas populações da Loja XPTO - Capital e Interior:

$\mu 1$ = média da população 1: (média da idade de todos os clientes que compram na Capital.

$\mu 2$ = média da população 2: (média da idade de todos os clientes que compram no Interior.

Estimação da diferença entre médias de duas Populações: Amostras Independentes

Como x1 é o ponto de estimação de μ1 e x2 é o ponto de estimação de μ2, o ponto de estimação da amostra é expressado como.

Ponto de Estimação da Diferença entre Médias da duas Populações:

$$\bar{X}_1 - \bar{X}_2 \qquad (1)$$

Estimação da diferença entre médias de duas Populações: Amostras Independentes

A diferença entre as médias das duas Populações é: **μ1 - μ2**

Para estimar μ1 - μ2, vamos selecionar uma amostra simples de n1 clientes da população 1, e uma amostra simples de n2 clientes da população 2.

Como a amostra simples de clientes n1 é selecionada independente da amostra simples de clientes n2, elas são chamadas: **Amostras Aleatórias Simples e Independentes.**

Amostra Aleatória Simples Clientes da Capital n1

$\overline{X}_1$ = média de idade da amostra de clientes da Capital

Amostra Aleatória Simples Clientes do Interior n2

$\overline{X}_2$= média de idade da amostra de clientes da Capital

Comparações Envolvendo Médias

População 1
Clientes de uma Loja da Capital da XPTO

μ1 = média da idade dos clientes da capital

População 2
Clientes de uma Loja do Interior da XPTO

μ2 = média da idade dos clientes do Interior

μ1 - μ2 = diferença entre a média de idade das **populações**

Amostra Aleatória Simples Clientes da Capital n1

$\overline{X}_1$ = média de idade da amostra de clientes da Capital

Amostra Aleatória Simples Clientes do Interior n2

$\overline{X}_2$= média de idade da amostra de clientes da Capital

$\overline{X}_1$ - $\overline{X}_2$ = Ponto de Estimação de μ1 - μ2 (1)

Distribuições de Amostragem

Distribuições da Amostragem de $\overline{X}_1$ - $\overline{X}_2$

Valor Esperado: E ($\overline{X}_1$ - $\overline{X}_2$) = μ1 - μ2 **(2)**

Desvio Padrão: $$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (3)$$

Com:

σ1 = Desvio padrão da população 1

σ2 = Desvio padrão da população 2

n1 = tamanho da amostra aleatória da população 1

n2 = tamanho da amostra aleatória da população 2

Distribuições de Amostragem

Distribuições da Amostragem de $\overline{X}_1 - \overline{X}_2$

Valor Esperado: $E(\overline{X}_1 - \overline{X}_2) = \mu 1 - \mu 2$ **(2)**

Desvio Padrão: $$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (3)$$

Com:

σ1 = Desvio padrão da população 1

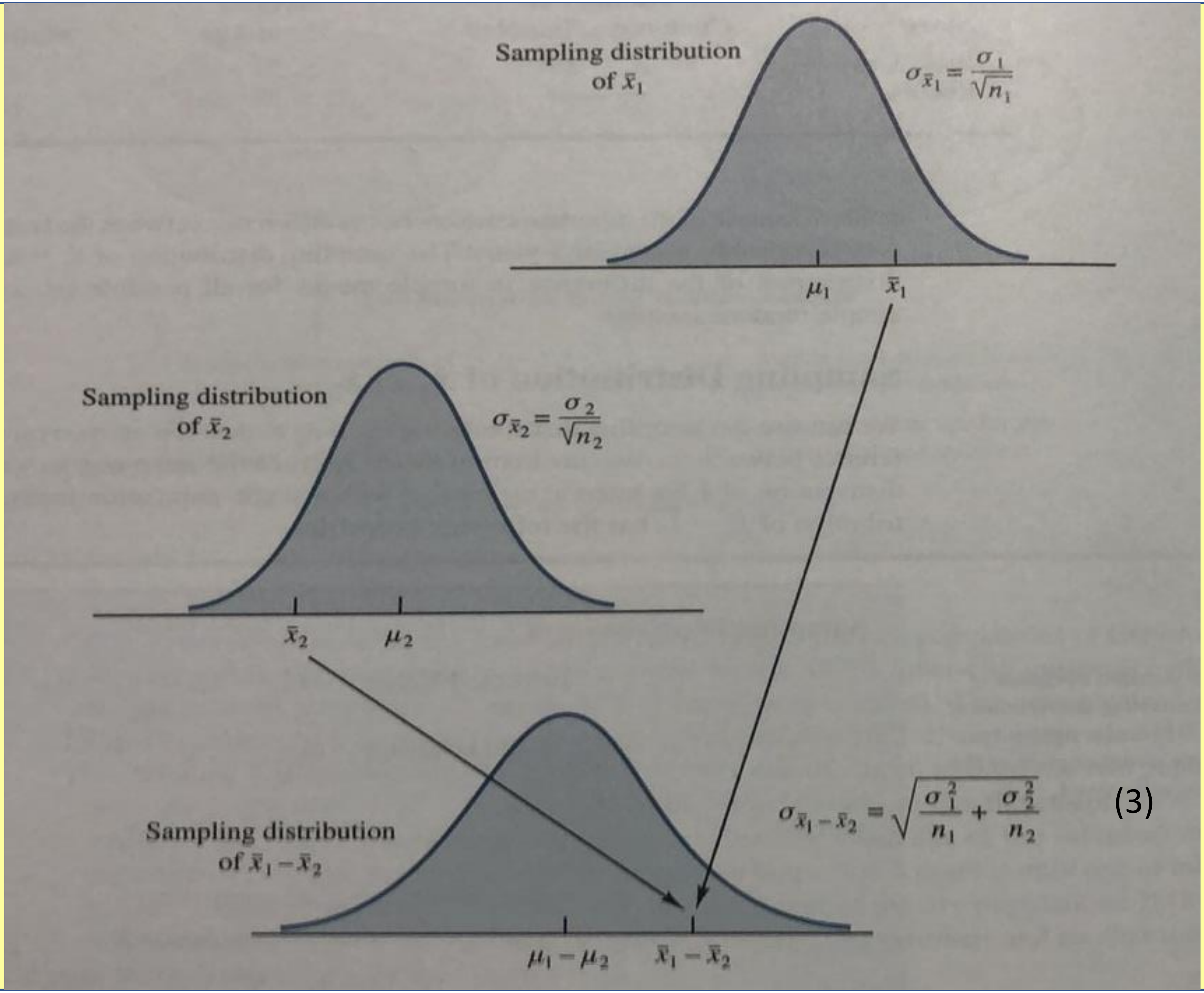
σ2 = Desvio padrão da população 2

n1 = tamanho da amostra aleatória da população 1

n2 = tamanho da amostra aleatória da população 2

Forma de Distribuição: Se o tamanho das amostras forem ambas grandes (n1 e n2 ≥ 30), a distribuição da amostra de $X_1 - X_2$ pode ser aproximada a uma distribuição de probabilidade normal.

Distribuições da Amostra $\overline{x_1} - \overline{x_2}$ em Relação às Distribuições das Amostras $\overline{x_1}$ e $\overline{x_2}$.



(3)

Estimação da diferença entre médias de duas Populações: Amostras Independentes

Em casos de amostras grandes (**n ≥ 30**) pode-se aplicar o teorema do limite central e a distribuição da amostra $\bar{x}1 - \bar{x}2$ pode ser aproximada pela distribuição normal de probabilidade.

Com essa aproximação pode-se usar a seguinte expressão para estimar o intervalo de diferença entre as médias de duas populações:

Interval Estimate of the Difference Between the Means of Two Populations: Large-Sample Case ($n_1 \geq 30$ and $n_2 \geq 30$) with σ_1 and σ_2 Assumed Known

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sigma_{\bar{x}_1-\bar{x}_2} \qquad (4)$$

Obs: No Excel $Z_{\alpha/2}$ = INV.NORMP(1- $\alpha/2$), sendo α é o Nível de Significância (sig.) estatística e **1- α** é o coeficiente de confiança. Se α = 0.05, o Coef. De Confiança é 95%

Estimação da diferença entre médias de duas Populações: Amostras Independentes

Note que na equação (3) é necessário conhecer os desvios padrões das Populações. Se os desvios padrões das populações são desconhecidos, um recurso é usar o desvio padrão das amostras para calcular o **ponto de estimação:**

Point Estimator of $\sigma_{\bar{x}_1-\bar{x}_2}$

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad (5)$$

Quando σ_1 e σ_2 são estimados por s_1 e s_2, podemos utilizar o seguinte intervalo de confiança para estimar a diferença entre a média das duas populações

Interval Estimate of the Difference Between the Means of Two Populations: Large-Sample Case ($n_1 \geq 30$ and $n_2 \geq 30$) with σ_1 and σ_2 Estimated by s_1 and s_2

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} s_{\bar{x}_1-\bar{x}_2} \qquad (6)$$

Obs: No Excel $Z_{\alpha/2}$ = INV.NORMP(1- $\alpha/2$), sendo α é o Nível de Significância (sig.) estatística e **1- α** é o coeficiente de confiança. Se α = 0.05, o Coef. De Confiança é 95%

Estimação da diferença entre médias de duas Populações: Amostras Independentes

Loja da XPTO	Amostra Simples	Média Idade Amostra	Desvio Padrão Amostra
Cidade	n_1=36	$\bar{x}_1$=40 anos	s_1=9 anos
Interior	n_2=49	$\bar{x}_2$=35 anos	s_2=10 anos

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{(9)^2}{36} + \frac{(10)^2}{49}} = 2.07 \qquad (5)$$

Com α=0.05 -> $z_{\alpha/2} = z_{.025}$ = 1,96 a expressão (6) resulta no Coef. de Confiança de 95%

$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} s_{\bar{x}_1-\bar{x}_2} \qquad (6)$$

$$5 \pm (1.96)(2.07)$$

$$5 \pm 4.06$$

Com o Coef. de Confiança de 95%, a margem de erro é de 4.06 anos e o intervalo de estimação de diferença das médias das populações é de 0.94 a 9.06 anos.