**Disciplina**: MAE0212: Introdução a Probabilidade e Estatística II
**Professora**: Silvia Ferrari
**Nome**: **Nota**:

Use os espaços em branco, inclusive o verso das páginas, para responder às questões. Mostre os desenvolvimentos, não apenas as respostas. **ATENÇÃO: Identifique quais suposições são necessárias para a resolução de cada item**.

## Provinha 3 - Gabarito

Para resolver as questões, você pode usar as informações abaixo. Se não encontrar exatamente o que precisa, pode dar o melhor resultado aproximado possível usando apenas o que está fornecido aqui. Lembre de justificar sua resposta!

1. Se $Z \sim N(0,1)$ então,

| | | | |
|---|---|---|---|
| $\mathbb{P}(Z \leq 3.03) = 0.999$ | $\mathbb{P}(Z \leq 1.96) = 0.975$ | $\mathbb{P}(Z \leq 1.64) = 0.95$ | $\mathbb{P}(Z \leq 2.60) = 0.995$ |
| $\mathbb{P}(Z \leq 1.28) = 0.90$ | $\mathbb{P}(Z \leq 2.33) = 0.99$ | $\mathbb{P}(Z \leq 1.67) = 0.953$ | $\mathbb{P}(Z \leq 1.47) = 0.93$ |

2. Se $T_\nu \sim t_{(\nu)}$, então

| | | | | |
|---|---|---|---|---|
| $\mathbb{P}(T_8 \leq 1.39) = 0.90$ | $\mathbb{P}(T_8 \leq 1.86) = 0.95$ | $\mathbb{P}(T_8 \leq 2.31) = 0.975$ | $\mathbb{P}(T_8 \leq 2.89) = 0.99$ | $\mathbb{P}(T_8 \leq 3.35) = 0.995$ |
| $\mathbb{P}(T_9 \leq 1.38) = 0.90$ | $\mathbb{P}(T_9 \leq 1.83) = 0.95$ | $\mathbb{P}(T_9 \leq 2.26) = 0.975$ | $\mathbb{P}(T_9 \leq 2.82) = 0.99$ | $\mathbb{P}(T_9 \leq 3.25) = 0.995$ |
| $\mathbb{P}(T_{10} \leq 1.37) = 0.90$ | $\mathbb{P}(T_{10} \leq 1.81) = 0.95$ | $\mathbb{P}(T_{10} \leq 2.23) = 0.975$ | $\mathbb{P}(T_{10} \leq 2.76) = 0.99$ | $\mathbb{P}(T_{10} \leq 3.17) = 0.995$ |
| $\mathbb{P}(T_{11} \leq 1.36) = 0.90$ | $\mathbb{P}(T_{11} \leq 1.79) = 0.95$ | $\mathbb{P}(T_{11} \leq 2.20) = 0.975$ | $\mathbb{P}(T_{11} \leq 2.71) = 0.99$ | $\mathbb{P}(T_{11} \leq 3.10) = 0.995$ |
| $\mathbb{P}(T_{35} \leq 1.31) = 0.90$ | $\mathbb{P}(T_{35} \leq 1.69) = 0.95$ | $\mathbb{P}(T_{35} \leq 2.03) = 0.975$ | $\mathbb{P}(T_{35} \leq 2.44) = 0.99$ | $\mathbb{P}(T_{35} \leq 2.72) = 0.995$ |
| $\mathbb{P}(T_{36} \leq 1.30) = 0.90$ | $\mathbb{P}(T_{36} \leq 1.69) = 0.95$ | $\mathbb{P}(T_{36} \leq 2.03) = 0.975$ | $\mathbb{P}(T_{36} \leq 2.43) = 0.99$ | $\mathbb{P}(T_{36} \leq 2.72) = 0.995$ |
| ... | | | | |
| $\mathbb{P}(T_{200} \leq 1.28) = 0.90$ | $\mathbb{P}(T_{200} \leq 1.65) = 0.95$ | $\mathbb{P}(T_{200} \leq 1.97) = 0.975$ | $\mathbb{P}(T_{200} \leq 2.34) = 0.99$ | $\mathbb{P}(T_{200} \leq 2.60) = 0.995$ |

3. Se $\chi^2_\nu \sim \chi^2_{(\nu)}$, então

| | | | | |
|---|---|---|---|---|
| $\mathbb{P}(\chi^2_9 \leq 14.68) = 0.90$ | $\mathbb{P}(\chi^2_9 \leq 16.92) = 0.95$ | $\mathbb{P}(\chi^2_9 \leq 19.02) = 0.975$ | $\mathbb{P}(\chi^2_9 \leq 21.66) = 0.99$ | $\mathbb{P}(\chi^2_9 \leq 23.59) = 0.995$ |
| $\mathbb{P}(\chi^2_{10} \leq 15.99) = 0.90$ | $\mathbb{P}(\chi^2_{10} \leq 18.31) = 0.95$ | $\mathbb{P}(\chi^2_{10} \leq 20.48) = 0.975$ | $\mathbb{P}(\chi^2_{10} \leq 23.21) = 0.99$ | $\mathbb{P}(\chi^2_{10} \leq 25.19) = 0.995$ |
| $\mathbb{P}(\chi^2_{35} \leq 46.06) = 0.90$ | $\mathbb{P}(\chi^2_{35} \leq 49.80) = 0.95$ | $\mathbb{P}(\chi^2_{35} \leq 53.20) = 0.975$ | $\mathbb{P}(\chi^2_{35} \leq 57.34) = 0.99$ | $\mathbb{P}(\chi^2_{35} \leq 60.27) = 0.995$ |
| $\mathbb{P}(\chi^2_{36} \leq 47.21) = 0.90$ | $\mathbb{P}(\chi^2_{36} \leq 51.00) = 0.95$ | $\mathbb{P}(\chi^2_{36} \leq 54.44) = 0.975$ | $\mathbb{P}(\chi^2_{36} \leq 58.62) = 0.99$ | $\mathbb{P}(\chi^2_{36} \leq 61.58) = 0.995$ |

4. Se $W_{p,q} \sim F_{p,q}$, então

| | | | | |
|---|---|---|---|---|
| $\mathbb{P}(W_{9,9} \leq 2.44) = 0.90$ | $\mathbb{P}(W_{9,9} \leq 3.18) = 0.95$ | $\mathbb{P}(W_{9,9} \leq 4.02) = 0.975$ | $\mathbb{P}(W_{9,9} \leq 5.35) = 0.99$ | $\mathbb{P}(W_{9,9} \leq 6.54) = 0.995$ |
| $\mathbb{P}(W_{9,14} \leq 2.12) = 0.90$ | $\mathbb{P}(W_{9,14} \leq 2.65) = 0.95$ | $\mathbb{P}(W_{9,14} \leq 3.21) = 0.975$ | $\mathbb{P}(W_{9,14} \leq 4.03) = 0.99$ | $\mathbb{P}(W_{9,14} \leq 4.72) = 0.995$ |
| $\mathbb{P}(W_{14,9} \leq 2.35) = 0.90$ | $\mathbb{P}(W_{14,9} \leq 3.03) = 0.95$ | $\mathbb{P}(W_{14,9} \leq 3.80) = 0.975$ | $\mathbb{P}(W_{14,9} \leq 5.01) = 0.99$ | $\mathbb{P}(W_{14,9} \leq 6.09) = 0.995$ |
| $\mathbb{P}(W_{10,15} \leq 2.06) = 0.90$ | $\mathbb{P}(W_{10,15} \leq 2.54) = 0.95$ | $\mathbb{P}(W_{10,15} \leq 3.06) = 0.975$ | $\mathbb{P}(W_{10,15} \leq 3.80) = 0.99$ | $\mathbb{P}(W_{10,15} \leq 4.42) = 0.995$ |
| $\mathbb{P}(W_{15,10} \leq 2.24) = 0.90$ | $\mathbb{P}(W_{15,10} \leq 2.86) = 0.95$ | $\mathbb{P}(W_{15,10} \leq 3.52) = 0.975$ | $\mathbb{P}(W_{15,10} \leq 4.56) = 0.99$ | $\mathbb{P}(W_{15,10} \leq 5.47) = 0.995$ |

Figura 1: Algumas informações sobre probabilidades envolvendo as distribuições normal, t-Student, $\chi^2$ e F.

1. A fim de diminuir o tempo médio em que um analgésico leva para penetrar na corrente sanguínea, um químico analista acrescentou certo componente à fórmula original, que acusava um tempo médio de 43 minutos. Em 36 observações com a nova fórmula, obteve um tempo médio amostral de 40 minutos e desvio padrão de 6 minutos.

   a) (1,0 ponto) Formule o problema como um teste de hipóteses. O que se pode concluir, no nível de 5% de significância, sobre a eficiência do novo componente?

   a) (0,5 ponto) Forneça um valor aproximado para o valor-p do teste.

   c) (1,0 ponto) Forneça um intervalo de 90% de confiança para a **redução** no tempo médio com que o analgésico penetra na corrente sanguínea após a introdução do novo componente à fórmula original.

   *Solução:*

   Considere a variável aleatória $X$ que representa o tempo que o analgésico (que tem a fórmula alterada) leva para penetrar na corrente sanguínea de um indivíduo selecionado ao acaso.
   **Suposições**:

   - $X \sim N(\mu, \sigma^2)$;
   - $X_1, \ldots, X_{36}$ é uma amostra aleatória simples de $X$.

   *Observe que não temos evidências a respeito do conhecimento da variância populacional.*

   a) *Passo 1*: Note que o interese é verificar se o tempo médio diminuiu. Dessa forma, consideremos as seguintes hipóteses:

   $$\mathcal{H}_0 : \mu = 43 \quad vs \quad \mathcal{H}_1 : \mu < 43.$$

   b) *Passo 2*: Considerando uma amostra aleatória simples de 36 indivíduos, temos que

   $$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}.$$

   E, sob $\mathcal{H}_0$, temos que

   $$T = \frac{\overline{X} - 43}{S/\sqrt{n}} \sim t_{(n-1)}.$$

   *Passo 3*: Fixe $\alpha = 5\%$ e queremos encontrar uma região crítica da forma $RC = \{t \leq t_c\}$, em que $t_c$ é tal que $\mathbb{P}(\text{erro I}) = 0.05$. Dessa forma, temos que

   $$\alpha = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \mu = 43) \Rightarrow \mathbb{P}(T \leq t_c) = 0.05 \Rightarrow t_c = -1.69,$$

   em que $T \sim t_{(35)}$. Assim, a região crítica é

   $$RC = \{t \leq -1.69\}.$$

*Passo 4*: O valor $t_{obs} = \sqrt{36}(40 - 43)/6 = -3$.

*Passo 5*: Como o valor $t_{obs}$ observado pertence à região crítica, temos evidências amostrais para rejeitar a hipótese nula, considerando $\alpha = 5\%$. Isto é, temos evidências amostrais para dizer que o tempo médio diminuiu.

b) Queremos obter o valor-p do teste. Note que, neste caso, o valor-p é definido como

$$\widehat{\alpha} = \mathbb{P}(T \leq t_{obs}) = \mathbb{P}(T \leq -3).$$

Pelas informações fornecidas, temos que $\mathbb{P}(T \leq -2.44) = 0.01$. Então $\mathbb{P}(T \leq -3) < 0.01$. Observe que o valor-p é bem pequeno ($\leq 0.01$), ou seja, há forte evidência amostral para se concluir que o tempo médio diminuiu.

c) Recorde que, nesta situação, o intervalo de confiança para $\mu$ com coeficiente de confiança $\gamma$ é da forma

$$\text{IC}(\mu; \gamma) = \left[\overline{X} - t_\gamma \frac{S}{\sqrt{n}}, \overline{X} + t_\gamma \frac{S}{\sqrt{n}}\right].$$

Assim, o intervalo de confiança observado para $\mu$, considerando $\gamma = 90\%$, é dado por

$$\text{IC}(\mu; \gamma) = \left[40 - 1.67 \frac{6}{\sqrt{36}}, 40 + 1.67 \frac{6}{\sqrt{36}}\right] = [38.33, 41.67].$$

Logo, um IC para a redução média, ou seja, para $43 - \mu$, é

$$\text{IC}(43 - \mu; \gamma) = [43 - 41.67; 43 - 38.33] = [1.33; 4,67].$$

2. (2,5 pontos) Uma das maneiras de medir o grau de satisfação dos empregados de uma mesma categoria quanto à política salarial é através da variabilidade de seus salários; menor variabilidade é considerada como maior coerência na política salarial. A fábrica A diz ser mais coerente na política salarial do que a fábrica B. Para verificar essa afirmação, sorteou-se uma amostra de 10 funcionários não especializados de A, e 15 de B, obtendo-se os desvios padrão $s_A$ = 1.0 sm (salário mínimo) e $s_B$ = 1.6 sm. Qual é a sua conclusão com base em um teste de hipóteses adequado, considerando um nível de significância de 5%?

*Solução:*

Considere as variáveis aleatórias $X_A$ que representa o salário (em número de salários mínimos) de um empregado selecionado aleatoriamente da fábrica A e $X_B$ representando o salário de um empregado selecionado aleatoriamente da fábrica B .

**Suposições**:

- $X_A \sim N(\mu_A, \sigma_A^2)$ e $X_B \sim N(\mu_B, \sigma_B^2)$;
- $X_{A_1}, \ldots, X_{A_{10}}$ é uma amostra aleatória simples de $X_A$ e $X_{B_1}, \ldots, X_{B_{15}}$ é uma amostra aleatória simples de $X_B$.
- As amostras são independentes.

*Observe que estamos no contexto de populações normais com amostras independentes.*

a) *Passo 1*: Note que o interese é verificar se a fábrica A é mais coerente na política salarial do que a fábrica B. Dessa forma, estamos interessados nas seguintes hipóteses:

$$\mathcal{H}_0 : \sigma_A^2 \geq \sigma_B^2 \quad vs \quad \mathcal{H}_1 : \sigma_A^2 < \sigma_B^2.$$

b) *Passo 2*: Neste caso,

$$\frac{S_B^2/\sigma_B^2}{S_A^2/\sigma_A^2} \sim F_{(15-1,10-1)}.$$

Se $\sigma_A^2 = \sigma_B^2$, temos que $W = S_B^2/S_A^2 \sim F_{(14,9)}$.

*Passo 3*: Fixe $\alpha = 5\%$ e queremos encontrar uma região crítica da forma $RC = \{W \geq f\}$, em que $f$ é tal que $\mathbb{P}(\text{erro I}) = 0.05$. Dessa forma, temos que

$$\alpha = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \sigma_A^2 = \sigma_B^2) \Rightarrow \mathbb{P}(W \geq f) = 0.05 \Rightarrow \mathbb{P}(W < f) = 0.95.$$

Dessa forma, $f = 2.64$. Assim, a região crítica é

$$RC = \{w : w \geq 3.03\}.$$

*Passo 4*: Considerando os dados da amostra, temos $w_0 = 1.6^2/1.0^2 = 2.56$.

*Passo 5*: Como o valor $w_0$ observado não pertence à região crítica, não temos evidências amostrais para rejeitar a hipótese nula, considerando $\alpha = 5\%$. Isto é, não temos evidências em favor da afirmação da fábrica A.

3. Um agricultor tem tido prejuízo considerável em sua produção de laranja devido a infestação por insetos. Decide, então, experimentar um novo tipo de inseticida conjecturando que este possa ser melhor do que aquele que usualmente utiliza, aumentando, assim, sua produção. O agricultor vaporiza 200 árvores com o novo produto e 200 com o produto que normalmente usa, obtendo os dados na tabela a seguir (em unidade adequada).

| | Inseticida novo | Inseticida usual |
|---|---|---|
| Produção média por árvore | 240 | 227 |
| Variância | 980 | 950 |

a) (1,5 ponto) Esses dados indicam evidência suficiente de que o inseticida novo é melhor? Considere $\alpha = 5\%$.

b) (1 ponto) Construa um intervalo de 90% de confiança para a diferença entre as produções médias sob o uso desses dois inseticidas.

*Solução:*

Considere as variáveis aleatórias $X_N$ que representa a produção (em unidade adequada) de uma árvore selecionada aleatoriamente após o uso do inseticida novo e $X_U$ representando a produção (em unidade adequada) de uma árvore selecionada aleatoriamente após o uso do inseticida usual. Observe que a variabilidade da produção, com base na amostra observada, não muda muito (as variâncias amostrais são próximas), indicando que é razoável que o uso dos inseticidas (novo e usual) não influencia na variabilidade.

**Suposições**:

- $X_N \sim N(\mu_N, \sigma^2)$ e $X_U \sim N(\mu_U, \sigma^2)$;
- $X_{N_1}, \ldots, X_{N_{200}}$ é uma amostra aleatória simples de $X_N$ e $X_{U_1}, \ldots, X_{U_{200}}$ é uma amostra aleatória simples de $X_U$.
- As amostras são independentes.

*Observe que estamos no contexto de populações normais com amostras independentes e variâncias iguais (desconhecidas).*

a) *Passo 1*: Note que o interese é verificar se o uso do inseticida novo faz a produção aumentar. Dessa forma, estamos interessados nas seguintes hipóteses:

$$\mathcal{H}_0 : \mu_N = \mu_U \quad vs \quad \mathcal{H}_1 : \mu_N > \mu_U.$$

b) *Passo 2*: Neste caso, a estatística de teste é

$$T = \frac{\overline{X}_N - \overline{X}_U}{S_p\sqrt{\dfrac{1}{n_N} + \dfrac{1}{n_U}}},$$

em que

$$S_p^2 = \frac{(n_N - 1)S_N^2 + (n_U - 1)S_U^2}{n_N + n_U - 2}.$$

E, sob $\mathcal{H}_0$, temos que $T \sim t_{(398)}$.

*Passo 3*: Fixe $\alpha = 5\%$ e queremos encontrar uma região crítica da forma $RC = \{T > t_c\}$, em que $t_c$ é tal que $\mathbb{P}(\text{erro I}) = 0.05$. Dessa forma, temos que

$$\alpha = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \mu_N = \mu_U) \Rightarrow \mathbb{P}(T > t_c) = 0.05.$$

em que $T \sim t_{(398)}$. Note que não temos na Tabela de valores referentes a 398 graus de liberdade. Por outro lado, notamos que para $\nu = 200$, os quantis da distribuição t-Student são bem próximos dos quantis da distribuição $N(0,1)$. Dessa forma, aproximamos o quantil $t_c$ pelo valor correspondente a normal padrão, ou seja, tomamos $t_c$ tal que

$$\mathbb{P}(Z \leq t_c) = 0.95 \Rightarrow t_c = 1.64.$$

Assim, a região crítica é

$$RC = \{t : t > 1.64\}.$$

*Passo 4*: Considerando os dados da amostra, temos

$$t_{obs} = \frac{240 - 227}{31.06\sqrt{\dfrac{1}{200} + \dfrac{1}{200}}} = \frac{13}{31.06/10} = 4.18.$$

*Passo 5*: Como o valor $t_{obs}$ observado pertence à região crítica, temos evidências amostrais para rejeitar a hipótese nula, considerando $\alpha = 5\%$. Isto é, temos evidências para dizer que o inseticida novo aumenta a produção.

Observação: Como os tamanhos das amostras são grandes, a estatística $T$ tem distribuição aproximadamente normal padrão, mesmo sem a suposição de normalidade de $X_N$ e $X_U$. Uma solução alternativa para o problema é omitir a suposição de normalidade e tratar $T$ como tendo distribuição aproximadamente normal. Nesse caso, seria obtido também $t_c$ = 1.64.

b) Seja $\Delta = \mu_N - \mu_U$, isto é, a diferença entre as produções médias sob o uso desses dois inseticidas. O intervalo de confiança para $\Delta$ com coeficiente de confiança $\gamma$ é dado por

$$\text{IC}(\Delta;\gamma) = \left[(\overline{X}_N - \overline{X}_U) - t_\gamma S_p\sqrt{\frac{1}{n_N} + \frac{1}{n_U}}, (\overline{X}_N - \overline{X}_U) + t_\gamma S_p\sqrt{\frac{1}{n_N} + \frac{1}{n_U}}\right].$$

Assim, o intervalo de confiança observado para o novo tempo médio, considerando $\gamma = 90\%$, é dado por

$$\text{IC}(\Delta; 90\%) = \left[13 - 1.64\frac{31.06}{10}, 13 - 1.64\frac{31.06}{10}\right] = [7.91, 18.10].$$

4. (2,5 pontos) Uma grande rede de lojas deseja avaliar o efeito de uma campanha promocional. A tabela a seguir apresenta o volume de vendas para cada uma de 10 lojas selecionadas ao acaso, antes e depois da campanha. Teste se houve **diferença** no volume médio de vendas antes e depois da campanha com base em um teste de hipóteses adequado. Use um nível de significância de 1%.

| Loja ($i$) | Depois da promoção ($x_i$) | Antes da promoção ($y_i$) |
|---|---|---|
| 1 | 53 | 45 |
| 2 | 30 | 27 |
| 3 | 25 | 30 |
| 4 | 82 | 72 |
| 5 | 28 | 19 |
| 6 | 65 | 67 |
| 7 | 58 | 51 |
| 8 | 37 | 38 |
| 9 | 31 | 33 |
| 10 | 43 | 40 |
| Total | 452 | 422 |

Se necessário, use $\sum_{i=1}^{10} x_i = 452$, $\sum_{i=1}^{10} y_i = 422$ e $\sum_{i=1}^{10} (x_i - y_i)^2 = 346$.

*Solução:*

Considere a variável aleatória $X$ que representa a quantidade de vendas de uma loja selecionada aleatoriamente antes de uma promoção e $Y$ que representa a quantidade de vendas desta mesma loja depois de uma promoção. Seja $(X_1, Y_1), \ldots, (X_{10}, Y_{10})$ uma amostra aleatória dos volumes de venda depois e antes da campanha.

**Suposições**:

- Não há dependência entre as lojas do estudo com relação à quantidade de vendas;
- A variável aleatória $D = X - Y \sim N(\mu_D, \sigma_D^2)$.

*Observe que temos duas medidas para uma mesma loja ⇒ amostras dependentes.*

a) *Passo 1*: Note que o interese é verificar se há diferença no volume de vendas antes e depois da promoção. Dessa forma, consideremos as seguintes hipóteses:

$$\mathcal{H}_0 : \mu_D = 0 \quad vs \quad \mathcal{H}_1 : \mu_D \neq 0.$$

b) *Passo 2*: Considerando uma amostra aleatória simples de 10 indivíduos, temos que

$$\frac{\overline{D} - \mu_D}{S_D/\sqrt{10}} \sim t_{(10-1)},$$

em que $\overline{D} = \sum_{i=1}^{n} D_i/n$, com $D_i = X_i - Y_i$. Sob $\mathcal{H}_0$, temos que

$$T = \frac{\overline{D}}{S_D/\sqrt{10}} \sim t_{(10-1)}.$$

Note que, agora, estamos trabalhando com a amostra $D_1, \ldots, D_{10}$.

*Passo 3*: Fixe $\alpha = 1\%$ e queremos encontrar uma região crítica da forma $RC = (-\infty, -t_c) \cup (t_c, \infty)$, em que $t_c$ é tal que $\mathbb{P}(\text{erro I}) = 0.01$. Dessa forma, temos que

$$\alpha = \mathbb{P}(\text{rejeitar } \mathcal{H}_0 \mid \mu_D = 0) \Rightarrow \mathbb{P}(|T| > t_c) = 0.01 \Rightarrow \mathbb{P}(|T| < t_c) = 1 - 0.01,$$

$$\Rightarrow t_c = 3.25,$$

em que $T \sim t_{(9)}$. Assim, a região crítica é

$$RC = (-\infty, -3.25) \cup (t_c, 3.25).$$

*Passo 4*: O valor

$$t_{obs} = \frac{3}{5.33/\sqrt{10}} = 1.778.$$

*Passo 5*: Como o valor $t_{obs}$ observado não pertence à região crítica, não temos evidências amostrais para rejeitar a hipótese nula, considerando $\alpha = 1\%$. Isto é, não temos evidências amostrais para dizer que a produção da loja mudou com a campanha.