

# Curso de Especialização em *Data Science* e Estatística Aplicada

## Módulo II - Análise estatística de várias populações

### Atividade Avaliativa

Profa. Dra. Tatiane F. N. Melo

28/09/2024 até 25/10/2024 às 23h59

### Instruções

- O desenvolvimento desta atividade deve ser realizada de forma individual.
- Deve-se completar o arquivo Rmd enviado na atividade.
- A entrega da atividade deve ser realizada em Rmd e um arquivo no formato pdf, gerado a partir do RMarkdown.
- É preciso manipular o conjunto de dados, a exclusão dos *NA* 's deve ser realizada considerando apenas dados faltantes nas variáveis de interesse;
- Valor da atividade: 10 pontos.

**Data máxima para entrega: 25/10/2024**

### Conjunto de dados - Base de dados do Sistema de Informações sobre Nascidos Vivos (SINASC) 2023

Para a atividade avaliativa o conjunto de dados que será analisado é a base de dados do Sistema de Informações sobre Nascidos Vivos (SINASC) para o município de São Paulo, referente ao ano de 2023. Os dados foram obtidos no site Secretaria Municipal da Saúde da cidade de São Paulo: [https://www.prefeitura.sp.gov.br/cidade/secretarias/saude/epidemiologia\\_e\\_informacao/nascidos\\_vivos/index.php?p=306422](https://www.prefeitura.sp.gov.br/cidade/secretarias/saude/epidemiologia_e_informacao/nascidos_vivos/index.php?p=306422) (acesso e download em 23/09/2024). O conjunto de dados e seu respectivo dicionário foram postados no MOODLE Ipê.

**Questão 1:** A pontuação Apgar é uma avaliação rápida da condição do recém-nascido, realizada nos primeiro e quinto minuto de vida. As variáveis analisadas serão APGAR1 e APGAR5, que representam, respectivamente, as pontuações de Apgar nesses momentos. Verifique se houve uma mudança significativa na condição dos recém-nascidos entre o 1º e o 5º minuto após o nascimento. Em outras palavras, teste as hipóteses:

- $H_0$  : Não há diferença significativa entre as pontuações de Apgar no 1º e no 5º minuto, ou seja:

$$H_0 : \mu_1 = \mu_2,$$

onde

- $\mu_1$  é a média das pontuações de Apgar no 1º minuto após o nascimento;
- $\mu_2$  é a média das pontuações de Apgar no 5º minuto após o nascimento.

- $H_1$  : Há uma diferença significativa entre as pontuações de Apgar no 1º e no 5º minuto, ou seja,

$$H_1 : \mu_1 \neq \mu_2.$$

Como estamos comparando as condições de um mesmo recém-nascido em dois momentos, os dados são considerados pareados.

Se for observada uma diferença significativa entre as pontuações APGAR1 e APGAR5, realize um teste t unilateral adequado para verificar se a pontuação de Apgar no 1º minuto é significativamente menor (ou maior) do que no 5º minuto. Interprete todos os resultados com um nível de significância de  $\alpha = 1\%$ .

**Questão 2:** Verifique se há uma diferença significativa no peso médio ao nascer entre recém-nascidos do sexo masculino e feminino. Os dados disponíveis incluem informações sobre o peso ao nascer (em gramas) e o sexo dos recém-nascidos. As variáveis consideradas para a análise são:

- PESO: peso ao nascer (em gramas).
- SEXO: sexo dos recém-nascidos (masculino ou feminino)

O objetivo é testar a hipótese de que não há diferença no peso médio ao nascer entre os dois grupos.

Hipóteses de interesse:

- $H_0$ : Não há diferença significativa entre os pesos médios ao nascer de recém-nascidos do sexo masculino e feminino, ou seja, as médias são iguais.
- $H_1$ : Existe uma diferença significativa entre os pesos médios ao nascer de recém-nascidos do sexo masculino e feminino.

Ou seja,

$$H_0 : \mu_M = \mu_F \text{ contra } H_1 : \mu_M \neq \mu_F,$$

onde  $\mu_M$  é o peso médio dos recém-nascidos do sexo masculino e  $\mu_F$  é o peso médio dos recém-nascidos do sexo feminino.

Caso seja observada uma diferença significativa entre os pesos médios, realize um teste  $t$  unilateral adequado para verificar se o peso médio entre meninos é significativamente maior ou menor do que entre as meninas. Todos os resultados deverão ser interpretados ao nível de significância de  $\alpha = 1\%$ .

**Questão 3:** Usando os dados do SINASC, deseja-se verificar se existe uma associação entre a duração da gestação (GESTACAO) e o tipo de gravidez (GRAVIDEZ). A variável GESTACAO indica o período gestacional em semanas, enquanto GRAVIDEZ refere-se ao tipo de gravidez, como única, dupla, ou tripla. Para isso, será realizado um teste de independência entre essas duas variáveis categóricas. Para simplificar a análise, algumas categorias das variáveis deverão ser agrupadas: na variável GRAVIDEZ, as categorias “dupla” e “tripla ou mais” serão unidas e formarão a nova categoria “dupla ou mais”; na variável GESTACAO, as categorias “menos de 22 semanas” e “de 22 a 27 semanas” serão agrupadas e formarão uma nova categoria chamada “menos de 27 semanas”.

O objetivo é testar se a duração da gestação está associada ao tipo de gravidez.

Hipóteses de interesse:

- $H_0$ : As variáveis GESTACAO (semanas de gestação) e GRAVIDEZ (tipo de gravidez) são independentes, ou seja, não há associação entre o período gestacional e o tipo de gravidez.
- $H_1$ : As variáveis GESTACAO e GRAVIDEZ não são independentes, ou seja, há uma associação entre o período gestacional e o tipo de gravidez.

Interprete detalhadamente todos os resultados obtidos. Considere um nível de significância de  $\alpha = 1\%$ .

**Questão 4:** Verifique se há uma diferença estatisticamente significativa entre os pesos médios ao nascer (em gramas) dos recém-nascidos, de acordo com a raça/cor das mães. As variáveis utilizadas são: PESO, representando o peso ao nascer, e RACACORMAE, cujos níveis são: 1 – Branca, 2 – Preta, 3 – Amarela, 4 – Parda e 5 – Indígena.

Para essa análise, aplique a Análise de Variância (ANOVA) de um fator (adequada), verificando se o peso médio dos recém-nascidos difere significativamente entre os grupos de raça/cor das mães.

Caso a ANOVA aponte uma diferença significativa entre os grupos, realize um teste de comparações múltiplas adequado. Utilize o teste de Tukey se as variâncias forem homogêneas, ou o teste de Games-Howell caso as variâncias não sejam homogêneas, para identificar quais grupos raciais apresentam diferenças significativas entre os pesos médios dos recém-nascidos.

Interprete detalhadamente todos os resultados obtidos. Considere um nível de significância de  $\alpha = 1\%$ .