

ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS

TAREFA 06

Bernardo Chrispim Baron

Exercício 1

1. -

2. Acerca do tempo de gestação, a aplicação do teste T para a média de dias de gestação de toda a amostra, tendo como hipótese nula (H_0) uma média $\mu = 280$, não permite rejeitar a hipótese nula com um nível de significância de 5%, em favor de uma hipótese alternativa em que $\mu \neq 280$ (p-valor = 0,13). A amostra corrobora, portanto, os estudos prévios segundo os quais o tempo médio de gestação em uma população é de 280 dias.

Para as gestantes fumantes, por sua vez, o teste T mostra traz resultados altamente significativos para rejeição de H_0 ($\mu = 280$), em favor de uma hipótese alternativa segundo a qual $\mu < 280$ (p-valor = 0,0187). Ou seja, a pesquisa favorece a conclusão de que o hábito de fumar durante a gravidez reduz o tempo médio de gestação.

A comparação da fração da amostra que fumou durante a gravidez com aquela que não fumou indica resultados semelhantes. Como a distribuição das frequências de observações dos tempos de gravidez é semelhante entre os dois grupos, mas não normal (teste de Shapiro-Wilk: p-valor < 0,001), optou-se por uma comparação utilizando o teste de Wilcoxon para igualdade de centralidade. Nesse caso, tendo por hipótese nula $\mu_{\text{Fumantes}} = \mu_{\text{Não fumantes}}$, os resultados são extremamente significativos (p-valor = 0,0005) no sentido de rejeitar a hipótese nula e acatar uma hipótese alternativa segundo a qual a média do tempo de gestação de mulheres que fumam é menor que a de mulheres que não fumam ($\mu_{\text{Fumantes}} < \mu_{\text{Não fumantes}}$).

3. A comparação das médias das massas dos recém-nascidos entre mães que fumaram durante a gravidez e mães que não fumaram pode ser feita pelo teste de Wilcoxon para igualdade da centralidade, já que as duas distribuições de frequências são semelhantes, mas não são necessariamente normais (para os filhos de grávidas que fumaram, o teste de Shapiro-Wilk retorna p-valor < 0.001, indicando não normalidade). Dada uma hipótese nula de que o peso dos recém-nascidos cujas mães fumaram durante a gestação é igual ao daqueles cujas mães não fumaram ($\mu_{\text{Fumantes}} = \mu_{\text{Não fumantes}}$), o teste de Wilcoxon indica um p-Valor ínfimo ($2,2 \times 10^{-16}$), corroborando a rejeição dessa hipótese em

favor de uma hipótese alternativa segundo a qual o peso dos recém-nascidos cujas mães fumaram durante a gestação é menor que o daqueles cujas mães não fumaram ($\mu_{\text{Fumantes}} < \mu_{\text{Não fumantes}}$).

Exercício 2

1. Dada uma hipótese nula segundo a qual a mediana do salário dos CEOs no universo considerado é igual a US\$ 220.000,00 ($m = 22$, em dezenas de milhares de dólares estadunidenses), o teste de sinais para a mediana permite rejeitar com nível de significância de 1% ($p\text{-valor} = 0,0085$) a hipótese nula em favor da hipótese alternativa de que a mediana é superior aos mencionados US\$ 220.000,00 ($m > 22$), corroborando a hipótese da pesquisa.
2. Aplicando-se uma transformação logarítmica sobre as observações, a distribuição das frequências adquire aspecto simétrico. É o caso, portanto, de utilizar o teste Wilcoxon para mediana, com hipótese nula de que $m = \log(22)$. Também nesse caso, é possível rejeitar com extrema significância a hipótese nula em favor da hipótese alternativa de que $m > \log(22)$, com $p\text{-valor}$ ainda menor ($p\text{-valor} = 0,0041$).

Exercício 3

Assumindo como hipótese nula que a média do desgaste dos sapatos é idêntica entre os materiais A e B ($\mu_A = \mu_B$) e considerando que ambas as distribuições são normais (teste de Shapiro-Wilk para a diferença dos valores observados com $p\text{-valor}$ igual a 0,80) e dependentes, é possível utilizar o teste T para comparação de médias de amostras pareadas. Por esse teste, é possível rechaçar com nível de significância de 1% a hipótese nula, em favor da hipótese alternativa de que o material A sofre, em média, menos desgaste do que o material B ($\mu_A < \mu_B$; $p\text{-valor} = 0,0042$). Portanto, segundo esse critério, o material A é superior ao material B.