

Curso: MCDE / Unidade Curricular: Fundamentos Estatísticos
Profs. Sandra Nunes & Sandra Oliveira
Atividade 7

Testes de Hipóteses

Data: 29/11/2023

1. Uma instituição do ensino superior realizou um inquérito com o objetivo de analisar algumas características dos estudantes que concluíram o 1º ciclo. Um dos problemas consistia em averiguar se em média existem diferenças nos salários oferecidos (u.m.) para o primeiro trabalho após a conclusão, relativamente ao curso de licenciatura. Foram analisados os estudantes de 4 licenciaturas (A, B, C, D). Os resultados obtidos encontram-se na tabela seguinte.

Considere um nível de significância igual a 0,01.

Teste <i>de Levene</i>	Teste Shapiro-Wilk	Teste ANOVA
Variável: “Salário (u.m.)” Variável de agrupamento: “Curso” $F = 3,643$ $p - value = 0,017$	Variável: “Salário (u.m.)” <div> Curso A $df = 14$ $Estatística = 0,887$ $p - value = 0,074$ </div> <hr/> <div> Curso B $df = 18$ $Estatística = 0,914$ $p - value = 0,101$ </div> <hr/> <div> Curso C $df = 18$ $Estatística = 0,958$ $p - value = 0,564$ </div> <hr/> <div> Curso D $df = 15$ $Estatística = 0,895$ $p - value = 0,081$ </div>	Variável: “Salário (u.m.)” Variável de agrupamento: “Curso” $F = 171,19$ $p - value = 0,000$

a. Escreva as hipóteses do teste realizado para responder à questão colocada?

As hipóteses são:

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

$H_1: \exists_{i \neq j} \mu_i \neq \mu_j \text{ com } i, j = A, B, C, D$

Onde μ_A representa o salário médio populacional estudantes do curso A, μ_B representa o salário médio populacional estudantes do curso B, μ_C representa o salário médio populacional estudantes do curso C e μ_D representa o salário médio populacional estudantes do curso D.

b. Os pressupostos do teste foram validados? Justifique convenientemente.

A variável em estudo, o salário, é quantitativa contínua.

Analisando o resultado do teste de Shapiro-Wilk constatamos que para as quatro amostras o p-value (curso A: 0,074; curso B: 0,101; curso C: 0,564 e cursos D: 0,081) é superior ao nível de significância (0,01), concluindo-se que os dados têm distribuição normal. Está validado o pressuposto da normalidade.

No que respeita ao teste de Levene obteve-se um p-value igual a 0,017 > 0,01, não se rejeita H_0 e conclui-se que as variâncias das quatro amostras são homogêneas, isto é, está validado o pressuposto da homogeneidade das variâncias.

c. O que pode concluir relativamente aos salários auferidos pelos estudantes das diferentes licenciaturas?

Analisando o resultado do teste ANOVA constatamos que se obteve um p-value igual a $0 < 0,01$, pelo que se rejeita H_0 , concluindo-se que existem evidências estatísticas que apontam para a existência de pelo menos um curso onde o salário médio é significativamente diferente dos restantes.

d. Perante o resultado obtido que tipo de testes seria importante realizar?

Uma vez que se rejeita H_0 no teste ANOVA, apenas conseguimos afirmar que existe de pelo menos um curso onde o salário médio é significativamente diferente dos restantes, não conseguindo precisar onde residem as diferenças estatisticamente significativas.

Neste tipo de situação devem realizar os testes de comparações múltiplas, de que são exemplo o teste de Tukey e o teste de Scheffe, os quais permitem saber onde estão as diferenças detetadas no teste ANOVA.

e. Se tivesse optado por uma alternativa não paramétrica, qual o teste que escolheria?

A alternativa não paramétrica ao teste ANOVA é o teste de Kruskal-Wallis.

2. Pretende-se avaliar se a distribuição do consumo doméstico de energia elétrica por habitante nas regiões A e B é semelhante. Recolheram-se duas amostras independentes referentes a consumos anuais por habitante (milhares de Kwh) e os resultados que se obtiveram são os que se apresentam na tabela seguinte.

Considere um nível de significância igual a 0,05.

Teste Mann-Whitney
Variável: "Consumo de energia (Kwh)" Região A: $n = 10$ Região B: $n = 18$ $Estatística = 27,0$ $p - value = 0,274$

- a. Escreva as hipóteses do teste realizado para responder à questão colocada?

As hipóteses são:

H0: as duas amostras são provenientes de populações com a mesma distribuição

H1: as duas amostras são provenientes de populações com distribuições distintas

- b. O que pode concluir relativamente ao consumo de energia nas duas regiões analisadas?

Analisando o resultado do teste de Mann-Whitney constatamos que se obteve um p-value igual a $0,274 > 0,05$, pelo que não se rejeita H0, concluindo-se que existem evidências estatísticas que permitem afirmar que as amostras são provenientes de populações com distribuição idêntica, ou seja, não existem diferenças significativas no consumo de energia entre as duas regiões.

- c. Na sua opinião qual a razão da utilização de um teste não paramétrico?

Como as amostras têm dimensão muito pequena (inferior a 30) e uma vez que nada se sabe sobre a normalidade dos dados, faz sentido a opção pelo teste não paramétrico.

- d. Se pudesse utilizar um teste paramétrico, qual escolheria?

Escolheria o teste t para duas amostras independentes.

3. Um conjunto de investigadores realizou um estudo para comparar o número médio de crianças por mulher em países da OCDE e países de África. Os resultados obtidos são apresentados de seguida.

Considere um nível de significância igual a 0,05.

Teste de Levene	Teste Shapiro-Wilk	Teste t
Variável: “Nº médio de crianças por mulher – índice de fertilidade” Variável de agrupamento: “Grupo de países” $F = 18,19$ $p - value = 0,000$	Variável: “Nº médio de crianças por mulher – índice de fertilidade” Países da OCDE $df = 21$ $Estatística = 0,947$ $p - value = 0,296$ <hr/> Países de África $df = 19$ $Estatística = 0,936$ $p - value = 0,227$	Variável: “Nº médio de crianças por mulher – índice de fertilidade” Variável de agrupamento: “Grupo de países” $t = -16,32$ $p - value = 0,000$ <hr/> 95% Confidence Interval for the Difference = $[-4,89 ; -3,78]$

- a. Escreva as hipóteses do teste realizado para responder à questão colocada?

Foi realizado um teste t para duas amostras independentes cujas hipóteses são:

$$H_0: \mu_O = \mu_A \Leftrightarrow \mu_O - \mu_A = 0$$

$$H_1: \mu_O \neq \mu_A \Leftrightarrow \mu_O - \mu_A \neq 0$$

Onde μ_O representa o índice de fertilidade médio populacional dos países da OCDE e μ_A representa o índice de fertilidade médio populacional dos países da África.

- b. Os pressupostos do teste foram validados? Justifique convenientemente.

A variável em estudo, o índice de fertilidade: nº médio de crianças por mulher, é quantitativa contínua.

Analisando o resultado do teste de Shapiro-Wilk constatamos que para as duas amostras o p -value (países da OCDE: 0,296 e países de África: 0,227) é superior ao nível de significância (0,05), concluindo-se que os dados têm distribuição normal. Está validado o pressuposto da normalidade.

No que respeita ao teste de Levene obteve-se um p-value igual a $0 < 0,05$, rejeita-se H_0 e conclui-se que as variâncias das duas amostras não são homogêneas, isto é, o pressuposto da homogeneidade das variâncias não foi validado.

c. O que pode concluir relativamente ao índice de fertilidade nos dois grupos de países?

Analisando o resultado do teste t constatamos que se obteve um p-value igual a $0 < 0,05$, pelo que se rejeita H_0 , concluindo-se que existem evidências estatísticas que permitem afirmar existem diferenças significativas no índice de fertilidade médio entre os dois grupos de países, OCDE vs África.

d. Nesta situação deveria optar por uma alternativa não paramétrica? Se sim, qual?

Uma vez que o pressuposto da homogeneidade das variâncias não foi validado, fará todo o sentido optar por uma alternativa não paramétrica, neste caso seria o teste de Mann-Whitney.