

**Curso: MCDE / Unidade Curricular: Fundamentos Estatísticos**  
**Profs. Sandra Nunes & Sandra Oliveira**  
**Atividade 5**

**Testes de Hipóteses**

**Data: 29/11/2023**

1. Foi realizado um estudo relativo a uma cadeia de supermercados num país da Europa. As variáveis presentes na análise são as seguintes:

**Valor das vendas de uma loja da cadeia de supermercados (u.m.);**

**Despesa total na loja (u.m.);**

**Número de funcionários da loja;**

**Localização da loja, onde o valor 0 representa uma cidade do litoral e 1 uma cidade do interior;**

**Zona de implementação da loja, onde 1 representa loja de bairro, 2 loja isolada e 3 loja associada a um centro comercial.**

Com os dados observados foram realizados alguns testes de hipóteses, cujos resultados são apresentados de seguida. **Considere um nível de significância igual a 0,05.**

Teste $t$ – uma amostra	Teste Shapiro-Wilk	Teste $t$ – duas amostras independentes	Teste de Kruskal-Wallis
<b>Variável:</b> “Valor das vendas” Valor de referência: 45000 u.m. $t = -3,16$ $p - value = 0,002$ Mean Difference = -3804,28	<b>Variável:</b> “Valor das vendas” em cidade do litoral $n = 230$ $Estatística = 0,205$ $p - value = 0,000$ <hr/> <b>Variável:</b> “Valor das vendas” em cidade do interior $n = 167$ $Estatística = 0,259$ $p - value = 0,000$	<b>Variável:</b> “Valor das vendas”  <b>Variável de agrupamento:</b> “Localização da loja”  Teste de Levene $F = 116,16$ $p - value = 0,000$  Teste $t$ $t = 8,88$ $p - value = 0,000$  Mean Difference = 19819,41  95% Confidence Interval for the Difference = [15433,04 ; 24205,77]	<b>Variável:</b> “Valor das vendas”  <b>Variável de agrupamento:</b> “Zona de implementação” Loja de Bairro $n = 302$ Loja isolada $n = 15$ Loja associada a um grande centro comercial $n = 80$  $Estatística = 215,78$ $p - value = 0,000$

- a. Realizou-se um teste t-student para uma amostra para averiguar se, em média, o valor das vendas é igual a 45000 u.m.? O que pode concluir?

As hipóteses do teste t para uma amostra são:

$$H_0: \mu = 45000 \Leftrightarrow \mu - 45000 = 0$$

$$H_1: \mu \neq 45000 \Leftrightarrow \mu - 45000 \neq 0$$

Onde  $\mu$  representa o valor médio das vendas.

Analisando o resultado do teste t constatamos que se obteve um p-value igual a  $0,002 < 0,05$ , pelo que se rejeita  $H_0$ , concluindo-se que existem evidências estatísticas que apontam para a existência de diferenças significativas entre o valor médio das vendas e o valor de referência de 45000.

- b. Realizou-se também um teste de hipóteses t-student para averiguar se em média o valor das vendas de uma loja localizada numa cidade do litoral é diferente de outra que se encontra localizada numa cidade do interior.

- i. Indique as hipóteses adequadas ao problema

As hipóteses são:

$$H_0: \mu_L = \mu_I \Leftrightarrow \mu_L - \mu_I = 0$$

$$H_1: \mu_L \neq \mu_I \Leftrightarrow \mu_L - \mu_I \neq 0$$

Onde  $\mu_L$  representa o valor médio populacional das vendas das lojas do Litoral e  $\mu_I$  representa o valor médio populacional das vendas das lojas do Interior.

- ii. Verifique se os pressupostos do teste de hipóteses utilizado estão a ser verificados? Justifique.

Analisando o resultado do teste à normalidade (Teste de Shapiro-Wilk), cujas hipóteses são:

$H_0$ : Os dados têm distribuição normal

$H_1$ : Os dados não têm distribuição normal

Constatamos que para ambas as amostras o valor do p-value é igual a  $0 < 0,05$ , rejeitando-se  $H_0$  e concluindo-se que os dados não têm distribuição normal. No entanto também constatamos que ambas as amostras têm dimensão superior a 30 (230 observações para o Litoral e 167 para o Interior), pelo que o teorema do limite central garante a normalidade aproximada, permitindo a validação do pressuposto da normalidade.

No que respeita ao pressuposto da homogeneidade das variâncias, o teste de Levene, cujas hipóteses são:

$$H_0: \sigma_L^2 = \sigma_I^2$$

$$H_1: \sigma_L^2 \neq \sigma_I^2$$

(onde  $\sigma_L^2$  representa a variância populacional das vendas das lojas do Litoral e  $\sigma_I^2$  representa a variância populacional das vendas das lojas do Interior)

Produziu um p-value igual a zero  $< 0,05$ , rejeitando-se  $H_0$  e concluindo que existem diferenças significativas entre a variâncias dos dois grupos, ou seja, o pressuposto da homogeneidade das variâncias não foi validado.

iii. O que pode concluir relativamente ao valor médio das vendas face à localização?

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 apontam para a existência de diferenças significativas entre o valor médio das vendas face à localização da loja.

iv. Interprete o intervalo de confiança.

O intervalo de confiança a 95% para a diferença de médias é o seguinte:

[15433,04 ; 24205,77]

Observe-se que o zero não pertence a este intervalo, quer isto dizer que existem evidências estatísticas para afirmar que existem diferenças significativas entre o valor médio das vendas das lojas do Litoral face às Lojas do Interior.

v. Em alternativa ao teste utilizado, que outro teste de hipóteses poderia ter sido realizado?

Dado que o pressuposto da homogeneidade de variâncias não foi validado, a alternativa terá de ser não paramétrica, pelo que escolheríamos o teste de Mann-Whitney, cujas hipóteses são:

$H_0$ : as amostras são provenientes de populações com distribuições idênticas

$H_1$ : as amostras são provenientes de populações com distribuições distintas

c. Tendo em conta os resultados obtidos, verifique se existem diferenças no valor das vendas da loja relativamente à zona de implementação.

O teste utilizado para dar resposta a esta questão foi o teste não paramétrico de Kruskal-Wallis, cujas hipóteses são:

$H_0$ : as amostras são provenientes de populações com distribuições idênticas

$H_1$ : existe pelo menos uma amostra que é proveniente de uma população com distribuição distinta das restantes

O valor do p-value foi de  $0 < 0,05$ , rejeitando-se  $H_0$  e concluindo que existe pelo menos uma zona de implementação onde o valor médio das vendas é estatisticamente e significativamente diferente dos restantes.

- d. Em alternativa ao teste utilizado na alínea anterior, que outro teste de hipóteses poderia ter sido realizado?

Caso os pressupostos fossem validados (normalidade dos dados das três amostras e homogeneidade da variância das três amostras) poderíamos utilizar o teste paramétrico ANOVA.

2. Realizou-se um estudo, num país da União Europeia, com o intuito de analisar características da sua população. Algumas dessas características foram as seguintes:

Saúde – Despesas anuais com saúde (euros)

Rendimento – Escalão de Rendimentos (Baixo=1/Médio=2/Alto=3)

Género (Masculino/ Feminino)

Idade (anos)

**Considere um nível de significância igual a 0,05.**

- a. Para avaliar se existem diferenças significativas entre as idades médias dos homens e das mulheres utilizou-se um teste  $t$  cujos resultados são apresentados de seguida.

Teste Kolmogorov-Smirnov	Teste $t$ – duas amostras independentes
<b>Variável:</b> “Idade” Género feminino $n = 14$ <i>Estatística</i> = 0,186 $p - value = 0,20$	<b>Variável:</b> “Idade”  <b>Variável de agrupamento:</b> “Género”  Teste de Levene $F = 0,037$ $p - value = 0,848$
<hr/> <b>Variável:</b> “Idade” Género masculino $n = 22$ <i>Estatística</i> = 0,116 $p - value = 0,20$	Teste $t$ $t = -0,275$ $p - value = 0,785$ Mean Difference= $-1,727$ 95% Confidence Interval for the Difference = $[-14,48 ; 11,03]$

i. O pressuposto da Normalidade verifica-se? Justifique.

Sim.

Analisando o resultado do teste à normalidade (Teste de Kolmogorov-Smirnov), cujas hipóteses são:

H0: Os dados têm distribuição normal

H1: Os dados não têm distribuição normal

Constatamos que para ambas as amostras o valor do p-value é igual a  $0,2 > 0,05$ , não se rejeitando H0 e concluindo que os dados têm distribuição normal.

ii. Averigue se existe homogeneidade de variâncias.

No que respeita ao pressuposto da homogeneidade das variâncias, o teste de Levene, cujas hipóteses são:

H0:  $\sigma_F^2 = \sigma_M^2$

H1:  $\sigma_F^2 \neq \sigma_M^2$

(onde  $\sigma_F^2$  representa a variância populacional da idade das mulheres e  $\sigma_M^2$  representa a variância populacional da idade dos homens)

Produziu um p-value igual a  $0,848 > 0,05$ , não se rejeita H0 e conclui-se que não existem diferenças significativas entre as variâncias dos dois grupos, ou seja, o pressuposto da homogeneidade das variâncias é validado.

iii. Defina as hipóteses associadas ao teste t.

As hipóteses são:

H0:  $\mu_F = \mu_M \Leftrightarrow \mu_F - \mu_M = 0$

H1:  $\mu_F \neq \mu_M \Leftrightarrow \mu_F - \mu_M \neq 0$

Onde  $\mu_F$  representa o valor médio populacional da idade dos indivíduos do sexo feminino e  $\mu_M$  representa o valor médio da idade dos indivíduos do sexo masculino.

iv. Pode concluir que existem diferenças significativas entre as idades médias dos homens e das mulheres? Justifique.

Analisando o resultado do teste t constatamos que se obteve um p-value igual a  $0,785 > 0,05$ , pelo que não se rejeita H0, concluindo-se que existem evidências estatísticas que apontam para a inexistência de diferenças significativas entre a média das idades das mulheres e dos homens.

v. Interprete o intervalo de confiança.

O intervalo de confiança a 95% para a diferença de médias é o seguinte:

$$[-14,48 ; 11,03]$$

Observe-se que o zero pertence a este intervalo, quer isto dizer que existem evidências estatísticas para afirmar que não existem diferenças significativas entre a média das idades das mulheres e dos homens.

- vi. Que outro Teste de Hipóteses poderia ter sido utilizado para testar se existem diferenças significativas entre as idades dos homens e das mulheres?

Pensando numa alternativa não paramétrica, poderíamos utilizar o teste de Mann-Whitney.

- b. Um dos objetivos do estudo realizado foi avaliar se existem diferenças significativas entre as despesas anuais de saúde (euros) segundo o escalão de rendimentos (Baixo/Médio/Alto). Aplicou-se o teste de hipóteses ANOVA, cujos resultados se apresentam de seguida:

$$F = 0,255$$

$$p - value = 0,777$$

- i. Defina as hipóteses associadas ao teste.

As hipóteses são:

$$H_0: \mu_B = \mu_M = \mu_A$$

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

Onde  $\mu_B$  representa o valor médio populacional das despesas anuais de saúde dos indivíduos do escalão Baixo,  $\mu_M$  representa o valor médio populacional das despesas anuais de saúde dos indivíduos do escalão Médio e  $\mu_A$  representa o valor médio populacional das despesas anuais de saúde dos indivíduos do escalão Alto.

- ii. Pode afirmar que existem diferenças significativas das despesas de saúde segundo o escalão de rendimentos? Justifique.

Analisando o resultado do teste ANOVA constatamos que se obteve um p-value igual a  $0,777 > 0,05$ , pelo que não se rejeita  $H_0$ , concluindo-se que existem evidências estatísticas que apontam para a inexistência de diferenças significativas na despesa média anula em saúde face ao escalão de rendimentos.

- iii. Quais os pressupostos deste teste que deveriam ter sido validados?

Deveriam ter sido validados os dois seguintes pressupostos:

- normalidade dos dados das três amostras
- homogeneidade das variâncias das três amostras

- iv. Que outro teste de hipóteses poderia ter sido utilizado.

Pensando numa alternativa não paramétrica, poderíamos utilizar o teste de Kruskal-Wallis.