E-fólio B | Folha de resolução para E-fólio

ANO LETIVO: 2018/19

UNIDADE CURRICULAR: Elementos de Bioestatística

CÓDIGO: 21036

DOCENTE: M. Rosário Ramos

A preencher pelo estudante

NOME: Joana Cristina Alves Quitério Russo

N.º DE ESTUDANTE: 1601289

CURSO: Matemática e Aplicações

DATA DE ENTREGA: 19 de janeiro de 2019

TRABALHO / RESOLUÇÃO:

a) Determine um intervalo de 95% de confiança para a proporção total de preços (i.e., juntando as amostras dos 3 tipos de cabaz) que são inferiores 120 euros. Assuma os pressupostos de que necessitar e justifique.

Resolução:

Juntando as amostras dos 3 tipos de cabaz, inferiores a 120 euros, temos:

$$p = \frac{20}{33} = 0.60$$

Variância $\theta^2 = p(1-p) \leftrightarrow \theta^2 = 0.24$

Desvio Padrão $\theta = 0.49$

$$Z_{a/2} \times \frac{\theta}{\sqrt{n}} = 1.96 \times \frac{0.49}{\sqrt{33}} = 0.17$$

O intervalo de confiança é 0.60 ± 0.17, ou seja] 0.43; 0.77 [.

b) **(0,5 v - Apenas para os estudantes da Lic. Matemática e Aplicações)**. Tomando o intervalo de confiança calculado em **a)**, encontre o valor de n que lhe permite reduzir a margem de erro para metade da obtida inicialmente. Indique os pressupostos que assume e justifique os resultados.

Resolução:

I.C =] 0.43; 0.77 [

Se mantivermos o grau de confiança, mantém-se z=1.96

O erro é de 0.17 neste intervalo e queremos reduzir para metade, que é 0.085.

Tomando o mesmo intervalo de confiança, penso que podemos pegar na fórmula

$$Z_{a/2} \times \frac{\theta}{\sqrt{n}} = 1.96 \times \frac{0.49}{\sqrt{n}} = 0.085 \leftrightarrow \frac{0.49}{\sqrt{n}} = 0.04 \leftrightarrow n = 150$$

Utilizamos o valor positivo de n, pois a amostra não suporta valores negativos.

Se aumentarmos o tamanho da amostra, n, a raiz quadrada vai tomar valores cada vez mais pequenos, pois n encontra-se no denominador da fração, dentro da raiz quadrada. Assim, se a proporção amostral ficar fixa em $\hat{p}$= 0.60 na nova amostra ampliada, o intervalo final vai ter menor amplitude. Fica "mais apertado", i.e., mais preciso, com menor margem de erro.

Em termos genéricos, para reduzir a amplitude do intervalo temos de aumentar o tamanho da amostra substancialmente.

c) **(1 v)** Realize um teste de hipóteses paramétrico para averiguar se o preço médio do cabaz de Tipo I é diferente do preço médio do cabaz de Tipo II. O que conclui? Indique os pressupostos, justifique o teste que escolheu (amostras independentes/dependentes) e os resultados.

Resolução:

Pretende-se a realização de um teste de hipóteses para comparar as médias dos preços do cabaz de tipo I e do cabaz do tipo II.

- Os registos são preços em euros, ou seja, dados numéricos (de uma variável quantitativa contínua).
- As duas amostras são independentes. Cada valor pertence a um tipo de cabaz distinto.
- O número de cabazes é 10. O tipo 1 tem n1=10 e o tipo 2 tem n2=11.

$H_0: \mu_1 = \mu_2 \leftrightarrow H_0: \mu_1 - \mu_2 = 0$ → não há diferenças significativas entre as duas médias.

$H_1: \mu_1 > \mu_2 \leftrightarrow H_1: \mu_1 - \mu_2 > 0 \leftrightarrow H_1: \mu_d > 0$ → Em média, o preço do cabaz do tipo 1 é maior do que o preço do cabaz do tipo 2.

d) **(1 v)** A tabela ANOVA em anexo apresenta parcialmente os resultados dos testes de comparação simultânea dos preços médios dos cabazes de compras dos três tipos em estudo. Formalize as hipóteses do estudo na linguagem estatística. Complete a tabela ANOVA justificando os valores apresentados. Conclua se existem diferenças significativas. Considere um nível de confiança de 95% e indique os pressupostos que assume. Justifique as suas afirmações e resultados.

Resolução:

A hipótese experimental consiste numa hipótese que envolve a mesma variável, avaliada em mais do que duas situações distintas (sendo as 3 situações aplicadas) sobre os mesmos cabazes.

Então, o teste paramétrico indicado é a Análise de Variância a 1 fator (ANOVA) para amostras emparelhadas.

O fator em estudo é o Tipo de cabaz: O Tipo I de cabaz de compras é constituído na sua maioria por produtos de marca branca (marca da própria do supermercado); o Tipo II de cabaz é constituído na sua maioria (90%) por marcas tradicionais bem implantadas no mercado; o Tipo III de cabaz contém cerca de metade dos produtos com origem na produção biológica/orgânica.

Tabela ANOVA

SUMÁRIO

Grupos	*Contagem*	*Soma*	*Média*	*Variância*	*Desvio Padrão*
Tipo I	10	1069,10	106,91	196,07	14,00
Tipo II	11	1306,30	118,75	330,07	18,17
Tipo III	12	1557,40	129,78	408,55	20,21

ANOVA

Fonte de variação	*SQ*	*gl*	*MQ*	*F*	*valor P*	*F crítico*
Entre grupos	2856,69	2,00	1428,34	4,48	0,02	3,32
Dentro de grupos	9559,37	30,00	318,65			
Total	12416,06	32,00				

Vamos agora estabelecer um grau de confiança para a nossa decisão estatística.

No enunciado é dada a indicação, pelo que vamos estabelecer o grau de confiança de 95% (0.95). O correspondente nível de significância α=5% (0.05) (confiança+ significância=100%).

Um dos objetivos da aplicação ANOVA é o de realizar um teste estatístico para verificar se há diferença entre distribuição de uma medida entre três ou mais grupos. No nosso caso, vamos definir as hipóteses do teste como:

H_0: Não existe diferença entre os preços de cada cabaz.

H_1: Há pelo menos um cabaz com preço diferente.

Rejeitamos H_0 se o valor p for inferior ao nível de significância (0.05). Neste caso verifica-se, portanto, rejeitamos H_0, concluindo que existe pelo menos dois tipos de cabazes com valor significamente diferentes ao avaliar o valor p.

A conclusão da ANOVA pode ser feita também com base na estatística F. A estatística F tem distribuição F de Fisher-Snedecor com k-1 e n-k grays de liberdade, onde k é o número de grupos (k=3) e n é o número de observações (n=33). Neste caso, obteríamos através da tabela F=3,32 e como a estatística F (4,48) foi maior que o F tabelado (3,32), concluiu-se que existe pelo menos dois tipos de cabaz com valores significativamente diferentes.

Após alguma pesquisa, verifiquei que podemos saber quais os tipos de cabaz com valor diferente através da utilização do teste de comparação múltipla, como por exemplo, o teste de Tukey, no entanto penso que sai fora do âmbito da cadeira.

e) **(0,5 v)** Suponha agora que foi realizado um teste que verifica a Normalidade dos dados e rejeitou-se a normalidade nas amostras de preços dos Tipo I e II de cabaz de compras. Utilize o teste de Mann-Whitney para comparar os preços dos cabazes de Tipo I e Tipo II. Considere um nível de confiança de 95%. O que conclui? Justifique o teste escolhido e os resultados.

Resolução:

As amostras são independentes.

H_0: As medianas dos preços de cada cabaz são iguais.

H_1: As medianas dos preços dos cabazes são diferentes.

O primeiro passo para elaborar o teste é ordenar do menor para o mais pequeno os dados que temos e dar-lhes uma ordem.

Tipo I	Rank	Tipo II	Rank
85.7	1	98.2	4
90.8	2	100.3	5
97	3	100.8	7
100.5	6	108.9	9
104.2	8	115.6	13
110.1	10	117.3	14
112.6	11	118.3	15
114.1	12	122.2	16
125.8	17	129.8	19
128.3	18	134.7	20
		160.2	21
n1=10	w1=88	n2=11	w2=143

$$u_1 = w_1 - n_1 \times \frac{(n_1 + 1)}{2} = 33$$

$$u_2 = w_2 - n_2 \times \frac{(n_2 + 1)}{2} = 77$$

Para n1=10, n2=11, a=0.05 e teste bilateral, rejeita-se H_0 se $u = min\{u_1, u_2\} \leq 26$

Como $u = 33 > 26$, não se rejeita H_0.

Conclui-se que os dois tipos de cabaz são equivalentes quanto ao preço.