

Universidade de São Paulo
Faculdade de Filosofia, Letras e Ciências Humanas Departamento de Ciência
Política

FLS 5028- Métodos Quantitativos e Técnicas de Pesquisa em Ciência Política
FLP 0406 - Métodos e Técnicas de Pesquisa em Ciência Política
1º Semestre/ 2018

Profº Dr. Glauco Peres da Silva

LISTA DE EXERCÍCIO 08

Data de entrega: 28/05/2018 (noturno) e 30/05/2018 (vespertino)

Exercício 01 (2 pontos)

Nas alternativas abaixo marque “Verdadeiro” (V) ou “Falso” (F) e justifique as suas escolhas em, no máximo, 5 linhas. Quando marcar falso, justifique porque discorda da afirmação, e quando marcar verdadeiro, justifique o raciocínio no qual sua escolha se baseia.

() A associação entre variáveis de uma tabela de contingência existe quando as distribuições condicionais de cada uma das categorias são idênticas.

FALSA. Somente quando os valores observados são diferentes é que há evidência de dependência entre as variáveis numa tabela de contingência.

() Quando as frequências observadas são muito próximas às esperadas, o valor da estatística qui quadrado é pequeno. Mas, quando as divergências são grandes, consequentemente, o teste assume valores altos.

VERDADEIRO.

() A estatística qui quadrado é um estimador da discrepância entre frequências esperadas e observadas, e estabelece se as diferenças encontradas se devem ou não a uma relação de causalidade.

FALSO. O que o teste consegue estabelecer é a associação entre as variáveis, mas não é capaz de estabelecer causalidade.

() O teste de qui quadrado é especialmente útil para amostras pequenas e nos fornece evidências sobre a natureza ou a força da associação entre as variáveis.

FALSO. O teste qui quadrado assemelha-se ao teste z na medida em que é apropriado a amostras grandes. Mas o qui quadrado não mensura a força ou a natureza da associação entre as variáveis. Conforme Agresti e Finlay, a análise de resíduos padronizados permite uma análise “informal” do padrão de associação entre as variáveis. Não somente, medidas de associação ou diferença de proporções são mecanismos mais adequados para mensurar a força da associação.

Exercício 02 (4 pontos)

Uma empresa deseja saber se o consumo de seus produtos está associado ao gênero de seus consumidores a fim de melhorar suas campanhas publicitárias e vender mais produtos. Essa empresa selecionou aleatoriamente 150 consumidores para responder um questionário sobre a preferência por 3 de seus produtos. Cada consumidor respondeu qual seu produto preferido, dentre as opções A, B e C. Os resultados são mostrados na tabela de frequências a seguir:

Os resultados são mostrados a seguir:

Gênero	Produto A	Produto B	Produto C	Total
Feminino	20	40	20	80
Masculino	30	30	10	70
Total	50	70	30	150

- a) Quais são as hipóteses nula e alternativa que a empresa deve empregar para formular seu teste?

H_0 : a preferência por um produto não tem associação com o gênero

H_1 : a preferência está associada ao gênero

- b) Calcule a frequência esperada para cada uma das frequências observadas

Frequência esperada = soma linha * soma coluna / soma total = $(80*50)/150 = 26,67...$

Frequências esperadas:

Gênero	Produto A	Produto B	Produto C	Total
Feminino	26,67	37,33	16	80
Masculino	23,33	32,67	14	70
Total	50	70	30	150

- c) Calcule a estatística do teste qui quadrado para esse caso. É possível afirmar que há associação entre gênero e o produto consumido com 95% de confiança?

$$\text{Qui-quadrado} = \frac{\sum_i^N (\text{freq obs} - \text{freq esperada})^2}{\text{freq esperada}} = \sum \left[\frac{20-26,67}{26,67} \right]^2 + \dots + \left[\frac{10-14}{14} \right]^2 = 6,13$$

Graus de liberdade = (linhas - 1)*(colunas - 1) = 2

Na tabela de distribuição do qui quadrado (exemplo https://fenix.tecnico.ulisboa.pt/downloadFile/1970943312279876/Tabela_QuiQuadrado.pdf), para 2 graus de liberdade, o qui quadrado crítico é de 5,99:

n	P($\chi_n^2 \leq x$)									
	0,005	0,01	0,025	0,05	0,1	0,25	0,5	0,75	0,9	0,95
1	3,93E-05	0,000157	0,000982	0,003932	0,016	0,102	0,455	1,323	2,706	3,841
2	0,010	0,020	0,051	0,103	0,211	0,575	1,386	2,773	4,605	5,991
3	0,072	0,115	0,216	0,352	0,584	1,213	2,366	4,108	6,251	7,815

Como o qui quadrado crítico (5,99) é menor que o qui quadrado do teste (6,13), a hipótese nula é rejeitada. Assim, com 95% de confiança, rejeitamos a hipótese nula e podemos afirmar que o consumo dos produtos está associado ao gênero dos consumidores.

Exercício 3 (4 pontos)

Na tabela abaixo, estão representados os níveis de satisfação de indivíduos sobre a qualidade do serviço de um parque de diversões, indicando +5 (muito satisfeito) e +1 (pouquíssimo satisfeito), de acordo com os dias da semana. Acredita-se que a satisfação dos visitantes do parque é influenciada pelo dia da semana, já que o parque fica mais lotado nos finais de semana e que atrações diferentes funcionam em dias distintos.

Nível de Satisfação	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
+5	100	120	95	105	110	65	75
+4	80	85	75	50	90	105	115
+3	75	80	55	80	80	105	85
+2	35	50	60	40	25	90	80
+1	10	15	25	15	15	35	45

- a. Monte uma tabela com os valores esperados;

Nível de

Satisfação	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo	
5	84.8	98.9	87.6	82.0	90.5	113.1	113.1	670
4	75.9	88.6	78.5	73.4	81.0	101.3	101.3	600
3	70.9	82.7	73.2	68.5	75.6	94.5	94.5	560
2	48.1	56.1	49.7	46.5	51.3	64.1	64.1	380
1	20.3	23.6	20.9	19.6	21.6	27.0	27.0	160
	300	350	310	290	320	400	400	2370

b. Apresente uma hipótese nula e uma hipótese alternativa de acordo com o problema que se deseja resolver;

H_0 : As variáveis “nível de satisfação” e “dia da semana” são independentes

H_1 : As variáveis “nível de satisfação” e “dia da semana” não são independentes

c. Encontre o valor da estatística do Qui-quadrado;

O valor de cada termo do somatório da fórmula do qui-quadrado está expresso na tabela abaixo:

Nível de Satisfação	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo
5	2.72	4.48	0.62	6.46	4.22	20.44	12.82
4	0.22	0.15	0.15	7.47	1.00	0.14	1.86
3	0.24	0.09	4.55	1.92	0.25	1.16	0.96
2	3.57	0.67	2.13	0.91	13.49	10.43	3.92
1	5.19	3.15	0.79	1.07	2.02	2.37	11.99

O valor do somatório é igual a 133,626. Com 24 graus de liberdade, o p-valor é igual a 3,407e-17

d. Responda se é possível afirmar que o nível de satisfação varia de acordo com os dias da semana;

De acordo com o p-valor indicado acima, rejeitamos a hipótese nula de independência entre as variáveis.

e. Monte uma tabela com os resíduos padronizados para cada célula;

f. Determine quais são as células que mais afetam o resultado. A impressão de que nos finais de semana o nível de satisfação cai é sustentada por estes resultados?

Explique.

Resposta para os itens e) e f):

Resíduos Padronizados

2.084	2.707	0.996	3.204	2.608	-5.855	-4.638
0.575	-0.480	-0.488	-3.376	1.242	0.471	1.732
0.598	-0.368	-2.617	1.693	0.621	1.354	-1.228
-2.206	-0.965	1.709	-1.110	-4.310	3.866	2.371
-2.525	-1.991	0.989	-1.144	-1.582	1.748	3.933

t-score 2.064

As células em vermelho são aquelas que estão acima do valor do t-score para um alfa de 5%. As marcadas em azul são as que estão abaixo do valor negativo do t-score para o mesmo alfa. Os resultados sugerem que assim como o esperado pelos administradores do parque, nos dias da semana, a avaliação positiva tende a ser maior do que nos finais de semana. As células em azul estão na diagonal oposta, confirmando a suspeita inicial.

Exercício 4 – Pós-graduação (4 pontos)

Considere a tabela abaixo. Nela estão representadas as medidas da altura de indivíduos de dois países diferentes, a partir de uma amostra selecionada aleatoriamente:

Indivíduo	País	Altura (m)
1	A	2,01
2	B	1,80
3	B	1,74
4	A	1,85
5	A	1,80
6	A	1,90
7	A	1,88
8	B	1,75
9	A	1,92
10	B	2,00
11	B	1,65
12	B	1,56
13	A	1,83
14	B	1,62
15	B	1,58

Indivíduo	País	Altura (m)
16	B	1,77
17	A	1,80
18	B	1,63
19	B	1,83
20	A	1,80
21	B	1,79
22	B	1,64
23	B	1,70
24	A	1,93
25	A	1,87
26	B	1,65
27	A	1,78
28	B	1,79
29	A	1,68
30	A	1,72

Para os dados indicados acima:

- a. Estabeleça um intervalo de confiança para a altura média dos indivíduos de cada um dos países;

$$\text{IC - País A: } \left[1,841 - 2,145 * \frac{0,0868}{\sqrt{14}}; 1,841 + 2,145 * \frac{0,0868}{\sqrt{14}} \right] = [1,791; 1,891]$$

$$\text{IC - País B: } \left[1,719 - 2,12 * \frac{0,1126}{\sqrt{16}}; 1,719 + 2,12 * \frac{0,1126}{\sqrt{16}} \right] = [1,659; 1,779]$$

- b. Suponha que você deseje saber se os indivíduos de um país são mais altos do que do outro. Faça um teste de médias, indicando a hipótese nula e a hipótese alternativa;

$$H_0: \mu_a - \mu_b = 0$$

$$H_1: \mu_a - \mu_b \neq 0$$

Sabemos que $\bar{X}_a = 1,841$ e $\bar{X}_b = 1,719$, então $\bar{X}_a - \bar{X}_b = 0,122$.

$$\text{ep}(\bar{X}_a - \bar{X}_b) = 0,0365$$

$$\text{Teste-t: } \frac{0,122-0}{0,0365} = 3,3415$$

- c. Monte um intervalo de confiança para a diferença de médias;

Ao nível de confiança de 95%

$$\text{IC: } [0,122 - 2,048 * 0,0365; 0,122 + 2,048 * 0,0365] = [0,047; 0,219]$$

- d. Encontre o valor do teste t para o item b;

$$\text{t-teste} = 3,3415$$

$$\text{t-crítico} = 2,048$$

- e. Os indivíduos de um dos países são mais altos do que do outro? Justifique (máx. 5 linhas).

Sim, pois com o valor do teste-t, rejeitamos a hipótese nula.