

FLS 5028
Métodos Quantitativos e Técnicas de Pesquisa em Ciência Política

FLP0406
Métodos e Técnicas de Pesquisa em Ciência Política
1º semestre / 2016

Glauco Peres da Silva

LISTA DE EXERCÍCIOS 08 (GABARITO)

Data de entrega: 23/05/2016 (noturno) e 25/05/2016 (vespertino).

Exercício 01 (2 pontos)

Encontre os erros nas sentenças abaixo e os corrija sempre justificando sua resposta.

1- A análise das variáveis categóricas pode ser apresentada em tabelas de contingência. Os totais das linhas e colunas são denominados distribuições condicionais e a distribuição dos dados amostrais condicionada a uma das variáveis categóricas de interesse é denominada distribuição marginal.

Falsa. Segundo Agresti e Finlay (2012), a denominação se dá ao contrário. Os totais das linhas e colunas são denominados distribuições marginais enquanto a distribuição dos dados amostrais condicionada a uma das categorias de interesse é, por sua vez, a distribuição condicionada.

2- A comprovação de dependência estatística entre duas variáveis categóricas é demonstrada a partir da comprovação de que as distribuições condicionais da população são idênticas a de cada categoria.

Falsa. A independência estatística é identificada pela identidade entre as distribuições condicionais e de cada categoria. A não identidade é um indicio de dependência.

3- O teste qui-quadrado compara frequências observadas entre si para determinar se a hipótese nula pela dependência entre as variáveis se satisfaz.

Falsa. O teste qui-quadrado compara frequências esperadas e observadas entre si de modo a por a teste a hipótese nula que preconiza a independência entre as variáveis categóricas em questão.

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

Falsa. Segundo Agresti e Finlay (2012), o teste qui-quadrado assemelha-se aos testes z na medida em que é especialmente apropriado a amostras grandes. Como a distribuição do qui-quadrado deve aproximar-se da distribuição da estatística-teste χ^2 , isso somente ocorrerá se a amostra for grande ou, segundo os autores, se a frequência esperada for maior que 5 em cada célula. O teste qui-quadrado, ademais, não mensura a força ou a natureza da associação entre as variáveis. Como destacam os mesmos autores, 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.

5- A medida de associação denominada *gama* é especialmente útil para mensurar a associação entre variáveis não ordinais. Sua fórmula amostral divide a soma do número de pares concordantes e discordantes pela sua diferença. O valor de *gama* varia entre 0 e 1, de modo que apenas pode assumir valores positivos. Ademais, quanto menor seu valor absoluto, maior a associação indicada.

Falsa. Segundo Agresti e Finlay (2012), a medida de associação *gama* serve especialmente para testar a associação entre variáveis ordinais, com fórmula amostral definida pela equação $\hat{\gamma} = \frac{C-D}{C+D}$, onde C indica o número de pares concordantes e D de pares discordantes. Pela equação deduz-se que o valor de *gama* varia entre -1 e +1 e quanto maior seu valor absoluto, maior a associação.

Exercício 02 (4 pontos)

Em um estudo recente Merlo e Moreira (2015)¹ procuraram analisar se o gênero do eleitor impacta em suas preferências por políticas públicas. Fazendo uso um *survey* de opinião pública sobre a política nacional do Brasil realizado em 2014 as autoras tentaram encontrar evidências de que a preferência por políticas é dada de forma diferente entre os homens e as mulheres.

a-) Tendo essa breve elucidação do estudo em mente, você consegue pensar em qual é a hipótese nula e a hipótese alternativa com a qual as autoras estão trabalhando? Quais seriam essas hipóteses?

Sim.

¹ MERLO, M.; MOREIRA, N. **O que querem as mulheres? Investigação do comportamento eleitoral e formulação de prioridades sob a perspectiva do gênero.** . In: V SEMINÁRIO DISCENTE DA PÓS-GRADUAÇÃO DO DEPARTAMENTO DE CIÊNCIA POLÍTICA DA USP. São Paulo, Brasil: 2015

H_0 : A variável gênero e preferência por políticas são independentes.

H_0 : A variável gênero e preferência por políticas são dependentes.

b-) Uma das perguntas do *survey* utilizada pelas autoras nesse trabalho foi a seguinte:

Questão: “As pessoas às vezes falam sobre quais objetivos o país deveria seguir para os próximos dez anos. Neste cartão estão listados alguns dos objetivos que diferentes pessoas dariam prioridade. Você poderia, por favor, dizer qual destes você considera o mais importante?”

As alternativas disponíveis para os respondentes foram as seguintes:

Crescimento Econômico	Ordem da Nação
Forças Armadas Fortes	Participação no Governo
Participação Local	Combate à inflação
Embelezar o país	Liberdade de expressão

Para cada uma dessas opções, cada respondente deveria responder se considerava o objetivo como “não é o mais importante” ou o “é o mais importante”. Analisando somente a opção “Crescimento Econômico”, tem-se que dos 1417 entrevistados, 245 homens e 439 mulheres responderam que consideram o crescimento econômico como o objetivo que “não é o mais importante”. Por outro lado, 298 homens e 435 mulheres, consideraram esse objetivo como o “mais importante”.

De acordo com os dados descritos acima, construa uma tabela de contingência com os valores absolutos e relativos. Mostre seus cálculos. Você consegue explicar por que a construção de uma tabela de contingência é importante? Ainda, de acordo com os dados de sua tabela, você consegue identificar se as duas variáveis são dependentes ou independentes?

	Não é importante	É o mais importante	Totais
Homem	245 (45%)	298 (55%)	543 (100%)
Mulher	439 (50%)	435 (50%)	874 (100%)
Totais	684 (48%)	733 (52%)	1417 (100%)

De acordo com Agresti e Finlay (2012, p.253) construir uma tabela de contingência de um arquivo de dados é o primeiro passo para a investigação de uma associação entre duas variáveis categóricas. Não é possível somente com a tabela criada acima identificar se temos dependência ou independência entre as variáveis. Somente seria possível fazer tal afirmação se estivéssemos falando de uma população, contudo os dados apresentados são de uma amostra.

c-) Levando em consideração a hipótese das autoras, é possível afirmar que há uma relação entre gênero de preferência por políticas públicas no que diz respeito ao crescimento econômico? Qual teste você usaria para comprovar ou rejeitar a hipótese das autoras? Apresente esse teste, bem como encontre o valor- p associado a ele. Assuma $\alpha = 5\%$. Apresente seus cálculos e justifique sua resposta.

Como estamos lidando com a possível associação entre variáveis categóricas em uma amostra, o teste mais apropriado seria o teste qui-quadrado. O teste qui-quadrado compara frequências observadas na tabela de contingência com os valores que satisfaçam a hipótese nula de independência (Agresti e Finlay, 2012, p.256)

Cálculo do qui -quadrado

$$f_e = \frac{(\text{Total da Linha}) * (\text{Total da Coluna})}{\text{Tamanho total da amostra}}$$

$$f_e = \frac{(543) * (684)}{1417} = 262,11$$

$$f_e = \frac{(874) * (684)}{1417} = 421,88$$

$$f_e = \frac{(543) * (733)}{1417} = 280,88$$

$$f_e = \frac{(874) * (733)}{1417} = 452,11$$

$$X^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

$$X^2 = \frac{(245 - 262,11)^2}{262,11} + \frac{(298 - 280,88)^2}{280,88} + \frac{(439 - 421,88)^2}{421,88} + \frac{(435 - 452,11)^2}{452,11} =$$

$$X^2 = 1,11 + 1,04 + 0,69 + 0,64 = 3,48$$

$$\text{Graus de Liberdade} = (\text{linha}-1) * (\text{coluna}-1)$$

Pela regra colunas 2 x 2 como a nossa tem Graus de Liberdade = 1

Olhando a tabela, temos que a probabilidade associada a 3,48 com 1 grau de liberdade é de 0,06.

Logo não é possível rejeitar H_0 de que as variáveis são independentes, ou seja, a variável gênero e preferência por políticas públicas quando o assunto é crescimento econômico parece não variar de acordo com o gênero.

d-) Levando em consideração a resposta do item anterior (letra “c”) é possível afirmar que que o gênero causa preferências distintas sobre políticas públicas? Para além disso, mediante a realização do teste que você escolheu realizar no item anterior, é possível medir a força da associação entre gênero e preferência por políticas? (Máximo de 10 linhas)

Primeiro ponto da resposta: os alunos devem destacar que associação não significa causação. A resposta deve ser montada em cima dessa discussão. Eles devem ser capazes de perceber que o teste que eles realizaram no item anterior apenas mede associação entre as variáveis. Para além disso, também não estamos medindo a força da associação entre as variáveis. Agresti e Finlay (2012, p. 260 e 261) destacam que o teste qui-quadrado não nos diz nada sobre a força e a natureza da associação entre duas variáveis. O teste não indica se todas as células se desviam muito da independência, ou talvez somente uma ou duas células se desviam.

Exercício 03 (4 pontos)

O ministro dos transportes, preocupado com o aumento do número de acidentes de carros envolvendo crianças encomendou uma pesquisa a respeito da segurança no trânsito. O instituto responsável por realizar a pesquisa, por sua vez, realizou uma análise a respeito dos ferimentos graves que crianças sofrem em acidentes envolvendo carros e chegou aos seguintes resultados:

FERIMENTOS GRAVES?			
USANDO CINTO?	SIM	NÃO	TOTAIS
SIM	240	50	290
NÃO	107	16	123
TOTAIS	347	66	413

a-) A partir dos resultados acima expostos você consegue pensar qual é a variável dependente e independente? Com quais hipóteses (nula e alternativa) o instituto está trabalhando? Justifique sua resposta (Máximo de 8 linhas)

A variável dependente é a ocorrência ou não de ferimentos graves em crianças e a variável explicativa é o uso ou não do cinto de segurança. É esperado que o aluno indique que

chegou à essa resposta por notar que na tabela de contingência a variável dependente sempre ocupa o lugar das colunas.

H_0 : A variável ferimento grave e uso do cinto de segurança são independentes.

H_0 : A variável ferimento grave e uso do cinto de segurança são dependentes.

b-) Calcule a razão de chance de ocorrência de ferimentos graves de acordo com os resultados apresentados pela pesquisa. O que podemos interpretar? Apresente seus cálculos e justifique sua resposta.

$$\text{Chance de ferimento}_{\text{cinto}} = \frac{16/123}{107/123} = 0,14$$

$$\text{Chance de ferimento}_{\text{s/cinto}} = \frac{50}{240} = 0,2083$$

$$\text{Razão de Chance} = \frac{0,14}{0,2083} = 0,718$$

A partir da razão de chance calculada podemos afirmar que usar o cinto de segurança reduz em cerca de 28% a chance de ferimentos graves em acidentes automobilísticos com crianças.

Exercício 04: Pós-Graduação (5 pontos)

É possível teoricamente imaginar que existiria uma associação entre o número efetivo de partidos e o sistema de financiamento de campanha adotado em um determinado país. O tipo de financiamento de campanha institucionalmente escolhido (privado ou público), por exemplo, poderia afetar o número efetivo de partidos de modo que quanto maior a disponibilidade de financiamento público (direto – como o Fundo Partidário – ou indireto – como o tempo de TV –, ou ambos), desconsiderando outras variáveis explicativas, maior seria a possibilidade de entrada de partidos na arena eleitoral e, portanto, maior o número efetivo de partidos. Considerando estas expectativas teóricas acima, vamos trabalhar com o banco de dados do “Democracy Barometer”, disponível no Moodle (dataset e codebook), e com as variáveis “Número efetivo de partidos” (Enep) e “Financiamento de Campanha” (Funding).

- a) Proponha ao menos uma relação ou expectativa de associação entre as variáveis Enep e Funding que possa ser testada por **teste de diferença de médias**. Explique porque este teste é pertinente.

Pensando no teste de diferença de médias, segundo Kellstedt e Whitten (2014) testamos a associação entre uma variável dependente contínua, que no caso será o

número efetivo de partidos; e uma variável independente categórica, o tipo de financiamento público de campanha. Espera-se que a média do número efetivo de partidos para os casos em que estão disponíveis todas as formas de financiamento público seja maior do que a média do número de partidos para os casos de ausência de financiamento público.

- b) Avalie o banco e as duas variáveis propostas. Existem células em branco ou “missing” para estas variáveis? Em caso afirmativo, como você pretende tratá-las? Existem 134 registros “missing” para a variável ENP e 243 para a variável “Funding”. Do total de casos (1770), estes valores “missing” representam 16% dos registros (linhas) do banco. Esta quantidade de “missing data”, se tratada como valor “0”, pode causar viés para a determinação das médias, de modo que é possível que se retire estas informações da análise para não influenciar suas medidas descritivas e inferência. Mesmo assim, importa reconhecer que ao desconsiderar estes registros existe perda de informação.

Para retirar, saída Stata:

```
. drop if Funding == "."
(243 observations deleted)

. drop if Enep == "."
(46 observations deleted)
```

- c) Como as duas variáveis são mensuradas? Seus possíveis valores estão compreendidos em que intervalo? Apresente as estatísticas descritivas destas variáveis e sua natureza (se categórica ou contínua).

A variável Enep, como define o Codebook do banco, indica o número efetivo de partidos. Esta variável varia de 1,12 a 16,57, seguindo fórmula apresentada por *Nohlen, Dieter/ Florian Grotz/ Christof Hartmann (2001). Elections in Asia and the Pacific. A Data Handbook. Vol. I+II. Oxford, New York Auckland: Oxford University Press.* (Pode-se criticar que a fórmula de cálculo não esteja presente no próprio codebook, mas exija remissão a outro texto). Trata-se de uma variável contínua.

Estatística Descritiva “Enep”:

```
. summarize Enep
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Enep	1480	4.610458	1.944593	1.347737	16.57013

A variável “Funding” é de natureza categórica, e assume apenas três “valores”: 0, para os casos de países que não oferecem nenhuma forma de financiamento público direto ou indireto; 1, se existe alguma forma de acesso direto ou indireto a financiamento público; e 2, quando existe acesso a financiamento direto e indireto concomitantemente.

Frequências absoluta e relativa: “Funding”

```
. tabulate Funding
```

Funding	Freq.	Percent	Cum.
0	124	8.38	8.38
1	327	22.09	30.47
2	1,029	69.53	100.00
Total	1,480	100.00	

Descrevendo “Funding” e “Enep” juntas:

```
Summary for variables: Enep
by categories of: Funding (Funding)
```

Funding	mean	sd	se (mean)	variance
0	3.895329	1.604951	.1441289	2.575868
1	4.513278	1.679093	.092854	2.819353
2	4.727518	2.039544	.0635807	4.159738
Total	4.610458	1.944593	.0505473	3.781443

- d) Formule a hipótese desenvolvida em “a” dentro do esquema de teste de hipóteses, estabelecendo uma hipótese nula e uma alternativa.

Hipótese Nula: A média do numero efetivo de partidos para os casos em que a variável Funding for igual a 2 não é estatisticamente diferente da média para os casos em que a variável Funding for igual a 0.

Hipótese Alternativa: A diferença das médias é diferente de 0.

Assim:

$$H_0: \hat{Y}_2 - \hat{Y}_0 = 0$$

$$H_A: \hat{Y}_2 - \hat{Y}_0 \neq 0$$

Outra forma de formular estas hipóteses é já sugerir uma direção específica para a diferença.

Hipótese Nula: A média do numero efetivo de partidos para os casos em que a variável Funding for igual a 2 não é estatisticamente maior que média para os casos em que a variável Funding for igual a 0.

Hipótese Alternativa: A média para o primeiro caso é maior que para o segundo.

Assim:

$$H_0: \hat{Y}_2 - \hat{Y}_0 \leq 0$$

$$H_A: \hat{Y}_2 - \hat{Y}_0 > 0$$

Vamos optar nesta resolução pelo teste bicaudal, pois a saída do Stata fornece informações para todos os testes. Mas vocês pode testar em casa o teste monocaudal.

- e) Desenvolva o teste de hipóteses e apresente seus achados. Adote um p-valor de 0,05.

Existem duas formas de realizar o teste de médias, dado que Funding assume três valores. Em primeiro lugar é necessário abandonar todos os valores missing, para que eles não sejam considerados um quarto grupo. Para Funding igual a “1”, é possível: (a) abandonar as observações com esta informação, e apenas trabalhar com os grupos “0” e “2”, como abaixo:

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	124	3.895329	.1441289	1.604951	3.610035	4.180624
2	1029	4.727518	.0635807	2.039544	4.602755	4.852281
combined	1153	4.63802	.0592929	2.013341	4.521685	4.754354
diff		-.8321884	.1898929		-1.204763	-.4596134

diff = mean(0) - mean(2) t = -4.3824
 Ho: diff = 0 degrees of freedom = 1151
 Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
 Pr(T < t) = 0.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 1.0000

Uma segunda forma é agrupar os casos “1” e “2” para Funding em uma mesma categoria (como uma nova variável Fund12), para não se perder informação:

```
. ttest Enep, by (Fund12)
```

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
0	124	3.895329	.1441289	1.604951	3.610035	4.180624
1	1356	4.675854	.0532331	1.960252	4.571425	4.780282
combined	1480	4.610458	.0505473	1.944593	4.511306	4.70961
diff		-.7805243	.1813684		-1.136291	-.4247574

```
diff = mean(0) - mean(1)                                t = -4.3035
Ho: diff = 0                                             degrees of freedom = 1478
```

```
Ha: diff < 0                                             Ha: diff != 0                                             Ha: diff > 0
Pr(T < t) = 0.0000                                       Pr(|T| > |t|) = 0.0000                                       Pr(T > t) = 1.0000
```

Para ambos os casos, a diferença entre as médias é estatisticamente diferente de 0, como se pode observar pelo alto valor da estatística “t” para 1478 graus de liberdade. Neste caso, é possível abandonar a hipótese nula de que a diferença da média do número efetivo de partidos seria igual a zero entre países que adotam alguma forma de financiamento público e aqueles não adotam nenhuma.

Para a exposição detalhada dos cálculos, confira a seguir, para a segunda forma de teste de hipótese acima.

Para se obter a estatística “t” que permitirá determinar a plausibilidade da hipótese nula, tem-se: $t = \frac{\hat{Y}_1 - \hat{Y}_2}{se(\hat{Y}_1 - \hat{Y}_2)}$, ou seja, a diferença das médias da variável dependente e independente, dividida pelo erro padrão da diferença entre estas médias, dado pela equação: $se(\hat{Y}_1 - \hat{Y}_2) = \sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$. Assim, com as médias e desvios padrões da planilha acima, no item “c”, tem-se que:

$$se(\hat{Y}_1 - \hat{Y}_2) = \sqrt{\left(\frac{1,60}{124} + \frac{1,96}{1356}\right)} = 0,0505$$

Logo:

$$t = \frac{3,89 - 4,675}{0,050} = -4,303$$

Nossos graus de liberdade são mensurados pelo n total menos as dois (ou seja, as duas estatísticas que utilizamos para esta mensuração, a média do grupo 1 e a média

do grupo 2). Logo – 1478 graus de liberdade. Também é possível utilizar a forma como Kellstedt e Whitten (2014) calculam os graus de liberdade, pela redução em um grau do “n” do menor grupo – assim: para $F_{un12} = 0$, ou seja, sem financiamento público, temos $n=124$, ou seja, $gl=124-1=123$.

Olhando para a tabela da distribuição da estatística t , temos que para 123 graus de liberdade, e p-valor de 0,025 (teste bicaudal, já que se trata da diferença das médias diferente de 0), temos que a estatística t da tabela é igual a -1,96 ou seja, maior que a estatística t encontrada na nossa análise, -4,303. Ou seja, considerando a simetria dos dois lados da distribuição, para tornar a comparação mais intuitiva, temos que:

$$t > t_{table}, \text{ ou seja, } 4,303 > 1,96,$$

Deste modo é altamente improvável (ou seja, uma probabilidade menor que 5%) que, considerando a hipótese nula como verdadeira (ou seja, que não existe diferença entre as médias para os dois grupos), encontre-se esta diferença entre médias. Deste modo é possível descartar a hipótese nula.

- f) A partir do item acima, como explicar seus resultados a luz das suas expectativas teóricas? Existe alguma explicação alternativa aos seus achados? O que significa desconsiderar o pressuposto acima adotado de que não haveria mais nenhuma variável relevante a esta relação?

Como realizamos um teste bicaudal, sabemos que, desconsiderando outras variáveis explicativas, existe uma associação estatisticamente relevante entre as duas variáveis – número efetivo de partidos e financiamento público de campanha. O teste acima fornece informações sobre a possível direção desta diferença. A diferença tende a ser estatisticamente menor que zero, o que sugere uma relação contrária àquela que esperávamos, ou seja, parece apontar que mais financiamento público tende a estar associado a menos partidos. Para de fato analisar esta relação, contudo, é necessário incluir outras variáveis de controle que podem afetar esta relação, mas que foram desconsideradas nesta análise. Tais variáveis poderiam ser: possíveis restrições legais ao financiamento privado de campanha; restrições ao próprio acesso aos recursos públicos a partir do número de cadeiras de votos recebidos por partido; o sistema eleitoral e as regras eleitorais, que condicionam o formato da competição e, assim, como candidatos e partidos buscam e distribuem recursos; etc. Em suma, o teste de diferença de médias ajuda a esclarecer aspectos da relação entre duas

variáveis, mas a ausência de controles impede que se utilizem seus resultados de forma inconteste.