

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 10

Data de entrega: 18/06/2018 (noturno) e 20/06/2018 (vespertino)

Exercício 01 (2 pontos)

Marque “Verdadeiro” (V) ou “Falso” (F) para as sentenças que serão apresentadas a seguir e **justifique** cada uma das suas escolhas em no máximo 5 linhas.

(F) A seguinte equação: $y = \alpha + \beta * x$ expressa as observações em y como uma função linear em x , em que α representa o valor que y assume quando x é diferente de zero e β representa a mudança ocasionada em x , para o aumento de uma unidade em y .

Falso. De acordo com Agresti e Finlay (2012, p.289), na equação, α representa o valor que y assume quando x é igual a zero e β representa a mudança em ocasionada em y , para o aumento de uma unidade em x .

(V) Erros de previsão em modelos de regressão linear são chamados de resíduo. Esse valor é obtido fazendo a diferença entre o valor $-y$ real (ou valor observado) e o valor previsto representado por \hat{y} . Enquanto um resíduo positivo é obtido quando o valor

observado de y é maior do que o valor previsto \hat{y} , um resíduo negativo é obtido quando o valor observado de y é menor do que o valor previsto \hat{y} .

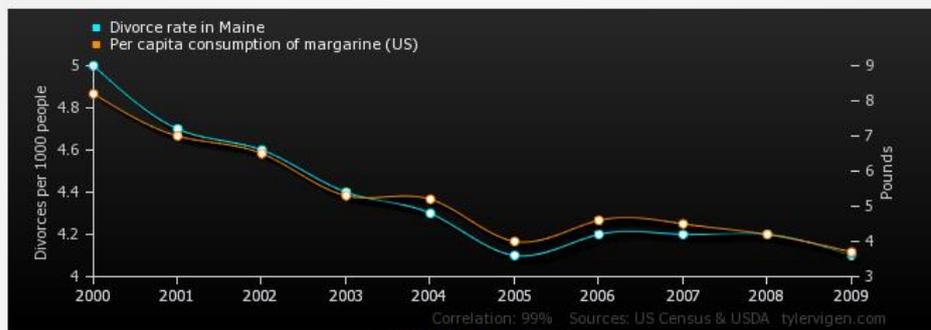
Verdadeiro. De acordo com Agresti e Finlay (2012, p.296), O erro de previsão, também chamado de resíduo, é a diferença entre o valor observado de y e o valor previsto \hat{y} . Um resíduo positivo resulta quando o valor observado de y é maior do que o valor previsto \hat{y} , assim $y - \hat{y} > 0$. Um resíduo é negativo quando o valor observado de y é menor do que o valor previsto \hat{y} , assim $y - \hat{y} < 0$.

(F) O modelo de regressão linear usa uma linha reta para descrever e mostrar o relacionamento entre duas variáveis – uma variável que chamamos de dependente (y) e uma outra variável que chamamos de explicativa (x). A inclinação b da equação que prevê a relação entre essas duas variáveis traz informações, justamente, sobre a direção da associação entre essas duas variáveis – se positiva ou negativa – e também sobre a força dessa associação – se é forte, fraca ou moderada.

Falso. Agresti e Finlay (2012, p.301) pontuam bem essa questão. A inclinação b da equação que prevê a relação entre essas duas variáveis traz informações sobre a direção da associação entre duas variáveis – se positiva ou negativa. No entanto, a inclinação b da equação não nos informa sobre a força dessa associação. A correlação entre duas variáveis é que a responsável por indicar sobre essa força de associação entre duas variáveis.

(V) O gráfico abaixo apresenta a associação e a correlação entre a taxa de divórcios em Maine e o consumo de margarina nos EUA. A correlação apresentada é positiva e no valor de 0,99. No entanto, podemos afirmar que essa associação se trata de uma associação espúria entre as variáveis taxa de divórcio e consumo de margarina, dado que as relações entre as variáveis não são dependentes entre si, ou seja, não apresentam uma relação causal entre si.

Divorce rate in Maine correlates with Per capita consumption of margarine (US)



| | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|------|------|------|------|------|------|------|------|------|------|
| Divorce rate in Maine Divorces per 1000 people (US Census) | 5 | 4.7 | 4.6 | 4.4 | 4.3 | 4.1 | 4.2 | 4.2 | 4.2 | 4.1 |
| Per capita consumption of margarine (US) Pounds (USDA) | 8.2 | 7 | 6.5 | 5.3 | 5.2 | 4 | 4.6 | 4.5 | 4.2 | 3.7 |

Correlation: 0.992558

Fonte: <http://www.tylervigen.com/spurious-correlations>

Verdadeiro. Agresti e Finlay (2012, p.345) explicam que relações espúrias são aquelas em que a associação entre x e y não guardam relações de causa-consequência entre elas, mas acabam sendo fruto da influência de uma terceira variável que impacta as duas primeiras (que seguem sem relação direta entre si).

Exercício 02 (4 pontos)

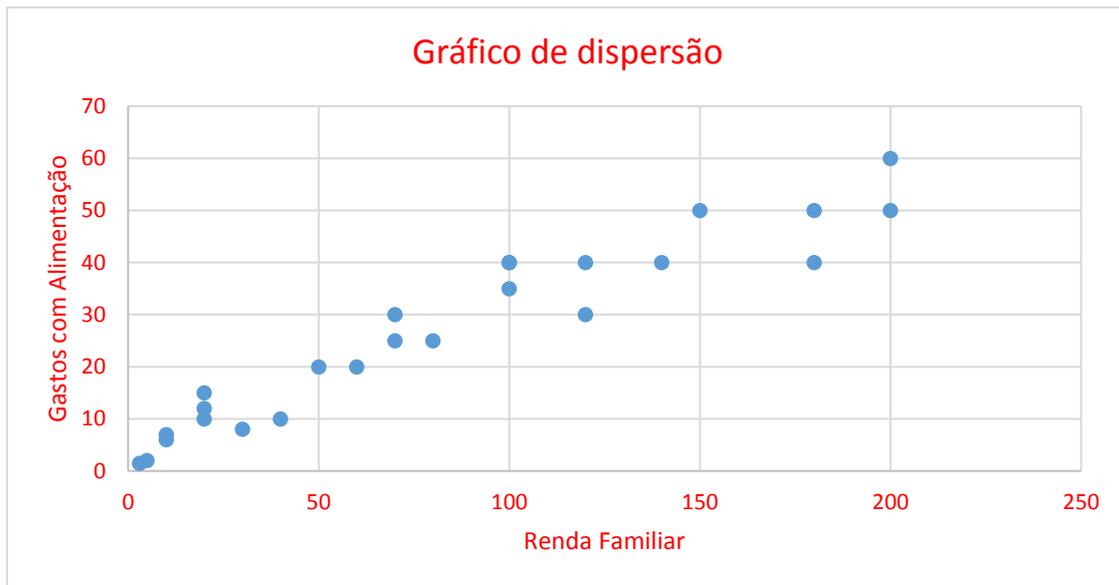
Os dados a seguir¹ correspondem à variável renda familiar e gasto com alimentação (em unidades monetárias) para uma amostra de 25 famílias.

| Renda Familiar (X) | Gasto com Alimentação (Y) |
|--------------------|---------------------------|
| 3 | 1,5 |
| 5 | 2,0 |
| 10 | 6,0 |
| 10 | 7,0 |
| 20 | 10,0 |
| 20 | 12,0 |

¹ Os dados também podem ser encontrados no arquivo em Excel disponível no moodle.

| | |
|-----|------|
| 20 | 15,0 |
| 30 | 8,0 |
| 40 | 10,0 |
| 50 | 20,0 |
| 60 | 20,0 |
| 70 | 25,0 |
| 70 | 30,0 |
| 80 | 25,0 |
| 100 | 40,0 |
| 100 | 35,0 |
| 100 | 40,0 |
| 120 | 30,0 |
| 120 | 40,0 |
| 140 | 40,0 |
| 150 | 50,0 |
| 180 | 40,0 |
| 180 | 50,0 |
| 200 | 60,0 |
| 200 | 50,0 |

a-) Construa o diagrama de dispersão da variável gasto com alimentação (y) em função da renda familiar (x). O que é possível interpretar com o gráfico?



O gráfico mostra que as variáveis renda familiar (x) e Gastos com alimentação (y) guardam uma relação entre si. Quanto maior a renda familiar, maiores são os gastos com alimentação. A relação também se mostra linear.

b-) Calcule e interprete o coeficiente de correlação entre essas variáveis.

$$\bar{X} = 83,120 \quad \bar{Y} = 26,660 \quad \sum_{i=1}^{25} X_i^2 = 271934 \quad \sum_{i=1}^{25} Y_i^2 = 24899,250$$

$$\sum_{i=1}^{25} Y_i X_i = 80774,500$$

$$r = \frac{S_{XY}}{\sqrt{S_X S_Y}} = \frac{\sum_{i=1}^{25} X_i Y_i - 25 \bar{X} \bar{Y}}{\sqrt{S_X S_Y}} = 0,954$$

c-) Obtenha a equação de regressão do gasto com alimentação em função da renda familiar.

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{i=1}^{25} X_i Y_i - 25\bar{X}\bar{Y}}{S_{XX}} = \frac{80774,5 - 25(83,12)(26,66)}{271934 - 25(83,12)^2} = 0,256$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = 26,66 - 0,256(83,120) = 5,380$$

A reta de regressão estimada da variável Gasto de alimentação (Y) em função da Renda familiar (X) é

$$\hat{Y} = 5,380 + 0,256X$$

d-) Qual o significado prático do valor da inclinação da reta de regressão do item (c)?

O valor $\beta_1=0,256$ significa que estima-se que para cada aumento de uma unidade monetária da renda familiar ocorre um acréscimo em média de 0,256 unidades no gasto com alimentação.

Exercício 3 (4 pontos)

Uma das relações causais mais estudadas na ciência política tenta entender se e como o desempenho da economia afeta a aprovação presidencial em um determinado ano. Considere os seguintes dados referentes ao Brasil para responder aos itens:

| Presidente | Ano | Crescimento PIB em % | Aprovação do presidente |
|------------|------|----------------------|-------------------------|
| Lula | 2003 | 1,3 | 42,8 |
| Lula | 2004 | 5,7 | 39,3 |
| Lula | 2005 | 3,2 | 32,1 |
| Lula | 2006 | 4,0 | 45,2 |
| Lula | 2007 | 6,1 | 48,6 |
| Lula | 2008 | 5,2 | 63,0 |
| Lula | 2009 | -0,2 | 68,0 |
| Lula | 2010 | 7,6 | 78,0 |
| Dilma | 2011 | 3,9 | 48,0 |
| Dilma | 2012 | 1,8 | 61,6 |
| Dilma | 2013 | 2,7 | 40,5 |

| | | | |
|-------|------|------|------|
| Dilma | 2014 | -0,1 | 37,5 |
| Dilma | 2015 | -3,8 | 12,3 |
| Temer | 2016 | -3,6 | 14,0 |
| Temer | 2017 | 1 | 6,0 |

Aprovação do presidente: média anual das pesquisas Datafolha com os entrevistados que responderam Ótimo/Bom Crescimento do pib: IBGE

Representação gráfica das variáveis aprovação do presidente e crescimento do PIB



a-) Considerando o PIB como a variável independente que explica a aprovação presidencial, calcule a covariância e o coeficiente de correlação (“r de Pearson”) entre as variáveis X e Y. O que você pode inferir destes cálculos? Você consegue relacionar os resultados dos cálculos com a interpretação visual do gráfico acima?

Para o caso em análise, trata-se de uma amostra. Portanto, a fórmula da covariância a ser utilizada é:

$$cov_{xy} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n-1} = \frac{601,362}{14} = 42,95$$

A partir das fórmulas de variância e covariância amostrais, temos que a correlação é de:

$$\rho = \frac{cov_{xy}}{\sqrt{var_x var_y}} = \frac{42,95}{\sqrt{11,14 \times 426,74}} = \frac{42,95}{\sqrt{4753,88}} = \frac{42,95}{68,95} = 0,62$$

Tendo em vista que a fórmula adotada por Kellstedt e Whitten (p. 183) é a da covariância populacional, também aceitamos as respostas dadas a partir dessa fórmula. Porém, o resultado da correlação é o mesmo, uma vez que são usadas as fórmulas de variância populacionais.

Usando a fórmula adotada por Kellstedt e Whitten (p. 183) o resultado seria o seguinte:

A análise superficial do gráfico já indicava uma tendência de correlação positiva entre as variáveis X (crescimento do PIB) e Y (aprovação presidencial), embora possamos verificar que algumas observações fogem ao esperado (baixo crescimento do PIB com relativa alta popularidade, por exemplo). A tendência geral parece ser a de uma relação positiva. Quanto maior o PIB, maior tende a ser a aprovação presidencial. Os dados calculados de covariância e coeficiente de correlação indicam um grau de correlação entre a variável X e Y significativa (40,01 e 0,62 respectivamente). A covariância limitado poder de interpretação porque é influenciada pela magnitude e pela escala de mensuração das variáveis, mas o coeficiente de correlação (r) indica de forma mais padronizada que, de fato, há uma correlação positiva entre as variáveis (0,62).

Exercício 4 Pós-Graduação (5 pontos)

Em um artigo publicado e disponível no moodle, Marques, Leite, Mendes e Ferreira (2009)² analisam o impacto do programa de transferência de renda condicionada Bolsa Família nos resultados eleitorais das eleições presidenciais de 2006. Com base na leitura do artigo³, responda as questões abaixo:

a-) Qual o problema de pesquisa abordado pelo texto? Qual o fenômeno de interesse dos autores e quais as principais variáveis explicativas que os autores mobilizam para responder ou explicar esse fenômeno? Como os autores mensuram essas variáveis? (Máximo de 10 linhas)

² MARQUES, R. M. et al. Discutindo o papel do Programa Bolsa Família na decisão das eleições presidenciais brasileiras de 2006. **Brazilian Journal of Political Economy**, v. 29, n. 1, p. 114–132, mar. 2009.

³ Dica: Dê ênfase em sua leitura para as seções de Metodologia da Análise e para os Resultados Obtidos

O texto se propõe a investigar a relação entre o impacto do programa Bolsa Família nas eleições presidenciais de 2006. Os autores pretendem explicar o resultado do sucesso eleitoral do candidato à presidência Luiz Inácio Lula da Silva, mensurado através da proporção de votos válidos recebido pelo candidato nos municípios. Como principal variável explicativa os autores mobilizam a cobertura do programa bolsa família, mensurada pela proporção/ parcela da população nos municípios que era atendida/beneficiada pelo programa.

b-) Na seção do artigo “Metodologia de Análise” os autores explicam:

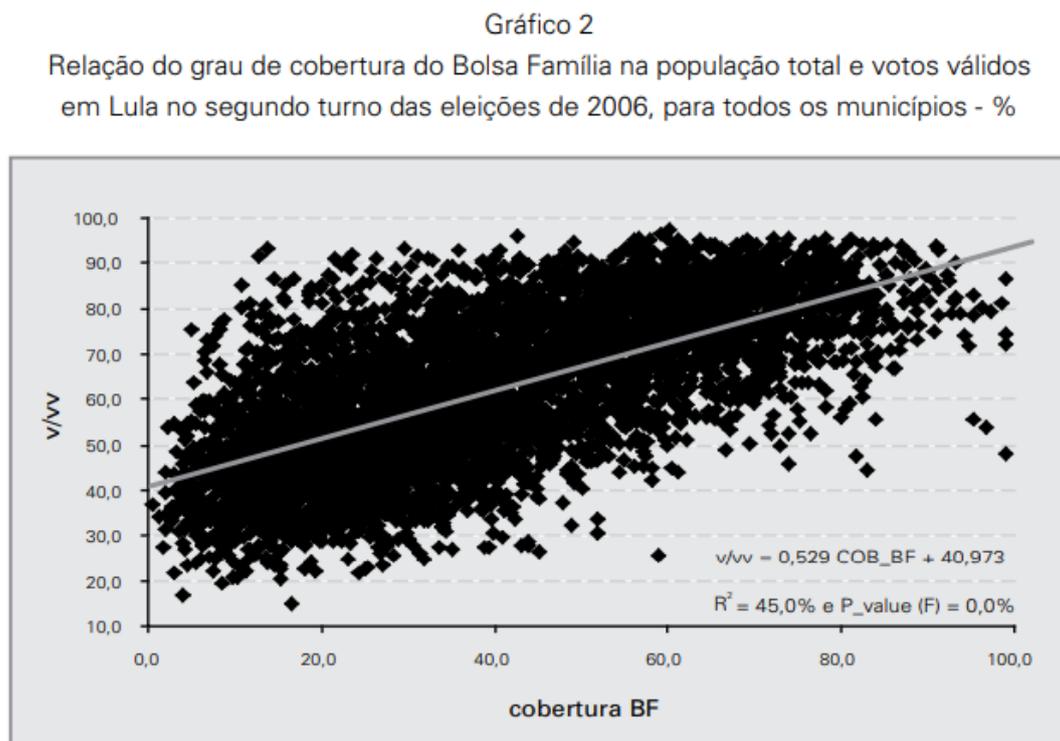
A relação do grau de cobertura (sobre a população total) com a porcentagem dos votos válidos recebidos por Lula foi analisada em três níveis: municipal, estadual e por grandes regiões geográficas. A fim de verificar se este resultado poderia ser influenciado por outras características dos municípios brasileiros, que não o nível de pobreza (posto que o PBF é dirigido exatamente ao segmento mais pobre da população), isto é, se não existe **correlação espúria**, foram consideradas algumas dessas possíveis características. A investigação se deu inicialmente através de regressões simples entre cada uma destas características e o grau de cobertura, a fim de identificar a existência de uma relação estatisticamente aceitável ou não entre elas. (MARQUES et al., 2009, p. 121 e 122, grifo nosso)

O que é uma correlação espúria? Por que os autores demonstram tanta preocupação com ela? Quais as correlações espúrias que os autores verificam a possibilidade de existência em sua pesquisa? Como os autores realizaram tais verificações? (Máximo de 20 linhas)

Segundo Agresti e Finlay (2012, p.345) associações espúrias são aquelas nas quais a relação entre x e y é dependente de uma terceira variável, dada por x_2 , e quando essa mesma relação de dependência desaparece quando x_2 é controlada. No caso do texto de Marques, Leite, Mendes e Ferreira (2009), os autores realizam testes para a verificação da ocorrência de correlações espúrias ou não entre as seguintes características: (i) distribuição do coeficiente de participação do eleitorado na população com a finalidade de observar se não existiam diferenças significativas entre os estados e municípios da amostra; (ii) se o comportamento do relacionamento entre os dois indicadores não estaria associado ao porte populacional dos municípios e (iii) se a caracterização dos municípios como rural ou urbano, bem como sua distribuição entre a capital, a região metropolitana e o interior não poderiam induzir a uma relação sistemática entre o grau de cobertura do Bolsa Família e os votos recebidos por Lula.

A verificação da consistência dessas relações espúrias ou não foram feitas com o uso e aplicação de regressões lineares simples. No primeiro caso, a relação entre a participação do eleitorado na população e a proporção de votos no candidato Lula não se mostra estatisticamente significativa. No segundo caso, apesar de existir uma correlação entre a participação do eleitorado na população total e o grau de cobertura do Bolsa Família, a mesma relação não é encontrada quando com a porcentagem de votos no candidato Lula. No terceiro caso, os valores do grau de cobertura do PBF na população total e da proporção dos votos válidos obtidos por Lula, associados à média dos municípios predominantemente rurais e os urbanos, são diferentes. Contudo, ao se testar a significância dessas diferenças (observadas tanto para a cobertura como para a proporção de votos em Lula), concluiu-se que, a 5% de significância, os valores são indiferentes, isto é, iguais.

c-) Observe o gráfico abaixo extraído do artigo:



Quais informações o gráfico mostra? (Dica: não responda resumidamente e interprete todos os elementos que compõem o gráfico. Máximo de 10 linhas)

O gráfico mostra uma associação positiva entre a cobertura do Bolsa Família e a proporção de votos no candidato Lula nos municípios brasileiros. Trata-se de um gráfico

de dispersão, onde cada ponto do Gráfico corresponde a um município brasileiro. A relação é positiva na medida em que o aumento da cobertura do Bolsa Família, também ocasiona um aumento na proporção de votos para o candidato Lula. Há também uma equação da reta que mostra que a proporção de votos no candidato Lula aumenta em 0,529 com o aumento de uma unidade da cobertura do Bolsa Família e que quando a cobertura do programa Bolsa Família é igual a zero, a proporção de votos no candidato Lula é igual a 40,97. Há também a informação do R^2 que nos indica o quanto a variável cobertura do Bolsa Família está explicando a proporção de votos no candidato Lula. Nota-se que o valor é alto, o modelo explica 45,0% da variação dos votos no candidato Lula nas eleições de 2006.

d-) Discorra sobre os resultados encontrados pelos autores e o problemas de pesquisa apresentado inicialmente por eles. Vocês acham que os resultados respondem à questão central levantada pelos autores? Os autores seguiram o passo a passo para a verificação de associação entre as variáveis? Vocês conseguem pensar em problemas ou sugestões para a análise dos autores? (Máximo 15 linhas)

A resposta esperada é um pouco mais solta. A princípio os alunos podem responder que os autores respondem à questão inicial do texto que é explicar os votos no candidato Lula em 2006 pela distribuição do Bolsa Família. Também podem concordar que os autores investigaram cada uma das possibilidades de associação entre as variáveis e a correlação entre elas. E que, finalmente, estimaram a equação da regressão linear. Contudo, aqui os alunos também podem apontar alguns problemas e sugestões como:

- Os autores não apresentam os resultados do modelo multivariado, ou seja, com todas as variáveis de controle juntas;
- Os autores não analisam a distribuição de votos para os outros candidatos;
- Entre outros;

Boa Lista!