

**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
FLP0406 Métodos e Técnicas de Pesquisa em Ciência Política**

**1º semestre / 2018
Prof. Glauco Peres da Silva**

LISTA DE EXERCÍCIOS 04

Data de entrega: 09/04/2018 (noturno) e 11/04/2018 (vespertino).

Exercício 01 (2 pontos)

Marque “Verdadeiro” (V) ou “Falso” (F) para as sentenças que serão apresentadas a seguir. Se você escolher “Falso” **justifique** sua escolha em, no máximo, 5 linhas.

- () Chamamos de escore-z a distância entre y (média da distribuição da probabilidade) e μ (valor da variável observada) medida em desvios padrão.
- () As distribuições de probabilidade das variáveis contínuas atribuem probabilidades a intervalos de números. Assim, a probabilidade de que o valor de uma variável esteja em algum dos intervalos é 1, enquanto que a probabilidade de o intervalo conter todos os valores possíveis que a variável assume varia de 0 a 1.
- () O erro padrão fica menor quanto maior é o tamanho da amostra dado por n .
- () Somente as variáveis discretas apresentam uma distribuição de probabilidade para os valores que elas assumem.
- () O Teorema do Limite Central afirma que para uma amostra aleatória com n grande, a distribuição amostral da média da amostra dada por \bar{y} é aproximadamente uma distribuição normal com média $\mu = 0$ e desvio padrão $\sigma = 1$.

Exercício 2 (4 pontos)

Para este exercício, utilize o banco de dados “CPDS_1960-2013” (e codebook), disponível no Moodle. Vamos trabalhar com a variável “effpar_ele”, discutida em sala de aula.

- a) Descreva brevemente o conceito que a variável está mensurando e como foi operacionalizada.

- b) Calcule a média, variância e desvio padrão desta variável. Reporte seus cálculos e resultados.
- c) É possível afirmar que os parâmetros obtidos acima correspondem aos parâmetros da população? Estamos trabalhando com uma população ou uma amostra?

Exercício 3 (4 pontos)

Para este exercício vamos utilizar o mesmo banco de dados do exercício acima e a mesma variável discutida acima.

- a) Seleccionamos uma amostra aleatória de 10 observações da variável “effpar_ele”. Calcule a média e o desvio padrão desta amostra.

2013	USA	2,192516
2013	Slovakia	4,387273
2012	Luxembourg	4,257221
1996	Luxembourg	4,728691
1970	Japan	3,392061
1981	Germany	3,097414
2011	New Zealand	3,164247
1975	Italy	4,073303
2002	United Kingdom	3,337483
1979	Denmark	4,995479

- b) Agora seleccionamos mais quatro amostras aleatórias da mesma variável, com o mesmo número de 10 observações. Calcule a média, variância e desvio padrão para cada uma delas. Os valores encontrados são os mesmos que aqueles do item “a”? Por que?

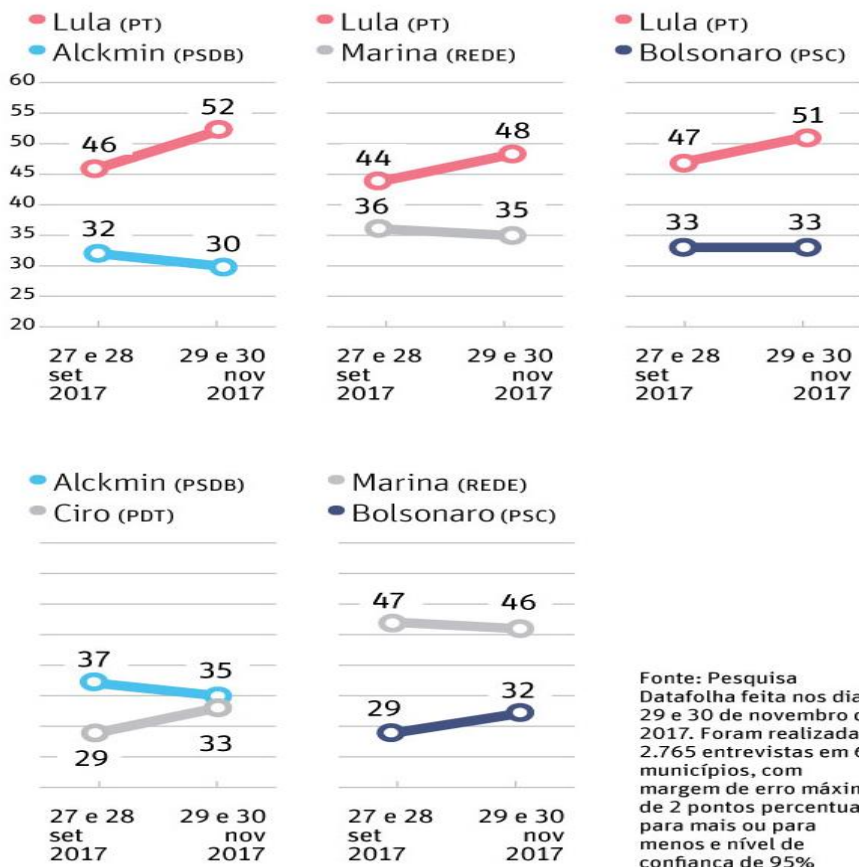
	Amostra 1	
1972	New Zealand	2,4309
2007	Poland	3,32656
1999	Netherlands	5,15413
1978	Austria	2,26771
1992	Denmark	4,85956
1979	Spain	4,29952
1986	Denmark	5,25285
2007	Bulgaria	5,83077
1969	Norway	3,51985
1994	Denmark	4,76665
	Amostra 2	
1994	Switzerland	7,42363
1984	Japan	3,66913
1990	Slovakia	5,81625
1965	Germany	3,14971
2007	Australia	3,01367
1977	Norway	3,7595
2001	Austria	3,81984
1997	Japan	4,09068
1989	Sweden	3,90643
1986	Netherlands	3,77926
	Amostra 3	
2011	Portugal	3,67381
2008	France	4,09311
1999	Austria	3,81984
2003	Bulgaria	3,95079
1965	Norway	3,83054
1969	Ireland	2,82462
2012	Estonia	4,77929
1981	Spain	4,29952
1978	Ireland	2,75849
1995	Denmark	4,76665
	Amostra 4	
1992	Germany	3,7359
1993	Switzerland	7,42363
2005	Austria	3,01952
1995	Sweden	3,65473
1998	Romania	6,12156
1979	Australia	3,11708
1996	Ireland	3,95026
2009	Croatia	4,10374
1995	Germany	3,74792
2003	Slovenia	5,16177

- c) Anote em uma planilha diferente as cinco médias que você encontrou a partir das cinco amostras dos itens “a” e “b”. Calcule a média deste conjunto de médias e compare-a com a média encontrada no exercício 2. À luz do Teorema Central do Limite, explique seus achados.
- d) Agora selecionamos novamente uma segunda amostra da mesma variável, desta vez com 100 observações (veja a seleção na planilha 2 do banco de dados, intitulada AMOSTRA). Calcule a média e o desvio padrão desta nova amostra. Existe diferença em relação a primeira? Estamos diante de uma demonstração do Teorema Central do Limite?

Exercício 4 Pós-Graduação (5 pontos)

Na lista 01, foi apresentado a você uma pesquisa que o Data Folha publicou no dia 02 de dezembro de 2017 sobre o possível resultado das eleições presidenciais de 2018. Para relembrar, segue novamente a divulgação dos resultados da pesquisa criada a partir de disputas hipotéticas para o segundo turno.

Lula lidera também nas **simulações para 2º turno**



Lula ganha em todos os cenários de segundo turno. Ele ampliou em quatro pontos percentuais sua vantagem, em relação à pesquisa feita no fim de setembro, no confronto com Alckmin (52% a 30%), Marina (48% a 35%) e Bolsonaro (51% a 33%).

O tucano empata tecnicamente com Ciro (35% a 33%) e Marina ganharia de Bolsonaro (46% a 32%).

(Fonte: <http://www1.folha.uol.com.br/poder/2017/12/1940171-lula-lidera-e-bolsonaro-se-consolida-em-2-aponta-datafolha.shtml>)

a-) Repare que para as simulações do 2º turno, o candidato Luiz Inácio Lula da Silva lidera as pesquisas em todos os cenários. Você, enquanto pesquisador, está interessado em entender melhor a intenção de votos dada a esse candidato. Quais variáveis você mobilizaria para tentar explicar esse fenômeno de interesse? Qual seria sua variável dependente e suas variáveis explicativas? Como as variáveis selecionadas ajudam a explicar o fenômeno de interesse? Qual teoria você mobilizaria para justificar essas escolhas? Justifique todas as suas escolhas. (Máximo de 20 linhas)

b-) Olhando apenas para o cenário colocado pelo Data Folha em que o candidato Lula (PT) e o candidato Bolsonaro (PSC) disputam o 2º turno das eleições temos o seguinte resultado:

Candidato	Intenção de Votos
Lula	51%
Bolsonaro	33%
Branco e Nulos	16%

Você, enquanto pesquisador, continua interessado em medir a probabilidade de sucesso do candidato Luiz Inácio Lula da Silva. Tendo em vista o cenário descrito acima, calcule a média e o desvio padrão para a amostra dessa pesquisa. Lembre-se de que estamos interessados no sucesso de um candidato, apenas. Demonstre seus cálculos e interprete os resultados (Máximo de 20 linhas)

c-) Calcule agora o erro-padrão e em seguida calcule o intervalo de confiança para 68%, 95% e 99% de confiança. O que o erro-padrão significa? O que é possível interpretar a partir dos intervalos de confiança calculados? (Máximo de 15 linhas)

d-) Em um cenário hipotético, suponha que nos resultados oficiais das eleições o candidato Lula tenha ganhado com 52,7% dos votos. Como podemos explicar a diferença entre os resultados da pesquisa do Data Folha e o resultado oficial das eleições? Essa diferença era prevista pelo Data Folha? Justifique sua resposta e aponte sugestões para a melhora na precisão das pesquisas de intenção de voto. (Máximo de 15 linhas)