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 06

Data de entrega: 23/04/2018 (noturno) e 25/04/2018 (vespertino)

Para essa lista, o termo "calcule" sempre implicará na apresentação no corpo do texto de todos os cálculos necessários para a obtenção dos valores especificados nas fórmulas correspondentes. Isso pode ser feito sem desconto de nota:

a) à mão, ou seja, sem o uso de ferramentas ou softwares estatísticos;

b) em ferramentas ou softwares de cálculo ou estatísticos, desde que todas as etapas estejam claras no desenvolvimento do exercício. A aluna pode, caso julgue necessário, complementar textualmente a resolução.

Exercício 1 (2 pontos)

Marque verdadeiro (V) ou falso (F) para cada uma das alternativas abaixo e justifique todas as suas respostas.

( ) – Em quaisquer distribuições assimétricas, aproximadamente 68%, 95% e 99,7% dos valores estarão a menos de – respectivamente – um, dois e três desvios-padrão de distância da média (tamanho máximo: 3 linhas).

(F). A chamada regra 68-95-99,7 se aplica somente a distribuições simétricas contínuas unimodais (Ver Sharpe et al, p. 247-248).

( ) – A distribuição t de Student converge para a distribuição normal padrão conforme aumenta o tamanho da amostra (tamanho máximo: 4 linhas).

(V). A distribuição t de Student tem n-1 graus de liberdade, sendo n o tamanho da amostra. Conforme aumenta o número de graus de liberdade da distribuição t, ela converge para a normal padrão (ver Agresti e Finlay, p. 146-147).

(   ) – Um exemplo de estimativa pontual[1] é o intervalo de confiança – ao nível de significância de 95% – de uma proporção estimada (tamanho máximo: 3 linhas).

(F). Uma estimativa pontual é uma avaliação de um parâmetro como um número único. O intervalo de confiança é um conjunto de números e, portanto, é uma estimativa intervalar (ver Agresti e Finlay, p. 131).

(   ) – Caso o pesquisador estime a média amostral $\bar{x}$ de uma amostra pequena e não aleatória, a fórmula $\bar{x} \pm 1{,}96\frac{s}{\sqrt{n}}$ (em que s é o erro-padrão estimado e n é o tamanho da amostra) é adequada para calcular o intervalo de 95% de confiança da média (tamanho máximo: 3 linhas).

(F). Em primeiro lugar, na fórmula acima, s é o desvio-padrão da amostra e $s/\sqrt{n}$ é o erro-padrão da média. Mais importante que isso, porém, é que a fórmula apresentada no enunciado para o intervalo de 95% de confiança da média é adequada para amostras grandes e aleatórias (ver Agresti e Finlay, p. 141). Obs.: caso a amostra seja não aleatória, a fórmula correta para o intervalo de confiança depende da técnica de amostragem e pode mesmo ser desconhecida, caso a amostra não seja representativa. Para amostras aleatórias pequenas, é necessário substituir 1,96 pelo nível crítico da distribuição t a 95% de confiança e n-1 graus de liberdade.

( ) – A interpretação intuitiva do intervalo de confiança de 95% da média de uma variável qualquer é que 95% dos valores observados para aquela variável na população pertencerão ao intervalo de confiança (tamanho máximo: 5 linhas).

(F). A interpretação do intervalo de confiança de 95% é que se fosse possível calcular a média amostral de cada uma de infinitas amostras aleatórias do mesmo tamanho, retiradas da mesma população, 95% dessas médias amostrais pertenceriam ao intervalo de confiança (ver Agresti e Finlay, p. 139). Obs.: essa interpretação é dita frequentista porque remete à frequência de médias amostrais possíveis que pertenceriam ao intervalo de confiança.

<u>Questão 2 (4 pontos)</u>

O *Latinobarómetro* é uma pesquisa de opinião pública realizada desde 1995 e que investiga diversos temas em 18 países latino-americanos. Um desses temas é a satisfação com a democracia. No ano de 2015, a seguinte pergunta constava do *survey*: *"En general, ¿diría Ud. que está muy satisfecho, más bien satisfecho, no muy satisfecho o nada*

[1] Estimativa pontual é um sinônimo de estimativa por ponto.

satisfecho con el funcionamiento de la democracia en (país)?"[2]. As respostas para o Brasil encontram-se na tabela abaixo.

Tabela 1 – Satisfação com o funcionamento da democracia (Brasil, 2015)

	Nº de casos	%/Total
Muy satisfecho	23	1,8%
Más bien satisfecho	239	19,1%
No muy satisfecho	546	43,7%
Nada satisfecho	359	28,7%
No responde	5	0,4%
No sabe	78	6,2%
(N)	(1.250)	100%

Muestras seleccionadas: Brasil (1250)

Fonte: *Latinobarómetro* (2015)

Responda as seguintes perguntas sobre a pesquisa descrita acima:

a) **(2 pontos)** Assumindo que esta é uma amostra aleatória da população brasileira, calcule a proporção de cidadãos que se consideram muito satisfeitos com a democracia e seu intervalo de confiança, aos níveis de confiança de 95% e de 99%.

Seguimos na resolução a notação de Agresti e Finlay.

n = 1250

Muito satisfeitos = 0,018

$$ep = \sqrt{\hat{\pi}(1-\hat{\pi})/n} = \sqrt{0{,}018(1-0{,}018)/1250} = \sqrt{0{,}0177/1250} =$$

$$= \sqrt{0{,}00001416} = 0{,}004$$

$$IC_{95\%} = 0{,}018 \pm 1{,}96 * 0{,}004 = 0{,}018 \pm 0{,}00784 = [0{,}01; 0{,}026]$$

$$IC_{99\%} = 0{,}018 \pm 2{,}576 * 0{,}004 = 0{,}018 \pm 0{,}0103 = [0{,}008; 0{,}028]$$

b) **(2 pontos)** A partir da leitura de Agresti e Finlay (2012) indicada, comente sobre a tendenciosidade[3] e eficiência deste estimador.

Neste caso, queremos estimar a proporção da população que responde que está muito satisfeita com o funcionamento da democracia no Brasil. Conforme Agresti e Finlay (2012, p. 135), a proporção amostral é uma média amostral e sua distribuição está

[2] "Em geral, você diria que está muito satisfeito, satisfeito, pouco satisfeito ou totalmente insatisfeito com o funcionamento da democracia no (país)?" (tradução nossa).

[3] Viés é um sinônimo de tendenciosidade comumente usado.

centrada na média (proporção) populacional. Dessa forma, pode-se afirmar que este é um estimador não tendencioso. Além disso, trata-se de um estimador eficiente, uma vez que seu erro padrão diminui conforme aumenta o tamanho da amostra. Como a amostra é aleatória relativamente grande (mais de 1000 entrevistados), temos motivos para supor que a estimativa da proporção é confiável e sua distribuição se concentra em torno do parâmetro populacional verdadeiro, dado que o estimador é eficiente e não tendencioso.

Exercício 3 (4 pontos)

A Tabela 2 apresenta o número de brasileiros adultos entrevistados pelo *Latinobarómetro* em 2015 por total de anos de estudo.

Tabela 2 – Anos de estudo dos entrevistados (Brasil, 2015)

Anos de estudo	Nº de casos	% do Total
Sem estudos	111	8,90%
2 anos	38	3,00%
3 anos	26	2,10%
4 anos	53	4,20%
5 anos	66	5,30%
6 anos	22	1,80%
7 anos	39	3,10%
8 anos	32	2,60%
9 anos	117	9,40%
10 anos	60	4,80%
11 anos	65	5,20%
12 anos	385	30,80%
Superior incompleto	114	9,10%
Superior completo	122	9,80%
Tamanho da amostra	1.250	100%

Fonte: *Latinobarómetro* (2015).

a) **(1 ponto)** Usando os dados da Tabela 2, calcule a escolaridade média da amostra (em anos de estudos) e o erro padrão da média. Para isso, suponha que:

- Os indivíduos sem estudo têm zero anos de escolaridade;
- Os indivíduos com superior incompleto têm 14 anos de estudo;
- Os indivíduos com superior completo têm 16 anos de estudo;
- A amostra da pesquisa é aleatória simples.

Para calcular a média, fazemos as substituições dos valores categóricos pelos valores numéricos indicados no enunciado, e fazemos a média dos anos de estudo ponderada pelo percentual de indivíduos em cada categoria, como mostramos na tabela R1. A escolaridade média dos entrevistados é, aproximadamente, 9 anos e meio.

Tabela R1 – Cálculo da média amostral

Anos (categórica)	Anos	% do Total	Anos * % do Total
Sem estudos	0	8,90%	0,000
2 anos	2	3,00%	0,060
3 anos	3	2,10%	0,063
4 anos	4	4,20%	0,168
5 anos	5	5,30%	0,265
6 anos	6	1,80%	0,108
7 anos	7	3,10%	0,217
8 anos	8	2,60%	0,208
9 anos	9	9,40%	0,846
10 anos	10	4,80%	0,480
11 anos	11	5,20%	0,572
12 anos	12	30,80%	3,696
Superior incompleto	14	9,10%	1,274
Superior completo	16	9,80%	1,568
Média (somatório da quarta coluna)			**9,525**

Obs.: Alternativamente e de maneira mais precisa, é possível multiplicar o valor de cada categoria (Anos) pelo número de casos daquela categoria, fazer o somatório desses produtos e dividir o resultado pelo tamanho da amostra. A média amostral calculada dessa forma seria **9,5144**. A diferença provavelmente decorre de arredondamentos nos percentuais do total apresentados pelo *Latinobarómetro*.

A tabela R2 apresenta esquematicamente o cálculo do erro-padrão da média. Para calcular o erro padrão da média, primeiramente precisamos achar o desvio-padrão amostral. Para calcular o **desvio-padrão amostral**, calculamos os quadrados dos desvios para cada categoria, multiplicamos cada um deles pelo número de casos na categoria e fazemos o somatório desses produtos (SQD, na tabela). Em seguida, dividimos esse somatório pelo número de observações (1250) menos um e teremos a variância amostral (s^2). A raiz da variância amostral é o desvio-padrão amostral (s). Por fim, dividimos o desvio padrão amostral pela raiz do número de observações e teremos o **erro padrão da média**.

Tabela R2 – Cálculo do erro padrão da média

Anos	Nº de casos	Desvio (Anos - 9,525)	Desvio²	Desvio² * Nº de casos
0	111	-9,525	90,726	10070,544
2	38	-7,525	56,626	2151,774
3	26	-6,525	42,576	1106,966
4	53	-5,525	30,526	1617,858
5	66	-4,525	20,476	1351,391
6	22	-3,525	12,426	273,364
7	39	-2,525	6,376	248,649
8	32	-1,525	2,326	74,420
9	117	-0,525	0,276	32,248
10	60	0,475	0,226	13,538
11	65	1,475	2,176	141,416
12	385	2,475	6,126	2358,366
14	114	4,475	20,026	2282,921
16	122	6,475	41,926	5114,926
Soma dos quadrados dos desvios (SQD)				26838,381
s² = Variância amostral (SQD/(n-1))				21,488
s = Desvio-padrão amostral (Raiz de s²)				4,636
Erro-padrão da média ($s/\sqrt{n}$)				**0,131**

b) **(1 ponto)** Usando os resultados do item anterior, calcule o intervalo de confiança da média aos níveis de 95% e 99%. Qual dos intervalos de confiança é maior? Por quê? Baseado nessas estimativas, qual o nível de ensino que representa a tendência central da população brasileira adulta? Para isso, suponha que:

- A amostra do *Latinobarómetro* é aleatória simples;
- A escolaridade em anos de estudos é uma variável contínua.

Usamos na resolução a notação de Agresti e Finlay.

$IC_{\alpha} = \bar{y} \pm z_{\alpha} \frac{s}{\sqrt{n}}$, em que α é o nível de confiança

$IC_{95\%} = 9{,}525 \pm 1{,}96 * 0{,}131 = 9{,}525 \pm 0{,}257 = [9{,}268; 9{,}782]$

$IC_{99\%} = 9{,}525 \pm 2{,}576 * 0{,}131 = 9{,}525 \pm 0{,}338 = [9{,}187; 9{,}863]$

O intervalo de 99% de confiança é maior que o intervalo de 95%. Isso faz sentido, pois você tem mais confiança que uma média amostral possível estará em um intervalo mais vasto de valores que em um intervalo mais restrito.

Os intervalos de confiança estão contidos entre 9 e 10 anos de escolaridade. Assim, de acordo com esses dados, o nível médio incompleto representaria a tendência central da população brasileira.

c) **(1 ponto)** Observe a tabela de frequência dos anos de estudos dos brasileiros da amostra de brasileiros do *Latinobarómetro* em 2015 com atenção. Qual a moda da distribuição amostral da escolaridade? Ela parece simétrica? Baseado em seus conhecimentos sobre a população brasileira, você acreditaria que essa amostra é realmente aleatória simples? Dica: você não precisa procurar dados precisos sobre a escolaridade no Brasil para responder essa pergunta (tamanho máximo: 10 linhas).

A moda da distribuição amostral da escolaridade é 12 anos (que corresponde a mais de 30% das observações). Essa distribuição não é simétrica, pois quase 50% das observações tem mais de 12 anos de escolaridade e pouco mais de 50% tem entre 0 e 11 anos de estudo. Sabemos que a quase universalização do ensino fundamental no Brasil é um fenômeno recente (dos anos 2000). Assim, não parece provável que uma amostral aleatória simples contivesse tantos brasileiros com 12 anos ou mais de escolaridade. Parece mais provável que a amostra tenha sido estratificada para conter um mínimo de indivíduos altamente escolarizados.

d) **(1 ponto)** Você decidiu que não é conveniente supor que a amostra do *Latinobarómetro* é aleatória simples e não conseguiu descobrir como fazer a correção amostral adequada para as pesquisas disponíveis. Por isso, você decidiu criar a sua própria pesquisa sobre satisfação com a democracia no Brasil. Suponha que você tem uma lista de todos os brasileiros adultos e fará uma amostra aleatória simples. Sua pergunta na pesquisa permitirá apenas duas respostas (satisfeito ou insatisfeito) e você tem como garantir que todos os indivíduos pesquisados responderão a pesquisa. Se você deseja uma margem de erro de 3 pontos percentuais e um nível de confiança de 98%, qual o tamanho mínimo que sua amostra deverá ter? Esse tamanho de amostra depende do número de brasileiros adultos? Dica: Use a abordagem conservadora ou "segura" do intervalo de confiança da proporção (tamanho máximo: 15 linhas).

Como só há duas respostas possíveis, a variável estudada será uma proporção de satisfeitos ou de insatisfeitos com a democracia no Brasil. Seja ϵ a margem de erro da proporção. Dado que a amostra será aleatória simples e que não teremos problemas com indivíduos que não responderão à pergunta, sabemos que

$$\epsilon = z_{98\%}\sqrt{\frac{\pi(1-\pi)}{n}}$$

Rearranjando os termos,

$$n = \frac{(z_{98\%})^2 \pi(1-\pi)}{\epsilon^2}$$

Sabemos que $\pi(1-\pi)$ atinge o valor máximo quando $\pi = 0{,}5$. Assim, a abordagem conservadora é substituir $\pi(1-\pi)$ por 0,25 (ver Agresti e Finlay, p. 149-150). Substituindo $\pi(1-\pi)$ por 0,25, $z_{98\%}$ por 2,326 e ϵ por 0,03, temos:

$$n = \frac{(2{,}326)^2 \times 0{,}25}{0{,}03^2} = \frac{5{,}429 \times 0{,}25}{0{,}0009} \cong 1503$$

Assim, a amostra deverá conter, no mínimo, **1503** indivíduos brasileiros adultos. Obs.: Caso se arredonde $z_{98\%}$ por 2,33, o tamanho mínimo obtido será **1508**.

Observe que o tamanho mínimo da amostra depende apenas do nível de confiança, da margem de erro e do erro padrão da proporção (que, no caso, é fixo, pois adotamos a postura conservadora). Assim, o tamanho da amostra não depende do tamanho da população de brasileiros adultos. De fato, se quiséssemos obter uma amostra, nas condições (margem de erro e nível de confiança) do enunciado, para estudar a população brasileira, a população portuguesa ou a população de qualquer outro país, concluiríamos que o tamanho mínimo da amostra deveria ser sempre 1503. Obs.: Se estivéssemos tratando de uma população muito pequena, menor que 1503 pessoas, não poderíamos usar essas técnicas de amostras grandes (ou infinitas) e teríamos de buscar técnicas de amostragem não cobertas nessa disciplina.

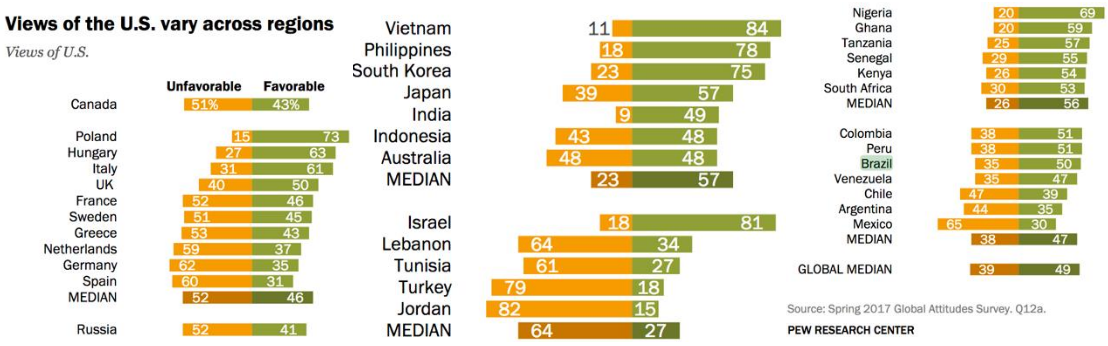
Questão 4 – Pós-Graduação (5 pontos)

Diversos analistas políticos vêm apontando para a possibilidade da presidência de Donald Trump ter afetado a visão que povos de outros países do mundo têm acerca dos Estados Unidos. A pesquisa *Global Attitudes Survey* do Pew Research Center, um instituto de pesquisas de opinião pública norte-americano, pode oferecer algumas pistas nesse sentido. Uma das perguntas incluídas neste questionário é: “12.a Please tell me if you have a very favorable, somewhat favorable, somewhat unfavorable or very unfavorable opinion of the United States”.

a) (0,5 ponto) Qual é a variável que a pergunta apresenta busca medir? Qual é o tipo desta variável? (tamanho máximo: 6 linhas)

A variável buscar medir a opinião positiva ou negativa de cidadãos em diferentes países do mundo sobre os Estados Unidos. Trata-se de uma variável categórica com 5 respostas possíveis: muito favorável (very favorable), um pouco favorável (somewhat favorable), um pouco desfavorável (somewhat unfavorable), muito desfavorável (very unfavorable) e não sabe ou não respondeu (DK/refused).

Na página 17 de seu relatório (disponível no Moodle), porém, os resultados para o ano de 2017 são apresentados somente em duas categorias.



b) **(0,5 ponto)** Como esta variável foi transformada para se alcançar os resultados apresentados no quadro acima? (tamanho máximo: 5 linhas)

Os resultados de muito favorável e um pouco favorável foram somados como *favoráveis* e os resultados de pouco desfavorável e muito desfavorável foram agregados em *desfavoráveis*.

A seguir, vamos trabalhar somente com os resultados para o Brasil. A tabela abaixo apresenta os resultados dessa pesquisa para os anos de 2014, 2015 e 2017.

	Favorável (%)	**Um pouco Favorável (%)**	**Um pouco Desfavorável (%)**	**Desfavorável (%)**	**Não sabe/ Não respondeu (%)**	**Total (%)**
Brasil, 2017	9	41	30	5	15	100
Brasil, 2015	18	55	17	6	5	100
Brasil, 2014	8	57	22	5	9	100

Fonte: adaptado de Pew Research Center (2017, p. 66)

No site http://www.pewresearch.org/methodology/international-survey-research/international-methodology/ é possível encontrar o desenho e tamanho da amostra.

c) **(2 pontos)** Utilizando os conceitos já estudados, explique como foi construída a amostra para esta pesquisa. Sabemos que diversas pesquisas de opinião pública no Brasil são realizadas a partir de pontos de fluxo para coleta dos dados. Contraponha os impactos dessas duas formas de desenho amostral (aquela utilizada na Global Attitudes Survey e a de pontos de fluxo) em seu impacto sobre a construção de intervalos de confiança. (tamanho máximo: 20 linhas)

Nesses três anos, a amostragem é probabilística em múltiplos estágios. A unidade primária de amostragem foram municípios segundo estratificação por estado. Nessa etapa, foram selecionados 5 municípios. Dentro desses municípios, estratificou-se por meio dos distritos escolhidos pelo censo. Nesses distritos são selecionadas residências onde se escolhem indivíduos, por meio do método de último aniversário. Não fica claro pela descrição se houve amostragem por conglomerados, mas é razoável se supor que sim. Trata-se portanto de uma amostragem probabilística (ou, ao menos, quase-aleatória) (Bulfarine e Bussab, p. 15-16), mas limitada pela idade (somente maiores de 18 anos). A amostragem por pontos de fluxo não é probabilística ou aleatória, mas sim uma amostra de conveniência.

A utilização de amostras aleatórias é fundamental para a construção de intervalos de confiança para médias de uma distribuição amostral, visto que é a aleatoriedade que permite a utilização do teorema do limite central (Agresti e Finlay, 2012; Kellstedt e Whitten, 2015). As informações advindas de amostras não aleatórias, como os pontos de fluxo, não são confiáveis para inferências sobre a população.

d) **(2 pontos)** Utilizando os resultados de 2017 e 2015, transforme as variáveis em uma variável dicotômica (favorável/desfavorável), de forma semelhante à realizada pelo instituto de pesquisa. A partir dessa nova variável, calcule a proporção estimada do percentual de opiniões favoráveis sobre os Estados Unidos e seu intervalo de confiança, a partir de um nível de confiança de 95%.

2015:

N = 1000

Favorável = 0,73

$ep = \sqrt{\hat{\pi}(1-\hat{\pi})/n}$

$ep = \sqrt{0{,}73(1-0{,}73)/1000} = \sqrt{0{,}1971/1000} = \sqrt{0{,}0001971} = 0{,}014$

$IC_{95} = 0{,}73 \pm 1{,}96 * 0{,}014 = 0{,}73 \pm 0{,}027 = [0{,}703; 0{,}757]$

2017:

N = 1008

Favorável = 0,50

$ep = \sqrt{\hat{\pi}(1-\hat{\pi})/n}$

$ep \ = \sqrt{0{,}50(1-0{,}50)/1008} = \sqrt{0{,}25/1008} = 0{,}016$

$IC_{95} = 0{,}50 \pm 1{,}96 * 0{,}016 = 0{,}50 \pm 0{,}031 = [0{,}469; 0{,}531]$