

Gabarito - Avaliação

PRI - 5003

Métodos Quantitativos para Política Comparada e Relações Internacionais

17/05/2018

Questão 1 [1 ponto]. Identifique, nos resumos abaixo, as variáveis dependentes, as variáveis independentes e as variáveis condicionais (quando houver).

- (a) **Roessler, P., & Ohls, D. (2018). *Self-Enforcing Power Sharing in Weak States. International Organization*, 72(2)**

Power sharing, in which elites from rival societal groups agree to share control of the central government, is a key source of domestic peace, enabling states to escape devastating cycles of exclusion and civil war. Yet the conditions giving rise to inclusive governance are not well understood. In contrast to existing scholarship that emphasizes the importance of external third-party mediation or strong formal institutions, we point to the structural roots of power sharing in which political inclusion stems from the distribution of societal power and the balance of threat capabilities it produces. Only when both the ruling group and a given rival group possess strong mobilizational capabilities, such that each could credibly threaten to recapture state power if excluded from the central government, does self-enforcing power sharing emerge. A strong rival induces the ruler to commit to power sharing and to reluctantly accept coup risk over civil war risk. The ruling group's own threat capabilities, in turn, constrain rivals from trying to convert their share of power into absolute power. Supported by extensive quantitative and qualitative evidence with particular reference to weak states in sub-Saharan Africa, we shed light on the conditions under which the distribution of violence within a state underwrites a peaceful and productive equilibrium. In doing so, we rethink how scholars approach the study of civil war. Rather than conceiving of it in terms of effective resistance, we model civil war as a contest for state power shaped by groups' capabilities to project force in the capital.

- (b) **Tajima, Y. (2018). *Political Development and the Fragmentation of Protection Markets: Politically Affiliated Gangs in Indonesia. Journal of Conflict Resolution*, 62(5)**

Why do gangs proliferate during democratization and decline in number during authoritarian consolidation? I utilize primary evidence of two Indonesian gangs to inform a model of protection gangs under varying states of political development.

Modeling gangs as territorial firms under different regulatory conditions, I attribute their number and political affiliation to the interaction between state capacity and political fragmentation. In weak states, gangs will lack political affiliations and their number will be determined by the scalability of their coercive capacities. In countries where states have the capacity to significantly constrain gangs, but lack significant costs for politicians to associate with them, gangs will seek political affiliation, trading coercive services for lax law enforcement. In such contexts, their number will be determined by state factionalization. Thus, gangs proliferate during democratization due to more political actors sharing state control. I assess the theory examining Indonesia's history of statebuilding and political transition.

- (c) **Dassonneville, R. & McAllister, I. (2018), *Gender, Political Knowledge, and Descriptive Representation: The Impact of Long-Term Socialization*. *American Journal of Political Science*, 62**

Successive studies have found a persistent gender gap in political knowledge. Despite much international research, this gap has remained largely impervious to explanation. A promising line of recent inquiry has been the low levels of women's elected representation in many democracies. We test the hypothesis that higher levels of women's elected representation will increase women's political knowledge. Using two large, comparative data sets, we find that the proportion of women elected representatives at the time of the survey has no significant effect on the gender gap. By contrast, there is a strong and significant long-term impact for descriptive representation when respondents were aged 18 to 21. The results are in line with political socialization, which posits that the impact of political context is greatest during adolescence and early adulthood. These findings have important implications not only for explaining the gender knowledge gap, but also for the impact of descriptive representation on political engagement generally.

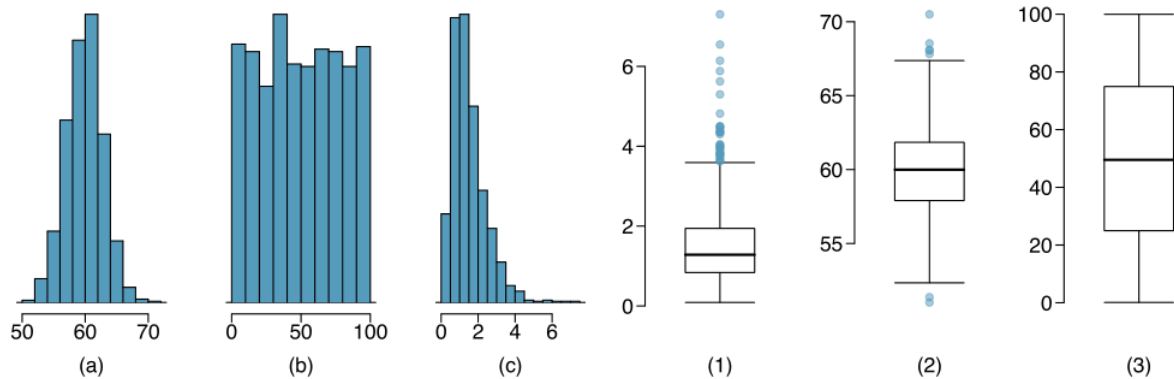
- (a) Variável dependente: divisão do poder político
Variável independente: capacidade de mobilização dos atores políticos
Variável condicional: fraqueza do Estado
- (b) Variável dependente: número de gangues com afiliações políticas
Variável independente: capacidade de coerção do Estado
Variável condicional: fragmentação do poder
- (c) Variável dependente: conhecimento político
Variável independente: representação feminina políticos eleitos
Variável condicional: não há

Questão 2 [1 ponto]. A respeito dos desenhos de pesquisa de metodologia comparada, responda:

- (a) Quais são os papéis das variáveis de “unificação” e “diferenciação” em MSSD e MDSD?
- (b) Identifique mecanismos que mitigam a limitação de estudos de caso para efeito de teste de hipóteses

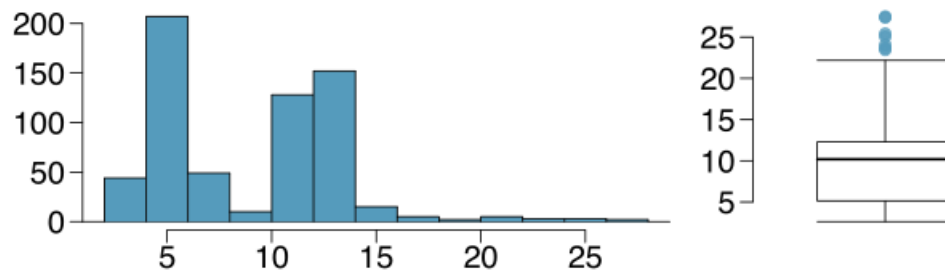
Cf. comentários na prova.

Questão 3 [0,5 ponto]. Descreva as distribuições (a), (b) e (c) como simétricas, assimétricas à direita ou assimétricas à esquerda. Em seguida, associe cada histograma ao *boxplot* correspondente.



- (a) é uma distribuição simétrica, e corresponde ao *boxplot* 2
- (b) é uma distribuição simétrica, e corresponde ao *boxplot* 3. Note que a diferença entre os *boxplots* 2 e 3 é a maior dispersão do último.
- (c) é uma distribuição assimétrica à direita, e corresponde ao *boxplot* 1

Questão 4 [0,5 ponto]. A imagem abaixo mostra duas representações gráficas de uma mesma variável. Responda:



- (a) Que características da distribuição são aparentes no histograma, mas não aparecem no *boxplot*?
- (b) Que características da distribuição são aparentes no *boxplot*, mas não aparecem no histograma?
- (c) Qual medida de tendência central seria mais adequada para representar uma observação típica dessa variável: a média ou a mediana? Justifique.
- (d) Qual medida de dispersão seria mais adequada para representar a dispersão dessa variável: o desvio-padrão ou o intervalo interquartil? Justifique.

- (a) No histograma podemos observar que a variável tem dois picos de valores. Essa característica bimodal não aparece de maneira explícita no *boxplot*
- (b) O *boxplot* permite distinguir claramente os valores dos *outliers* da variável
- (c) A mediana seria a representação mais adequada de um valor típico dessa variável, já que ela não é afetada pelos *outliers*. A média dessa variável seria puxada para cima devido à grande quantidade de pontos acima do Limite Superior do *boxplot*
- (d) De maneira semelhante à média, o desvio-padrão é afetado por valores extremos. Por ser uma medida de posição que não sofre a influência de *outliers*, o intervalo interquartil representa de maneira mais fiel a dispersão de variáveis altamente assimétricas.

Questão 5 [0,5 ponto]. A figura abaixo apresenta a renda dos brasileiros de acordo com os últimos dados disponíveis da PNAD. Com base nessas informações, responda:



- Por que não podemos dizer que o salário médio do brasileiro é de R\$ 1307?
- O gráfico sugere que o salário médio é maior ou menor do que o mediano? Justifique.
- Um trabalhador que ganha um salário mínimo está em que quartil da população?
- Qual é o valor do intervalo interquartil?
- Um colega diz que a figura traz boas notícias, pois demonstra que há mais pessoas ganhando entre R\$ 2000 e R\$ 3000 do que pessoas ganhando entre R\$ 1000 e R\$ 2000. Você concorda? Justifique.

- R\$ 1307 é o salário mediano do brasileiro, e esse valor só corresponderia à média se a distribuição de riqueza fosse perfeitamente simétrica. Como sabemos que salário é uma variável altamente assimétrica à direita, a média é significativamente maior do que a mediana
- O fato de a renda aumentar exponencialmente nos vintis mais altos sugere fortemente que essa variável tem uma longa cauda à direita. Nesses casos, a média apresenta valores mais altos do que a mediana
- Segundo quartil
- $IIQ = 3Q - 1Q = 2298 - 937 = 1361$

- (e) A interpretação do enunciado é baseada em uma leitura equivocada do gráfico: o eixo vertical não se refere à frequência dos salários, e sim a seus valores em cada vintil. Assim, a interpretação correta deveria levar em consideração o fato de que há 7 vintis (35%) da população ganhando entre R\$ 1000 e R\$ 2000, e 3 vintis (15%) ganhando entre R\$ 2000 e R\$ 3000. Portanto, há muito mais gente na primeira faixa salarial do que na segunda.

Questão 6 [1 ponto]. A tabela abaixo foi produzida a partir de uma *survey* nos Estados Unidos que procurou entender a relação entre a ideologia dos eleitores e suas posições a respeito dos direitos dos imigrantes. Os entrevistados foram perguntados se imigrantes que tenham entrado ilegalmente no país devem (i) ter o direito de trabalhar e solicitar cidadania americana, (ii) ter o direito de trabalhar com uma licença especial para visitantes, mas sem direito a pedir cidadania, ou (iii) ser extraditados. Com base nessas informações, responda:

	<i>Political ideology</i>			Total
	Conservative	Moderate	Liberal	
<i>Response</i> (i) Apply for citizenship	57	120	101	278
(ii) Guest worker	121	113	28	262
(iii) Leave the country	179	126	45	350
(iv) Not sure	15	4	1	20
Total	372	363	175	910

- (a) Qual é a probabilidade de um entrevistado selecionado aleatoriamente se identificar como conservador?
- (b) Se selecionarmos um entrevistado aleatoriamente entre os moderados, qual é a probabilidade de ele ser favorável à solicitação de cidadania por parte dos imigrantes ilegais?
- (c) Qual é a probabilidade de um entrevistado selecionado aleatoriamente ser moderado e a favor da extradição de imigrantes ilegais?
- (d) O que é mais provável: selecionar aleatoriamente um entrevistado pró-extradição entre os liberais, ou um entrevistado pró-cidadania entre os conservadores?

(a)

$$P(\text{Conservador}) = \frac{372}{910} = 0,4088 \rightarrow 40,88\%$$

(b)

$$P(\text{Favoravel} \mid \text{Moderado}) = \frac{120}{363} = 0,3306 \rightarrow 33,06\%$$

(c)

$$P(\text{Moderado e a favor da extradicao}) = \frac{126}{910} = 0,1385 \rightarrow 13,85\%$$

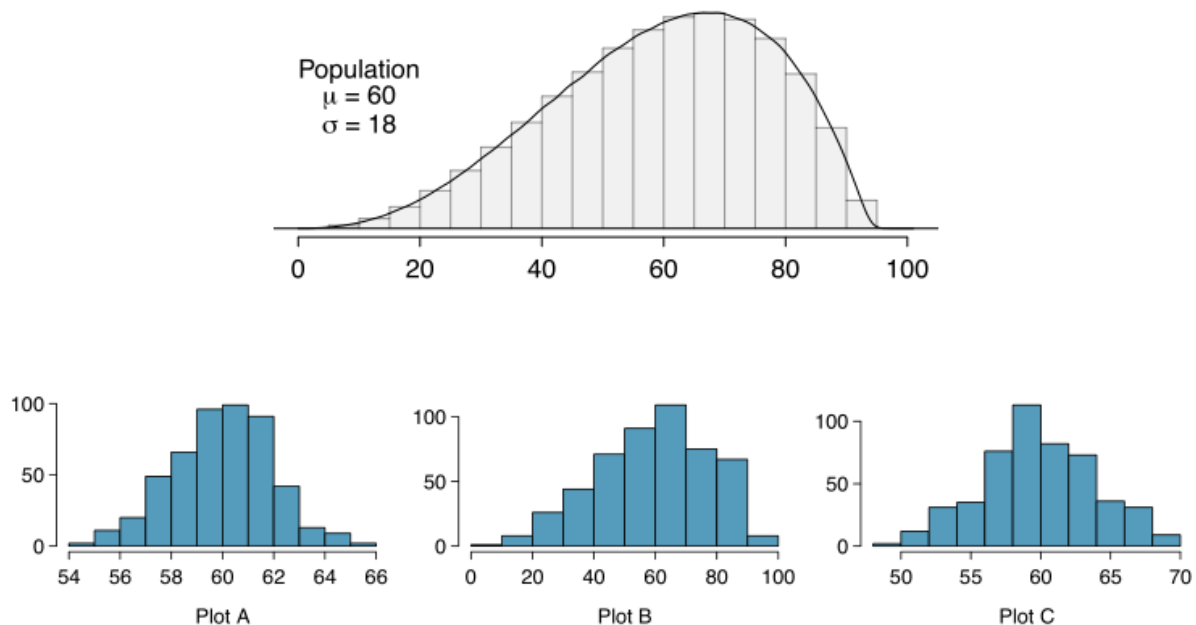
(d)

$$P(\text{Extradicao} \mid \text{Liberal}) = \frac{45}{175} = 0,2571 \rightarrow 25,71\%$$

$$P(\text{Cidadania} \mid \text{Conservador}) = \frac{57}{372} = 0,1532 \rightarrow 15,32\%$$

Questão 7 [1 ponto]. A imagem abaixo apresenta 4 distribuições. A superior é uma distribuição populacional com média 60 e desvio-padrão 18. As três inferiores representam, cada uma, uma das seguintes distribuições (em nenhuma ordem particular):

- Uma amostra única de 500 observações retiradas aleatoriamente da distribuição populacional
- Uma distribuição de 500 médias amostrais, em que cada amostra tinha 18 observações
- Uma distribuição de 500 médias amostrais, em que cada amostra tinha 81 observações



Associe as distribuições descritas acima às figuras A, B e C da parte inferior da imagem. Em cada caso, justifique sua classificação à luz do Teorema do Limite Central e da Lei dos Grandes Números.

- (Plot A) Trata-se da distribuição de 500 médias amostrais com $n = 81$. O Teorema do Limite Central explica por a distribuição dessas estatísticas tem um formato aproximadamente normal, mesmo que a distribuição populacional seja assimétrica. A Lei dos Grandes Números explica por que essa distribuição tem dispersão menor do que a do gráfico C.
- (Plot B) Trata-se da distribuição de uma única amostra de 500 observações. O gráfico B é, portanto, uma distribuição de observações, e não de estatísticas, e isso pode ser confirmado de a amostra aleatória ter distribuição semelhante à variável populacional. Não há convergência pontual (Lei dos Grandes Números) nem convergência em distribuição (Teorema do Limite Central).
- (Plot C) Trata-se da distribuição de 500 médias amostrais com $n = 18$. O Teorema do Limite Central explica por a distribuição dessas estatísticas tem um formato aproximadamente normal, mesmo que a distribuição populacional seja assimétrica. A Lei dos Grandes Números explica por que essa distribuição tem dispersão maior do que a do gráfico A.

Questão 8 [1,5 pontos]. A tabela abaixo cruza a preferência ideológica de uma amostra de eleitores americanos com suas identidades raciais auto-declaradas. A escala ideológica é tratada como um contínuo de 1 a 7 e segue a classificação habitualmente utilizada nos Estados Unidos, na qual indivíduos extremamente liberais se encontram no ponto 1, indivíduos moderados se encontram no ponto 4, e indivíduos extremamente conservadores estão no ponto 7.

Response	Race		
	Black	White	Other
1. Extremely liberal	10	36	1
2. Liberal	21	109	13
3. Slightly liberal	22	124	13
4. Moderate, middle of road	74	421	27
5. Slightly conservative	21	179	9
6. Conservative	27	176	7
7. Extremely conservative	11	28	2
	$n = 186$	$n = 1073$	$n = 72$

Uma pesquisadora quer testar a hipótese de que, em média, os eleitores que se identificam como brancos não são moderados: eles teriam uma tendência conservadora. Ela nota que a ideologia desses eleitores tem uma distribuição normal com média $\bar{y} = 4,15$ e variância $s^2 = 8$. Com base nessas informações, responda os itens abaixo:

- (a) Formalize as hipóteses e indique o valor de α .
- (b) Calcule a estatística-teste. Trata-se um teste t ou um teste z ?

- (c) Encontre o p-valor com base na tabela apropriada.
- (d) Apresente um desenho aproximado com os valores da estatística-teste, do valor crítico, do p-valor e da região crítica.
- (e) Decida se rejeita ou não rejeita a hipótese nula, explicitando o critério para tomada de decisão.
- (f) Apresente uma conclusão formal adequada (em palavras) a ser enviada à imprensa.

- (a) O primeiro passo neste exercício é definir um valor para a hipótese nula. Como a pesquisadora quer avaliar a tendência ideológica dos entrevistados, vamos fixar a hipótese nula no centro do espectro: eleitores moderados, com valor 4 na escala. O enunciado também afirma que a pesquisadora está trabalhando com a hipótese de que eleitores brancos têm tendência conservadora, então ela já tem uma noção de direção para o teste. Portanto, faremos um teste *unilateral* para descobrir se a ideologia média dos eleitores brancos é *maior* do que 4.

Podemos formalizar as hipóteses da seguinte maneira:

$$H_0 : \mu = 4$$

$$H_a : \mu > 4$$

Neste gabarito continuaremos o teste com um α de 0,05. Provas que utilizarem valores diferentes de α não perderão pontos, desde que as conclusões estejam coerentes com o valor escolhido.

- (b) Como o n é grande, os valores de t e z são convergentes, e por praticidade utilizaremos a última. Como a escala diz respeito a um contínuo ideológico, faremos um teste de hipótese para médias.

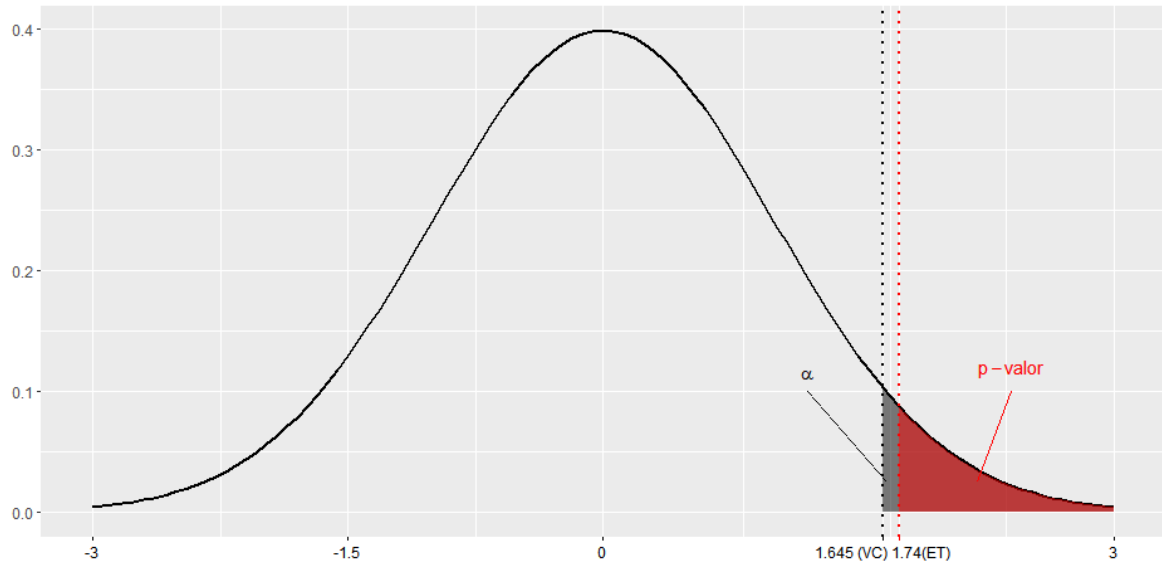
$$z = \frac{\bar{y} - \mu_0}{ep}, \text{ onde } ep = \frac{s}{\sqrt{n}}$$

$$z = \frac{4,15 - 4}{ep}, \text{ onde } ep = \frac{2,83}{32,76}$$

$$z = \frac{4,15 - 4}{0,0864}$$

$$z = 1,74$$

- (c) 0,0409
- (d) Será avaliado se o aluno identificou corretamente o valor crítico e a estatística-teste como *pontos* na distribuição, e o α e o p-valor como *áreas*.



(e) A decisão pode ser tomada por dois critérios

- Decisão por pontos: estatística-teste $>$ valor crítico
- Decisão por áreas: $p\text{-valor} < \alpha$

Por ambos os critérios rejeitamos H_0 a 5% de significância

(f) Ao escrever conclusões sobre testes estatísticos para um público mais amplo, é importante comunicar a incerteza sobre o resultado. A pesquisa nos mostra que temos indícios de que os eleitores brancos têm tendência conservadora, e esse resultado está associado a uma margem de erro e a um nível de confiança. Será descontada a nota de respostas contundentes e categóricas

Questão 9 [1 ponto]. Para cada um dos testes de hipótese abaixo, encontre o p-valor e determine se temos evidência para rejeitar H_0 .

- (a) $H_a : \mu > \mu_0; z = 1,91; \alpha = 0,05$
- (b) $H_a : \mu < \mu_0; z = -3,5; \alpha = 0,01$
- (c) $H_a : \mu \neq \mu_0; z = 0,83; \alpha = 0,1$
- (d) $H_a : \mu > \mu_0; z = 1,88; \alpha = 0,03$
- (e) $H_a : \mu \neq 3; z = 2,485; \alpha = 0,01$

(a) $p = 0,0281$. Rejeita-se H_0

(b) $p = 0,000233$. Rejeita-se H_0

(c) $p = 0,2033$. Não se rejeita H_0

(d) $p = 0,0301$. Não se rejeita H_0

(e) $p = 0,0065$. Rejeita-se H_0

Questão 10 [1 ponto]. A imagem abaixo mostra a estimação de um intervalo de confiança para a variável *expectativa de vida média feminina*, disponível no banco de dados *world95.dta*.

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
lifeexpf	100			67.82621	72.15379

(a) Escreva a interpretação correta para esse intervalo de confiança

(b) Encontre a média e do desvio-padrão da amostra, demonstrando passo a passo o raciocínio. Considere que todas as suposições estão sendo satisfeitas e arredonde os valores para duas casas decimais.

(a) Se sorteássemos amostras aleatórias de países repetidas vezes, em 95% delas a expectativa de vida feminina estaria contida nos limites do intervalo de confiança

(b) Primeiramente, encontraremos a média: como o intervalo de confiança é simétrico em torno dela, basta achar o ponto médio do intervalo de confiança.

$$\bar{y} = \frac{67,83 + 72,15}{2} = 69,99$$

Sabemos que o intervalo de confiança é composto da estimativa pontual \pm margem de erro. Agora que sabemos o valor da média, é simples concluir que a margem de erro é

$$M = 72,15 - 69,99 = 2,16$$

.

O próximo passo é utilizar a fórmula da margem de erro para descobrir o desvio-padrão. A margem de erro é definida por:

$$M = z \frac{s}{\sqrt{n}}$$

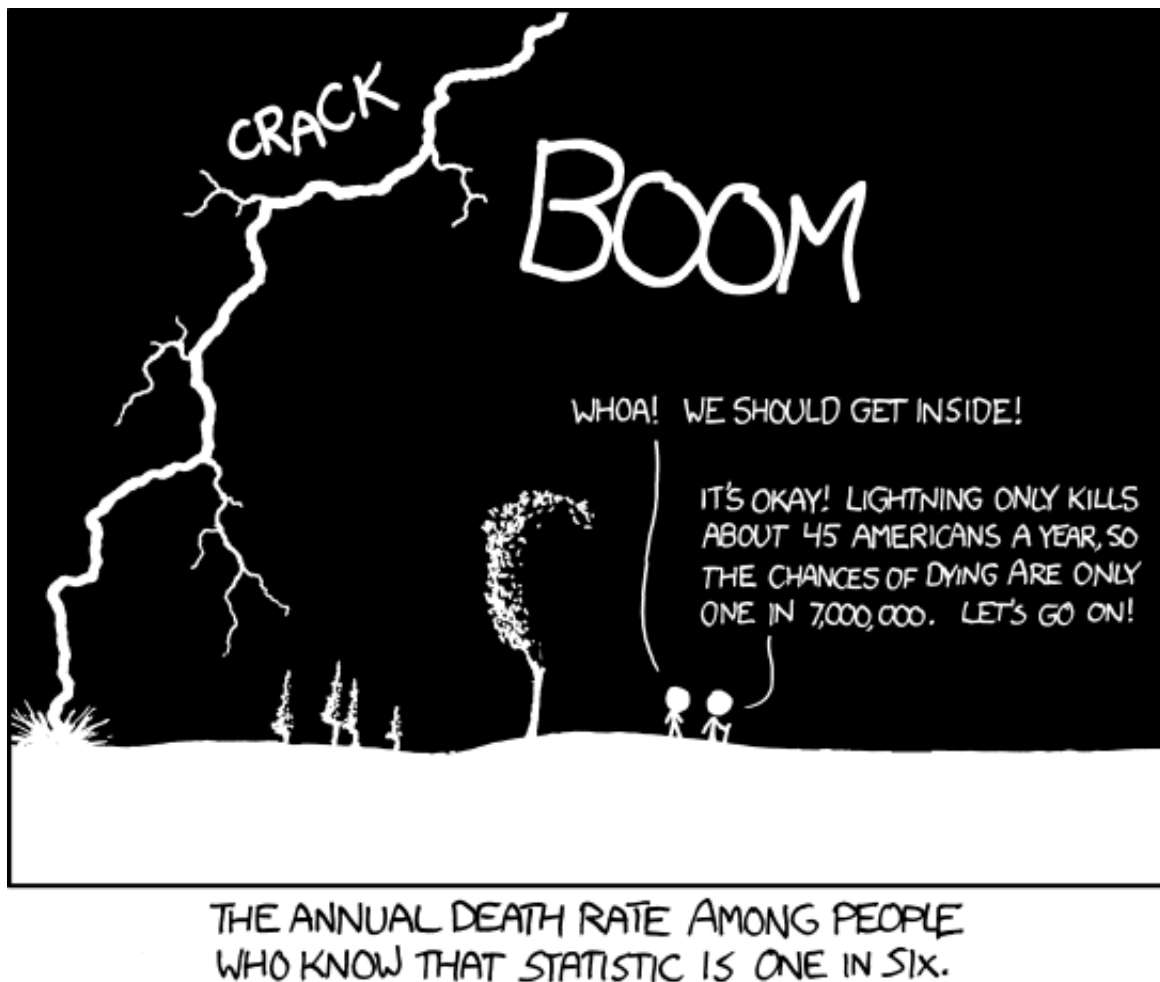
Nós já temos o valor de M (2,16). Também sabemos que o valor de z para um intervalo de confiança de 95% é 1,96. O valor de n, por sua vez é dado na imagem: 100 observações. Para achar o desvio-padrão, basta substituir os valores acima e isolar s:

$$2,16 = 1,96 \frac{s}{\sqrt{100}}$$

$$21,6 = 1,96s$$

$$s = 11,02$$

Questão 11 [1 ponto]. A tirinha abaixo ilustra um dos conceitos fundamentais para a inferência estatística. Identifique qual é esse conceito, indique como ele é ilustrado na imagem e aponte como ele é utilizado no teste de hipóteses.



O conceito ilustrado na tirinha é o de **probabilidade condicional**. O personagem diz que a probabilidade de uma pessoa ser atingida por um raio é de $1/7000000$ nos Estados Unidos. Portanto, não haveria motivo para se alarmar.

O comentário abaixo da imagem, porém, diz que a probabilidade de uma pessoa ser atingida **condicional a seu conhecimento da estatística nacional** é muito maior: $1/6$. Presumivelmente, pessoas que conhecem a estatística nacional se tranquilizariam em saber que a taxa de mortes é tão baixa e seriam menos cuidadosas. Essa negligência, por sua vez, aumenta a exposição desse subgrupo e, conseqüentemente, a probabilidade de que um de seus integrantes seja atingido.

Probabilidade condicional é um dos conceitos mais importantes da inferência estatística, e um dos lugares em que ele aparece com mais destaque é no teste de hipóteses. Quando calculamos o p-valor, o que temos é a probabilidade de aleatoriamente encontrarmos uma estimativa igual ou mais extrema do que a encontramos, **assumindo que**

H_0 é **verdadeira**. Em outras palavras, estamos respondendo à seguinte pergunta: se H_0 é verdadeira, qual é a probabilidade de eu observar dados tão extremos quanto os que coletei? (Lembrando: a amostra que coletamos é apenas uma de infinitas que poderiam ser feitas).

O detalhe de que a decisão em um teste de hipóteses é sempre condicional a H_0 ser verdadeira é relativamente sutil, e frequentemente esquecido mesmo por pesquisadores experientes, mas fundamental para a interpretação substantiva de testes estatísticos nas ciências sociais.