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 de 2024

Prof. Dr. Glauco Peres da Silva

Lista 3 - gabarito

Instruções

Esta lista faz parte da composição do sistema de avaliação desta disciplina.

Cada discente deve, assim, respondê-la individualmente e entregá-la no link destinado para coletar estas informações no *moodle* na data divulgada para este fim.

Não será tolerada cópia de respostas em nenhuma circunstância, embora seja encorajado que os alunos discutam entre si a respeito desta atividade.

Apresente respostas claras e justificadas, de forma a evidenciar o seu raciocínio.

Há perguntas que devem ser respondida apenas por alunos matriculados na pós-graduação. Elas estão indicadas no início do enunciado. Não serão computadas as respostas a estas perguntas para os alunos da graduação.

Questão 1

Para as afirmações abaixo, indique V para as sentenças verdadeiras e F para as sentenças falsas. Em caso de a sentença ser considerada falsa, justifique sua resposta.

a) Quanto maior o tamanho de uma amostra, maior o intervalo de confiança de um parâmetro estimado e assim, maior a precisão da estimativa;
FALSA. Expressamos um intervalo de confiança para a estimativa da média populacional da seguinte forma: $IC: \bar{x} \pm Z\frac{s}{\sqrt{n}}$. Pode-se notar que a sua amplitude depende inversamente do tamanho *n* da amostra considerada: quanto maior a amostra, menor o intervalo.

b) Quanto menor é o p-valor menor é a possibilidade de se rejeitar a hipótese nula, para um dado nível de significância;
FALSA. O p-valor indica a probabilidade de encontrarmos um valor igual ou mais extremo ao encontrado em nossa amostra, se o valor atribuído ao parâmetro de interesse pela hipótese nula for correto. Isto implica dizer que se o p-valor é um número muito baixo, maior a nossa chance de concluir que a hipótese nula está incorreta e, assim, rejeitá-la.

c) O intervalo de confiança também é uma medida sujeito ao erro amostral e, assim, pode não conter o parâmetro populacional correto;
VERDADEIRA.

Questão 2

Em maio de 2021, o instituto Datafolha divulgou uma pesquisa a respeito da avaliação popular sobre o mandato do atual presidente, Jair Bolsonaro. De acordo com esta pesquisa[1], foram ouvidas 2.071 pessoas em 146 municípios e a margem de erro máxima é de "2 pontos percentuais, para mais ou para menos, dentro do nível de confiança de 95%" (p. 2). Diz ainda a pesquisa que: "A aprovação ao governo Jair Bolsonaro (sem partido) caiu de 30% em março deste ano para 24% em maio, índice mais baixo registrado desde o início do seu mandato. No mesmo período, a reprovação oscilou de 44% para 45%, dentro da margem de erro, e a aprovação regular subiu de 24% para 30%. Há ainda 1% que não opinaram sobre o desempenho da gestão Bolsonaro." (p. 3)
Diante do exposto, responda às perguntas abaixo:

a) Por que o instituto informa que a "margem de erro máxima" é de 2 pontos percentuais? Mostre como se atinge este resultado nas condições apresentadas nesta pesquisa.

A margem de erro é a expressão que estabelece os limites inferior e superior de intervalo de confiança. Novamente, indicamos aqui a expressão de um intervalo de confiança: $IC: \bar{x} \pm Z\frac{s}{\sqrt{n}}$. O intervalo será tão maior quanto maior for *s*, para uma amostra de tamanho *n* dado e para dado nível de significância. Quando lidamos com proporção, o maior valor para *s* ocorrerá quando nossa estimativa média for igual a 50%. Neste caso, para um nível de significância de 5%, a margem de erro se torna igual a:

$$1{,}96 * \sqrt{\frac{50\% * (1 - 50\%)}{2071}} \cong 2{,}15\%$$

[1] Disponível em http://media.folha.uol.com.br/datafolha/2021/05/13/574c277a171a64f166dee28d083f08cfavbc.pdf

b) Por que não se deve apenas comparar o resultado da estimativa de março, quando a aprovação era de 30%, com o resultado obtido em maio, quando a aprovação foi de 24% (chegando à conclusão de que a aprovação se reduziu) ao buscar a identificação da opinião da população a respeito do mandato do presidente? Explique o(s) conceito(s) que afeta(m) esta comparação.

Não se deve comparar diretamente duas estimativas por conta do erro amostral. A diferença das médias, no caso, 30% e 24%, pode se dar apenas por conta deste erro, e não por alguma mudança na população de referência. É preciso incorporar a incerteza em ambas as estimativas para que se tenha condição de afirmar que há uma variação maior do que a que seria esperado pela existência de erro amostral.

c) **Pós-graduação**: para o exercício acima, qual é a maneira adequada de comparação entre os dois resultados das pesquisas? Explique.

A maneira de se comparar as médias está em supor que a sua diferença é zero da seguinte forma: Definimos uma variável *d* que seria igual a diferença entre as médias das duas populações:

$$d = \mu_1 - \mu_2$$

A partir daí, definimos o teste de hipóteses:

$$H_0: d = 0$$
$$H_1: d \neq 0$$

Como não temos os valores dos parâmetros populacionais, devemos considerar nossas estimativas e a incerteza em decorrência do erro amostral em ambas. A dificuldade está em estabelecer a estimativa do desvio padrão. Assim, a forma mais completa de se fazer isso usando o teste T a partir da seguinte expressão:

$$t = \frac{d - 0}{se(\hat{\pi}_1 - \hat{\pi}_2)}$$

Em que $\hat{\pi}_1$ e $\hat{\pi}_2$ são as estimativas da proporções médias no primeiro e no segundo grupo.
A fórmula para cálculo do erro padrão *se* da expressão acima se dá por:

$$se(\hat{\pi}_1 - \hat{\pi}_2) = \sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}}$$

No caso acima, não nos foi informado o tamanho da amostra na primeira pesquisa. Consideraremos o mesmo tamanho da segunda, apenas para efeitos didáticos. Então, temos que:

$$se(\hat{\pi}_1 - \hat{\pi}_2) = \sqrt{\frac{0{,}3(1-0{,}3)}{2071} + \frac{0{,}24(1-0{,}24)}{2071}} = 0{,}01377$$

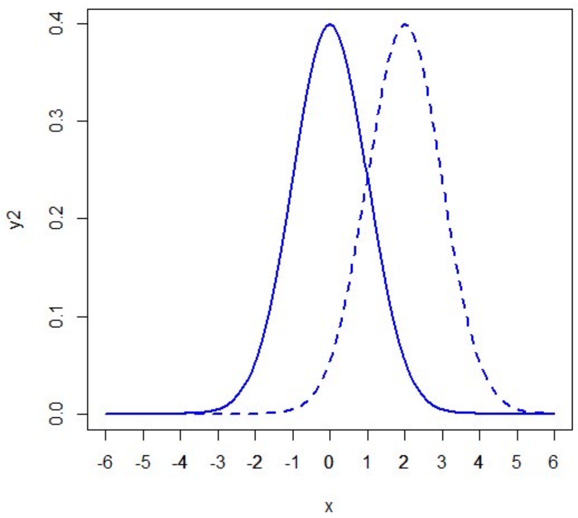
Então, temos que:

$$t = \frac{d - 0}{se(\hat{\pi}_1 - \hat{\pi}_2)} = \frac{0{,}06}{0{,}1377} = 4{,}359$$

Rejeitamos Ho, ou seja, há diferença entre as duas médias..

Questão 3

Considere o gráfico a seguir:



Ambos os gráficos pressupõem distribuições normais com mesma variância (1), mas com médias diferentes. Suponha que a linha sólida represente a distribuição dos dados que um pesquisador acredita refletir a verdadeira distribuição da população e que a linha pontilhada represente de fato a real distribuição. Considere que o pesquisador colheu uma amostra de 2.000 observações e que o valor do teste Z seja igual a 1,5.

Considerando um teste monocaudal, indique a alternativa que melhor representa o valor de α para o qual o pesquisador ainda não rejeitará a hipótese de que a média é igual a 0 e o valor de β para o qual o pesquisador cometerá o Erro Tipo II se não rejeitar a hipótese nula. Justifique sua resposta.

a) α=13,36% e β=15,43%

b) α = 6,68% e β = *30,85%*

c) α = 6,68% e β = *15,43%*

d) α = 13,36% e β = *30,85%*

A curva de linha cheia está situada em torno da média igual a zero, enquanto a curva de linha pontilhada se coloca em torno da média igual a 2, de acordo com o gráfico. Se as variâncias são iguais em ambas com valor 1, o teste Z que podemos fazer para cada um dos tipos de erro é:

Erro tipo I:

$$Z_{critico} = \frac{1{,}5 - 0}{1} = 1{,}5$$

Então:

$$\alpha = P(X \geq Z_{critico}) = 0{,}5 - 0{,}4332 = 6{,}68\%$$

Erro tipo II:

$$Z_{critico} = \frac{1{,}5 - 2}{1} = -0{,}5$$

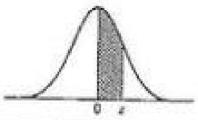
Então:

$$\beta = P(X \leq Z_{critico}) = 0{,}5 - 0{,}1915 = 30{,}85\%$$

Alternativa b).

AREAS UNDER THE NORMAL CURVE

An entry in the table is the proportion under the entire curve which is between $z = 0$ and a positive value of z. Areas for negative values of z are obtained by symmetry.



Second decimal place of z

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2703	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

From Paul G. Hoel, *Elementary Statistics*, 3rd ed., © 1971, John Wiley and Sons, Inc., New York, p. 287.

Questão 4

Antes de uma eleição, um determinado partido está interessado em estimar a proporção p de eleitores favoráveis ao seu candidato. Uma amostra piloto de tamanho 100 revelou que 60% dos eleitores manifestaram-se favoráveis ao candidato em questão.

a) Determine o tamanho da amostra necessária para que o erro cometido na estimação seja de, no máximo, 0,01 com probabilidade 80%;

b) Se na amostra final, com o tamanho obtido em (a), observou-se que 55% dos eleitores eram favoráveis ao candidato de interesse, construa um intervalo de confiança para a proporção p. Utilize $\alpha = 5\%$

O erro em uma estimação é o termo que, em um intervalo de confiança, adicionamos e subtraímos da média estimada para a determinação dos valores mínimo e máximo do intervalo. Podemos chamar esse termo de *E*.

Sendo $IC: \bar{x} \pm Z\frac{s}{\sqrt{n}}$, temos que: $E = Z\frac{s}{\sqrt{n}}$.

O enunciado diz que $\alpha = 20\%$, o que implica termos $Z = 1{,}28$ em um teste bicaudal. Sendo $E = 1\%$, nos falta apenas encontrar o valor de *s*. Este será máximo quando a proporção for igual a 50%. Neste caso, $s = 0{,}5$.

Assim, temos que:

$$E = Z\frac{s}{\sqrt{n}}$$

$$0{,}01 = 1{,}28\frac{0{,}5}{\sqrt{n}}$$

$$\sqrt{n} = 1{,}28\frac{0{,}5}{0{,}01} = 64$$

$$n = 4096$$

* Aceitei como correto o caso em que os alunos utilizaram a proporção de 60% como referência. Neste caso, a resposta é $n = 3.933$.

O intervalo de confiança a ser estimado é:

$$55\% \pm 1{,}96\sqrt{\frac{55\% * (1-55\%)}{4096}} = 55\% \pm 1{,}52\% = [53{,}48\%; 56{,}52\%]$$