

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 2019
Profº. Dr. Glauco Peres da Silva
LISTA DE EXERCÍCIOS 07

Lista de fim de módulo

Gabarito

Data de entrega: 13/05 (noturno) / 15/05 (vespertino)

Nome:

Graduação () Pós-Graduação ()
Período: () Vespertino, () Noturno

Exercício 01 (4 pontos)

Com base na bibliografia obrigatória indicada, marque “Verdadeiro” (V) ou “Falso” (F) para as afirmações a seguir e, quando falsas, justifique sua escolha em 5 linhas, no máximo.

1- (**V**) Uma suposição importante feita pelos que assumem uma Distribuição de Poisson é a de que, para pequenos intervalos de tempo, a probabilidade de uma chegada (evento) é proporcional ao tempo de espera. Essa distribuição possui apenas o parâmetro λ , que representa ao mesmo tempo a média e a variância.

2- (**V**) A Distribuição Beta atribui probabilidade 1 para um intervalo finito, dessa forma, ela é geralmente utilizada para modelar proporções, uma vez que elas são representadas por valores restritos entre 0 e 1.

3- (**F**) A Distribuição Binomial Negativa conta o número de sucessos em um número fixo de provas de Bernoulli.

Essa definição corresponde à Distribuição Binomial comum, a Binomial Negativa conta o número de provas de Bernoulli necessário para obter um número fixo de falhas.

4- (F)A Distribuição Lognormal é uma transformação feita para aproximar da normal variáveis aleatórias cujo logaritmo não é normalmente distribuído. Ela é muito interessante para modelar variáveis de interesse assimétricas à direita.

Essa distribuição se aplica a variáveis aleatórias cujo logaritmo é normalmente distribuído. Ela é muito interessante para modelar variáveis de interesse assimétricas à direita

5 (V)- A Distribuição Binomial é uma distribuição para variáveis discretas e necessita somente, dois resultados possíveis para ser executada.

Exercício 2 (6 pontos)

Em relação à distribuição de probabilidade normal, responda aos seguintes itens.

1- Uma pesquisa Ibope divulgada em 30 de março de 2016 apresentou os seguintes resultados sobre a confiança dos brasileiros na presidenta Dilma: 80% dos entrevistados afirmaram não confiar na presidenta; 18% afirmaram confiar e; 2% não souberam ou não opinaram. Considere como sucesso o percentual de respondentes que não confia na presidenta e, como fracasso, o percentual de entrevistados que confia somado ao percentual daqueles que não sabem ou não responderam. Observando o sucesso e o fracasso, qual é a distribuição de probabilidade apropriada tendo em vista estritamente os valores que a variável dependente pode assumir? O que permite aproximar essa distribuição à distribuição normal, como fazem os institutos de pesquisa?

A distribuição apropriada tendo em vista os valores que a variável dependente pode assumir é a Bernoulli (confia ou não confia na presidenta). O que permite que os institutos tratem a variável como tendo uma distribuição normal é o Teorema do Limite Central. Como a variável busca analisar a proporção de entrevistados que não confia na presidenta (conforme sucesso definido acima), trata-se de uma média, o que permite que a variável seja modelada a partir distribuição normal. Em resumo, o Teorema do Limite Central demonstra que as médias de uma variável aleatória possuem distribuição em forma de sino.

2- Sobre a pesquisa Ibope acima mencionada, assuma que ela possui uma distribuição normal. Suponha também que ela tenha sido realizada com 1500 pessoas, das quais 80% disseram não confiar na presidente e 20% disseram confiar, não responderam ou não sabiam. Descubra o erro amostral da pesquisa e, dado o erro amostral, informe qual é a probabilidade de que a real proporção de indivíduos na população que não confia na presidenta seja de 75%. Dica: i. utilize o desvio padrão da amostra para calcular o erro amostral; ii. desenhe a distribuição normal, indicando a média e os valores correspondentes a aproximadamente 68%, 95% e 99% da distribuição.

Sucesso: 0,8; Fracasso: 0,2

Variância: $\sigma^2 = p(1-p)$

$\sigma^2 = 0,8(0,2)$

$\sigma^2 = 0,16$

Desvio padrão: $\sqrt{0,16} = 0,4$

Erro padrão: $\sigma/\sqrt{n} = 0,4/\sqrt{1500} = 0,010$

Como se trata de uma distribuição normal, sabemos que a quase totalidade das observações está a até 3 erros padrão da média. Desse modo, com nível de confiança de 99% (três erros padrão da média), a proporção amostral daqueles que não confiam na presidenta está entre 77% e 83%. Assim, a probabilidade de a real proporção ser de 75% é menor que 1%.

3- Suponha que um(a) pesquisador(a) tenha selecionado, por meio de uma amostra aleatória simples, 2000 moradores do estado de São Paulo que possuam carteira assinada e tenha lhes perguntado seu salário, em reais. Para fins didáticos, assumamos também que a distribuição de salários obedece a uma distribuição normal. O(A) pesquisador(a) obteve os seguintes resultados: média igual a 1500 reais e desvio padrão igual a 1000 reais. O mesmo procedimento foi realizado no estado do Espírito Santo, com os seguintes resultados: média igual a 1200 reais e desvio padrão igual a 500 reais. Imagine dois trabalhadores, um cujo salário seja de 2000 reais em São Paulo (trabalhador A) e outro cujo salário seja de 1800 no Espírito Santo (trabalhador B). Relativamente ao seu estado, qual dos trabalhadores possui maior salário em relação aos seus conterrâneos? Eles possuem salário superior a qual porcentagem da população? Dica: desenhe a distribuição normal para cada um dos estados, indicando a média, o valor do salário do trabalhador e o escore-z correspondente ao salário deste (não se esqueça de interpretar o escore-z).

São Paulo

escore-z: $y_i - \bar{y}/\sigma = 2000 - 1500/1000 = 0,5$

O trabalhador A possui um salário que está a 0,5 desvio padrão da média. De acordo com a tabela Z, o referido trabalhador possui salário superior a 69,15% da população de seu estado – São Paulo.

Espírito Santo

escore-z: $y_i - \bar{y}/\sigma = 1800 - 1200/500 = 1,2$

O trabalhador B possui um salário que está a 1,2 desvios padrão da média. De acordo com a tabela Z, o referido trabalhador possui salário superior a 88,49% da população de seu estado – Espírito Santo.

Portanto, o trabalhador que possui maior salário em comparação aos seus conterrâneos é o B, do estado do Espírito Santo.

Exercício 03

Com base em pesquisa eleitoral online, Alex Kunrath divulgou a seguinte informação às vésperas das eleições presidenciais de 2018:

Em 29/08 publicamos um artigo onde demonstrávamos a ferramenta Trends de pesquisa online e a confrontávamos com os dados das pesquisas eleitorais tradicionais.

Os dados encontrados nos permitiram averiguar que a metodologia utilizada por institutos de pesquisa como Ibope e Datafolha não condiziam com a realidade. Ao mesmo tempo, as pesquisas online, com amostras muito mais significativas, apontavam resultados mais próximos da realidade.

[...]

Caso repetirmos a taxa de votos brancos, nulos e abstenções, que na última eleição chegou à casa de 20%, essa métrica nos remete ao provável resultado do primeiro turno (votos válidos):

Bolsonaro (51%)

Ciro Gomes (19%)

Fernando Haddad (15%)

Alckmin (7%)

Marina Silva (5%)

[...]

Ressaltamos que este artigo trata da apresentação de um método novo e não oficial de pesquisa eleitoral e que o mesmo está em fase de testes em nossa empresa de consultoria.

No dia 7 de outubro teremos uma resposta sobre a validade desse modelo.

Fonte:

KUNRATH, Alex. *Eleições 2018: Pesquisas online indicam vitória de Bolsonaro em 1o turno*. Portal RIC MAIS, 05.10.2018. URL:

<https://pr.ricmais.com.br/colunistas/noticias/eleicoes-2018-pesquisas-online-indicam-vitoria-de-bolsonaro-em-1o-turno#gref> [último acesso: 20.04.2019].

Tendo em vista apenas o trecho apresentado acima, indique qual é o principal problema do desenho de pesquisa. Justifique (máximo de 5 linhas).

Resposta: A amostra não é aleatória. Para uma pesquisa ter validade externa para toda a população, deve haver aleatoriedade na amostragem. Pesquisas online são limitadas aos segmentos da população que têm acesso à internet e, portanto, não têm validade

externa. Trata-se de uma "amostragem de conveniência", não representativa da população brasileira.

Exercício 4

Nesse exercício vamos retornar à nossa investigação da importância do tamanho das coligações nas eleições para os governos estaduais na corrida presidencial. Para fazer esse exercício você utilizará o arquivo “dados.csv” que foi criado para essa atividade. O arquivo traz informações sobre a proporção de votos obtidos por PT e PSDB no primeiro turno das eleições presidenciais de 2014 por estado (PROP_VOTOS), bem como o número de partidos que faziam parte da coligação em que PT e PSDB concorreram ao governo nesses estados (TAMANHO_COLIGACAO) e se a coligação incluía o partido do governador (GOV_ALIADO). Cada linha, isto é, unidade de observação, corresponde a uma combinação de estado (SIGLA_UF) e partido (SIGLA_PARTIDO). Em todos os cálculos, arredonde o resultado até a segunda casa decimal.

Lembra-se que concluímos que, para o PT, coligações maiores estavam associadas a votações mais expressivas, enquanto para o PSDB, acontecia o contrário? Nossa conclusão, contudo, não levou em conta se a coligação incluía o partido do governador. Podemos imaginar que, por diversas razões, onde os partidos do governador e do candidato à presidência são aliados, o segundo pode esperar uma votação mais expressiva. É o que a ciência política chama de efeitos “coattail” reversos. Por isso, precisamos incluir essa variável em nossa análise. Vamos seguir a mesma estratégia de segmentar a nossa principal variável explicativa em faixas. Como agora precisamos calcular as faixas por partido e valor da variável GOV_ALIADO, vamos fazê-lo a partir da mediana em vez dos quartis, caso contrário teríamos muitas faixas para poucas observações. Sendo assim, calcule a mediana de TAMANHO_COLIGACAO para cada combinação de SIGLA_PARTIDO e GOV_ALIADO (isto é, calcule a mediana para todas as observações do PT em que GOV_ALIADO é igual a 1, na sequência para todas as observações do PT em que GOV_ALIADO é igual a 0, e assim por diante). Crie uma nova variável TAMANHO_COLIGACAO_CAT que indica se o valor de TAMANHO_COLIGACAO de cada observação está “acima da mediana” ou “abaixo da mediana”. Depois, calcule a média de PROP_VOTOS para cada combinação de SIGLA_PARTIDO, GOV_ALIADO e TAMANHO_COLIGACAO (isto é, calcule a média para todas as observações do PT em que GOV_ALIADO é igual a 1 e TAMANHO_COLIGACAO_CAT igual a “abaixo da mediana”, na sequência para todas as observações do PT em que GOV_ALIADO é igual a 1 e TAMANHO_COLIGACAO_CAT igual a “acima da mediana”, e assim por diante). Apresente o resultado em uma tabela com as variáveis SIGLA_PARTIDO, GOV_ALIADO, TAMANHO_COLIGACAO_CAT e PROP_VOTOS (que registra as médias calculadas). Depois de controlar para GOV_ALIADO, TAMANHO_COLIGACAO ainda está associada a diferentes valores de PROP_VOTOS? Caso afirmativo, em que condições?

SIGLA_PARTIDO	GOV_ALIADO	TAMANHO_COLIGACAO_CAT	PROP_VOTOS
PSDB	0	abaixo da mediana	0,31
PSDB	0	acima da mediana	0,27
PSDB	1	abaixo da mediana	0,32
PSDB	1	acima da mediana	0,32
PT	0	abaixo da mediana	0,34
PT	0	acima da mediana	0,54
PT	1	abaixo da mediana	0,48
PT	1	acima da mediana	0,5

Depois de controlar para GOV_ALIADO, a associação entre TAMANHO_COLIGACAO e PROP_VOTOS muda. Vemos que conforme o tamanho da coligação aumenta de abaixo da mediana para acima da mediana, PROP_VOTOS aumenta apenas para o PT e, além disso, o maior aumento (0,2) ocorre quando a coligação não inclui o partido do governador. Quando a coligação inclui o partido do governador, o aumento é menor (0,02). Para o PSDB, PROP_VOTOS diminui conforme o tamanho da coligação aumenta de abaixo da mediana para acima da mediana, mas apenas para os casos em que a coligação não inclui o partido do governador. Em suma, a associação entre TAMANHO_COLIGACAO e PROP_VOTOS é mais aparente quando a coligação não inclui o partido do governo, e é fortemente positiva para o PT e fracamente negativa para o PSDB.

Exercício 5

Um deck de baralho comum possui 52 cartas que assumem 13 valores (A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K) em quatro diferentes naipes (paus, ouros, copas e espadas). Posto isto, responda as próximas questões.

A) Qual a probabilidade teórica de se retirar um “valet” (J), independente do naipe, de forma aleatória?

Como foi informado no enunciado, não existe nenhuma marca de identificação das cartas. Dessa forma, a probabilidade teórica de se retirar uma carta “x” pode ser expressa por:

$$p = 1/n$$

sendo “n” os valores possíveis que as cartas podem assumir. Com isso, a probabilidade de se extrair um “valet” é expressa por:

$$p = 1/13 \approx 0,77$$



Assim, a probabilidade de se retirar um “valete” de forma aleatória é de aproximadamente 0,77.

B) Agora, calcule a probabilidade de se retirar de forma aleatória uma carta que seja ou uma “dama” ou um “ás” ou do naipe de copas.

Agora, temos três diferentes probabilidades que são cumulativas, ou seja, o “ou” adiciona diferentes valores que a carta pode assumir. Pode ser uma dama ou um ás, independente do naipe, ou uma carta de copas (que inclui a dama e o ás). Dessa forma, temos quatro opções de ás e quatro de damas, mas é importante lembrar que uma de cada fazem parte do naipe de copas. Podemos representar a probabilidade como:

$$p = p_a + p_d + p(c-a-d)$$

na qual, p_a indica a probabilidade de ser um ás, p_d a probabilidade de ser uma dama e $p(c-a-d)$ a probabilidade de ser do naipe de copas, excetuando o ás e a dama. De outra forma:

Com n_a representando o número de possibilidade de assumir o valor de ás, n_d o de damas e n_c o de copas excetuando-se as duas cartas já incluídas anteriormente (ás e damas). Assim:

Assim, a probabilidade de se retirar uma dama **OU** um ás **OU** uma carta e copas é de aproximadamente 0,365.

C) Os gráficos abaixo apresentam as probabilidades de se retirar cada valor de carta em quatro diferentes amostras independentes feitas seguindo os mesmos procedimentos em um baralho sem marcas. É possível dizer que as amostras possuem viés na seleção? Justifique em até 5 linhas.

Com as informações apresentadas e os gráficos apresentados, não é possível dizer que existe viés nas amostras. Como visto com a Lei dos Grandes Números, a média amostral tende a se aproximar da média da população à medida em que o n aumenta.

D) Qual a diferença entre a Lei dos Grandes Números e o Teorema Central do Limite na explicação da aproximação das médias dos gráficos acima?

Diferente do que ocorre nas amostras utilizadas nos gráficos, o Teorema Central do Limite não explica a aproximação da média amostral à média da população (ou teórica, como no exercício). Ele indica que a distribuição das médias amostral é aproximadamente normal com a média tendendo à média da população. Em outras palavras, a Lei dos Grandes números está relacionada com o tamanho da amostra, enquanto o Teorema Central do Limite se relaciona com as médias das amostras.