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 / 2024

Prof. Glauco Peres da Silva

2ª Lista de Exercícios

Instruções

1) A lista de exercícios é individual. Você pode discutir a resposta com seus colegas, mas de forma nenhuma deverá copiar ou permitir que copiem alguma de suas respostas, sob pena de punição por plágio;
2) Para todas as questões, apresente seu raciocínio. Não serão consideradas respostas em que não se possa compreender o raciocínio seguido;
3) Há perguntas específicas para a pós-graduação e devem ser respondidas exclusivamente por quem está na pós-graduação;
4) A interpretação do enunciado faz parte da avaliação. Em caso de dúvida que altere o raciocínio, explicite as considerações necessárias utilizadas em sua resposta;
5) As respostas deverão ser submetidas em um arquivo via o link destinado a isto no *Moodle* da disciplina até o final do dia 11/06. O arquivo deverá ser nomeado da seguinte forma: *Lista 1_Seu Nome.docx*. Não serão aceitas respostas de outra forma;
6) Bom trabalho!

Questão 1

(Adaptado de Bussab e Morettin) Comente quais seria os tipos de problemas que surgiriam nos seguintes planos amostrais:

(a) Para investigar a proporção dos operários de uma fábrica favoráveis à mudança do início das atividades das 7h para as 7h30, decidiu-se entrevistar os 30 primeiros operários que chegassem à fábrica na quarta-feira.

As pessoas que chegam em horários próximos tendem a ser mais homogêneas em relação ao próprio horário de chegada. Pode haver alguma variação de opiniões entre eles, mas tendem a ser menores do que se considerarmos aqueles que chegam em horários distintos. A possibilidade de viés é bastante evidente neste caso.

(b) Mesmo procedimento, só que o objetivo é estimar a altura média dos operários.

Já quanto altura, não deve haver problemas de viés. O horário de chegada e a altura dos funcionários não devem estar correlacionadas de maneira a produzir algum tipo de viés na estimativa.

(c) Para estimar a porcentagem média da receita municipal investida em lazer, enviaram-se questionários a todas as prefeituras, e a amostra foi formada pelas prefeituras que enviaram as respostas.

A auto-seleção dos respondentes é uma potencial fonte de viés. Não sabemos qual seria a real razão de uma prefeitura decidir responder ou não ao questionário, mas o fato de caber a cada uma decidir se responde ou não implica em grandes chances de que haja algum aspecto que produza viés.

(d) Para verificar o fato de oferecer brindes nas vendas de sabão em pó, tomaram-se quatro supermercados na zona sul e quatro na zona norte de uma cidade. Nas quatro lojas da zona sul, o produto era vendido com brinde, enquanto nas outras quatro era vendido sem brinde. No fim do mês, compararam-se as vendas da zona sul com as da zona norte.

O primeiro problema está na seleção dos mercados. Por não sabermos como foram escolhidos, pode haver viés apenas pela seleção da área da cidade e dos próprios supermercados. Considerando que estas decisões tenham sido aleatórias, o passo seguinte – o de decidir qual loja ofereceria brinde e qual não – também precisaria ser aleatório. A coincidência das quatro lojas com e sem brindes serem das mesmas regiões sobrepõe as características regionais à da promoção, podendo gerar algum tipo de confundidor na avaliação final.

Questão 2

Nas últimas eleições, um determinado partido está avaliando a performance final de seus candidatos. Considere que a votação final de cada um deles foi a seguinte, em mil votos: 70, 72, 74, 76, 80 e 114.

Diante destes resultados, um analista sugere que a mediana seja escolhida como a medida de tendência central adequada para estes dados. O que justificaria esta decisão? Você concordaria com a proposição deste analista? Por que?

A justificativa desta decisão é o fato de haver uma observação destoante das demais. Uma das candidaturas recebeu quantidade de votos muito maior do que as demais, o que acaba puxando a média para cima. No caso, $\bar{x} = 81$ e este valor é maior do que 5 das 6 observações. Já a mediana seria igual a 75, medida mais indicada pela assimetria provocada pelo outlier. Assim, neste caso, concordo com o analista e consideraria a mediana como medida de tendência central destes dados.

Questão 3

A seguir, serão apresentados três problemas de pesquisa. Leia atentamente cada um dos problemas, de modo a responder corretamente os itens propostos.

I. Determinado(a) pesquisador(a) deseja saber se um prefeito será ou não removido do cargo em decorrência de um processo de cassação.

II. Determinado(a) pesquisador(a) pretende analisar o número de projetos apresentados pelos parlamentares na Câmara dos Deputados em determinado ano. Dica: assuma que as ocorrências da variável são independentes.

III. Tendo em mãos uma lista de prefeitos cassados nos últimos 20 anos no estado de São Paulo, um(a) pesquisador(a) deseja conhecer o tempo médio, em meses, que um prefeito cassado se mantém no cargo. Dica: o tempo no cargo é contabilizado em dias, desse modo, 45 dias equivalem a 1,5 mês (assuma que todos os meses possuem 30 dias).

a-) Para cada uma das situações, indique: I. qual é a variável dependente; II. quais são os possíveis valores que a variável pode assumir; III. se a variável é contínua ou discreta.

Problema de pesquisa I: VD: cassação de um prefeito; valores 1 – no caso de haver cassação; 0 – para caso contrário. Variável *dummy*, portanto – uma variável discreta.

Problema de pesquisa II: VD: número de projetos apresentados por deputado por ano. Seu valor se inicia em 0 e não há limite superior óbvio, a não ser que haja alguma restrição legal que impeça um deputado de propor um número maior de projetos do que um certo limite. A variável é discreta.

Problema de pesquisa III: VD: tempo médio de mandato de um prefeito; Seu valor se inicia em 0 e seu limite superior é igual a 4*365 = 1460. A variável é discreta, mas pode ser aproximada para contínua.

b-) Para cada uma das situações indique uma distribuição probabilística apropriada, demonstrando por que ela se adequa a cada um dos problemas de pesquisa. Dica: verifique se a variável é discreta ou contínua, conforme item anterior;

Problema I – Binomial; Problema II – Poisson ou Binomial Negativa; Problema III – Normal, mas neste caso, a variável poderia estar truncada à direita, já que nenhum prefeito pode ficar mais do que o limite superior. Por se tratar de prefeitos cassados, nenhum deles deve ter atingido este limite.

c-) **Pós-graduação:** Em relação ao problema de pesquisa esboçado no item I, assuma que a probabilidade de afastamento do prefeito do cargo em decorrência de uma cassação seja conhecida e equivalha a 0,68. Qual é a probabilidade da não ocorrência do afastamento do prefeito durante seu mandato? Indique a distribuição de probabilidade apropriada, sua fórmula matemática correspondente e substitua os valores fornecidos na questão na fórmula para obter o resultado.

Distribuição binomial pode ser utilizada nesta situação.

X = cassação do prefeito ao longo de seu mandato ➔ x = 1 em caso de cassação; x = 0 em caso contrário. P(X = 1) = 0,68 = p. P(X = 0) = 1- p = 0,32. Em São Paulo, há 645 cidades e, podemos, assim, nos perguntar a respeito da chance de vermos a cassação de todos eles ao mesmo tempo. Como a pergunta não especificava sobre quantos investigar, esta etapa da solução será diferente entre as respostas oferecidas pelos alunos e não trará problemas, desde que coerente com a situação.

A fórmula de sua PMF é dada por: $P(X = x) = \binom{n}{k} p^k (1 - p)^{n-k}$.

No caso em questão, temos: $P(X = x) = \binom{645}{0} 0{,}32^0 (1 - 0{,}32)^{645-0} = 0\%$.

Questão 4

Para ser aprovado em uma disciplina, você deve realizar uma prova teste do tipo "verdadeiro ou falso": para cada questão, você deve apenas informar corretamente se a afirmação é verdadeira ou falsa. Suponha que você não saiba absolutamente nada a respeito do conteúdo desta prova e ainda suponha que você deva acertar no mínimo 2/3 das suas respostas para ser aprovado. O professor está em dúvida sobre quantas questões elaborar e pergunta à turma, dando duas opções: uma prova com um total de 3 ou de 36 questões. Você prefere a prova com 3 questões ou com 36 questões? Ou você seria indiferente a extensão da prova? Explique usando os conceitos trabalhados na disciplina até aqui.

O aluno que nada sabe a respeito do conteúdo da matéria tem exatamente 50% de chance de acertar algo em uma questão do tipo "verdadeiro ou falso". Além disto, este aluno precisa acertar 2/3 (ou 66,7%) das respostas para ser aprovado. Essas informações mostram para ele que se sua estratégia for aquela que o aproxima da média ele tenderá a obter um resultado mais próximo de 50% e isto o reprovaria. Ele precisa, assim, estar em uma situação em que haja maior variação no resultado final ou, o que é dizer o mesmo, em que a convergência para a média não ocorra. Sabemos que a convergência para a média ocorre quando o número de observações aumenta. Neste sentido, quanto maior o número de observações (questões da prova) maior a chance de convergência para a média, ou seja, maior a chance de reprovação. Assim, é melhor para este aluno uma prova com menos questões.

Isto pode ser apresentado com uma distribuição binomial.

Vamos definir nossa variável X como sendo: X = número de questões acertadas na prova.

Caso I – prova com 3 questões

Neste caso, temos: n = 3; p = 50%.

Para a aprovação, o aluno precisa atingir 2/3 de acerto da prova. Neste caso, acertar 2 de 3 questões. Ou seja, ele precisa saber qual a probabilidade disto ocorrer. Podemos representar o problema assim:

$$P(X \geq 2) = P(X = 2) + P(X = 3)$$

Aplicando a distribuição binomial, temos como resultado:

$$P(X = 2) = \binom{3}{2} 0{,}5^2 (1 - 0{,}5)^{3-2} = 3 * 0{,}25 * 0{,}5 = 37{,}5\%$$

$$P(X = 3) = \binom{3}{3} 0{,}5^3 (1 - 0{,}5)^{3-3} = 1 * 0{,}125 * 1 = 12{,}5\%$$

Assim, temos que:

$$P(X \geq 2) = 50\%$$

Caso II – prova com 36 questões

Neste caso, temos: n = 36; p = 50%.

Para a aprovação, o aluno precisa atingir 2/3 de acerto da prova. Neste caso, acertar 24 de 36 questões. Ou seja, ele precisa saber qual a probabilidade disto ocorrer. Podemos representar o problema assim:

$$P(X \geq 24) = P(X = 24) + P(X = 25) + P(X = 26) + \cdots + P(X = 36)$$

Aplicando a distribuição binomial, temos como resultado:

$$P(X = 24) = \binom{36}{24} 0{,}5^{24}(1 - 0{,}5)^{36-24} = 1.251.677.700 * 0{,}000000006 * 0{,}000244 = 1{,}82\%$$

$$P(X = 25) = \binom{36}{25} 0{,}5^{25}(1 - 0{,}5)^{36-25} = 600.805.296 * 0{,}000000003 * 0{,}000488 = 0{,}87\%$$

$$P(X = 26) = 0{,}37\%$$

$$P(X = 27) = 0{,}14\%$$

$$P(X = 28) = 0{,}04\%$$

$$P(X = 29) = 0{,}01\%$$

$$\ldots$$

$$P(X \geq 24) = 3{,}26\%$$

Assim, é claramente preferível para o aluno uma prova com apenas 3 questões nesta situação.

Questão 5

Quando mensagens codificadas são enviadas, por vezes há erros na transmissão que confundem a sua decodificação. Considerando o uso do código Morse, sabe-se que os "pontos" e os "traços" utilizados ocorrem na proporção 3:4. Isto significa que para um dado símbolo qualquer:

$$P(enviar\ um\ ponto) = \frac{3}{7} \qquad \text{e} \qquad P(enviar\ um\ traço) = \frac{4}{7}.$$

Suponha que haja uma interferência na linha de transmissão que torne as mensagens erráticas. Sabe-se por experiência que, por conta destas interferências, há uma probabilidade de $\frac{1}{4}$ de um "ponto" na mensagem enviada se tornar erroneamente um "traço" na mensagem recebida e, também, que há probabilidade de $\frac{1}{3}$ de um "traço" na mensagem original transformar-se equivocadamente num "ponto" na mensagem recebida. Assim, responda as questões a seguir:

a) Se um "traço" é recebido, qual é a probabilidade de que de fato um "traço" tenha sido enviado?

Seja S uma variável aleatória que assuma valor igual a *1* quando o símbolo enviado é um traço e assuma valor *0* quando o símbolo enviado é um ponto.

Seja R uma variável aleatória que assuma valor igual a *1* quando o símbolo recebido é um traço e assuma valor *0* quando o símbolo recebido é um ponto.

A pergunta do item a) pode ser transformada na seguinte relação:

$$P(S = 1 \mid R = 1)$$

Usando o Teorema de Bayes, temos que:

$$P(S = 1 \mid R = 1) = \frac{P(R = 1 \mid S = 1) * P(S = 1)}{P(R = 1)}$$

Vamos decompor cada termo aqui.

$P(R = 1 \mid S = 1)$ é a probabilidade de recebermos um traço quando o símbolo enviado foi mesmo um traço. Como há 1/3 de chance de haver uma inversão de traço na mensagem original virar um ponto no recebimento, esta probabilidade aqui é de 2/3.

$P(S = 1)$ é dada pela proporção de traços enviados: 4/7.

$P(R = 1)$ é dada pela proporção de traços recebidos. Esta depende tanto da proporção de traços enviados, quanto dos erros que ocorrem invertendo os traços enviados em pontos e dos pontos enviados em traços. Com isso, essa probabilidade é dada por:

$$P(R = 1) = P(R = 1 \mid S = 1) * P(S = 1) + P(R = 1 \mid S = 0) * P(S = 0)$$
$$= \frac{2}{3} * \frac{4}{7} + \frac{1}{4} * \frac{3}{7} = \frac{8}{21} + \frac{3}{28} = \frac{8 * 4 + 3 * 3}{3 * 4 * 7} = \frac{32 + 9}{84} = \frac{41}{84}$$

Com isso, temos que:

$$P(S = 1 \mid R = 1) = \frac{P(R = 1 \mid S = 1) * P(S = 1)}{P(R = 1)} = \frac{\frac{2}{3} * \frac{4}{7}}{\frac{41}{84}} = \frac{\frac{8}{21}}{\frac{41}{84}} = \frac{8}{21} * \frac{84}{41} = \frac{32}{41} = 78\%$$

Pós-graduação

b) Assumindo independência entre os sinais, se a mensagem “ponto”-“ponto” foi recebida, qual é a probabilidade de cada uma das quatro mensagens possíveis que teriam sido enviadas?

As quatro mensagens possíveis de terem sido enviadas são: “ponto”-“ponto”, “ponto”-“traço”, “traço”-“ponto” e “traço”-“traço”.

Devemos nos perguntar, então, qual a probabilidade de que a mensagem recebida seja igual à mensagem enviada. Vamos manter a hipótese de que o envio de cada um dos símbolos seja independente do símbolo enviado anteriormente. Então, temos de encontrar a probabilidade de que de fato tenha sido enviado um ponto quando se recebeu um ponto e a de que tenha sido enviado um traço, quando se recebeu um ponto.

Ou seja, $P(S = 0 \mid R = 0)$ e $P(S = 1 \mid R = 0)$.

Vamos encontrar primeiro

$$P(S = 0 \mid R = 0)$$

Usamos Bayes outra vez:

$$P(S = 0 \mid R = 0) = \frac{P(R = 0 \mid S = 0) * P(S = 0)}{P(R = 0)}$$

Daqui, temos que:

$P(R=0\mid S=0)$ é a probabilidade de termos recebido um ponto quando se enviou um ponto de fato. Neste caso, é ¾.

$P(S=0)$ é a probabilidade de se enviar um ponto. Aqui é de 3/7.

$P(R=0)$ é a proporção de pontos recebidos. Depende da proporção de pontos enviados corretamente recebidos e da proporção de traços enviados que mudam para pontos. Temos assim que:

$$P(R=0)=\ P(R=0\mid S=0)*P(S=0)+P(R=0\mid S=1)*P(S=1)$$

$$=\frac{3}{4}*\frac{3}{7}+\frac{1}{3}*\frac{4}{7}=\frac{9}{28}+\frac{4}{21}=\frac{9*3+4*4}{3*4*7}=\frac{27+16}{84}=\frac{43}{84}$$

Assim, temos que:

$$P(S=0\mid R=0)=\frac{P(R=0\mid S=0)*P(S=0)}{P(R=0)}=\frac{\frac{3}{4}*\frac{3}{7}}{\frac{43}{84}}=\frac{\frac{9}{28}}{\frac{43}{84}}=\frac{9}{28}*\frac{84}{43}=\frac{27}{43}\cong 62{,}8\%$$

Nos falta agora encontrar a probabilidade de ter sido enviado um traço quando recebemos um ponto. O que queremos é $P(S=1\mid R=0)$. Usando Bayes, temos a seguinte igualdade:

$$P(S=1\mid R=0)=\frac{P(R=0\mid S=1)*P(S=1)}{P(R=0)}$$

Sabemos que:

$P(R=0\mid S=1)$ é a probabilidade de recebermos um ponto quando o símbolo enviado foi um traço. Pelo enunciado acima, essa chance é de 1/3.

$P(S=1)$ é a probabilidade de que um traço seja enviado. É igual 4/7.

$P(R=0)$ já encontramos no item anterior.

Então, temos que:

$$P(S=1\mid R=0)=\frac{P(R=0\mid S=1)*P(S=1)}{P(R=0)}=\frac{\frac{1}{3}*\frac{4}{7}}{\frac{43}{84}}=\frac{\frac{4}{21}}{\frac{43}{84}}=\frac{16}{43}\cong 37{,}2\%$$

Com isso, podemos informar as probabilidades pedidas:

Ponto-Ponto = $\frac{27}{43} * \frac{27}{43} \cong 39{,}43\%$