**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 06**

**"Probabilidade**"

Data de entrega: 22/04/2019 (noturno) e 24/04/2019 (vespertino)

**Nome: Kaue Oliveira Almeida**

**Período: (X) Vespertino, ( ) Noturno**

**Exercício 01**

**Para cada uma das afirmações abaixo, indique se é Verdadeira (V) ou Falsa (F).**

**Justifique apenas as falsas.**

1. (F) O Espaço Amostral consiste na listagem de todos os resultados obtidos em uma

amostragem. Se lançarmos uma moeda 10 vezes, com 8 resultados sendo ‘Cara’ e 2

resultados sendo ‘Coroa’, nosso espaço amostral será: Ω = {Cara, Cara, Cara, Cara, Cara,

Cara, Cara, Cara, Coroa, Coroa}.

Justificativa: O Espaço Amostral consiste na em todos os resultados do experimento. Nessa caso, o espaço amostral seria todas as possibilidades entre cara e coroa para cada resultado.

2. (F) Dois estudantes decidem jogar cara-ou-coroa para testar seus conhecimentos de

probabilidade. Após jogarem 1000 moedas, verificaram que obtiveram 10 Coroas e 990 Caras. Caso desejassem lançar mais uma moeda, a chance desse resultado ser Coroa é mais provável do que ser Cara. Esse resultado é explicado pela Teoria Central do Limite, que garante que, para populações grandes, a frequência dos resultados tende a se aproximar do valor esperado.

Justificativa: A chance de resultado de mais um lançamento ser Cara seria mais provável. Pois, se C indicar Cara e R indicar coroa, então as probabilidades de lançamento podem ser definidas a partir do experimento conduzido: P(R) = 10/1000 = 0.01 e P(C) = 990/1000 = 0.99. Desse modo, a probabilidade de um próximo resultado ser Cara é maior. A Teoria Central do Limite define que para uma amostragem aleatória com n grande, a distribuição amostral da média é aproximadamente uma distribuição normal.

3. (F) Os monitores da disciplina de Métodos decidem selecionar ao acaso um aluno para

responder uma questão. A turma do período noturno consiste em 20 alunos, 8 homens e 12

mulheres. A turma do vespertino consiste em 10 alunos, 7 homens e 3 mulheres. A chance de os monitores selecionarem aleatoriamente um homem do período noturno é maior do que a de selecionarem uma mulher.

Justificativa: No período noturno podemos calcular a probabilidade de uma seleção de um indivíduo pela proporção de indivíduos pelo sexo. Assim, calculamos:

P(Homem) ∩ P(Noturno) = P(Homem) \* P(Noturno|Homem) = 15/30 \* (20/30 / 8/30) = 0.5 \* (0.66 / 0.26) = 0.2

P(Mulher) = (12+3)/30 = 0.5

A chance de selecionar uma mulher é maior.

4. (F) A Distribuição Normal é uma distribuição muito importante dado, por exemplo, quão

comum ela é encontrada em eventos naturais ou sociais. A Regra do ‘68-95-99’ indica a

porcentagem aproximada de valores que estão dentro de 1, 2 e 3 desvios-padrões,

respectivamente.

Justificativa: Esse tipo de distribuição não é comumente encontrada no mundo real. Mas é importante, pois descreve muito bem histogramas e também é muito utilizada para inferência estatística.

5. (F) O erro-padrão é o desvio-padrão de uma amostra. Este tende a ser menor quanto maior for o tamanho da amostra coletada.

Justificativa: O erro-padrão é o desvio-padrão da distribuição amostral e não de apenas uma amostra.

6. (F) O escore-z é a distância entre a média de distribuição da probabilidade (y) e o valor da

variável observada (μ) medida em desvios-padrão.

Justificativa: O símbolo da média de distribuição da probabilidade é μ e o símbolo do valor da variável observada é y.

**Exercício 02**

Para este exercício vamos utilizar o banco ‘CPDS-1960-2016.xlsx’ disponível no Moodle. Caso julgar necessário, utilize também seu Codebook. Trabalharemos com o número efetivo de partidos, “effpar\_ele”.

A) Em ‘AMOSTRA\_5’ dividimos o banco de dados em cinco amostras. Calcule a média,

variância e desvio-padrão de effpar\_ele para cada um desses 5 grupos. Lembre-se de filtrar

seus dados para cada um desses grupos.

A)

|  |  |  |  |
| --- | --- | --- | --- |
| AMOSTRA\_5 | Média | Variância | Desvio-Padrão |
| 1 | 4.3791612868056 | 3.71677240143048 | 1.92789325467737 |
| 2 | 4.29000099603543 | 2.61571004821426 | 1.61731569219317 |
| 3 | 4.30043795378521 | 2.4226663482625 | 1.55649167947102 |
| 4 | 4.33701379293447 | 2.78387144776261 | 1.66849376617433 |
| 5 | 4.26934041321326 | 2.73920868755687 | 1.65505549379979 |

B) Os valores de suas médias, variâncias e desvios-padrão calculados no item anterior são os mesmos? Caso negativo, explique o que estaria gerando essa diferença.

B) Os valores não são os mesmos, pois as amostras são diferentes, assim cada amostra apresenta valores diferentes que refletem nas medidas estatísticas.

C) Em uma planilha separada, anote o valor de todas as médias obtidas no item A. Calcule a

média dessas médias, e compare com a média total de effpar\_ele. À luz do Teorema Central

do Limite, explique seus achados.

C) A média das médias do item A é igual a 4.31519088855479

A média total é igual a 4.31634736996723

As média das médias e a média total são extremamente próximas. Segundo o Teorema Central do Limite, ao calcularmos a distribuição amostral para um n grande, teríamos uma distribuição aproximadamente normal em torno da média da população. Assim, no nosso caso, a média da distribuição amostral nos dá um resultado muito aproximado da média da população.

D) A coluna ‘AMOSTRA\_100’ divide o banco de dados em 100 amostras. Filtre essa coluna

somente para o valor ‘99’. Calcule A média, desvio-padrão e variância dessa amostra. Agora

faça o mesmo procedimento escolhendo todos os valores exceto 99. Em que medida o

tamanho da amostra afeta nossos resultados? Pensando novamente no Teorema Central do

Limite, de que maneira o tamanho da amostra afetaria a média das médias?

Dica: Pense na média das médias de amostras com n pequeno em comparação com amostras

de n grande.

D)

Para o valor 99:

|  |  |  |
| --- | --- | --- |
| Média | Variância | Desvio-Padrão |
| 4.57398313037516 | 2.50728077364674 | 1.58343953899312 |

Para todos os valores, exceto 99:

|  |  |  |
| --- | --- | --- |
| Média | Variância | Desvio-Padrão |
| 4.31236659493002 | 2.87086425628242 | 1.69436249258605 |

Segundo o Teorema Central do Limite, quanto maior for o tamanho da amostra, menor será o erro amostral. Este erro corresponde ao quanto as estimativas da nossa amostra corresponde com os valores reais da população. Assim, em nosso caso, a amostra de valor 99 resultou em uma estimativa menos precisa da população. Enquanto a estimativa das outras amostras com exceção da amostra 99 resultou em uma precisão maior de estimação da população.

Script R feito para a resolução dos exercícios

|  |
| --- |
| *#Lista 6* **library**(tidyverse) **library**(readxl)  *#Importando a planilha* CPDS <- read\_excel("Lista 6/CPDS-1960-2016.xlsx")  *#Exercício A)* CPDS\_AMOSTRA\_5 <- CPDS %>%   group\_by(AMOSTRA\_5) %>%   na.omit() %>%   summarise("Média" = mean(effpar\_ele), "Variância" = var(effpar\_ele), "Desvio-Padrão" = sd(effpar\_ele))  *#Exercício C)* MEDIA\_AMOSTRA\_5 <- mean(CPDS\_AMOSTRA\_5$Média) MEDIA\_TOTAL <- mean(na.omit(CPDS$effpar\_ele))  *#Exercício D)* CPDS\_AMOSTRA\_99 <- CPDS %>%   subset(AMOSTRA\_100 == 99) %>%   na.omit() %>%   summarise("Média" = mean(effpar\_ele), "Variância" = var(effpar\_ele), "Desvio-Padrão" = sd(effpar\_ele))  CPDS\_AMOSTRA\_100 <- CPDS %>%   subset(AMOSTRA\_100 != 99) %>%   na.omit() %>%   summarise("Média" = mean(effpar\_ele), "Variância" = var(effpar\_ele), "Desvio-Padrão" = sd(effpar\_ele)) |