

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: 15/04/2019 (noturno) e 17/04/2019 (vespertino)

Nome:

Período: () 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: $\Omega = \{\text{Cara, Cara, Cara, Cara, Cara, Cara, Cara, Cara, Coroa, Coroa}\}$.

Falso. Bussab e Morettin, Capítulo 3, Pág. 104: "Um espaço amostral, Ω , que consiste, no caso discreto, da enumeração (finita ou infinita) de todos os resultados possíveis do experimento em questão". Para nosso experimento do lançamento de moedas, o espaço amostral seria a listagem de todas as possibilidades de os lançamentos de moedas. Pensando teoricamente no lançamento de uma moeda, temos dois resultados possíveis: Cara ou Coroa. Assim, para o lançamento de 10 moedas, $\{\text{Cara, Cara, Cara, Cara, Cara, Cara, Cara, Cara, Coroa, Coroa}\}$ é apenas um dos elementos de nosso espaço amostral. Poderíamos igualmente ter obtido, por exemplo, 10 Caras, 10 Coroas, ou qualquer combinação entre Cara e Coroa. O tamanho do espaço amostral é dado por:

(Nº Resultados Possíveis para um Lançamento)^(Nº de Lançamentos)

Ou seja, para 10 lançamentos, cada qual podendo ser um de 2 resultados (Cara ou Coroa), vemos que nosso espaço amostral seria composto por 2^{10} elementos.

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.

Falso. A chance de um evento ocorrer é independente do resultado de eventos anteriores. Logo, a chance do evento dar 'Cara' ou 'Coroa' permanece constante. Ademais, o Teorema Central do Limite indica que a distribuição das médias das amostras tende a se aproximar de uma distribuição Normal quanto mais amostras forem feitas. A média dessa distribuição de médias tenderá ainda a ser o valor da média populacional (Agresti e Finlay, Págs. 118-119, e Agresti e Finlay, Pág. 121).

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.

Falso. Seguindo o modelo de Bussab & Moretim (Pág. 107), podemos criar a seguinte tabela de dupla-entrada:

	Homem	Mulher	Total
Vespertino	7	3	10
Noturno	8	12	20
Total	15	15	30

A chance dos monitores selecionarem uma Mulher será $P(\text{Mulher}) = 15/30 = 50\%$.

A chance dos monitores selecionarem um Homem do período Noturno será $P(\text{Homem} \& \text{Noturno}) = (8/30) = 26\%$.

4. (**V**) 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.

Verdadeiro. Agresti e Finlay, Capítulo 4, Pág. 99: Cerca de 68% dos resultados de uma distribuição Normal estão dentro do intervalo de 1 desvio-padrão, 95% no intervalo de 2 desvios-padrões e cerca de 99% estão no intervalo de 2 desvios-padrões.

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

Verdadeiro. O erro-padrão é o desvio-padrão de uma distribuição amostral, sendo inversamente proporcional ao tamanho da amostra. De acordo com Agresti & Finlay (Pág. 113), temos que:

$$\text{Erro Padrão} = \frac{\text{desvio padrão populacional}}{\sqrt{\text{tamanho da amostra}}}$$

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.

Falso. De acordo com Agresti e Finlay, Pág. 103, o escore-z representa a distância entre y (valor observado) e μ (média do parâmetro dada uma grande quantidade de repetições) medida em desvios padrão.

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 2 grupos. Lembre-se de filtrar seus dados para cada um desses grupos.

Resposta:

Grupo	Média	Desvio-Padrão	Variância
1	4.38	1.93	3.72
2	4.29	1.62	2.62
3	4.30	1.56	2.42
4	4.34	1.67	2.78
5	4.27	1.66	2.74

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

Para este exercício, os dados foram separados aleatoriamente em 5 grupos de tratamento. Ao calcular as medidas de dispersão, verificamos que os valores encontrados não são os mesmos. São dois princípios que estariam por detrás dessa diferença. Em primeiro lugar, as amostras possuem observações distintas, que por sua vez trazem distintos valores e, portanto, os valores da média, da variância e do desvio-padrão sofrem alterações. Além disso, o tamanho da amostra afeta o valor encontrado dessas 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.

Média das Médias: 4.315

Média dos dados: 4.316

A média das médias (aprox. 4.315) possui um valor muito semelhante ao da média populacional (4.316). De acordo com o Teorema Central do Limite, quando repetimos amostras inúmeras vezes, a média das médias tenderá a ser igual à média populacional. Logo, caso continuássemos coletando novas amostras, esse resultado seria cada vez mais próximo.

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.

	Média	Desvio-Padrão	Variância
Só 99	4.57	1.58	2.51
Todos menos 99	4.312	1.69	2.87

Perceba que a média da amostra que possui todos os grupos exceto '99' é bem mais próxima da média de `effpar_ele` do que a média da amostra que contém somente o grupo 99. Isso se dá porque uma amostra de tamanho maior tende a ter uma média mais precisa do que uma amostra com poucas observações. Como a média dessas amostras é mais precisa, pensando no Teorema Central do Limite, vemos que caso usemos amostras mais largas, precisaríamos de uma menor quantidade de amostras para que a média das médias chegue a um resultado equivalente ao da média populacional.