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 05**

**"Estatística Descritiva"**

**Data de entrega: 08/04/2019 (noturno) e 10/04/2019 (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.**

**A. (V) O desvio padrão (s) é uma medida de dispersão que informa o quanto o conjunto de observações está próximo ou distante da média de uma distribuição. Seu valor é relativamente baixo quando os dados se concentram próximos à média e relativamente alto quando estão mais dispersos. (0,5)**

**B. (V) A Regra Empírica é adequada somente para distribuições aproximadamente simétricas, em forma de sino. Segundo a regra, cerca de 68% das observações estão a um desvio padrão acima e abaixo da média, 95% a dois desvios acima e abaixo, e quase todas (99%) a três desvios acima e abaixo. (0,5)**

**C. (V) A cada dez anos, o Instituto Brasileiro de Geografia e Estatística (IBGE) realiza o censo buscando estimar estatísticas da população com base em amostras de municípios e domicílios. (0,5)**

**D. (F) A mediana é uma medida de centro mais apropriada que a média quando lidamos com distribuições muito assimétricas ou com dados discretos que assumem poucos valores distintos. (0,5)**

Justificativa: A mediana é mais apropriada quando lidamos com distribuições muito assimétricas, porém apresenta problemas quando lidamos com dados discretos que assumem poucos valores diferentes. Isso decorre do fato de que padrões diferentes de dados podem ter a mesma mediana.

**E. (V) O intervalo interquartil (IIQ) é utilizado para descrever a dispersão da metade das observações e auxiliar na identificação dos valores atípicos da distribuição. Todos os valores que estão a mais do que 1,5 vezes (IIQ) acima do percentil 75º ou a mais do que 1,5 vezes (IIQ) abaixo do percentil 25º são considerados atípicos. (0,5)**

**F. (F) A correlação sempre indica a força de uma associação causal entre duas variáveis quantitativas. (0,5)**

Justificativa: Nem sempre uma correlação indica a força de uma associação causal. Uma situação muito comum é a interferência de uma ou mais variáveis nas variáveis resposta e explicativa analisadas. Assim, a variável explicativa também estaria sofrendo influência desta outra variável, tornado a associação entre as duas variáveis que apresentam correlação espúria.

**G. (V) A medida de posição escore-z nos permite identificar a quantos desvios padrões uma determinada observação se encontra da média da distribuição. (0,5)**

**H. (F) Tabelas de contingência são úteis para exibir combinações em que os resultados de uma variável independente categórica são *contingentes* em valores numéricos de uma variável dependente categórica. (0,5)**

Justificativa: Nas tabelas de contingência os resultados de uma variável dependente (resposta) são *contingentes* na categoria de uma variável independente (explicativa).

**Exercício 02**

**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). 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.**

**A. É comum se ouvir que as coligações nas eleições para os governos estaduais são importantes para a corrida presidencial. Grandes coligações, em particular com partidos capilarizados, funcionariam como cabos eleitorais do candidato a presidente, supostamente impulsionando sua votação nos estados em que ele consegue articulá-las. Como há dois diferentes partidos em nosso banco de dados e é plausível que a importância do tamanho da coligação varie entre eles, precisamos levar isso em conta em nossas análises. Antes de tudo, calcule a média e o desvio-padrão de PROP\_VOTOS para cada partido. Isto é, calcule a média e o desvio-padrão de PROP\_VOTOS apenas para observações de um partido e, na sequência, apenas para observações do outro. Isso será útil mais tarde. (1,0)**

**Média PROP\_VOTOS (PT):**

(0.2797074+0.4993196+0.5451080+0.5108227+0.6143224+0.6827801+0.2303367+0.3310038+0.3208983+0.6955433+0.4347059+0.3749889+0.3952318+0.5316740+0.5559624+0.4421253+0.7059569+0.3252248+0.3560452+0.6004049+0.4169182+0.3317468+0.4319166+0.3073847+0.5490634+0.2580751+0.5021028)/27 = **0.45**

**Variância PROP\_VOTOS (PT):**

[(0.2797074-0.45)²+(0.4993196-0.45)²+(0.5451080-0.45)²+(0.5108227-0.45)²+(0.6143224-0.45)²+(0.6827801-0.45)²+(0.2303367-0.45)²+(0.3310038-0.45)²+(0.3208983-0.45)²+(0.6955433-0.45)²+(0.4347059-0.45)²+(0.3749889-0.45)²+(0.3952318-0.45)²+(0.5316740-0.45)²+(0.5559624-0.45)²+(0.4421253-0.45)²+(0.7059569-0.45)²+(0.3252248-0.45)²+(0.3560452-0.45)²+(0.6004049-0.45)²+(0.4169182-0.45)²+(0.3317468-0.45)²+(0.4319166-0.45)²+(0.3073847-0.45)²+(0.5490634-0.45)²+(0.2580751-0.45)²+(0.5021028)²]/27 - 1 = **0.02**

**Desvio-padrão PROP\_VOTOS (PT):**

s = **0.02**= **0.14**

**Média PROP\_VOTOS (PSDB):**

(0.29090157+0.22122883+0.19430295+0.25442827+0.18281660+0.14983620+0.36114255+0.35137816+0.41550605+0.11624528+0.39754290+0.41326926+0.44482338+0.27569519+0.23392387+0.05930625+0.13844181+0.49803127+0.26953312+0.19836653+0.44907345+0.43625208+0.41427191+0.52899165+0.22697979+0.44237408+0.27690058)/27 = **0.31**

**Variância PROP\_VOTOS (PSDB):**

[(0.29090157-0.31)²+(0.22122883-0.31)²+(0.19430295-0.31)²+(0.25442827-0.31)²+(0.18281660-0.31)²+(0.14983620-0.31)²+(0.36114255-0.31)²+(0.35137816-0.31)²+(0.41550605-0.31)²+(0.11624528-0.31)²+(0.39754290-0.31)²+(0.41326926-0.31)²+(0.44482338-0.31)²+(0.27569519-0.31)²+(0.23392387-0.31)²+(0.05930625-0.31)²+(0.13844181-0.31)²+(0.49803127-0.31)²+(0.26953312-0.31)²+(0.19836653-0.31)²+(0.44907345-0.31)²+(0.43625208-0.31)²+(0.41427191-0.31)²+(0.52899165-0.31)²+(0.22697979-0.31)²+(0.44237408-0.31)²+(0.27690058-0.31)²]/27 - 1 = **0.016**

**Desvio-padrão PROP\_VOTOS (PSDB):**

s = **0.016**= **0.13**

**B. Para ver se a hipótese em que estamos interessados faz sentido, vamos observar como a proporção de votos recebida por um candidato à presidência varia de acordo com o tamanho da coligação para governador de que seu partido participa. Como nossa variável explicativa é contínua, uma maneira de fazer isso é observar as médias de proporção de votos em diferentes faixas de níveis de tamanho da coligação. Para tanto, calcule os quartis de TAMANHO\_COLIGACAO por partido (isto é, calcule os quartis de TAMANHO\_COLIGACAO apenas para observações de um partido e, na sequência, apenas para observações do outro). Crie uma nova variável TAMANHO\_COLIGACAO\_CAT (para categórica) que registra dentro de qual quartil cada observação se encontra (1, 2, 3 ou 4). Depois, calcule a média de PROP\_VOTOS por partido em cada nível de TAMANHO\_COLIGACAO\_CAT (isto é, calcule a média de PROP\_VOTOS para as observações do PT no quartil de número 1, na sequência para as observações do PT no quartil de número 2, e assim por diante). Apresente o resultado em uma tabela com as variáveis SIGLA\_PARTIDO, TAMANHO\_COLIGACAO\_CAT e PROP\_VOTOS (que registra as médias calculadas). (1,0)**

**PT**

1º quartil (Q1) = 4

2º quartil (Q2) = 7

3º quartil (Q3) = 11

4º quartil (Q4) = 18

**PSDB**

1º quartil (Q1) = 9

2º quartil (Q2) = 14

3º quartil (Q3) = 16.5

4º quartil (Q4) = 21.0

**Tabela**

|  |  |  |
| --- | --- | --- |
| **SIGLA\_PARTIDO** | **TAMANHO\_COLIGACAO\_CAT** | **PROP\_VOTOS** |
| PSDB | 1 | 0.285 |
| PSDB | 2 | 0.371 |
| PSDB | 3 | 0.278 |
| PSDB | 4 | 0.268 |
| PT | 1 | 0.375 |
| PT | 2 | 0.393 |
| PT | 3 | 0.575 |
| PT | 4 | 0.453 |

**C. Uma maneira de se entender melhor a variação dos valores de uma variável é normalizando-os. Uma normalização muito empregada é a pelo desvio-padrão. Para normalizar um valor pelo desvio padrão, nós subtraímos a média e dividimos pelo desvio-padrão. Isso nos permite observar a quantos desvios padrão da média um determinado valor se encontra, nos dando uma melhor ideia de quão baixo ou quão alto ele é. Normalize os valores calculados no exercício anterior, usando as médias e os desvios-padrão calculados em “A” (a média e o desvio-padrão do PT para os valores do PT e igualmente para o PSDB.) e os registre em uma nova variável PROP\_VOTOS\_NORM. Apresente seus resultados em uma tabela com as variáveis SIGLA\_PARTIDO, TAMANHO\_COLIGACAO\_CAT, PROP\_VOTOS e PROP\_VOTOS\_NORM. (1,0)**

|  |  |  |  |
| --- | --- | --- | --- |
| **SIGLA\_PARTIDO** | **TAMANHO\_COLIGACAO\_CAT** | **PROP\_VOTOS** | **PROP\_VOTOS\_NORM** |
| **PSDB** | **1** | **0.285** | **-0.157** |
| **PSDB** | **2** | **0.371** | **0.517** |
| **PSDB** | **3** | **0.278** | **-0.212** |
| **PSDB** | **4** | **0.268** | **-0.291** |
| **PT** | **1** | **0.375** | **-0.565** |
| **PT** | **2** | **0.393** | **-0.433** |
| **PT** | **3** | **0.575** | **0.892** |
| **PT** | **4** | **0.453** | **-0.00196** |

**D. Agora podemos tentar tirar algumas conclusões a partir dos valores que calculamos. De que forma a proporção de votos obtidas por cada partido varia conforme aumenta o tamanho das coligações de que participa? Ela varia da mesma forma para ambos os partidos? A variação é grande (compare os valores normalizados)? Para ambos os partidos? Diante desses apontamentos, a nossa hipótese inicial faz sentido? Comporta alguma qualificação? (1,0)**

D. A partir da análise da tabela construída no exercício C, é possível verificar que esses partidos possuem tendências diferentes sobre a relação entre o tamanho da coligação e os votos obtidos. No caso do PSDB, é possível verificar que coligações maiores obtiveram resultados piores, e em coligações menores, obteve-se melhores resultados. Com o PT, ocorreu o contrário, seus piores resultados se deram em coligações menores, e os melhores resultados, em coligações maiores. Outro fato também se verifica. Apesar de haver uma tendência de correlação diretamente proporcional no caso do PT, e uma correlação inversamente proporcional no caso do PSDB, a partir de um certo ponto essa relação perde força, indicando que essa relação é verdadeira até certo ponto.

Assim, pelos dados analisados, podemos verificar que a hipótese acerca do fenômeno pode ser rejeitada, pois a correlação diretamente proporcional entre tamanho de coligação e desempenho eleitoral apenas é verificado com o PT, e ainda assim com ressalvas. Os dados indicam que há um certo tamanho de coligação que apresenta melhores resultados, porém é dependente de outros fatores, tendo em vista que esse tamanho ideal divergiu entre os partidos.

**Script para realização dos exercícios na linguagem R**

|  |
| --- |
| *#Lista 5* **library**(tidyverse) **library**(readxl)  *# Obtendo os dados* dados <- read\_delim('Lista 5/dados.csv', col\_names = T, quote = "\"", delim = ';') dados$PROP\_VOTOS <- as.numeric(sub(",", ".", dados$PROP\_VOTOS, fixed = TRUE))  *# Exercício 2*  *# A)* stats <- dados %>%   group\_by(SIGLA\_PARTIDO) %>%   summarise("Média" = mean(PROP\_VOTOS), "Variância" = var(PROP\_VOTOS) ,"Desvio-padrão" = sd(PROP\_VOTOS))  dados %>%   subset(SIGLA\_PARTIDO == "PSDB") %>%   select(PROP\_VOTOS) %>%   print.data.frame()  *# B)* *#Medianas* dados %>%   group\_by(SIGLA\_PARTIDO) %>%   summarise("Mediana" = median(PROP\_VOTOS))  *#Quartis TAMANHO\_COLIGACAO - PT* quartis\_pt <- dados %>%   subset(SIGLA\_PARTIDO == "PT") %>%   .$TAMANHO\_COLIGACAO %>%   quantile()  *#Quartis TAMANHO\_COLIGACAO - PSDB* quartis\_psdb <- dados %>%   subset(SIGLA\_PARTIDO == "PSDB") %>%   .$TAMANHO\_COLIGACAO %>%   quantile()  *# Função para classificar o quartil* getQuartile <- **function**(x, vector) {  **if** (x <= vector["25%"]){  **return**(1)  } **else** **if** (x <= vector["50%"]){  **return**(2)  } **else** **if** (x <= vector["75%"]){  **return**(3)  } **else** {  **return**(4)  } }  *# Criando a variável TAMANHO\_COLIGACAO\_CAT* dados <- ungroup(dados) dados\_col <- dados %>%   rowwise %>%   mutate("TAMANHO\_COLIGACAO\_CAT" = if\_else(SIGLA\_PARTIDO == "PT", true = getQuartile(TAMANHO\_COLIGACAO, quartis\_pt),  false = getQuartile(TAMANHO\_COLIGACAO, quartis\_psdb)))  *# Criando a tabela por Categoria de Quartil e Sigla* dados\_coligacao <- dados\_col %>%  group\_by(SIGLA\_PARTIDO, TAMANHO\_COLIGACAO\_CAT) %>%   summarise("PROP\_VOTOS" = mean(PROP\_VOTOS))  *# Exercício C* dados\_colig\_norm <- dados\_coligacao %>%   mutate("PROP\_VOTOS\_NORM" = if\_else(SIGLA\_PARTIDO == "PT", true = (PROP\_VOTOS - stats$Média[stats$SIGLA\_PARTIDO == "PT"]) / stats$`Desvio-padrão`[stats$SIGLA\_PARTIDO == "PT"],  false = (PROP\_VOTOS - stats$Média[stats$SIGLA\_PARTIDO == "PSDB"]) / stats$`Desvio-padrão`[stats$SIGLA\_PARTIDO == "PSDB"])) |