

**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/ 2018**

**Profº Dr. Glauco Peres da Silva**

**LISTA DE EXERCÍCIO 03**

**Data de entrega: 02/04/2018 (noturno) e 04/04/2018 (vespertino)**

**Exercício 1 (2 pontos)**

Marque verdadeiro (V) ou falso (F) para cada uma das alternativas abaixo, justificando todas as suas respostas.

( ) – Apesar de semelhantes, diagramas de colunas e histogramas são adequados para diferentes tipos de variáveis. Os diagramas de colunas são utilizados para descrever variáveis categóricas e os histogramas para variáveis contínuas.

( ) – A mediana não é considerada uma medida de posição.

( ) – A medida de tendência central mais adequada para distribuições assimétricas é a mediana. Distribuições assimétricas à direita geralmente possuem um valor de mediana menor que o da média. Podemos ainda comparar os valores da distribuição em termos de desvios-padrão da média para diagnosticar assimetria.

( ) – Valores atípicos podem afetar análises estatísticas e devem ser analisados com cuidado. Uma boa maneira de apresentá-los é pelos diagramas de caixa e bigodes (boxplots). Uma regra de bolso arbitrária é considerar valores a mais que 1,5 intervalos interquartis (IIQ) da média ou a menos que 1,5 IIQ da média.

**Para as questões 2 e 3, considere o texto e a tabela a seguir.**

A Tabela 1 apresenta o rendimento médio dos trabalhos em 2014 em reais (R\$), calculado a partir dos dados da Pesquisa Nacional por Amostra de Domicílios (PNAD), realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). A variável está disponível no Ipeadata ([www.ipeadata.gov.br](http://www.ipeadata.gov.br)). Os valores foram deflacionados para preços de outubro de 2012. Segue abaixo a descrição da variável fornecida pelo Ipeadata.

“Média, por pessoa ocupada, dos rendimentos mensais brutos totais em dinheiro recebidos em todos os trabalhos no mês de referência da Pesquisa Nacional por Amostra de Domicílios (Pnad/IBGE). No caso de empregados, considera-se a remuneração obtida no mês de referência, tendo ou não trabalhado o mês completo. No caso de rendimento variável, média da remuneração mensal recebida. Para empregadores e trabalhadores por conta própria, retirada no mês de referência, ou seja, o rendimento bruto menos as despesas efetuadas com salários de empregados, matéria-prima, energia elétrica, telefone etc., sendo ainda considerado, no caso de rendimento variável, a retirada média mensal. Em nenhum caso são computadas a parcela referente ao 13o. salário (14o., 15o. etc.), nem a parcela referente à participação nos lucros paga pelas empresas aos empregados. Valores reais expressos aos preços vigentes no mês de referência da última Pnad disponível, calculados a partir dos microdados da pesquisa e atualizados conforme o deflator para rendimentos da Pnad apresentado pelo Ipeadata”.

**Tabela 1. Rendimento médio dos trabalhos em cada unidade federativa (2014)**

UF	Valor	UF	Valor
AC	1624,96	PB	1134,46
AL	1080,24	PE	1208,64
AM	1610,08	PI	984,93
AP	1695,26	PR	1882,98
BA	1203,25	RJ	2099,31
CE	1078,42	RN	1171,36
ES	1640,27	RO	1665,65
GO	1670,1	RR	1625,16
MA	963,82	RS	1844,38
MG	1592,85	SC	1957,21
MS	1882,35	SE	1080,57
MT	1856,6	TO	1516,92
PA	1287,58		

Fonte: Ipeadata.

Nas questões 2 e 3, trataremos o rendimento médio como uma variável contínua. Disponibilizamos no Moodle essa tabela em uma planilha Excel com o nome “Ipeadata – Rendimento médio” para facilitar a operacionalização.

### **Exercício 2 (4 pontos)**

**2.1** Você deve estudar as diferenças de rendimento entre as unidades federativas brasileiras. Com o auxílio de uma calculadora ou de um software, calcule as estatísticas descritivas da amostra da Tabela 1 solicitadas nos itens a seguir e **indique o passo-a-passo** que utilizou para o cálculo. Você não deve usar fórmulas prontas de software na sua explicação, tais como MÉDIA(A1:A25). Respostas que não descrevam o passo-a-passo do cálculo serão desconsideradas. Você pode optar por utilizar fórmulas para descrever seu cálculo, mas deverá explicar a que se refere cada uma das variáveis

indicadas na fórmula. Ex.: A fórmula da média é  $\sum y_i/n$ , em que  $y_i$  é o valor observado em cada unidade federativa  $i$  e  $n$  é o tamanho da amostra.

- a) Média. Tamanho máximo da resposta: 5 linhas. **0,5 ponto.**
- b) Mediana, quartil inferior, quartil superior e intervalo interquartil. Tamanho máximo da resposta: 10 linhas. **1 ponto.**
- c) A soma dos desvios. Você esperava esse resultado? Tamanho máximo da resposta: 5 linhas. **1 ponto.**
- d) A variância e o desvio-padrão. Tamanho máximo da resposta: 6 linhas. **1 ponto.**

2. 2 Você usaria a moda como medida de tendência central das variáveis da Tabela 1? Por quê? Tamanho máximo da resposta: 3 linhas. **0,5 ponto.**

### Exercício 3 (4 pontos)

Agora que você já tem algumas estatísticas descritivas, deseja analisar um pouco a dispersão dos seus dados.

- a) Faça um boxplot da amostra da Tabela 1. Qual a amplitude da dispersão da metade dos valores mais próximos da mediana? Existem valores atípicos nessa distribuição? Tamanho máximo da resposta: 4 linhas. **1 ponto.**
- b) Apenas com base na média e na mediana, você espera que a distribuição seja simétrica ou assimétrica? O boxplot desenhado no item anterior confirma essa hipótese? Se você estiver com dificuldades em interpretar o boxplot, tente fazer um diagrama de frequências ou um histograma. Tamanho máximo da resposta: 8 linhas. **1 ponto.**
- c) Você percebe que se esqueceu de incluir na amostra duas unidades federativas, apresentadas na Tabela 2.

**Tabela 2**

UF	Valor
SP	2203,28
DF	3230,25

Fonte: Ipeadata.

Junte na mesma amostra as observações da Tabela 1 e da Tabela 2 e recalcule a média e a mediana. Qual a diferença entre os valores obtidos agora e os valores da média e

da mediana calculados na Questão 2? Explique porque uma das medidas de tendência central mudou mais que a outra com a inclusão das observações SP e DF. Tamanho máximo da resposta: 8 linhas. **2 pontos**.

**Exercício 4 (5 pontos) – Apenas para alunos da pós-graduação**

Além da realização de pesquisas de opinião com amostras da população de uma cidade ou país acerca de temas relevantes para a opinião pública, é possível também a realização de *surveys* com outras populações de interesse, como representantes políticos. André e Depauw (2015) utilizam os resultados de um *survey* aplicado a membros de legislativos nacionais e regionais de diversos países europeus para analisar a relevância dada por esses políticos para a representação de pequenas porções geográficas (como o seu município de residência) ou da totalidade do distrito eleitoral, em diferentes contextos institucionais - dados principalmente pela magnitude do distrito (isto é, do número de cadeiras em disputa em um distrito eleitoral). A partir da leitura do artigo indicado, responda:

**a)** Na seção sobre os dados do trabalho (*data*), os autores discutem a amostra utilizada, a operacionalização e descrição das suas variáveis. Eles afirmam que possíveis vieses (tendenciosidades) da amostra não teriam ocorrido. Discuta um desses vieses, como ele poderia ter afetado a amostra e porque os autores afirmam que eles não são relevantes neste estudo. (Máximo 10 linhas) **(1 ponto)**

(Dica: não é necessário discutir índices específicos, somente ideias gerais sobre a composição da amostra).

**b)** Qual a variável independente (VI) deste trabalho? E sua variável dependente (VD)? Como os autores operacionalizaram cada uma delas? (Máximo 15 linhas) **(2 pontos)**

**c)** Quais as estatísticas descritivas utilizados pelos autores para caracterizar as variáveis independente e dependente do trabalho? Além disso, os autores também transformações nas variáveis para explorar uma diferente categorização dessas variáveis. Descreva como isto é feito. (Máximo 10 linhas) **(1,5 ponto)**

**d)** Segundo os autores, a tabela 2 apresenta algumas evidências obtidas a partir de estatísticas descritivas que sustentam a sua hipótese. Por que os autores não pararam a sua análise neste ponto, tendo em vista as evidências encontradas? (Máximo 5 linhas) **(0,5 ponto)**