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 11**

**Aula 12 – “Introdução à regressão e final do bloco sobre inferência estatística"**

**Data de entrega: 10/06/2019 (noturno) e 12/06/2019 (vespertino)**

**Nome:**

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

**Exercício 1**

Com base na bibliografia obrigatória indicada, marque “Verdadeiro” (V) ou “Falso” (F) para as afirmações a seguir e, quando falsas, justifique sua escolha em 5 linhas,  no máximo.

A. () Em uma função linear o intercepto-y indica o grau de inclinação de uma reta de regressão.

B. () O gráfico de dispersão pode ser útil para visualizarmos indícios de associação linear entre variáveis. No entanto, a fim de confirmarmos a existência ou ausência de associação, é imprescindível controlarmos por variáveis colineares.

C. () A inclinação da reta de regressão é significativamente afetada por valores atípicos (*outliers*), especialmente quando lidamos com uma amostra pequena. Por isso, em determinados casos é recomendável desconsiderá-los para a formulação da equação de regressão. Tais valores também podem estar relacionados a falhas no processo de mensuração. Contudo, se de fato estiverem corretos, merecem ser analisados em estudos de caso.

D. () Sempre que a correlação e a inclinação da reta de regressão são iguais ou próximas a zero podemos constatar que não há associação entre as variáveis.

E. () Uma associação linear forte é um indicativo de que a equação de regressão fornece previsões melhores do que a média amostral. Nesse sentido, a soma dos quadrados dos resíduos, correspondentes à diferença entre os valores de fato observados e os valores previstos pela equação, é menor do que a soma dos quadrados total dos valores-y em relação a sua média.

F. () Quando identificamos uma associação linear forte podemos constatar que encontramos uma relação causal.

G. () A correlação (r) é muito útil para comparações, pois, ao contrário da inclinação, ela não depende das unidades de mensuração, variando sempre entre -1 e 1. Basicamente, ela permite a padronização de diferentes variáveis, convertendo seus valores em escores-z.

H. () A estatística-teste para testar a hipótese nula de uma equação de regressão pode ser calculada tanto com base no coeficiente de correlação, associação linear e tamanho da amostra, quanto na inclinação amostral e seu erro padrão.

**Exercício 2**

Para esse exercício vamos utilizar o arquivo “dados.csv”. Vamos voltar à nossa investigação sobre a relação entre tamanho de coligação na disputa para o governo do estado e proporção de votos para presidente. Dessa vez vamos nos concentrar apenas nas observações do PT em estados nos quais o governador não era um aliado, que, como vimos, são os casos em que o tamanho da coligação parece ter uma relação positiva com a proporção de votos.

1. Queremos saber qual a magnitude dessa relação. Para tanto, calcule o α e o β do seguinte modelo de regressão:

PROP\_VOTOS = α + β\*TAMANHO\_COLIGACAO

O que significa o valor do β? Obs: o output de uma função de regressão não será aceito como resposta.

1. Também queremos saber se esse coeficiente reflete uma relação real ou é um artifício da nossa amostra. Para tanto, construa o intervalo de confiança de 95% para o β (assuma um t-score de 1,96). O β é estatisticamente significante no nível de 95% de confiança? Obs: o output de uma função de regressão não será aceito como resposta.

**Exercício 3**

A base de dados “Banco\_Horas\_Leitura” foi criada a partir da observação das horas de Leitura extra-classe que estudantes de Graduação e Pós-Graduação de um curso de estatística realizaram durante uma semana. Com as informações do banco, monte uma tabela com as médias e os dv padrão das amostras, posteriormente apresente o **erro padrão da diferença das médias** entre as duas amostras, a razão das médias e por fim, o intervalo de 99% de confiança para o tempo médio a mais gasto por estudantes de pós-graduação em seus estudos semanais.