

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
Prof. Dr. Glauco Peres da Silva

LISTA DE EXERCÍCIOS 12
“Modelos de Regressão II - Regressões Multivariadas”

Data de entrega: 17/06/2019 (noturno) e 19/06/2019 (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. (**V**). A causalidade é um conceito fundamental para as ciências de maneira geral. Para podermos afirmar que X causa Y, devemos comprovar que X e Y covariam, que X antecede temporalmente Y, e que os efeitos de X em Y não desaparecem caso incluamos explicações alternativas em nossos modelos.

Verdadeiro. A associação, a antecedência temporal e a eliminação de explicações alternativas são os elementos básicos para se identificar a causalidade entre dois ou mais fenômenos. Vide Agresti & Finlay, pp. 339-341.

2. (**V**). Suponha que queremos testar uma relação causal de que X causa Y. Ao incluirmos variáveis controle em nosso modelo, o que estamos fazendo é controlar os efeitos de explicações causais alternativas ao nosso modelo original.

Verdadeiro. De acordo com Agresti & Finlay (p. 341), “um componente fundamental para avaliar se X pode causar Y é procurar por uma explicação alternativa. Fazemos isto estudando se a associação entre X e Y continua quando removemos os efeitos de outras variáveis dessa

associação. Em uma análise multivariada, uma variável está ou é controlada quando sua influência é removida.”.

3. (**F**). Variáveis omitidas são variáveis que não foram incluídas em nossos modelos. Podemos descartar a hipótese da existência de variáveis omitidas caso nosso modelo de regressão mostre uma associação estatisticamente significativa entre nossa variável dependente e nossos regressores.

Falso. Modelos estatísticos reportam como os dados se comportam para as variáveis incluídas em um modelo. Caso haja alguma variável que foi omitida dos modelos, os coeficientes reportados serão enviesados. Por isso é tão importante a adição de variáveis controle aos modelos (Agresti & Finlay, p.341-343). Ademais, é através da investigação teórica que sabemos quais variáveis devem ser incluídas em nossos modelos estatísticos (Agresti & Finlay, p. 344).

4. (**F**). A associação espúria entre duas variáveis (X e Y) desaparece quando se controla por uma terceira variável (Z). Dizer que uma associação entre duas ou mais variáveis é espúria, todavia, não implica dizer que não há uma relação causal entre elas.

Falso. Associações espúrias não apresentam relações causais entre si. (Agresti & Finlay, p. 345).

5. (**F**). Suponha que tenhamos um modelo que utilize duas variáveis (X_1 e X_2) para explicar a variação de nossa variável de interesse Y . Dizemos que há uma interação estatística entre X_1 e X_2 quando o efeito de X_1 em Y varia conforme mantemos X_2 constante, e vice-versa.

Falso. Uma interação estatística pressupõe que o efeito de X_1 em Y muda de acordo com o valor de X_2 , e vice-versa. Ou seja, seus valores não permanecem constantes. Vide Agresti & Finlay, pp. 348-349 e p. 353.

Exercício 02

A taxa aprovação presidencial é um fator crucial para o sucesso de um governo. Para os exercícios abaixo, fazemos um modelo de regressão bem simplificado da taxa de aprovação mensal dos presidentes dos EUA, do período de 1953 a 2015. Os foram retirados do artigo “Motivated Reasoning, Public Opinion, and Presidential Approval, de Donovan et al (2019)¹. Nossa variável dependente é a taxa de aprovação, medida em porcentagem. Temos 3

¹ Artigo disponível em: <https://link.springer.com/article/10.1007/s11109-019-09539-8>. Dados podem ser obtidos em: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/MN4PT4>.

variáveis independentes: a taxa de inflação (medida em porcentagem), a taxa de desemprego (porcentagem), e o número de casualidades de guerra (contagem do número de mortos).

Ao rodarmos nossa regressão, obtivemos os seguintes valores:

Termo	Coefficiente	Erro Padrão	Estatística-F	P-Valor
Intercepto	69.4228717	1.6724416	41.509893	0.000
Desemprego	-1.8712192	0.2873718	-6.511491	0.0000
Inflação	-0.0260636	0.0069111	-3.771261	0.0001752
Casualidades	-0.1452615	0.0219313	-6.623488	0.000

a) Escreva a equação do modelo utilizado.

$$\text{Aprovação} = \alpha + \beta_1 * \text{Inflação} + \beta_2 * \text{Desemprego} + \beta_3 * \text{Casualidades de Guerra} + e.$$

Adaptando para os resultados que obtivemos, temos:

$$\text{Aprovação} = 69.42 - 1.87 * \text{Inflação} - 0.02 * \text{Desemprego} - 0.14 * \text{Casualidades de Guerra} + e.$$

Em que “ α ” é nosso intercepto, os “ β s” são os coeficientes angulares (também chamados de inclinação ou declive) e “ e ” é o termo de erro.

b) O que os valores dos termos significam? Explique o que significa o α e os β do modelo de regressão. Explique o que significa um coeficiente β ser positivo ou negativo. Descreva brevemente o resultado encontrado para cada um dos termos. **Dica:** tome cuidado com as unidades de análise utilizadas em cada variável.

O intercepto (α) é o valor médio da variável dependente assumindo que as variáveis independentes sejam constantes em 0. Para nosso presente caso, mantidas as demais variáveis constantes, temos uma taxa de aprovação base de 69% em no modelo que produzimos.

Já os coeficientes β medem o impacto da mudança de uma unidade de análise de seu sua variável independente (X) na sua variável dependente (Y). Um valor positivo de β implica que conforme a variável independente aumenta, a variável dependente também aumenta. O inverso se aplica para valores negativos de β , i.e., conforme se aumenta X se diminui Y.

Para o desemprego, vemos que o aumento de 1% do desemprego implica em diminuir a aprovação presidencial em aproximadamente 1.87%. Já para a inflação, vemos que o aumento de 1% na inflação corresponde a um decréscimo de 0.02 da taxa de aprovação presidencial.

Por fim, vemos que o aumento de 1 casualidade de guerra corresponde a um decréscimo de 0.14% da aprovação presidencial.

c) Interprete os p-valores para cada um dos coeficientes β . Qual é a hipótese nula sendo testada para cada um desses termos? Podemos rejeitar ou não tais hipóteses nulas em um nível de 95% de confiança? E para um nível de 99% de confiança?

A Hipótese-Nula para os β é a de que o valor do coeficiente é igual a 0, isto é, que a variável independente não tem nenhum efeito sobre a variável dependente. Os p-valores são muito inferiores a 0.05 (IC de 95%) e a 0.01 (IC de 99%), ou seja, podemos rejeitar a hipótese nula para o presente modelo.

d) A estatística R^2 para o modelo é $R^2 = 0.1672$. O que esse número nos diz sobre nosso modelo?

Nas palavras de Agresti & Finlay, p. 372, “o R^2 mensura a proporção da variação total em y que é explicada pelo poder de previsão de todas as variáveis explicativas, por meio do modelo da regressão múltipla”. Isto é, o modelo utilizado explica cerca de 16.72% da variação total da taxa de aprovação presidencial.

e) Cite dois exemplos de variáveis omitidas e explique o porquê de, na sua opinião, elas precisarem ser controladas. Defina que tipo de variável elas são (categóricas, contínuas, etc), bem como quais os valores possíveis que elas podem assumir. Pensando teoricamente, o que ocorre com nossos coeficientes caso não controlamos essas variáveis? Justifique sua resposta.

Caso haja variáveis omitidas, isto é, variáveis que deixamos de incluir em nossos modelos, estaremos distorcendo os valores de nossos coeficientes, bem como inflacionando nosso termo de erro (Agresti & Finlay, 341-344).

No limite, uma variável que pode influenciar a taxa de aprovação presidencial seria o próprio presidente. No limite, presidentes, enquanto indivíduos, possuem características pessoais próprias, que os podem tornar mais ou menos aprováveis pela população em geral. Seria uma variável categórica, contendo o nome cada presidente americano eleito de 1953 a 2015.

Podemos incluir mais variáveis econômicas no nosso modelo, como o PIB (medido em dólar); o salário mínimo corrigido (medido em dólar, assumindo valores positivos); a taxa de impostos (medido, por exemplo, como a porcentagem do salário mínimo); dentre outras variáveis.

Podemos modificar nossa variável de guerra, transformando-a em uma variável dicotômica, isto é, se o país está em guerra ou não. A vantagem de tal transformação é que se diminui a variabilidade dos dados. Pensando na teoria, estamos mais interessados se a taxa de aprovação se altera quando o país está em guerra, e não o efeito do aumento da morte de 1 pessoa em guerra.

Podemos incluir variáveis temporais, como o ano (variável discreta, 1953 a 2015) ou o mês (categórica, janeiro a dezembro). Podemos gerar outras variáveis a partir das variáveis

temporais, como, por exemplo, se a aprovação é do primeiro ano de mandato do presidente (variável dicotômica), testando-se, assim, o “efeito lua de mel”.

Podemos incluir variáveis relativas ao desempenho legislativo, como a proporção de cadeiras ocupadas pelo partido e/ou coligação do presidente, o número de propostas aprovadas, ou mesmo o número de promessas de campanha cumpridas.