

Universidade Federal do Paraná - UFPR

Especialização em Data Science & Big Data

Professor: José Luiz Padilha da Silva

Data de entrega: 31/08/18

Questão 1. Considere a função de produção de Cobb-Douglas:

$$Y_i = \beta_0 X_{i1}^{\beta_1} X_{i2}^{\beta_2} e^{\varepsilon_i},$$

em que Y é a produção, X_1 é o insumo de trabalho, X_2 é o insumo de capital, e ε é o termo de erro. A soma $(\beta_1 + \beta_2)$ informa a respeito dos *retornos de escala*, a resposta do produto a uma variação proporcional nos insumos. Se essa soma for igual a 1, haverá *retornos constantes de escala*, isto é, se dobrarmos os insumos, a produção dobrará, se o triplicarmos, a produção triplicará.

- (a) Escreva a função de produção de Cobb-Douglas na forma do modelo linear geral.
- (b) Escreva as hipóteses nula e alternativa para testar se há retornos constantes de escala. Qual é o procedimento estatístico a ser usado?
- (c) Escreva o modelo reduzido referente ao teste em (b).

Questão 2. Considere uma amostra de tamanho $n = 20$ de uma variável aleatória Y e de um conjunto de variáveis explicativas X_1, X_2, X_3 e X_4 , para a qual foram ajustados os seguintes modelos:

```
> fit1=lm(Y~X1+X2+X3+X4)
```

```
> fit2=lm(Y~X2+X1+X3+X4)
```

```
> summary(fit1)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.07555	1.91636	1.605	0.1294
X1	0.83243	0.13337	6.242	1.58e-05 ***
X2	0.19218	0.09452	2.033	0.0601 .
X3	0.80688	0.31582	2.555	0.0220 *
X4	-0.23702	0.15195	-1.560	0.1396

```
> summary(fit2)
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.07555	1.91636	1.605	0.1294
X4	-0.23702	0.15195	-1.560	0.1396
X2	0.19218	0.09452	2.033	0.0601 .
X1	0.83243	0.13337	6.242	1.58e-05 ***
X3	0.80688	0.31582	2.555	0.0220 *

- a) Escreva o modelo e interprete os coeficientes das variáveis significativas para o primeiro ajuste (Assuma $\alpha = 5\%$).
- b) Foi também realizado o teste F para verificação da significância das variáveis potencialmente preditoras, sendo obtido os resultados a seguir:

```
> anova(fit1)
```

Analysis of Variance Table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
X1	1	59.687	59.687	66.8585	6.573e-07 ***
X2	1	2.724	2.724	3.0519	0.101082
X3	1	11.351	11.351	12.7147	0.002817 **
X4	1	2.172	2.172	2.4331	0.139645
Residuals	15	13.391	0.893		

```
> anova(fit2)
Analysis of Variance Table

      Df Sum Sq Mean Sq F value    Pr(>F)
X4      1  5.358    5.358   6.0020  0.02705 *
X2      1  0.363    0.363   0.4062  0.53349
X1      1 64.386   64.386  72.1225 4.095e-07 ***
X3      1  5.827    5.827   6.5273  0.02199 *
Residuals 15 13.391    0.893
```

As conclusões obtidas pela ANOVA são bem distintas entre os ajustes *fit1* e *fit2*. Por exemplo, para o primeiro ajuste vemos que X_4 apresenta $p = 0.1396$ enquanto para o segundo ajuste temos $p = 0.027$. Compare com o $p = 0.1396$ do teste t para ambos os modelos. Explique a que se deve essa diferença.

Questão 3. Para os dados da questão anterior o software relata

```
Residual standard error: 0.9448 on 15 degrees of freedom
Multiple R-squared:  0.8501,    Adjusted R-squared:  0.8101
F-statistic: 21.26 on 4 and 15 DF,  p-value: 4.86e-06
```

- Interprete a estatística R^2 . Explique por que ele é maior que o R^2 ajustado e se tal fato é esperado.
- Escreva as hipóteses relacionadas com a estatística F . Com base no valor de p , qual é a conclusão?

Questão 4. A Zarthan Company vende um creme para a pele exclusivamente através de lojas de moda. Estão disponíveis dados de venda para quinze distritos. Vendas (em lotes) é tratada como a variável dependente Y , e população alvo (em milhares de pessoas) e renda per capita (em dólares) são as variáveis independentes X_1 e X_2 , respectivamente. Espera-se que o modelo $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i$, $i = 1, \dots, n$ com erros normais seja adequado. Os resultados do ajuste foram:

```
> fit=lm(Y~X1+X2,data=dados)
> summary(fit)

      Estimate Std. Error t value Pr(>|t|)
(Intercept) 19.374572   11.786833   1.644    0.126
X1           0.479601    0.026852  17.861 5.2e-10 ***
X2           0.004429    0.004526   0.979    0.347
---
Residual standard error: 10.29 on 12 degrees of freedom
Multiple R-squared:  0.9734,    Adjusted R-squared:  0.9689
F-statistic: 219.3 on 2 and 12 DF,  p-value: 3.567e-10
```

- O R^2 obtido é bastante alto. Neste caso, por que devemos realizar análise de resíduos e de observações influentes?
- Assuma, por ora, que o modelo é adequado, e suponha que a companhia tenha interesse em estimar vendas em um distrito com população alvo $X_{h1} = 220$ mil pessoas e renda per capita $X_{h2} = 2500$ dólares. A estimativa pontual encontrada foi $\hat{Y}_h = 135.96$. Um intervalo de confiança de 95% calculado para a resposta média foi $(129.21 - 142.71)$, e o intervalo obtida para uma nova observação foi $(112.55 - 159.37)$. Explique a diferença entre eles.