Universidade Federal do Paraná - Departamento de Estatística
CE310 - Modelos de Regressão Linear
Prof. Cesar Augusto Taconeli
Lista de exercícios

1. Defina o modelo de regressão linear simples. Especifique cada um de seus componentes e as suposições assumidas para os erros.

Resposta:

$$y = \beta_0 + \beta_1 x + \epsilon, \quad \epsilon \sim \mathrm{N}(0, \sigma^2),$$

com $\mathrm{Cov}(\epsilon_i, \epsilon_j) = 0$ para $i \neq j$.

2. Qual o princípio da estimação por mínimos quadrados? Quais as principais propriedades dos estimadores de mínimos quadrados dos parâmetros de um modelo de regressão linear?

Resposta: Encontrar valores para os parâmetros tais que a soma dos quadrados dos erros seja mínima. Os estimadores de mínimos quadrados são não viciados, lineares (são combinações lineares dos $y's$) e, na classe dos estimadores lineares não viciados, são os mais eficientes (Teorema de Gauss Markov).

3. Qual o princípio da estimação por máxima verossimilhança? Qual a relação dos estimadores de mínimos quadrados e de máxima verossimilhança dos parâmetros do modelo de regressão linear se assumirmos que os erros são normalmente distribuídos?

Resposta: Encontrar valores para os parâmetros tais que a função de verossimilhança atinja seu máximo. Se os erros forem normalmente distribuídos, os estimadores de mínimos quadrados e de máxima verossimilhança dos $\beta's$ são idênticos.

4. Considere o modelo de regressão linear simples com $\beta_0 = 10$, $\beta_1 = 5$ e $\sigma = 4$. Assuma distribuição normal para os erros.

a) Apresente gráficos da distribuição de y condicional a (i) $x = 3$; (ii) $x = 5$ e (iii) $x = 10$;

b) Descreva o significado de β_0 e β_1. Suponha que $x = 0$ pertença ao escopo do modelo;

Resposta: $\beta_0 = 10$ corresponde ao valor esperado (média) de y quando $x = 0$. $\beta_1 = 5$ significa que para cada acréscimo de uma unidade em x espera-se um acréscimo de 5 unidades em y.

c) Calcule $P(20 < y < 30)$ para: (i) $x = 3$; $x = 5$.

Resposta:

Para $x = 3$, $y \sim \mathrm{N}(\mu = 10 + 5 \times 3 = 25, \sigma = 4)$. Então,

$$\begin{aligned} P(20 < y < 30) &= P(\frac{20-25}{4} < Z < \frac{30-25}{4}) \\ &= P(-1.25 < Z < 1.25) = 0.7887 \end{aligned}$$

Para $x = 5$, $y \sim \mathrm{N}(\mu = 10 + 5 \times 5 = 35, \sigma = 4)$. Então,

$$
\begin{aligned}
P(20 < y < 30) &= P(\frac{20-35}{4} < Z < \frac{30-35}{4}) \\
&= P(-3.75 < Z < -1.25) = 0.1055
\end{aligned}
$$

5. Considere o modelo de regressão linear simples com $\beta_0 = 10$; $\beta_1 = 0.5$ e $\sigma = 1$. Assuma que os erros sejam normalmente distribuídos.

a) Simule uma amostra de $n = 10$ observações para y locadas em x = -2, -1, 0, 1 e 2 (duas observações para cada valor de x);

Resposta:

```
x <- rep(-2:2, each = 2)
y <- rnorm(10, mean = 10 + 0.5*x, sd = 1)
y
```

```
##  [1]  8.841373  8.042524  9.700475  9.203350  8.267595  9.587911 10.699350
##  [8] 10.657847 12.321834 10.253720
```

b) Ajuste um modelo de regressão linear simples aos dados simulados no item a. Extraia as estimativas de β_0 e β_1.

Resposta:

```
fit <- lm(y ~ x)
coef(fit)
```

```
## (Intercept)           x
##   9.7575979   0.6918342
```

```
summary(fit)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.4900 -0.2910  0.1730  0.4131  1.1806
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.7576     0.2571  37.958 2.55e-10 ***
## x             0.6918     0.1818   3.806  0.00519 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8129 on 8 degrees of freedom
## Multiple R-squared:  0.6442, Adjusted R-squared:  0.5998
## F-statistic: 14.49 on 1 and 8 DF,  p-value: 0.005191
```

c) Repita os itens a e c 5000 vezes. Armazene as estimativas obtidas numa matriz com duas colunas (uma referente a cada parâmetro);

Resposta:

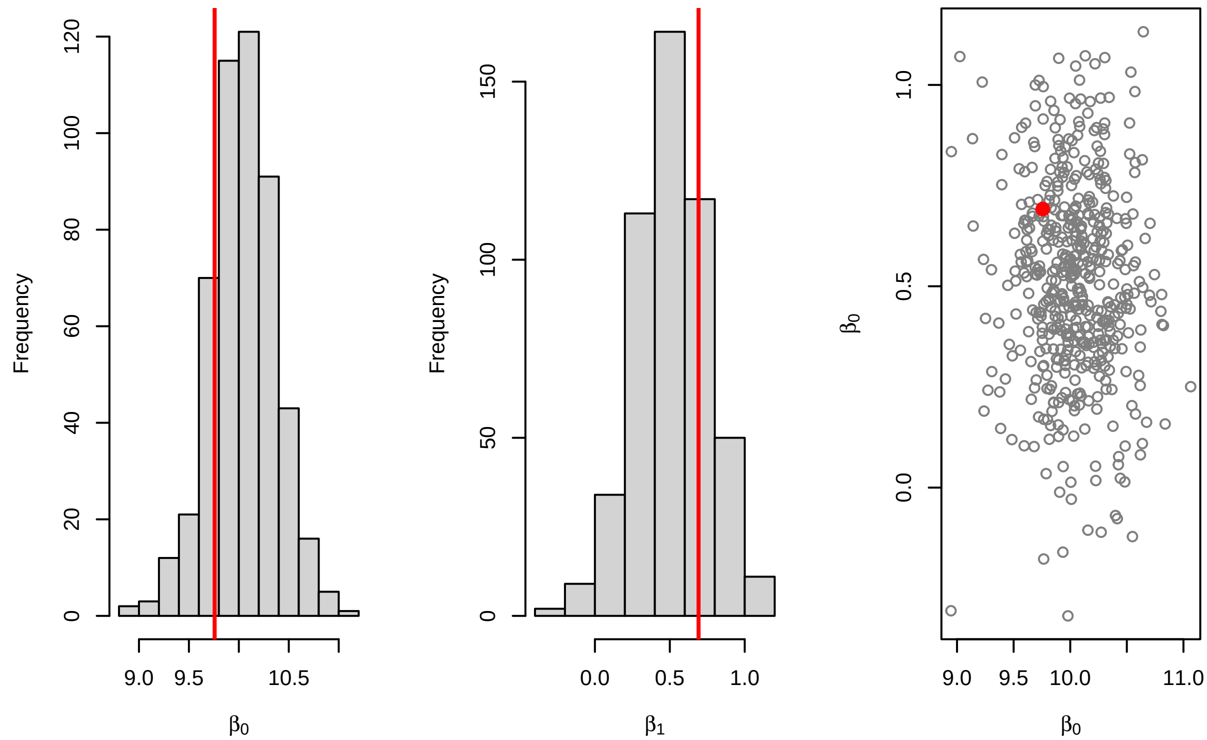
```
mat_estimat <- matrix(0, nrow = 500, ncol = 2)
mat_erros <- matrix(0, nrow = 500, ncol = 2)

for(i in 1:500){
    y <- rnorm(10, mean = 10 + 0.5*x, sd = 1)
    fit_sim <- lm(y ~ x)
    mat_estimat[i,] <- coef(fit_sim)
    mat_erros[i,] <- summary(fit_sim)$coefficients[,'Std. Error']}
head(mat_estimat, 10)
```

```
##            [,1]       [,2]
##  [1,] 10.493144  0.4455431
##  [2,]  9.690980  0.5467901
##  [3,] 10.047540  0.3376021
##  [4,]  9.711062  0.5419746
##  [5,]  9.619189  0.6405793
##  [6,] 10.085609  0.3281623
##  [7,] 10.305686  1.0678750
##  [8,] 10.071511  0.3756977
##  [9,]  9.935251 -0.1602444
## [10,] 10.352874  0.3877504
```

d) Construa histogramas e calcule média e variância das estimativas produzidas para cada parâmetro. Compare os resultados obtidos na simulação aos apresentados no item *b*;

Resposta:

```
par(mfrow = c(1,3))
hist(mat_estimat[,1], xlab = expression(beta[0]), main = "")
abline(v = coef(fit)[1], col = 'red', lwd = 2)
hist(mat_estimat[,2], xlab = expression(beta[1]), main = "")
abline(v = coef(fit)[2], col = 'red', lwd = 2)
plot(mat_estimat[,1], mat_estimat[,2], xlab = expression(beta[0]), ylab = expression(beta[0]), col = 'g
points(x = coef(fit)[1], y = coef(fit)[2], cex = 2, pch = 20, col = 'red')
```



e) Para cada uma das 5000 simulações, obtenha os intervalos de confiança 95% para β_0 e β_1. Qual proporção dos intervalos contêm os valores fixados para os respectivos parâmetros?

Resposta:

```
mat_ICs_beta0 <- matrix(0, nrow = 500, ncol = 2)
mat_ICs_beta1 <- matrix(0, nrow = 500, ncol = 2)

for(i in 1:500){
    mat_ICs_beta0[i,] <- c(mat_estimat[i,1] + qt(0.025, df = 10-2) * mat_erros[i,1],
                           mat_estimat[i,1] + qt(0.975, df = 10-2) * mat_erros[i,1])
    mat_ICs_beta1[i,] <- c(mat_estimat[i,2] + qt(0.025, df = 10-2) * mat_erros[i,2],
                           mat_estimat[i,2] + qt(0.975, df = 10-2) * mat_erros[i,2])
}

mean(mat_ICs_beta0[,1] < 10 & mat_ICs_beta0[,2] > 10)
```

```
## [1] 0.936
```

```
mean(mat_ICs_beta1[,1] < 0.5 & mat_ICs_beta1[,2] > 0.5)
```

```
## [1] 0.936
```

6. Considere o modelo de regressão linear simples sem intercepto:

$$y = \beta x + \epsilon,$$

com as suposições usuais para os erros para o modelo de regressão linear.

a) Mencione uma situação prática em que o modelo de regressão linear passando pela origem possa ser considerado;

b) Determine o estimador de mínimos quadrados de β;

c) Obtenha esperança e variância para o estimador deduzido no item b.

7. Neste exercício consideramos transformações lineares de x e y. Em todos os itens, considere o modelo de regressão linear simples conforme especificado em sala de aula. Sejam β_0, β_1, SQE e r os parâmetros do modelo, a soma de quadrados dos erros e o coeficiente de correlação, respectivamente.

a) Suponha que cada valor de x seja transformado usando $x' = x - 10$ e a regressão linear simples de y em x'. Como ficam β_0', β_1', $\mathrm{SQ}_{\mathrm{Res}}'$ e r'? O que acontece com essas quantidades quando $x' = 10x$? E quando $x' = 10(x-1) = 10x - 10$?

Resolução:

Seja $x' = x - 10$. Então:

$$
\begin{aligned}
\hat{\beta}_1' &= \frac{\sum(x_i' - \bar{x}')(y_i - \bar{y})}{\sum(x_i' - \bar{x}')^2} \\
&= \frac{\sum((x_i - 10) - (\bar{x} - 10))(y_i - \bar{y})}{\sum((x_i - 10) - (\bar{x} - 10))^2} \\
&= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \\
&= \hat{\beta}_1
\end{aligned}
$$

e

$$
\begin{aligned}
\hat{\beta}_0' &= \bar{y} - \hat{\beta}_1'\bar{x}' \\
&= \bar{y} - \hat{\beta}_1(\bar{x} - 10) \\
&= \bar{y} - \hat{\beta}_1\bar{x} - 10\hat{\beta}_1 \\
&= \hat{\beta}_0 - 10 \times \hat{\beta}_1
\end{aligned}
$$

e

$$
\begin{aligned}
\mathrm{SQ}_{\mathrm{Res}}' &= \sum(y_i - (\hat{\beta}_0' + \hat{\beta}_1' x_i'))^2 \\
&= \sum(y_i - (\hat{\beta}_0 - 10\hat{\beta}_1 + \hat{\beta}_1(x_i - 10))^2 \\
&= \sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \\
&= \mathrm{SQ}_{\mathrm{Res}}
\end{aligned}
$$

e

$$\begin{aligned}
r' &= \frac{\sum(x_i' - \bar{x}')(y_i - \bar{y})}{\sqrt{\sum(x_i' - \bar{x}')^2}\sqrt{\sum(y_i - \bar{y})^2}} \\
&= \frac{\sum((x_i - 10) - (\bar{x} - 10))(y_i - \bar{y})}{\sqrt{\sum((x_i - 10) - (\bar{x} - 10))^2}\sqrt{\sum(y_i - \bar{y})^2}} \\
&= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \\
&= r
\end{aligned}$$

Seja $x' = 10x$. Então:

$$\begin{aligned}
\hat{\beta}_1' &= \frac{\sum(x_i' - \bar{x})(y_i - \bar{y})}{\sum(x_i' - \bar{x}')^2} \\
&= \frac{\sum(10x_i - 10\bar{x})(y_i - \bar{y})}{\sum(10x_i - 10\bar{x})^2} \\
&= \frac{10\sum(x_i - \bar{x}')(y_i - \bar{y})}{10^2\sum(x_i - \bar{x})^2} \\
&= \frac{1}{10}\frac{\sum(x_i - \bar{x}')(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \\
&= \frac{1}{10} \times \hat{\beta}_1
\end{aligned}$$

e

$$\begin{aligned}
\hat{\beta}_0' &= \bar{y} - \hat{\beta}_1'\bar{x}' \\
&= \bar{y} - \frac{1}{10}\hat{\beta}_1 10\bar{x} \\
&= \bar{y} - \hat{\beta}_1\bar{x} \\
&= \hat{\beta}_0
\end{aligned}$$

e

$$\begin{aligned}
\mathrm{SQ}_{\mathrm{Res}}' &= \sum(y_i - (\hat{\beta}_0' + \hat{\beta}_1' x_i'))^2 \\
&= \sum(y_i - (\hat{\beta}_0 + \frac{1}{10}\hat{\beta}_1 10x_i)^2 \\
&= \sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \\
&= \mathrm{SQ}_{\mathrm{Res}}
\end{aligned}$$

e

$$
\begin{aligned}
r' &= \frac{\sum(x_i' - \bar{x}')(y_i - \bar{y})}{\sqrt{\sum(x_i' - \bar{x}')^2}\sqrt{\sum(y_i - \bar{y})^2}} \\
&= \frac{10\sum(x_i - \bar{x})(y_i - \bar{y})}{10\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \\
&= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \\
&= r
\end{aligned}
$$

Seja $x' = 10(x - 1)$. Então:

$$
\begin{aligned}
\hat{\beta}_1' &= \frac{\sum(x_i' - \bar{x})(y_i - \bar{y})}{\sum(x_i' - \bar{x}')^2} \\
&= \frac{\sum(10(x_i - 1) - 10(\bar{x} - 1))(y_i - \bar{y})}{\sum(10(x_i - 1) - 10(\bar{x} - 1))^2} \\
&= \frac{10\sum(x_i - 1 - \bar{x} + 1)(y_i - \bar{y})}{10^2\sum(x_i - 1 - \bar{x} + 1)^2} \\
&= \frac{1}{10}\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \\
&= \frac{1}{10} \times \hat{\beta}_1
\end{aligned}
$$

e

$$
\begin{aligned}
\hat{\beta}_0' &= \bar{y} - \hat{\beta}_1'\bar{x}' \\
&= \bar{y} - \frac{1}{10}\hat{\beta}_1 10(\bar{x} - 1) \\
&= \bar{y} - \hat{\beta}_1\bar{x} + \hat{\beta}_1 \\
&= \hat{\beta}_0 + \hat{\beta}_1
\end{aligned}
$$

e

$$
\begin{aligned}
\mathrm{SQ}_{\mathrm{Res}}' &= \sum(y_i - (\hat{\beta}_0' + \hat{\beta}_1' x_i'))^2 \\
&= \sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 + \frac{1}{10}\hat{\beta}_1 10(x_i - 1))^2 \\
&= \sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_1 x_i - \hat{\beta}_1))^2 \\
&= \sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \\
&= \mathrm{SQ}_{\mathrm{Res}}
\end{aligned}
$$

e

$$
\begin{aligned}
r' &= \frac{\sum(x_i' - \bar{x}')(y_i - \bar{y})}{\sqrt{\sum(x_i' - \bar{x}')^2}\sqrt{\sum(y_i - \bar{y})^2}} \\
&= \frac{\sum(10(x_i - 1) - 10(\bar{x} - 1))(y_i - \bar{y})}{\sqrt{\sum(10(x_i - 1) - 10(\bar{x} - 1))^2}\sqrt{\sum(y_i - \bar{y})^2}} \\
&= \frac{10\sum(x_i - \bar{x})(y_i - \bar{y})}{10\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \\
&= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \\
&= r
\end{aligned}
$$

b) Agora, suponha que os valores da variável resposta sejam transformados para $y' = y + 10$ e considere a regressão de y'. em x. Como ficam β_0', β_1', SQ_{Res}' e r'? O que acontece com essas quantidades quando $y' = 5y$? E quando $y' = 5(y + 2) = 5y + 10$?

Seja $y' = y + 10$. Então:

$$
\begin{aligned}
\hat{\beta}_1' &= \frac{\sum(x_i - \bar{x})(y_i' - \bar{y}')}{\sum(x_i - \bar{x})^2} \\
&= \frac{\sum((x_i - \bar{x})((y_i + 10) - (\bar{y} + 10))}{\sum(x_i - \bar{x})^2} \\
&= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \\
&= \hat{\beta}_1
\end{aligned}
$$

e

$$
\begin{aligned}
\hat{\beta}_0' &= \bar{y}' - \hat{\beta}_1'\bar{x} \\
&= \bar{y} + 10 - \hat{\beta}_1\bar{x} \\
&= \bar{y} - \hat{\beta}_1\bar{x} + 10 \\
&= \hat{\beta}_0 + 10
\end{aligned}
$$

e

$$
\begin{aligned}
\text{SQ}_{\text{Res}}' &= \sum(y_i' - (\hat{\beta}_0' + \hat{\beta}_1' x_i))^2 \\
&= \sum(y_i + 10 - (\hat{\beta}_0 + 10 + \hat{\beta}_1 x_i))^2 \\
&= \sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \\
&= \text{SQ}_{\text{Res}}
\end{aligned}
$$

e

$$
\begin{aligned}
r' &= \frac{\sum(x_i-\bar{x})(y'_i-\bar{y}')}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y'_i-\bar{y}')^2}} \\
&= \frac{\sum(x_i-\bar{x})(y_i+10-(\bar{y}+10))}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i+10-(\bar{y}+10))^2}} \\
&= \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i-\bar{y})^2}} \\
&= r
\end{aligned}
$$

Seja $y' = 5y$. Então:

$$
\begin{aligned}
\hat{\beta}'_1 &= \frac{\sum(x_i-\bar{x})(y'_i-\bar{y}')}{\sum(x_i-\bar{x})^2} \\
&= 5\frac{\sum((x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2} \\
&= 5\times\hat{\beta}_1
\end{aligned}
$$

e

$$
\begin{aligned}
\hat{\beta}'_0 &= \bar{y}'-\hat{\beta}'_1\bar{x} \\
&= 5\bar{y}-5\hat{\beta}_1\bar{x} \\
&= 5(\bar{y}-\hat{\beta}_1\bar{x}) \\
&= 5\times\hat{\beta}_0
\end{aligned}
$$

e

$$
\begin{aligned}
\mathrm{SQ}'_{\mathrm{Res}} &= \sum(y'_i-(\hat{\beta}'_0+\hat{\beta}'_1x_i))^2 \\
&= \sum(5y_i-5(\hat{\beta}_0+10+\hat{\beta}_1x_i))^2 \\
&= 5^2\sum(y_i-(\hat{\beta}_0+\hat{\beta}_1x_i))^2 \\
&= 25\times\mathrm{SQ}_{\mathrm{Res}}
\end{aligned}
$$

e

$$
\begin{aligned}
r' &= \frac{\sum(x_i-\bar{x})(y'_i-\bar{y}')}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y'_i-\bar{y}')^2}} \\
&= \frac{5\sum(x_i-\bar{x})(y_i\bar{y})}{\sqrt{25}\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i\bar{y})^2}} \\
&= \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i-\bar{y})^2}} \\
&= r
\end{aligned}
$$

Seja $y' = 5(y+2)$. Então:

$$
\begin{aligned}
\hat{\beta}_1' &= \frac{\sum(x_i - \bar{x})(y_i' - \bar{y}')}{\sum(x_i - \bar{x})^2} \\
&= \frac{\sum((x_i - \bar{x})(5(y_i + 2) - 5(\bar{y} + 2))}{\sum(x_i - \bar{x})^2} \\
&= \frac{5\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \\
&= 5 \times \hat{\beta}_1
\end{aligned}
$$

e

$$
\begin{aligned}
\hat{\beta}_0' &= \bar{y}' - \hat{\beta}_1'\bar{x} \\
&= 5(\bar{y} + 2) - 5\hat{\beta}_1\bar{x} \\
&= 5(\bar{y} - \hat{\beta}_1\bar{x}) + 10 \\
&= 5 \times \hat{\beta}_0 + 10
\end{aligned}
$$

e

$$
\begin{aligned}
\mathrm{SQ}_{\mathrm{Res}}' &= \sum(y_i' - (\hat{\beta}_0' + \hat{\beta}_1' x_i))^2 \\
&= \sum(5(y_i + 2) - (5\hat{\beta}_0 + 10 + 5\hat{\beta}_1 x_i))^2 \\
&= 5^2 \sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2 \\
&= 25 \times \mathrm{SQ}_{\mathrm{Res}}
\end{aligned}
$$

e

$$
\begin{aligned}
r' &= \frac{\sum(x_i - \bar{x})(y_i' - \bar{y}')}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i' - \bar{y}')^2}} \\
&= \frac{\sum(x_i - \bar{x})(5(y_i + 2) - (5(\bar{y} + 2)))}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(5(y_i + 2) - (5(\bar{y} + 2)))^2}} \\
&= \frac{5\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{25}\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \\
&= r
\end{aligned}
$$

c) Em geral, como os resultados da regressão linear simples ficam afetados por transformações lineares em x e em y?

8. Solicitado a especificar o modelo de regressão linear simples, um aluno escreveu o seguinte:

$$\mathrm{E}(y|x) = \beta_0 + \beta_1 x + \epsilon.$$

Você concorda com essa especificação? Justifique.

Resposta:

A especificação está incorreta, uma vez que $\mathrm{E}(\epsilon) = 0$.

9. Qual o impacto da ausência de normalidade dos erros nas propriedades dos estimadores de mínimos quadrados do modelo de regressão linear?

Resposta:

Satisfeitas as demais suposições, os estimadores ainda são não viciados e eficientes na classe de estimadores lineares. No entanto, não apresentam distribuição normal.

10. Mostre que:

a) $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i$;

$$\begin{aligned}
\sum \hat{y}_i &= \sum (\hat{\beta}_0 + \hat{\beta}_1 x_i) \\
&= \sum (\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 x_i) \\
&= n\bar{y} - n\hat{\beta}_1 \bar{x} + n\hat{\beta}_1 \bar{x} \\
&= n\bar{y} = \sum y_i
\end{aligned}$$

b) $\sum_{i=1}^{n} r_i = 0$;

$$\begin{aligned}
\sum r_i &= \sum (y_i - \hat{y}_i) \\
&= \sum y_i - \sum \hat{y}_i \\
&= \sum y_i - \sum y_i = 0
\end{aligned}$$

c) Para $x = \bar{x}$ tem-se $\hat{y} = \bar{y}$;

$$\begin{aligned}
\hat{y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \\
&= \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} \\
&= \bar{y}
\end{aligned}$$

d) $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$;

$$\sum (x_i - \bar{x}) = \sum x_i - n\bar{x} = n\bar{x} - n\bar{x} = 0$$

e) $\sum_{i=1}^{n} (x_i - \bar{x}) y_i = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$;

$$\begin{aligned}
\sum (x_i - \bar{x})(y_i - \bar{y}) &= \sum x_i y_i - \bar{x} \sum y_i - \bar{y} \sum x_i + n\bar{x}\bar{y} \\
&= \sum x_i y_i - \bar{x} \sum y_i - \bar{x} \sum y_i + \bar{x} \sum y_i \\
&= \sum x_i y_i - \bar{x} \sum y_i = \sum (x_i - \bar{x}) y_i
\end{aligned}$$

f) $\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}x_i^2-n\bar{x}^2$;

$$
\begin{aligned}
\sum(x_i-\bar{x})^2 &= \sum x_i^2-2\bar{x}\sum x_i+n\bar{x}^2\\
&= \sum x_i^2-2n\bar{x}^2+n\bar{x}^2\\
&= \sum x_i^2-n\bar{x}^2
\end{aligned}
$$

g) $\sum_{i=1}^{n}x_ir_i=0$;

$$
\begin{aligned}
\sum x_ir_i &= \sum(y_i-\hat{y}_i)x_i\\
&= \sum x_iy_i-\sum x_i\hat{y}_i\\
&= \sum x_iy_i-\sum x_i(\hat{\beta}_0+\hat{\beta}_1x_i)\\
&= \sum x_iy_i-\sum x_i(\bar{y}-\hat{\beta}_1\bar{x})-\hat{\beta}_1\sum x_i^2\\
&= \sum x_iy_i-n\bar{x}\bar{y}+n\bar{\beta}_1\bar{x}^2-\hat{\beta}_1\sum x_i^2\\
&= \sum x_iy_i-n\bar{x}\bar{y}-\bar{\beta}_1(\sum x_i^2-n\bar{x}^2)\\
&= \sum x_iy_i-n\bar{x}\bar{y}-\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}\sum(x_i-\bar{x})^2\\
&= \sum(x_i-\bar{x})(y_i-\bar{y})-\sum(x_i-\bar{x})(y_i-\bar{y})=0
\end{aligned}
$$

h) $\sum_{i=1}^{n}\hat{y}_ir_i=0$.

$$
\begin{aligned}
\sum r_i\hat{y}_i &= \sum(y_i-\hat{y}_i)\hat{y}_i\\
&= \sum y_i\hat{y}_i-\sum\hat{y}_i^2\\
&= \sum y_i(\hat{\beta}_0+\hat{\beta}_1x_i)-\sum(\hat{\beta}_0+\hat{\beta}_1x_i)^2\\
&= \hat{\beta}_0\sum y_i+\hat{\beta}_1\sum x_iy_i-n\hat{\beta}_0^2-2\hat{\beta}_0\hat{\beta}_1\sum x_i-\hat{\beta}_1^2\sum x_i^2\\
&= (\bar{y}-\hat{\beta}_1\bar{x})\sum y_i+\hat{\beta}_1\sum x_iy_i-n(\bar{y}-\hat{\beta}_1\bar{x})^2-2(\bar{y}-\hat{\beta}_1\bar{x})\hat{\beta}_1n\bar{x}-\hat{\beta}_1^2\sum x_i^2\\
&= n\bar{y}^2-n\hat{\beta}_1\bar{x}\bar{y}+\hat{\beta}_1\sum x_iy_i-n\bar{y}^2+2n\hat{\beta}_1\bar{x}\bar{y}-n\hat{\beta}_1^2\bar{x}^2-2n\hat{\beta}_1\bar{x}\bar{y}+2n\hat{\beta}_1^2\bar{x}^2-\hat{\beta}_1^2\sum x_i^2\\
&= \hat{\beta}_1(\sum x_iy_i-n\bar{x}\bar{y})-\hat{\beta}_1^2\sum(x_i-\bar{x})^2\\
&= \frac{\left[\sum(y_i-\bar{y})(x_i-\bar{x})\right]^2}{\sum(x_i-\bar{x})^2}-\frac{\left[\sum(y_i-\bar{y})(x_i-\bar{x})\right]^2}{\left[\sum(x_i-\bar{x})^2\right]^2}\sum(x_i-\bar{x})^2=0
\end{aligned}
$$

11. Um estudo foi conduzido para avaliar o efeito da temperatura na produção química de um processo. Os seguintes dados foram coletados:

Temp.	-5	-4	-3	-2	-1	0	1	2	3	4	5
Prod.	1	5	4	7	10	8	9	13	14	13	18

Vamos proceder a análise dos dados usando o modelo de regressão linear simples.

a) Determine as estimativas de mínimos quadrados de β_0 e β_1 e apresente a equação do modelo ajustado.

b) Apresente um intervalo de confiança (95%) para β_1;

c) Teste a hipótese $H_0 : \beta_1 = 0$ ao nível de significância de 5%;

d) Apresente os limites de confiança (95%) para a resposta média quando a temperatura é igual a 3;

e) Apresente limites de confiança (95%) para a diferença nas resposta média quando a temperatura é igual a 3 em relação à resposta média sob temperatura -2;

f) Sob qual temperatura se estima produção igual a 12?

12. A base de dados `Prestige` do pacote `car` apresenta dados referentes à percepção da população canadense quanto a 102 diferentes profissões. Vamos considerar, para ajuste de um modelo de regressão linear simples, as seguintes variáveis:

- `education`: Educação média dos profissionais (em anos de estudo);
- `prestige`: Escore de prestígio da profissão segundo a resposta dos entrevistados.

Considere o prestígio da profissão como a resposta e a escolaridade média como a variável explicativa.

a) Ajuste o modelo de regressão linear simples aos dados apresentados e apresente a equação do modelo ajustado;

b) Construa o diagrama de dispersão e adicione a reta de regressão ajustada. A reta obtida parece se ajustar bem aos dados?

c) Qual a predição para o escore de prestígio para uma profissão com escolaridade média de 12.5 anos?

d) Qual o valor ajustado pelo modelo para o prestígio dos administradores públicos (primeira linha da base)? Qual o correspondente resíduo?

e) Em quanto se estima a variação esperada no escore de prestígio para um ano a mais de escolaridade média entre os profissionais? E para três anos a mais?

f) O intercepto tem alguma interpretação prática nesta análise?

g) Apresente uma estimativa para σ^2;

h) Teste a hipótese $H_0 : \beta_1 = 0$ *vs* $H_1 : \beta_1 \neq 0$ ao nível de significância de 5% e apresente suas conclusões;

i) Teste a hipótese $H_0 : \beta_1 = 6$ *vs* $H_1 : \beta_1 \neq 6$ ao nível de significância de 5% e apresente suas conclusões;

j) Apresente intervalos de confiança (95%) para os parâmetros do modelo;

k) Apresente intervalos de confiança para a média e de predição considerando (i) $x = 9$; (ii) $x = 15$; (iii) $x = \bar{x}$;

l) Adicione ao diagrama de dispersão as bandas de confiança e de predição (95%);

m) Apresente o quadro de análise de variância e o teste F. Compare o p-valor desse teste ao teste da hipótese $H_0 : \beta_1 = 0$;

n) Calcule o valor de R^2 e interprete-o.

13. Neste exercício vamos analisar os dados de velocidade (x, em milhas por horas) e consumo de combustível (y, em milhas por galão) para $n = 28$ automóveis de certa marca. Os dados estão disponíveis na página

da disciplina no Moodle. O objetivo é ajustar um modelo de regressão linear simples para explicar o consumo de combustível em função da velocidade sem usar a função `lm` e suas dependências.

Dados:

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = -1184.39; \quad \sum_{i=1}^{n}(x_i - \bar{x})^2 = 7316.96; \quad \sum_{i=1}^{n}(y_i - \hat{(y)}_i)^2 = 11.20 \quad \bar{x} = 75.54; \quad \bar{y} = 12.89$$

a) Usando o R, faça o gráfico de dispersão;

b) Calcule as estimativas de mínimos quadrados para β_0 e β_1. Interprete-as.

c) Qual a variação estimada no consumo de combustível para $15mph$ a mais de velocidade?

d) Escreva a expressão do modelo ajustado. Calcule o consumo de combustível estimado sob velocidade $x = 75mph$;

e) Calcule os erros padrões de $\hat{\beta}_0$ e $\hat{\beta}_1$;

f) Estime as variâncias para (i) o consumo médio sob velocidade $x = 75mph$; (ii) o consumo predito para um particular automóvel sob velocidade $x = 75mph$. Compare os resultados;

g) Idêntico ao item anterior, mas para velocidade $x = 50mph$. Compare com os resultados do item anterior;

h) Apresente intervalos de confiança 95% para β_0 e para β_1;

i) Apresente intervalos de confiança 95% para o consumo médio e o consumo predito para um novo automóvel sob velocidades: (i) $x = 75mph$; (ii) $x = 50mph$;

j) Teste a significância do modelo de regressão, ou seja, teste a hipótese $H_0 : \beta_1 = 0$ *vs* $H_1 : \beta_1 \neq 0$;

k) Um especialista afirma que o consumo de combustível altera, em média, em $-0.18mpg$ para cada unidade a mais de velocidade (em mph). Teste essa hipótese.

Nota: Para as questões envolvendo testes de hipóteses, o seguinte procedimento deve ser aplicado:

(I) Formulação das hipóteses nula e alternativa;
(II) Apresentação e cálculo da estatística teste;
(III) Definição da regra de decisão para os níveis de significância de 5% e 1%;
(IV) Conclusão do problema baseada nas regras de decisão descritas no passo anterior;
(V) Cálculo do nível descritivo (p-valor) do teste.

14. Sejam y_1 e y_2 variáveis aleatórias com distribuição conjunta normal bivariada, conforme definido na sequência:

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim \text{Normal}\left(\boldsymbol{\mu} = \begin{bmatrix} 6 \\ 2 \end{bmatrix},\ \Sigma = \begin{bmatrix} 1 & 0.4 \\ 0.4 & 0.5 \end{bmatrix}\right) \tag{1}$$

a) Quais as distribuições marginais de y_1 e y_2;

$y_1 \sim \text{N}(6, 1)$ e $y_2 \sim \text{N}(2, 0.5)$

b) Usando o R, faça gráficos das distribuições (funções densidade de probabilidade) de y_1, y_2 e da conjunta de y_1 e y_2;

c) Qual a distribuição de probabilidades de:

i) $z_1 = y_1 + y_2$
ii) $z_2 = \frac{y_1+y_2}{2}$

iii) $z_3 = y_1 - y_2$

iv) $z_4 = 0.875y_1 - 2.278y_2$

Respostas

i) $z_1 \sim \mathrm{N}(8, 2.3)$;

ii) $z_2 \sim \mathrm{N}(4, 0.575)$;

iii) $z_3 \sim \mathrm{N}(4, 0.7)$;

iv) $z_4 \sim \mathrm{N}(0.694, 1.765)$.

Nota: Se desejado você pode verificar esses resultados por simulação.

15. Sejam $y_1, y_2, ..., y_{30}$ variáveis aleatórias independentes com distribuição $y_i \sim \mathrm{N}(i, i^2)$, $i = 1, 2, ..., 30$. Qual a distribuição de $z = y_1 + y_2 + ... + y_{30}$?

Respostas

$z \sim \mathrm{N}(\sum_{i=1}^{30} i, \sum_{i=1}^{30} i^2)$

Nota: $\sum_{i=1}^{30} i = \frac{30(30+1)}{2} = 465$ e $\sum_{i=1}^{30} i^2 = \frac{30(30+1)(2\times30+1)}{6} = 9455$