Lista 2, Exercício 8

Observaram-se os gastos mensais com alimentação (Y, em centenas de reais), a renda mensal (X_1, em centenas de reais) e a distância da residência ao supermercado mais próximo (X_2, em km) de quatro domicílios:

Y	X_1	X_2
4	10	2
2	20	3
3	20	3
6	30	2

Admite-se que a relação existente entre o gasto (Y), a renda familiar (X_1) e a distância ao supermercado (X_2) seja dada pelo modelo: $Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \mu_i$, em que os μ_i são erros independentes com distribuição normal de média zero e variância σ^2. À luz do modelo proposto, pede-se:

a) Estime e interprete os coeficientes de regressão do modelo proposto.
b) Construa a tabela ANOVA e conduza o teste F, interpretando o resultado obtido.
c) Calcule e interprete o coeficiente de determinação. Calcule também o coeficiente de determinação ajustado.

Teste de hipóteses envolvendo combinação linear dos parâmetros

d) Há evidência significativa, ao nível de 5%, para afirmar que o efeito <u>conjunto</u> de um aumento de uma centena de reais na renda familiar e de 1 km adicional de distância em relação ao supermercado mais próximo sobre os gastos mensais com alimentação seja negativo?

depois: $gasto_i = \alpha + \beta_1 (renda_i + 1) + \beta_2 (dist_i + 1) + \mu_i$

antes: $gasto_i = \alpha + \beta_1 renda_i + \beta_2 dist_i + \mu_i$

---

efeito: $\Delta gasto_i = \beta_1 + \beta_2$

Hipótese nula: $\beta_1 + \beta_2 = 0$

Hipótese alternativa: $\beta_1 + \beta_2 < 0$

Note que não é mesmo que testar H_0: $\beta_1 = \beta_2 = 0$ Não é um teste F!!!

O teste deve ser reescrito como

$$\begin{cases} H_0 : 0\cdot\alpha + 1\cdot\beta_1 + 1\cdot\beta_2 = 0 \\ H_A : 0\cdot\alpha + 1\cdot\beta_1 + 1\cdot\beta_2 < 0 \end{cases} \qquad \begin{cases} H_0 : \mathbf{c}'\boldsymbol{\beta} = \boldsymbol{\theta} \\ H_A : \mathbf{c}'\boldsymbol{\beta} < \boldsymbol{\theta} \end{cases}$$

A estatística de teste a ser calculada é $t = \dfrac{\mathbf{c}'\mathbf{b} - \boldsymbol{\theta}}{\sqrt{\hat{\mathrm{V}}(\mathbf{c}'\mathbf{b})}}$.

A matriz $\mathbf{c}'$ é formada pelas constantes que multiplicam α, β_1 e β_2 na hipótese de teste. Assim, $\mathbf{c}' = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}$.

Matriz $\mathbf{b} = \begin{pmatrix} 8 \\ 0{,}1 \\ -2{,}5 \end{pmatrix}$

Logo, $\mathbf{c}'\mathbf{b} = \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 8 \\ 0{,}1 \\ -2{,}5 \end{pmatrix} = -2{,}4$.

O valor de $\boldsymbol{\theta}$ corresponde à constante que aparece no membro direito da hipótese nula (ou da hip. alternativa, já que se trata do mesmo valor) depois que ela foi reorganizada. Como neste exercício $H_0:\ 0 \cdot \alpha + 1 \cdot \beta_1 + 1 \cdot \beta_2 = 0$, $\boldsymbol{\theta} = 0$.

Assim, temos $t = \dfrac{\mathbf{c}'\mathbf{b} - \boldsymbol{\theta}}{\sqrt{\hat{\mathrm{V}}(\mathbf{c}'\mathbf{b})}} = \dfrac{-2{,}4 - 0}{\sqrt{\hat{\mathrm{V}}(\mathbf{c}'\mathbf{b})}}$.

Para calcular $\hat{\mathrm{V}}(\mathbf{c}'\mathbf{b})$ devemos utilizar a fórmula $\hat{\mathrm{V}}(\mathbf{c}'\mathbf{b}) = \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} \cdot s^2$.

No item (a) do exercício obtivemos $(\mathbf{X}'\mathbf{X})^{-1} = \begin{pmatrix} 8,5 & -0,1 & -2,5 \\ -0,1 & 0,005 & 0 \\ -2,5 & 0 & 1 \end{pmatrix}$ e, no item (b), $\text{QMResíduos} = s^2 = 0,5$.

Logo, $\hat{\mathrm{V}}(\mathbf{c}'\mathbf{b}) = \mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c} \cdot s^2 = \underbrace{\begin{pmatrix} 0 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 8,5 & -0,1 & -2,5 \\ -0,1 & 0,005 & 0 \\ -2,5 & 0 & 1 \end{pmatrix}}_{\begin{pmatrix} -2,6 & 0,005 & 1 \end{pmatrix}} \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \cdot 0,5 = 0,5025$.

Assim, $t = \dfrac{\mathbf{c'b} - \boldsymbol{\theta}}{\sqrt{\hat{\mathrm{V}}\left(\mathbf{c'b}\right)}} = \dfrac{-2,4 - 0}{\sqrt{0,5025}} = -3,386$.

O valor crítico de t a 1 gl e $\alpha = 5\%$ é 6,314. É necessário fazer a consulta da tabela considerando 2α, pois o teste é unilateral.

Logo, $RC =]-\infty, -6,314]$.

Como a estatística de teste calculada não pertence à RC, o teste não é significativo e não se rejeita H_0.

Intervalo de confiança para um $E(Y_h)$

e) Estabeleça uma estimativa por intervalo com 90% de confiança para os gastos esperados de uma família com renda mensal de 50 (em centenas de reais) e que reside a dois quilômetros do supermercado mais próximo.

O modelo estimado no item (a) é $\widehat{gasto} = 8 + 0{,}1 \cdot renda - 2{,}5 dist$.

Então, para $renda_h = 50$ e $dist_h = 2$, temos $\widehat{gasto}_h = 8 + 0{,}1 \cdot 50 - 2{,}5 \cdot 2 = 8$.

O IC para essa estimativa é dado por $\hat{Y}_h \pm \underbrace{t_0 \cdot \sqrt{\hat{V}\left(\hat{Y}_h\right)}}_{\text{Erro padrão de } \hat{Y}_h}$.

Primeiramente calculemos $\hat{V}\left(\hat{Y}_h\right) = \mathbf{x}'_\mathbf{h}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_\mathbf{h} s^2$.

$$\mathbf{x}_\mathbf{h} = \begin{pmatrix} 1 \\ X_{1h} \\ X_{2h} \end{pmatrix} = \begin{pmatrix} 1 \\ 50 \\ 2 \end{pmatrix}$$

Assim, $$\hat{V}\left(\hat{Y}_h\right) = \mathbf{x}'_\mathbf{h}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{x}_\mathbf{h} s^2 = \underbrace{\begin{pmatrix} 1 & 50 & 2 \end{pmatrix} \cdot \begin{pmatrix} 8,5 & -0,1 & -2,5 \\ -0,1 & 0,005 & 0 \\ -2,5 & 0 & 1 \end{pmatrix}}_{\begin{pmatrix} -1,5 & 0,15 & -0,5 \end{pmatrix}} \cdot \begin{pmatrix} 1 \\ 50 \\ 2 \end{pmatrix} \cdot 0,5$$

$$\hat{\mathrm{V}}\left(\hat{Y}_h\right) = \begin{pmatrix} -1,5 & 0,15 & -0,5 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 50 \\ 2 \end{pmatrix} \cdot 0,5$$

$$\hat{\mathrm{V}}\left(\hat{Y}_h\right) = 5 \cdot 0,5 \qquad \therefore \hat{\mathrm{V}}\left(\hat{Y}_h\right) = 2,5.$$

Então, o IC será $8 \pm t_0 \cdot \sqrt{2,5}$.

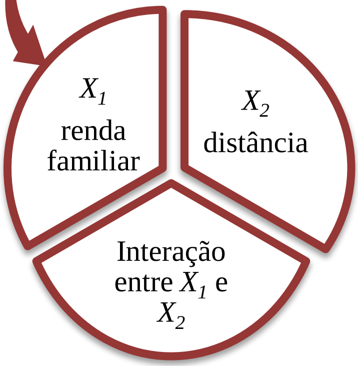
Considerando 1 gl (referente aos resíduos) e C=90% (o que implica que $\alpha = 10\%$), obtemos t_0=6,314.

Assim, o IC é $8 \pm 9,98$

Contribuição de um regressor para a SQRegressão

f) Determine o valor da contribuição da distância com relação ao supermercado sobre os gastos com alimentação.

Causas de Variação	Graus de liberdade (gl)	Soma de Quadrados (SQ)	Quadrados Médios (QM)
Regressão	2	8,25	4,125
Resíduo	1	0,50	0,50
Total	3	8,75	



Gasto com alimentação
Renda familiar
Distância

Como calcular o valor da contribuição da distância com relação ao supermercado sobre os gastos com alimentação?

$$\text{Contribuição da distância } \left(X_2\right) = \text{SQRegressão } Y \middle| X_1, X_2 \; - \text{SQRegressão } Y \middle| X_1$$

Regressão $Y|X_1$:

Y	X_1	y	x_1	$x_1 y$	x_1^2
4	10	+0,25	−10	−2,5	100
2	20	−1,75	0	0	0
3	20	−0,75	0	0	0
6	30	+2,25	10	22,5	100
$\bar{Y} = 3{,}75$	$\bar{X}_1 = 20$			$\sum = 20$	$\sum = 200$

$$b = \frac{\sum x_1 y}{\sum x_1^2} = \frac{20}{200} = 0{,}1$$

$$\therefore \text{ SQRegressão } Y|X_1 = b \cdot \sum x_1 y = 0{,}1 \cdot 20 = 2$$

Então,

$$\boxed{\text{Contribuição da distância}\left(X_2\right)=\underbrace{\text{SQRegressão } Y\middle|X_1,X_2}_{8{,}25}-\underbrace{\text{SQRegressão } Y\middle|X_1}_{2}=6{,}25}$$

A contribuição da renda familiar sobre os gastos com alimentação é dada por:

$$\text{Contribuição da renda familiar } (X_1) = \text{SQRegressão } Y \mid X_1, X_2 \; - \text{SQRegressão } Y \mid X_2$$

Regressão $Y \mid X_2$:

Y	X_2	y	x_2	$x_2 y$	x_2^2
4	2	+0,25	−0,50	−0,125	0,25
2	3	−1,75	+0,50	−0,875	0,25
3	3	−0,75	+0,50	−0,375	0,25
6	2	+2,25	−0,50	−1,125	0,25
$\bar{Y} = 3{,}75$	$\bar{X}_2 = 2{,}5$			$\sum = -2{,}5$	$\sum = 1$

$$b = \frac{\sum x_2 y}{\sum x_2^2} = \frac{-2{,}5}{1} = -2{,}5$$

$$\therefore \text{ SQRegressão } Y \mid X_2 = b \cdot \sum x_2 y = -2{,}5 \cdot (-2{,}5) = +6{,}25$$

Então,

$$\text{Contribuição da renda } (X_1) = \underbrace{\text{SQRegressão } Y|X_1, X_2}_{8,25} - \underbrace{\text{SQRegressão } Y|X_2}_{6,25} = 2$$

Para obter a contribuição que resulta da interação entre X_1 e X_2, descontamos da $\text{SQRegressão } Y|X_1, X_2$ a contribuição direta de X_1 e de X_2:

$$\text{Contribuição resultante da interação entre } X_1 \text{ e } X_2 =$$
$$\underbrace{\text{SQRegressão } Y|X_1, X_2}_{8,25} - \underbrace{\text{Contribuição de } X_1}_{2} - \underbrace{\text{Contribuição de } X_2}_{6,25} = 0$$

De fato, calculando o coeficiente de correlação de linear entre X_1 e X_2 neste exercício, verificamos que este é igual a zero. Assim, se X_1 e X_2 não guardam associação linear entre si, não há como X_1 afetar linearmente Y por meio de X_2, e nem de X_2 afetar linearmente Y por meio de X_1.

Como proceder caso o modelo tenha mais regressores?

(Lista 2 – Exercício 7) Admite-se que as variáveis estão relacionadas de acordo com o modelo $Y_j = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \beta_3 X_{3j} + \mu_j$, em que os μ_j são variáveis aleatórias independentes, homocedásticas, com média zero e distribuição normal.

g) Determine a contribuição de cada variável para a soma de quadrados de regressão.

$$\text{Contribuição de } X_1 = \text{SQRegressão } Y|X_1, X_2, X_3 - \text{SQRegressão } Y|X_2, X_3$$

$$\text{Contribuição de } X_2 = \text{SQRegressão } Y|X_1, X_2, X_3 - \text{SQRegressão } Y|X_1, X_3$$

$$\text{Contribuição de } X_3 = \text{SQRegressão } Y|X_1, X_2, X_3 - \text{SQRegressão } Y|X_1, X_2$$

Contribuição resultante da interação entre X_1, X_2 e $X_3 =$

$$\text{SQRegressão } Y|X_1, X_2, X_3 - \text{Contrib. de } X_1 - \text{Contrib. de } X_2 - \text{Contrib. de } X_3$$