Modelagem e Inferência Estatística -PES-310 — Semana: 2

Complemento dos exercícios de regressão linear do Prof. Fernando Berssaneti

Nota: Você deve ter percebido mas é importante reforçar que a reta usada pelo professor é $y = ax + b$ e devido à bibliografia base a reta adotada foi $y = \beta_0 + \beta_1 x$, ou seja $a = \beta_1$ e $b = \beta_0$.

Complemento Exercício 1
No exercício 1 foi obtida uma reta de regrssão linear $Y = 322,3149 + 9,80983x$ esta reta foi obtida com base nos seguintes valores $S_{xx} = 2247,5$, $S_{xy} = 22047,5$, $S_{yy} = 225363,46$ e coeficiente $r^2 = 0,9597$
Com esses dados calcule, a Soma de Quadrados dos Erros, a Soma de Quadrados Total, a Soma de Quadrados da Regressão e verifique se o valor do coeficiente de determinação fornecido confere. Finalmente, na tabela apresentam-se os pares de dados (7,389) e (8,381) para a sétima e oitava amostra respectivamente. Calcule qual o valor esperado de y com a curva de regressão obtida para essa amostra.
Solução Os dados importantes para realizar estes cálculos são: $S_{yy} = 225363,46$, $S_{xy} = 22047,5$ e β_1. Sendo que a equação da reta é $Y = 322,3149 + 9,80983x$ o valor de $\beta_1 = 9,80983$
Primero calcular a Soma de Quadrados dos Erros
$SQE = S_{yy} - \beta_1 S_{xy}$
$SQE = 225363,46 - (9,80983)(-341,9592)$
$SQE = 9081,23$
Agora a Soma de Quadrados Total
$SQT = S_{yy} = 225363,46$
$SQR = SQT - SQE = 225363,46 - 9081,23$
Com esses dados pode-se calcular a Soma de Quadrados da Regressão
$SQR = 216282.23$
O coeficiente de determinação: $r^2 = 1 - \frac{SQE}{SQT}$
$r^2 = 1 - \frac{9081,23}{225363,46}$
$r^2 = 0,9597$ valor que confere com o do enunciado.
Observe que existem dois coeficientes, se atente a não confundir eles. O coeficiente de correlação r indica quanto correlaciondos estão os dados x e y. O coeficiente de determinação r^2 indica se a reta de regressão obtida ajusta bem os dados da amostra.
Finalmente para o par de dados (7,389) observe que o valor 389 representa o valor observado de y (isto é y_8) e precisa calcular o valor esperado (isto é $\hat{y}_7$ ou $u_{y:7}$)
$u_{y:7} = \hat{y}_7 = 322,3149 + 9,80983x_7$

$u_{y:7} = 322,3149 + 9,80983(7)$
$u_{y:7} = 390,98$ (lembrando que o dado observado foi 389)
O mesmo procedimento se realiza para o par de dados (8,381), isto é, calcular $\hat{y}_8$ ou $u_{y:8}$)
$u_{y:8} = \hat{y}_8 = 322,3149 + 9,80983x_8$
$u_{y:8} = 322,3149 + 9,80983(8)$
$u_{y:8} = 400,79$ (lembrando que o dado observado foi 381)
Note que os resíduos $(y_i - \hat{y}_i)$ podem variar de uma amostra para outra.

Complemento Exercício 2

No exercício 2 foram analizados um conjunto de dados que relacionam o aumento de pressão sanguínea com o nível de pressão sonora, onde a variável preditora (x) é a pressão sonora e a variável resposta (y) é a pressão sanguínea. Com esse conjunto de 20 dados obteve-se uma reta de regresão $Y = -9,8131 + 0,1715x$ resultado de $S_{xx} = 3010,2$, $S_{xy} = 516,2$, $S_{yy} = 124,2$ e $r = 0,8442$ Com esses dados calcule o erro padrão estimado e o intervalo de confiança para un nível de 95%.

Solução Para calcular o erro padrão de β_1 usar a seguinte equação
$s_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{xx}}}$
S_{xx} é dado mas deve ser calculado o desvio padrão estimado s, sendo necessário calcular a variância s^2
$s^2 = \frac{SQE}{n-2}$
$SQE = S_{yy} - \beta_1 S_{xy}$, da reta $Y = -9,8131 + 0,1715x$, $\beta_1 = 0,1715$
$SQE = 124,2 - (0,1715)(516,2)$ Note que se β_1 fosse negativo deve manter o signo negativo, não esqueça.
$SQE = 35,6717$
Agora sim pode calcular a variância estimada
$s^2 = \frac{35,6717}{20-2} = 1,9818$
Portanto o desvio padrão esperado $s = \sqrt{s^2}$
$s = \sqrt{s^2} = \sqrt{1,9818} = 1,4078$
Com estes dados pode ser calculado o erro padrão de β_1
$s_{\hat{\beta}_1} = \frac{1,4078}{\sqrt{3010,2}} = 0,02566$
Para calcular o intervalo de confiânça (IC) deve ser usada a tabela t-student https://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf
$t_(\frac{\alpha}{2}, n-2)$ isto é $t_(0,025, 18)$ já que n=20.
Da tabela $t_(0,025, 18) = 2,101$
Finalmente o IC $\hat{\beta}_1 = \beta_1 \pm t_(\frac{\alpha}{2}, n-2)s_{\hat{\beta}_1}$
$\hat{\beta}_1 = 0,1715 \pm (2,101)(0,02566)$
$0,1176 < \hat{\beta}_1 < 0,2254$ assim o intervalo é $(0,1176, 0,2254)$