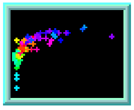
Regressão



Prof. Lorí Viali, Dr.
viali@pucrs.br
http://www.pucrs.br/~viali/

2

Exercício

O % de impurezas no gás oxigênio produzido por um processo de destilação supõem-se que esteja relacionado com o % de hidrocarbono no condensador principal do processador. Os dados de um mês de operação produziram a seguinte tabela

X	Y	X	Y
1,02	86,91	1,46	96,73
1,11	89,85	1,55	99,42
1,43	90,28	1,55	98,66
1,11	86,34	1,55	96,07
1,01	92,58	1,40	93,65
0,95	87,33	1,15	87,31
1,11	86,29	1,01	95,00
0,87	91,86	0,99	96,85
1,43	95,61	0,95	85,20
1,02	89,86	0,98	90,56

(a) Ajuste um modelo linear aos dados;

(b) Teste a existência da regressão, a 1% de significância;

(c) Determine o valor de R^2 para este modelo;

(d) Determine um IC, de 95%, para o valor da pureza, na hipótese do % de hidrocarbono ser 1,20% .

Solução

Dados

$n = 20$

$\sum X = 23{,}65$

$\sum X^2 = 29{,}0311$

$\sum Y = 1836{,}36$

$\sum Y^2 = 168992{,}0498$

$\sum XY = 2184{,}0635$

$\bar{X} = 1{,}1825$

$\bar{Y} = 91{,}8180$

$S_{XX} = 1{,}064975$

$S_{XY} = 12{,}5678$

$S_{YY} = 381{,}147320$

(a) A equação de regressão, será, então:

$$b = \frac{S_{XY}}{S_{XX}} = \frac{12{,}5678}{1{,}064975} = 11{,}8010 \cong 11{,}80$$

$$a = \bar{Y} - b\bar{X} = 91{,}8180 - 11{,}8010 . 1{,}1825 =$$

$$= 77{,}8633 \cong 77{,}86$$

$$\hat{Y} = 77{,}86 + 11{,}80\,x$$

Prof. Lorí Viali, Dr. – PUCRS – FAMAT: Departamento de Estatística

(b) Teste para a existência da regressão

Hipóteses:

H_0: $\beta = 0$

H_1: $\beta > 0$

Dados:

$n = 10$

$b = 11{,}80$

$\alpha = 1\%$

$$s = \sqrt{\frac{S_{YY} - b\,S_{XY}}{n-2}} = 3{,}5966$$

Prof. Lorí Viali, Dr. – PUCRS – FAMAT: Departamento de Estatística

$$t_{18} = \frac{11{,}8010 - 0}{\dfrac{3{,}5966}{\sqrt{1{,}064975}}} = 3{,}386$$

O valor crítico t_c é tal que: $P(T > t_c) = 1\%$ Então $t_c = 2{,}552$. Assim $RC = [2{,}552; \infty)$

Como $t = 3{,}386 > 2{,}552 = t_c$, Rejeito H_0, isto é , a 1% de significância pode-se afirmar que existe regressão.

Prof. Lorí Viali, Dr. – PUCRS – FAMAT: Departamento de Estatística

(c) O coeficiente de determinação

$$R^2 = \frac{S^2_{XY}}{S_{XX}\,S_{YY}} =$$

$$= \frac{12{,}5678^2}{1{,}0650.3811473} = 38{,}91\%$$

Ou seja, 38,91% das variações em Y são explicadas pelas variações em x

Prof. Lorí Viali, Dr. – PUCRS – FAMAT: Departamento de Estatística

(d) Intervalo para a previsão

O valor previsto é dado por:

$$\hat{Y}_{1,20\%} =$$

$$= 77{,}8633 + 11{,}8010 . 1{,}20 =$$

$$= 92{,}02\%$$

Prof. Lorí Viali, Dr. – PUCRS – FAMAT: Departamento de Estatística

(d) O intervalo é calculado por:

$$\hat{Y} \pm t_{n-2} S \sqrt{\frac{1}{n} + \frac{(X - \bar{X})^2}{S_{XX}}}$$

$$92{,}0245 \pm 2{,}101.3{,}5966 \sqrt{\frac{1}{20} + \frac{(1{,}20 - 1{,}1825)^2}{1{,}064975}}$$

$$92{,}0245 \pm 1{,}6945$$

[90,33%; 93,72%]