2012 Regression Summer Paper

Q2. Calculate Coefficients

$n = 5$
$\Sigma x_i = 2150$
$\Sigma x_i^2 = 931100$
$\Sigma y_i = 1430$
$\bar{x} = 430$
$\bar{y} = 286$
$\Sigma x_i y_i = 618500$
$\Sigma y_i^2 = 411900$

$b_0 = \bar{y} - b_1 \bar{x}$

$$b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}$$

$$= \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}$$

$$b_1 = \frac{618500 - \frac{2150(1430)}{5}}{931100 - \frac{(2150)^2}{5}} \qquad \frac{3600}{6600} = \frac{6}{11} = 0.55$$

$b_1 = 286 - \frac{6}{11}(430) = \frac{566}{11} = 51.454$

B $MSE = \frac{SSE}{N-5} = \frac{\Sigma(y_i - \hat{y}_i)^2}{n-2}$

$$\frac{320 - 320.96 + 250 - 260.46 + 300 - 287.96 + 270 - 293.40 + 290 - 279.69}{3}$$

$\frac{-12.53}{3}$ -4.177

$s^2 = -4.177$

C $R^2 = \frac{SSR}{SSTO} = \frac{S_{xy}}{S_x S_y} \quad \frac{3600}{\sqrt{6600}\sqrt{2920}} = 0.8200$

$r^2 = 0.8200^2 = 0.6724$

R^2 measures proportion of total variance about the mean $\bar{y}$ explained by the regression

I would be concerned for values below 70% around 20%, 20% represents a bad fit of model

- ε_i are independent
- ε_i are normally distributed
- mean of 0 and constant variance σ^2
- x_i is the i^{th} value of predictor variable which is a known constant for all i.
- ε_i are independent
- means of y_i can be joined by a straight line

  $E(y_i) = \beta_0 + \beta_1 x_i$

  where β_0 is and β_1 are unknown parameters such that β_1 is slope of r line and β_0 is intercept

$\Sigma(x_i - \bar{x})^2$

$\Sigma(x_i^2 + \bar{x}^2 - 2x_i\bar{x})$

$\Sigma x_i^2 + \Sigma \bar{x}^2 - 2\Sigma x_i \bar{x}$

$\Sigma x_i^2 + n\bar{x}^2 - 2n\bar{x}\,\Sigma x_i$ $\qquad \bar{x} = \frac{\Sigma x_i}{n}$

$\Sigma x_i^2 + \frac{\Sigma x_i^2}{n} - 2\frac{\Sigma x_i \Sigma x_i}{n}$

$\cancel{\Sigma x_i^2 + \Sigma x_i^2 - 2\Sigma x_i^2}$

$\Sigma x_i^2 - \Sigma x_i^2$

$\cancel{\Sigma x_i^2} - \frac{\Sigma x_i}{n}$

$\cancel{n\bar{x}^2} - 2\Sigma x_i \bar{x}$ $\qquad \Sigma x_i = n\bar{x}$

$\cancel{n\bar{x}^2} - 2n\bar{x}\frac{\bar{x}}{2}$

$\cancel{\Sigma\bar{x}^2} - 2\Sigma x_i \bar{x}$

$n\bar{x}^2 \quad -2 n\bar{x}\Sigma x_i \quad n\bar{x}^2$

$-2n\bar{x}$

$+\Sigma\bar{x}^2$	$-2\Sigma x_i \Sigma\bar{x}$
$+n\bar{x}^2$	$-2\Sigma x_i \bar{x}$
$+n\frac{\Sigma x_i \Sigma x_i}{n^2}$	$-2\Sigma x_i \frac{\Sigma x_i}{n}$
$+\frac{\Sigma x^2}{n}$	$-\frac{2\Sigma x_i^2}{n}$

$\Sigma x_i^2 - \frac{\Sigma x_i^2}{n}$

$\Sigma(x_i - \bar{x})(y_i - \bar{y})$

$\Sigma x_i\bar{y} \quad - \Sigma x_i\bar{y} \quad - \Sigma\bar{x}y_i \quad + \Sigma\bar{x}\bar{y}$

$\Sigma x_i y_i \quad - \Sigma x_i\bar{y} \quad - \Sigma\bar{x}y_i \quad + n\bar{x}\bar{y}$

$- n\bar{x}\bar{y} \quad - \bar{x}n\bar{y} \quad + n\bar{x}\bar{y}$

$\Sigma x_i y_i \quad - \quad n\bar{x}\bar{y}$

$\Sigma x_i y_i \quad - \quad \frac{\Sigma x_i \Sigma y_i}{n}$ ✓

2. $E(ax+by) = aE(x) + bE[y]$

$Var[aX+bY] = a^2 Var[X] + b^2 Var[Y]$

$Var[\Sigma x_i] = \Sigma Var[x_i]$

Corr $r = \dfrac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2}\sqrt{\Sigma(y_i-\bar{y})^2}}$

In statistical relationship it is the expected value of the dependent variable that is determined by the values of the independent variable through a functional relationship

B_1 is the slope of the regression line and indicates the change in the mean of the distribution of Y for one unit increase in X.

ASSUMPTIONS

1. X_i is the i^{th} ~~observed~~ value of the predictor variable, which is a known constant for all i
2. The observations Y_i (or ε_i) are independent
3. At any given X_i, Y_i (or ε_i) is normally distributed
4. The observations Y_i (or ε_i) have constant standard deviation
5. The means of Y_i can be joined by a straight line: $E[Y_i] = \beta_0 + \beta_1 X_i$ where β_0 and β_1 are unknown parameters

β_1 is the slope of the regression line and indicates the change in the mean of Y for a unit increase in X.

β_0 is the intercept of regression model. If it is sensible to think of a value of $X=0$ for a particular application, then β_0 gives the mean of the distribution of Y at $X=0$. So it is not always possible to have a physical explanation for this parameter.

$$\Sigma(y_i - B_0 - \beta_1 x_i)^2$$

$$\frac{dQ}{db_0} \quad -2\,\Sigma(y_i - b_0 - b_1 x_i) = 0$$

$$\Sigma y_i - \Sigma b_0 - \Sigma b_1 x_i = 0$$

$$\Sigma y_i - \Sigma b_1 x_i = \Sigma b_0$$

$$\Sigma y_i - b_1 \Sigma x_i = \Sigma b_0$$

$$n b_0 = \Sigma y_i - b_1 \Sigma x_i$$

$$b_0 = \frac{\Sigma y_i}{n} - \frac{b_1 \Sigma x_i}{n}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$\frac{dQ}{db_1} \quad -2\,\Sigma x_i(y_i - b_0 - b_1 x_i) = 0$$

$$\Sigma x_i y_i - \Sigma b_0 x_i - b_1 (\Sigma x_i)^2 = 0$$

$$b_1 \Sigma(x_i)^2 = \Sigma x_i y_i - B_0 \Sigma x_i$$

Sub in $b_0 \Rightarrow$ $$b_1 \Sigma(x_i)^2 = \Sigma x_i y_i - \left(\frac{\Sigma y_i}{n} - \frac{b_1 \Sigma x_i}{n}\right)\Sigma x_i$$

$$b_1 \Sigma(x_i)^2 = \Sigma x_i y_i - \frac{\Sigma y_i \Sigma x_i}{n} + \frac{b_1 (\Sigma x_i)^2}{n}$$

$$b_1 (\Sigma x_i)^2 + b_1 \frac{(\Sigma x_i)^2}{n} = \Sigma x_i y_i - \frac{\Sigma y_i \Sigma x_i}{n}$$

$$b_1 = \frac{\Sigma x_i y_i - \frac{\Sigma y_i \Sigma x_i}{n}}{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}$$

$$MSE = \frac{SSE}{n-2} = \frac{\Sigma(y_i - \hat{y})^2}{n-2}$$

B_1

$$b_1 \sim N\left(B_1, \frac{\sigma^2}{S_{xx}}\right) \qquad \frac{\sigma^2}{S_{xx}} = \frac{MSE}{S_{xx}} \quad \frac{MSE}{\Sigma(x_i - \bar{x})^2}$$

T-test $\frac{b_1 - B_1}{se(b_1)}$ $\xrightarrow{B_1 = 0}$ $\frac{b_1}{se(b_1)}$

$|t_{calc}| \leq t_{critical}$ accept H_0 $B_1 = 0$

The p-value measures the probability of observing a more extreme t-value (in either direction) than the one calculated in study

Bo

$E[B_0] = B_0$ $Var = \sigma^2 \frac{1}{n} + \frac{\bar{x}^2}{Sxx}$

$S.e(b_0) = \sqrt{MSE\left[\frac{1}{n} + \frac{\bar{x}^2}{Sxx}\right]}$

Var for a point

$Var[\hat{y}_p] = MSE\left[\frac{1}{n} + \frac{(x_p - \bar{x})^2}{Sxx}\right]$

Prediction interval

$se(x_p) = \sqrt{MSE\left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{Sxx}\right]}$

$\Sigma(y_i-\bar{y})^2 \quad \Sigma(y_i-\hat{y}_i)^2+\Sigma(\hat{y}_i-\bar{y})^2 + 2\Sigma(y_i-\hat{y}_i)^2(\hat{y}_i-\bar{y})$

$2\Sigma(y_i-\hat{y}_i)(\hat{y}_i-\bar{y}) \qquad \hat{y}_i = \bar{y}+b_1(x_i-\bar{x})$

$\Sigma\hat{y}_i(y_i-\hat{y}_i) - \Sigma\bar{y}(y_i-\hat{y}_i)$

$\Sigma(\bar{y}+b_1(x_i-\bar{x})(y_i-\hat{y}_i) \quad -\bar{y}\Sigma(y_i-\hat{y})$

$\bar{y}\Sigma(y_i-\hat{y}) + \Sigma b_1(x_i-\bar{x}) \quad -\bar{y}\Sigma(y_i-\hat{y})$

$\Sigma(y_i)=\Sigma(\hat{y}_i)$

$b_1\ \Sigma(x_i-\bar{x})(y_i-\hat{y}_i)$

$b_1\ \Sigma x_i(y_i-\hat{y}_i) \quad -\bar{x}\Sigma(y_i-\hat{y}_i) \qquad \Sigma(y_i-\hat{y}_i)=0$

$b_1\ \Sigma x_i(y_i-\hat{y}_i)$

$b_1\ \Sigma x_i(y_i-\bar{y}-b_1(x_i-\bar{x}) \qquad b_1(x_i-\bar{x}) = \hat{y}_i-\bar{y}$

$\Sigma x_i(y_i-\bar{y}) \quad -b_1\Sigma(x_i-\bar{x}) \qquad b_1 = \frac{\Sigma\hat{y}_i-\bar{y}}{\Sigma(x_i-\bar{x})}$

$-$

$\cancel{b_1\ \Sigma(x_i-\bar{x})(y_i-\hat{y}_i)}$

$\frac{\cancel{\Sigma(\hat{y}_i-\bar{y})} \quad \cancel{\Sigma(x_i-\bar{x})}\Sigma(y_i-\hat{y}_i)}{\cancel{\Sigma(x_i-\bar{x})}} \qquad b_1 = \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2}$

$\cancel{\Sigma(\hat{y}_i-\bar{y})}\Sigma$

$\Sigma x_i(y_i-\bar{y}) \quad -b_1\ \Sigma x_i(x_i-\bar{x}) \qquad -\bar{x}$ both from

$\Sigma(x_i-\bar{x})(y_i-\bar{y}) \quad -b_1\ \Sigma(x_i-\bar{x})^2$

$\Sigma(x_i-\bar{x})(y_i-\bar{y}) \quad - \frac{\Sigma(x_i-\bar{x})(y_i-\bar{y})}{\cancel{\Sigma(x_i-\bar{x})^2}}\cancel{\Sigma(x_i-\bar{x})^2}$

$1 \qquad -1 \qquad =0$

W

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\bar{y}-\hat{y}_i}{2\sigma^2}}$$

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\bar{y}}$$

$$\pi \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_i - b_0 - b_1 x_i}{\sigma}\right)^2}$$

$$\Sigma = \frac{1}{(2\pi\sigma)^{\frac{n}{2}}} e^{-\frac{1}{2}\frac{\Sigma(y_i - b_0 - b_1 x_i)^2}{\sigma}}$$

$$\log(L) = \log\left(\frac{1}{(2\pi\sigma)^{n/2}}\right) \frac{-1}{2\sigma^2}\Sigma(y_i - b_0 - b_1 x_i)^2$$

$$\log(1) - \log(2\pi\sigma)^{n/2} \quad -\frac{1}{2\sigma^2}\Sigma(y_i - b_0 - b_1 x_i)^2$$

$$-\frac{n}{2}\log 2\pi\sigma \quad -$$

$$-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\Sigma(y_i - b_0 - b_1 x)^2$$

$$\frac{dL}{d(b_0)} = \frac{1}{2\sigma^2} - 2\Sigma(y_i - b_0 - b_1 x_i)$$

$$\Sigma(y_i - b_0 - b_1 x_i) = 0$$

Same for b_1 $= \Sigma x_i(y_i - b_0 - b_1 x$

$$\frac{dL}{d\sigma} \quad -\frac{n}{2}\left(\frac{1}{\sigma^2}\right) + \frac{1}{2\sigma^4} \quad \Sigma(y_i - b_0 + b_1 x_i)^2 = 0.$$

$$\frac{1}{2\sigma^4}\Sigma(y_i - b_0 - b_1 x_i) = \frac{n}{2\sigma^2}$$

$$\Sigma(y_i - b_0 - b_1 x_i) = n\sigma^2$$

$$\sigma^2 = \frac{\Sigma(y_i - b_0 - b_1 x_i)}{n}$$

$$= \sigma^2 = \frac{\Sigma(y_i - \hat{y}_i)}{n}$$