2013 Regression Paper

Q1A. $b_1 = \dfrac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}$

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

$E[y_i] = b_0 + b_1 x_i$

$b_1 = \dfrac{225864 - \frac{(1118)(3220)}{16}}{81452 - \frac{(1118)^2}{16}} = \dfrac{3486}{13377} = 0.26057 = 0.26$

$b_0 = \frac{3220}{16} - \frac{1118}{16}(0.26) = 182.9724 = 182.97$

$E[y_i] = 182.97 + 0.26 x_i$

B b_1 - Slope: For each unit increase in the population density of the city, the mean distribution of robberies per 10 000 people increase by 0.26

b_0 - Intercept: Sometimes a physical explanation is not possible, it is the mean distribution of robbery rate for population density of 0. In this context it does not make sense

C $MSE = \dfrac{SSE}{N-2} = \dfrac{S_{yy} - \frac{S_{xy}^2}{S_{xx}}}{N-2}$

$\Sigma y_i^2 - n\bar{y}^2 - \dfrac{(\Sigma x_i y_i - n\bar{x}\bar{y})^2}{\Sigma x_i^2 - n\bar{x}^2}$ $\quad S_{yy} - \frac{S_{xy} S_{xy}}{S_{xx}}$

SSE formula $\Sigma(y_i - \hat{y}_i)^2 = SS_{yy} - \hat{\beta}_1 SS_{xy}$

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

$(649736 - 648025) - (0.26)[\text{illegible}]$

$1711 - 226.54 = \frac{148441}{N-2} = 106.029$

5. $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

$t_{calc} = \frac{b_1 - \beta_1}{se(b_1)} = \frac{b_1}{\sqrt{\frac{MSE}{S_{xx}}}} \quad \frac{0.26}{\sqrt{\frac{106.024}{3331.75}}} \quad \frac{0.26}{0.178}$

$t_{calc} = 1.457$

$t_{critical\ (95\%,\ 14)} = 2.145$

$t_{calc} < t_{critical}$ Fail to reject H_0.

There is not enough evidence to suggest higher robbery rate in more densely populated area

6. 95% CI for b_1 $\quad b_1 \pm t_{critical}(se(b_1))$

$0.26 \pm 2.145(0.178)$

0.26 ± 0.382

95% CI for b_0 $\quad b_0 \pm t_{critical}(se(b_0))$

$se(b_0) = \sqrt{MSE\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]} \quad \sqrt{106.029\left[\frac{1}{16} + \frac{1249424}{3331.75}\right]}$

$182.97 \pm 2.145(19.945)$

182.97 ± 42.784

2013 Regression Exam Paper

Q2

A.

Source	D.F.	SS	MS	F
Regression	1	49	49	49/11
Error	9	1.1	11/90	
Total (corrected)	10	60		

B. $F_{(0.95,\ 1,\ 9)} = F_{crit} = 5.3177$

$H_0: B_1 = 0$ vs $H_1: B_1 \neq 0$

$F_{critical} = 5.3177$

$F_{calc} = 40.090$

$F_{calc} >> F_{crit}$

reject H_0, relationship exists between advertising expenditure and sales revenue

C. NO we cannot say whether it is strictly increasing / decreasing. The F ratio / test is based on a ratio of two mean square values, this gives no indication about positive/negative relationship only the relationship

- The F figure can never be negative as it is a squared number

D. $SSE = \sum(y_i - \hat{y}_i)^2$ Error Sum of Squares

This is the squared deviation of the predicted value from the observed by value. By the method of least squares we aim to minimise this value (variance) so as to get as close to the observed value as possible

E. - r to explain a positive linear relationship

If correlation = 1 we have a perfect relationship.

The change in the 2 variables move opposite in same direction

No linear relationship corresponds to r = 0. The two variables are uncorrelated. This does NOT imply independence

Warning!:
- Does not capture nonlinear relationships
- A third variable may be factor in picture
- Can be coincidence
- Correlation does not imply causality
- It needs direction but not slope → job for regression

5.

2013 Regression Exam Paper

Q5 A Assumptions:

1. X_i is the i^{th} value of the predictor variable, which is 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 are normally distributed
4. The observations y_i or ε_i have constant Standard deviation
5. The mean of y can be joined by a straight line

$$E[y_i] = \beta_0 + \beta_1 x_i$$

β_1: Slope of regression line, indicates change in mean value of y distribution for one unit increase in x

β_0: intercept - value of $x=0$ for a particular application, the β_0 gives the mean distribution of y at $x=0$. Not always possible to have an explanation

B $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

$$T_{calc} = \frac{b_1 - \beta_1}{se(b_1)} \quad \frac{0.7 - 0}{\sqrt{\frac{MSE}{SS_x}}} \quad \frac{0.7}{\sqrt{\frac{0.6}{10}}} = 2.834$$

T_{crit} $t_{(0.025, 8)} = 2.080$

$T_{calc} > T_{critical}$ reject H_0

C $10 = b_0$

$0.5 = b_1$

There is a 0.5 increase in the mean value for the distribution for y for each one unit increase of x.

$\hat{y} = 10 + 0.5(10) = 10 + 5 = 15$ correct. ✓

D The mean do seem to be zero ✓

Constant variance is in question

Seems to be funnel shaped ≠ possible a mistake

The data seems to be funnel shaped indicating a non constant variance and non independent

Regression Exam Paper 2013

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

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

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

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

$$-2n\bar{x}\sum x_i$$

$$-2n\bar{x}\,\bar{x} + n\bar{x}^2$$

$$\sum x_i^2 - n\bar{x}^2$$

1 A. $E[y_i] = \hat{b}_0 + \hat{b}_1 x_i$

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

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

$$b_1 = \frac{225864 - 1118(3270)/16}{81452 - \frac{(1118)^2}{16}} \qquad \frac{-8737.875}{3331.75}$$

$$= 0.26157 \qquad = 0.26 = b_1$$

$$b_0 = \bar{y} + 0.26\bar{x}$$

$$\frac{3270}{16} + 0.26\left(\frac{1118}{16}\right) = 219.5275$$

$$= 219.53$$

$$y_i = 219.53 + 0.26x_i$$

b $\hat{B}_0$ can be interpted as the robbery rate when the population is zero, however this does not make sense in this case as we cannot rob if no one is in population

$\hat{B}_1$ is the increase in the mean distribution of robbery rate per (100,000) per unit increase in estd population

c MSE is $\sqrt{\frac{SSE}{n-2}}$ $\frac{\sum y_i - \hat{y}_i^2}{n-2}$

$\sum y_i = \sum \hat{y}_i$

$\frac{\sum (y_i^2 - 2 y_i \hat{y}_i + \hat{y}_i)}{n-2}$

$\sum y_i^2 - 2 \sum y_i \hat{y}_i + \sum \hat{y}_i$

$\sum y_i^2 - 2 \sum y_i^2 + \sum y_i^2$

$= \frac{\sum y_i - \sum y_i^2}{n-2}$

$= 3220 -$

$SSE = \sum y_i^2 - n\bar{y}^2 - \frac{(\sum x_i y_i - n\bar{x}\bar{y})^2}{\sum x_i^2 - n\bar{x}^2}$

$649736 - 16(40501.5625) - \frac{75951225}{3331.75}$

$1711 - 227.9619$

$= \frac{1483.038243}{14}$

$MSE = 105.9312888 = 105.93$

One tale

D $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$.

Test value $\frac{0.26}{\sqrt{\frac{MSE}{S_{xx}}}} = \frac{0.26}{\sqrt{\frac{105.93}{3331.75}}} = 0.46811 \approx 1.457$

$T_{critical}(0.05, 14)$ 2.7641

$|t| \leq t_{critical}$ no evidence against H_0

$0.26 \pm 1.761(0.5650)$

0.26 ± 1.212 0 lies on interval

3

Regression Exam Paper 2013

1\. E β_1: 0.26 ± 1.212

β_0: $21953 \pm 2.145 \left(\sqrt{MSE \left[\frac{1}{16} + \frac{4882.515625}{331.75} \right]} \right)$

$21953 \pm 2.145 \ (12.7222)$

21953 ± 27.289

Q2 A

Source	DF	SS	MS	F
Regression	1	4.9	4.9	40.0909
Error	9	1.1	0.1222	
Total	10	6.0		

B $F_{0.95,1,8}$ $F_{critical} = 5.3177$

$F_{calc} >> F_{critical}$ we reject H_0, there is a relationship between variables

C We cannot say, we can use of ANOVA table to calculate R^2, then the square root of this figure is the r-value, the correlation coefficient, which indicate either a strictly positive or negative relationship, because of the square root, we do not know if it is positive or negative

D SSE the error sum of squares. This is actually the error terms e_i, our model aims to minimize this value, it is the squared deviation of the predicted value from the actual value.
By method of least squares this value should be at a minimum

6. For correlation $r > 0$ we have positive linear correlation. If correlation $= 1$, we have perfect positive linear correlation. The changes in the 2 variable move approx in same direction

No linear relationship corresponds to $r = 0$. The two variables are uncorrelated. This does not imply that variable are independ.

- It does not capture nonlinear relationship
- It provides direction but not slop $\Rightarrow$ job is regression

Q5 a. Assumption:

ε_i is a random error term with properties:

- $E(\varepsilon_i) = 0$
- $Var(\varepsilon_i) = \sigma^2$
- ε_i and ε_j are uncorrelated so $Cov(\varepsilon_i, \varepsilon_j) = 0$ for all i, j $i \neq j$.
- ε_i's are normally distributed

b. $n = 23$ $b_1 = 0.7$ $MSE = 0.61$ $S_{xx} = 10$.

H_0 $\beta_1 = 0$ vs H_1 $\beta_1 \neq 0$

$$t_{calc} = \frac{0.7}{\sqrt{\frac{0.61}{10}}} = 2.83421 7$$

$t_{critical}$ $(0.95, 21)$ $= 2.080$

$|t| > t_{critical}$

$\Rightarrow$ reject H_0 there is a relationship

c. Correction: The mean value of the y distribution increase by 05 with each increase in one unit by x

But yes when x is 10 $y = 15$.

5.

Regression Exam Paper 2013

Q52: mean is zero ✓.
constant variance ✓ in question with 3 values close to ±30 range.

The data seems to be funnel shaped ✓ indicating a non-constant variance and non independent