STAT-UB 103 Homework 8

Ishan Pranav

April 23, 2023

1 Learning the mechanics

a. See below.

x_i	y_i	x_i^2	$x_i y_i$
7	2	49	14
4	4	16	16
6	2	36	12
2	5	4	10
1	7	1	7
1	6	1	6
3	5	9	15

$\sum_{i=0}^{6} x_i = 24.$ $\sum_{i=0}^{6} y_i = 31.$ $\sum_{i=0}^{6} x_i^2 = 116.$ $\sum_{i=0}^{6} x_i y_i = 80.$

b.

$$s_{x,y}^2 = \sum_{i=0}^{n-1} (x_i - \bar{x})(y_i - \bar{y}) \approx -26.2857\ldots$$

c.

$$s_x^2 = \sum_{i=0}^{n-1} (x_i - \bar{x})^2 \approx 33.7143\ldots$$

d.

$$b_1 = \frac{s_{x,y}^2}{s_x^2} \approx -0.7797\ldots$$

e.

$$\bar{x} \approx 3.4286\ldots$$
$$\bar{y} \approx 4.4286\ldots$$

f.

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

$$b_0 = \bar{y} - b_1 \bar{x} \approx 7.1017\ldots$$

g.

$$\hat{y} = b_1 x + b_0 = -0.7797x + 7.1017.$$

2 Forecasting movie revenues with Twitter

$$b_1 = 0.078767\ldots$$

Assuming that movie revenue and tweet rate are linearly related, we estimate a movie's opening weekend revenue increases by an average of 7,876,700 dollars as the tweet rate for the movie increases by an average of 100 tweets per hour.

3 Congress voting on women's issues

Let y represent a legislator's American Association of University Women score as a function of the number of daughters (x) that the legislator has.

$$y_i = \beta_1 x_i + \beta_0 + \epsilon_i.$$

a. If it is true that having a daughter influences voting on women's issues, the sign of β_1 will either be positive or negative. A positive linear coefficient (β_1) indicates that legislators with daughters are more likely to vote in favor of women's rights issues. A negative linear coefficient indicates that legislators with daughters are less likely to vote in favor of women's rights issues.

b.

$$n = 434.$$

$$\nu = n - 2 = 432.$$

$$b_1 = 0.27.$$

$$s_{b_1} = 0.74.$$

$$t^* = F_{432}^{-1}(0.025) \approx 1.9655\ldots$$

We can be ninety-five-percent confident that the true value of β_1 is in the interval:

$$(-1.18, 1.72)$$

4 RateMyProfessors.com

Let y_i represent the student evaluation of teaching and x_i represent the RateMyProfessors.com rating.

$$n = 426.$$

$$r_{x,y} = 0.68.$$

a. Let β_1 represent the linear coefficient of the RateMyProfessors.com rating, β_0 represent the constant coefficient, and ϵ_i represent the error term.

$$y_i = \beta_1 x_i + \beta_0 + \epsilon_i.$$

b. The sample correlation of 0.68 indicates a strong positive linear relationship between the RateMyProfessors.com rating and the student evaluation of teaching. Professors with high RateMyProfessors.com ratings also tend to have high student evaluations of teaching.

c. The estimated slope of the line is positive. A positive sample correlation indicates a positive linear relationship.

d. There is convincing evidence, at the 0.1% significance level, that RateMyProfessors.com ratings and student evaluations of teaching are correlated.

e.

$$r^2 \approx 0.46.$$

Approximately 46 percent of the variation in student evaluations of teaching can be explained by variations in RateMyProfessors.com ratings using a least-squares regression line.

5 Data on pricing of ladies' diamond rings

$$n = 48.$$

$$\nu = n - 2 = 46.$$

$$b_1 = 3721.0.$$

$$s_{b_1} = 81.8.$$

a. Based on the scatterplot of price and weight, a linear regression model appears very appropriate. The data resemble a straight line.

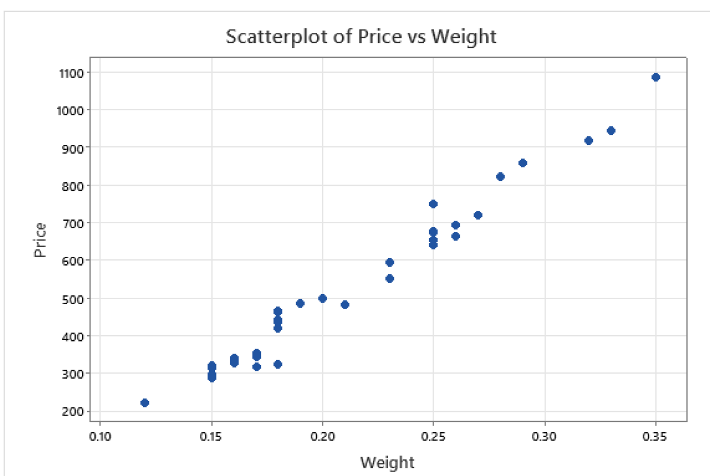

Figure 1: Scatterplot of price (vertical) and weight (horizontal).

b. See Figure 2.

Regression Analysis: Price versus Weight

Regression Equation

Price = -259.6 + 3721.0 Weight

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-259.6	17.3	-14.99	0.000	
Weight	3721.0	81.8	45.50	0.000	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
31.8405	97.83%	97.78%	97.63%

Figure 2: Linear regression analysis for price (response) and weight (predictor).

c. Let y represent the price of a ring and x represent the weight of its diamond.

$$y_i = \beta_1 x_i + \beta_0 + \epsilon_i.$$

$$y_i = b_1 x_i + b_0 + e_i.$$

$$\hat{y} = (\text{SGD\$3721.00})x - (\text{SGD\$259.60}).$$

$$y_{0.23} = \text{SGD\$596.23}.$$

d. Yes, there is evidence of a significant linear relationship between the price of a ring and the weight of its diamond. The weight coefficient (β_1) has a Student t-statistic of -14.99, with a P-value of approximately 0.0000. This probability is statistically significant even at the 0.01 percent significance level ($\alpha = 0.0001$), indicating that it is very unlikely that the null hypothesis ($H_0 : \beta_1 = 0$) is true. Therefore, there is sufficient evidence to reject the null hypothesis and conclude in favor of the alternative hypothesis ($H_1 : \beta_1 \neq 0$), that there is in fact a linear relationship between a ring's price and its diamond's weight.

e. On average, for each additional caret a ring's diamond weighs, its price is about \$3721.00 more expensive.

$$t^* = F_{46}^{-1}(0.025) \approx 2.0129\ldots$$

We can be ninety-five-percent confident that the true value of β_1 is in the interval:

$$(\$3556.35, \$3885.65)$$

f. Approximately 97.83 percent of the variation in ring prices can be explained by variations in diamond weight using a least-squares regression line.

g.

$$s_e \approx \$31.84\ldots$$

h.

$$H_0 : \beta_1 = \$3500.00.$$

$$H_1 : \beta_0 \neq \$3500.00.$$

$$\alpha = 0.01.$$

$$t^* = F_{46}^{-1}(0.005) \approx 2.6870\ldots$$

We can be ninety-nine-percent confident that the true value of β_1 is in this interval

$$(\$3501.20, \$3940.80)$$

Since the value $3500.00 is outside the interval, there is convincing evidence to reject the null hypothesis at the one-percent significance level. We can conclude that the true slope is different from $3500.00.

i. We can be ninety-five-percent confident that the expected price of rings which weigh 0.230000 carats is in this interval:

$$(\$586.03, \$606.39)$$

6 Data on stock returns and earnings per share

a. It is somewhat unreasonable to fit a linear regression model to these data.

Visually, Figure 3 indicates that it is plausible for a negative linear relationship to exist between a company's earnings per share and its stock returns. However, that relationship appears to be extremely weak, and a great deal of error is apparent.

Observation suggests that the underlying assumption that the data are linear may not hold true, and that it is therefore inappropriate to employ a linear regression model.

b. According to the regression analysis in Figure 4, the earnings per share metric is not necessarily useful for predicting stock returns.

$$H_0 : \beta_1 = 0.$$

$$H_1 : \beta_1 \neq 0.$$

$$\alpha = 0.05.$$

$$P(t = -1.59 \,|\, H_0) \approx 0.118.$$

Given that the null hypothesis is true, the probability of obtaining a Student t-statistic as extreme as (or more extreme than) -1.59 is approximately 11.8

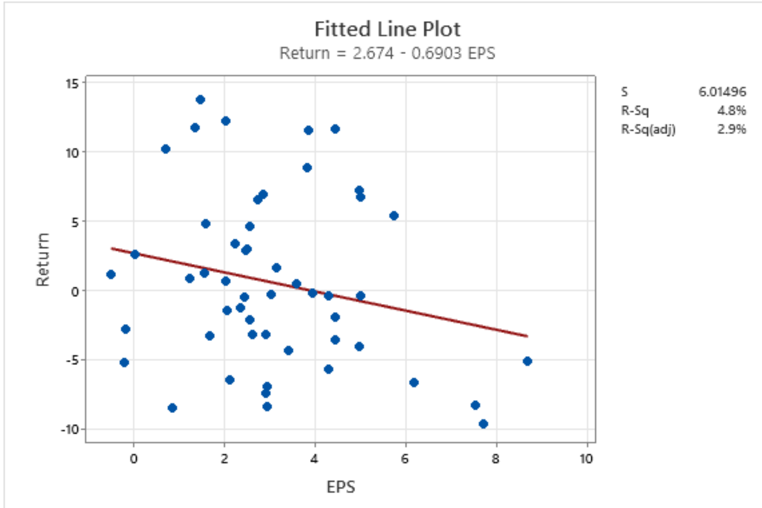

Figure 3: Fitted line for return (response) and earnings per share (predictor).

percent. Since 0.118 is greater than the significance level ($\alpha = 0.05$), there is insufficient evidence to reject the null hypothesis at the five-percent significance level. We cannot conclude beyond a reasonable doubt that a linear relationship exists between a company's earnings per share and its stock returns.

c. Only 4.82 percent of the variation in a company's stock returns can be explained by variations in its earnings per share using a least-squares regression line. This suggests that the linear relationship is very weak.

d. We can be ninety-five-percent confident that the stock return of companies with earnings per share of 6 is in this interval:

$$(-13.9263, 10.9902)$$

This is not a useful interval: It contains positive values, negative values, and zero. By straddling zero, the interval provides little to no information about whether a stock with earnings per share of 6 offers positive returns, negative returns, or no returns at all. The large margin of error makes this interval even less meaningful.

Regression Analysis: Return versus EPS

Regression Equation

Return = 2.67 - 0.690 EPS

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	2.67	1.58	1.70	0.096	
EPS	-0.690	0.434	-1.59	0.118	1.00

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
6.01496	4.82%	2.92%	0.00%

Figure 4: Fitted line for return (response) and earnings per share (predictor).