Stat 235

Lab 5

WOOSAREE, Arun

Lab EL12

TA: Jessa Marley

November 30, 2018

1

1 a)

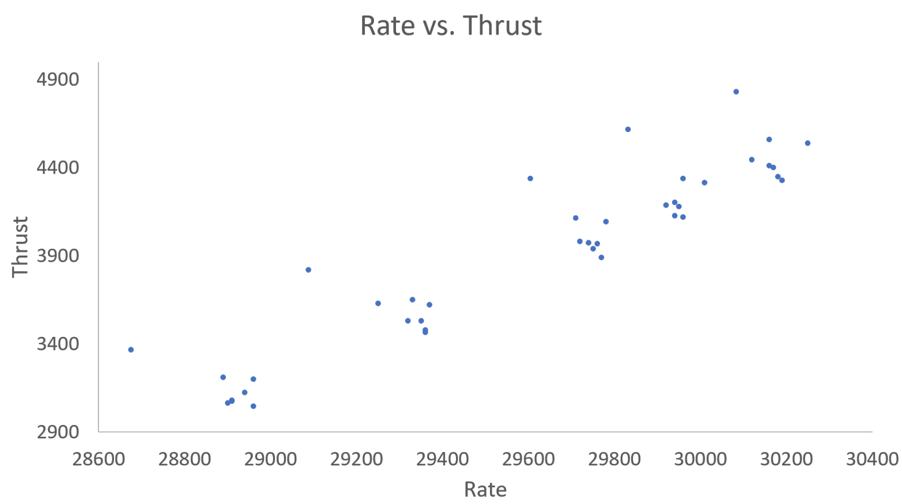

Figure 1: Fuel flow rate and the corresponding thrust

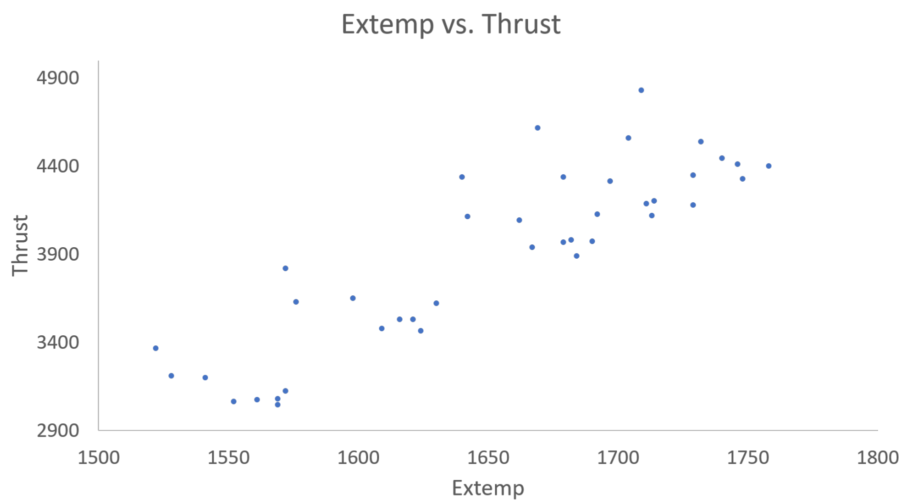

Figure 2: Exhaust temperature and the corresponding thrust

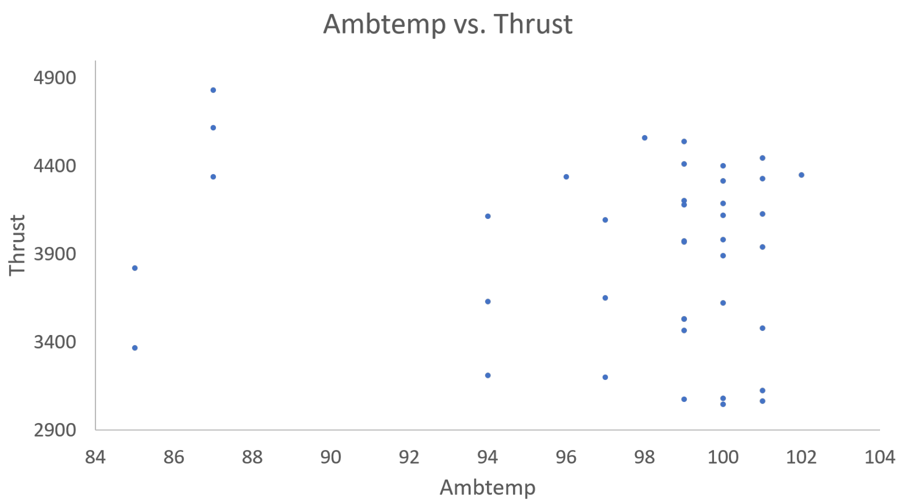

Figure 3: Ambient temperature and the corresponding thrust

1 b)

- The plot for rate vs. thrust above appears to exhibit a relatively strong positive linear relationship, with less scatter relative to the other plots.
- The plot for extemp vs. thrust above appears to exhibit a fairly strong positive linear relationship, but with a bit more scatter than Figure 1.
- The plot for ambtemp vs. thrust above does not appear to exhibit a strong linear relationship due to the points being very visually scattered. If there is a linear relationship, it looks like it would be almost flat, with a slightly negative slope.

Overall, the rate seems to have the stongest linear relationship with the thrust, since it has the least amount of scatter.

2

2 a)

	rate	extemp	ambtemp	thrust
rate	1			
extemp	0.975750317	1		
ambtemp	0.216465939	0.301764125	1	
thrust	0.928792499	0.87169251	-0.14744	1

Table 1: Correlation matrix for the variables

2 b)

The regressor variable with the largest magnitude of correlation with the response variable thrust is the rate, with a value of 0.928792499. The regressor variable with the lowest magnitude of correlation with the thrust is ambtemp, with a value of -0.14744. The values obtained in Table 1 agree with the conclusions reached in Question 1, with the rate having the largest correlation with a positive slope, and the ambient temperature having a weak correlation and slightly negative slope. Additionally, we see that the exhaust temperature also has a fairly strong linear relation with the thrust, and its correlation value is both positive and less than the rate variable, as predicted in Question 1 b).

3

The regression model is as follows:

$$y = \beta_0 + \beta_1 x + \epsilon \tag{1}$$

, where y is the thrust, x is the fuel flow rate, β_0 is the y-intercept, β_1 is the slope, and ϵ is the random error. So, Equation 1 becomes

$$thrust = \beta_0 + \beta_{rate} \times (rate) + \epsilon$$

The model's assumptions are as follows:

- The distribution of ϵ at any x has a mean of 0 ($\mu_\epsilon = 0$ to aid linearity)
- The standard deviation of ϵ is the same for any x (it's constant)
- The distribution of ϵ at any x is normal
- The random deviations $\epsilon_1, \epsilon_2, ..., \epsilon_n$ associated with different observations are independent of one another

4

4 a)

Using the regression tool in Excel, we find $\beta_0 = -25860.12772$, and $\beta_1 = 1.005348711$. So, Equation 1 becomes

$$thrust = -25860.12772 + 1.005348711 \times (rate)$$

The following points from the scatterplot in Figure 1 were visually far away from the rest of the data, and therefore may be influential on the fitted line:

(rate, thrust)

(28675, 3368),

(29088, 3820),

(29604, 4340),

(29831, 4617), and

(30083, 4833)

The following points on the residual plot had relatively large residuals which were visually far away from the rest of the data and therefore are probably outliers:

(rate, residual)

(28675, 399.7534461),

(29088, 436.5444286),

(29604, 437.784494),

(29831, 486.5703367), and

(30083, 449.2224616)

4 b)

The estimate of the model standard deviation is:

$$s = 189.466735$$

4 c)

The percentage of the variation in thrust which is explained by the fuel flow rate is the R^2 value.

$$R^2 = 0.862655506 \approx 86.27\%$$

The other predictor variables extemp and ambtemp may explain the remaining variation (probably the extemp variable moreso than the ambtemp variable).

4 d)

Let the null hypothesis be that the fuel flow rate does not have any value in explaining the thrust:

$$H_0 : \beta_1 = 0 \quad vs. \quad H_A : \beta_1 \neq 0$$

The distribution of the test statistic follows a null dustribution. Using Excel, we find that

$$F_0 = 15.44915999 \sim F^{1}_{39-2}$$

Using Excel, we find the p-value is:

$$p = 5.727\,62 \times 10^{-18}$$

Because the p-value obtained is extremely small, we reject H_0. i.e. the data strongly suggests that there is a relationship between the fuel flow rate and the thrust.

4 e)

The mean change in thrust as the fuel flow rate increases by 1 unit is the slope. From the Regression Tool output, we find that the 95% confidence interval for this is:

$$(0.873611965, 1.137085456)$$

4 f)

With a flow rate of 30 250, the predicted thrust is:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = -25860.12772 + 1.005348711 \times 30250 \approx 4551.671$$

In this case, the value of the residual is

$$e = y - \hat{y} = 4540 - 4551.671 = -11.67077$$

4 g)

Figure 4: Plot of residuals for the rate variable

The residuals appear randomly scattered about the horizontal 0 line. Because of this, the normality assumption for the residuals does seem appropriate. The constant variance assumption may not hold, however since the plot seems to pointed ever so slightly upwards.

5

5 a)

Using the regression tool in Excel, we find $\beta_0 = -6643.268654$, and $\beta_1 = 6.384931687$. So, Equation 1 becomes

$$thrust = -6643.268654 + 6.384931687 \times (extemp)$$

From the scatterplot in Figure 2, there don't seem to be any influential points which are far away from the rest of the data. Similarly, there are no obvious outliers in the residual plot.

5 b)

The estimate of the model standard deviation is:

$$s = 250.5362661$$

5 c)

The percentage of the variation in thrust which is explained by the fuel flow rate is the R^2 value.

$$R^2 = 0.759847831 \approx 75.98\%$$

The other predictor variables rate and ambtemp may explain the remaining variation (probably the rate variable moreso than the ambtemp variable).

5 d)

Let the null hypothesis be that the fuel flow rate does not have any value in explaining the thrust:

$$H_0 : \beta_1 = 0 \quad vs. \quad H_A : \beta_1 \neq 0$$

The distribution of the test statistic follows a null dustribution. Using Excel, we find that

$$F_0 = 10.9650813 \sim F^1_{39-2}$$

Using Excel, we find that the p-value is:

$$p = 2.481\,43 \times 10^{-13}$$

Because the p-value obtained is extremely small, we reject H_0. i.e. the data strongly suggests that there is a relationship between the exhaust temperature and the thrust.

5 g)

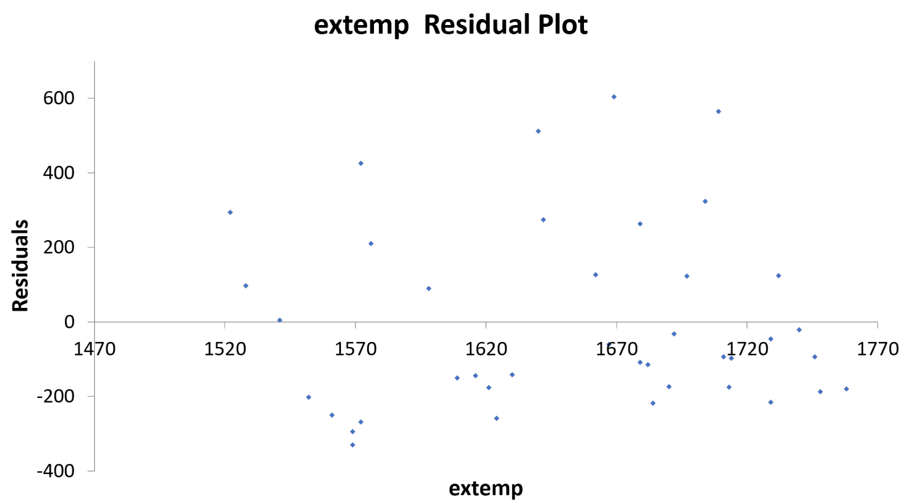

Figure 5: Residual plot for the exhaust temperature variable

The residuals appear randomly scattered about the horizontal 0 line. Because of this, the normality assumption for the residuals does seem appropriate. The constant variance assumption may not hold, since the plot appears to be pointed slightly upwards.

6

Thrust vs.	R^2	s	t-statistic	p-value
Rate	0.862655506	189.466735	15.44915999	$5.727\,62 \times 10^{-18}$
Extemp	0.759847831	250.5362661	10.9650813	$2.481\,43 \times 10^{-13}$

Table 2: Summary of regression results

The fuel flow rate appears to be the best predictor of thrust. It has a stronger correlation (larger R^2), and it also accounts for a larger percentage of the variation in thrust. It also has a smaller standard error relative to its respective mean, and a much smaller p-value.