**Regression Inference and Forecasting**

STAT-UB.0103 – Statistics for Business Control and Regression Models

## Inference

1. Here are the least squares estimates from the fit to model

$$\text{Price} = \beta_0 + \beta_1 \,\text{Size} + \varepsilon,$$

where price is measured in units of \$1000 and size is measured in units of 100 $\text{ft}^2$.

```
The regression equation is
price = 182 + 45.0 size

              Coef    SE Coef        T        P
Constant    182.27      62.43     2.92    0.010
size         44.95       4.37    10.29    0.000

S = 101.4    R-Sq = 86.9%    R-Sq(adj) = 86%
```

(a) Construct a 95% confidence interval for $\beta_1$.

(b) What is the meaning of the confidence interval for $\beta_1$?

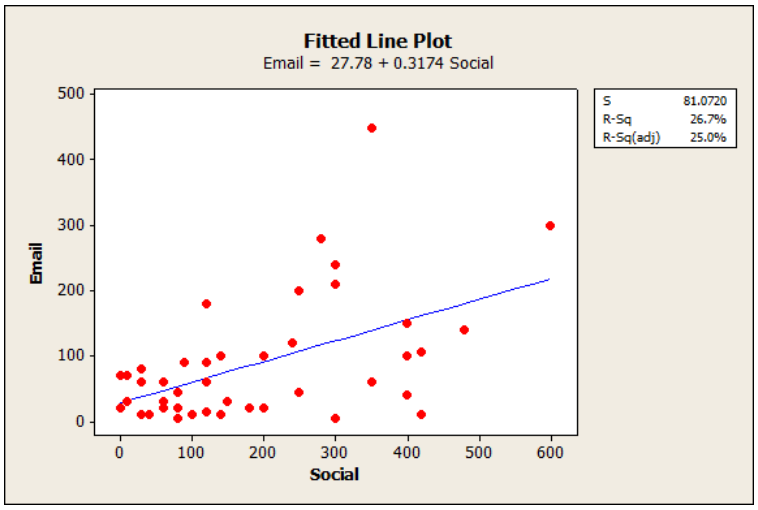
(c) What is the meaning of a 95% confidence interval for $\beta_0$? In the context of the housing data, is this useful?

(d) Perform a hypothesis test at level 5% of whether or not the is a linear relationship between Size and mean Price.

2. 44 NYU undergraduates reported the amount of time they spent communicating via email and via social media (in minutes per week). We will use this data to examine the relationship between email usage and social media usage. We fit the model

$$\text{Email} = \beta_0 + \beta_1 \text{Social} + \varepsilon.$$

using least-squares. The scatterplot and Minitab regression output follow.




The regression equation is
Email = 27.8 + 0.317 Social

| Predictor | Coef | SE Coef | T | P |
|---|---|---|---|---|
| Constant | 27.78 | 19.32 | 1.44 | 0.158 |
| Social | 0.31744 | 0.08109 | 3.91 | 0.000 |

S = 81.0720  R-Sq = 26.7%  R-Sq(adj) = 25.0%

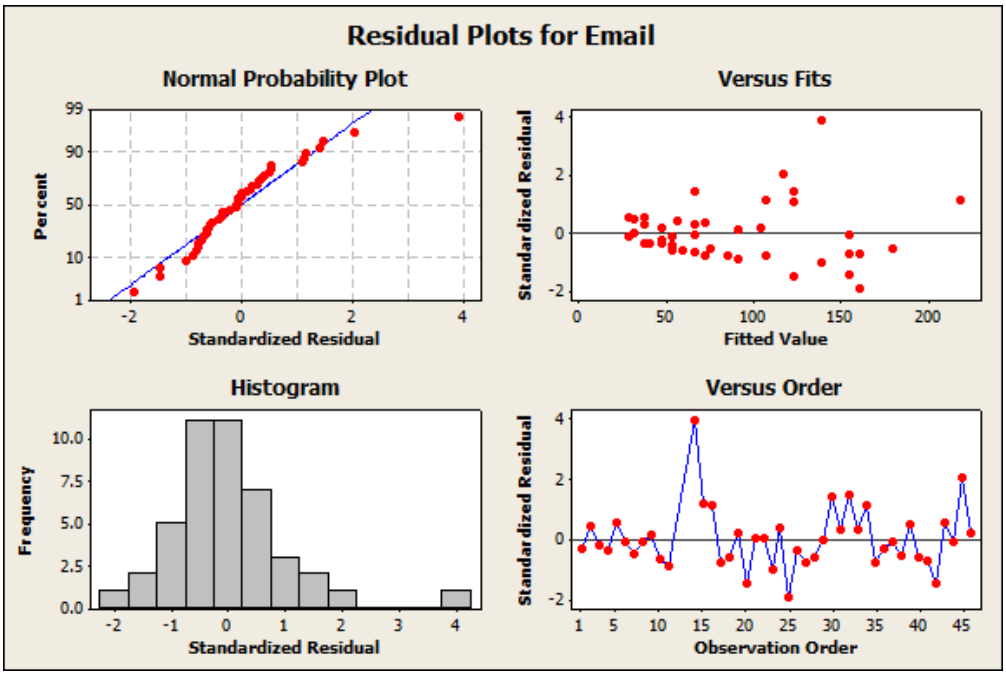
Analysis of Variance

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Regression | 1 | 100716 | 100716 | 15.32 | 0.000 |
| Residual Error | 42 | 276052 | 6573 | | |
| Total | 43 | 376768 | | | |

Unusual Observations

| Obs | Social | Email | Fit | SE Fit | Residual | St Resid |
|---|---|---|---|---|---|---|
| 14 | 350 | 450.0 | 138.9 | 18.1 | 311.1 | 3.94R |
| 16 | 600 | 300.0 | 218.2 | 35.8 | 81.8 | 1.12 X |
| 45 | 280 | 280.0 | 116.7 | 14.5 | 163.3 | 2.05R |

(a) Use the residual plots below to assess whether or not the regression assumptions hold.




(b) Is there a significant linear relationship between social media usage and email usage?

(c) Quantify the relationship between email usage and social media usage using a 95% confidence interval. (You will need the value $t_{.025} = 2.018$.)

## Forecasting

3. Here are the least squares estimates from the fit to model Price $= \beta_0 + \beta_1$ Size $+ \varepsilon$, where price is measured in units of \$1000 and size is measured in units of 100 ft$^2$, along with the result of using the model to predict the mean price at size 2000 ft$^2$.

```
The regression equation is
price = 182 + 45.0 size

              Coef    SE Coef        T        P
Constant    182.27      62.43     2.92    0.010
size         44.95       4.37    10.29    0.000

S = 101.4    R-Sq = 86.9%    R-Sq(adj) = 86%

Predicted Values for New Observations

NewObs     Fit  SE Fit       95% CI            95% PI
      1  1081.3    38.1  (1000.4, 1162.1)  (851.7, 1310.9)

Values of Predictors for New Observations

NewObs           size
      1           20.0
```

(a) Find a 95% confidence interval for the mean price of all apartments with size 2000 ft$^2$.

(b) Find a 95% prediction interval for the price of a particular apartments with size 2000 ft$^2$.

(c) Make a statement about the prices of 95% of all apartments with size 2000 ft$^2$.

(d) What is the difference between the confidence interval and the prediction interval?