Regression Inference and Forecasting
COR1-GB.1305 – Statistics and Data Analysis

Inference

1. Here are the least squares estimates from the fitting the model

$$\text{Price} = \beta_0 + \beta_1 \text{Size} + \varepsilon_1$$

for $n = 18$ apartments in Greenwich Village. Price is measured in units of \$1000 and size is measured in units of 100 ft^2.

```
Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
101.375  86.87%     86.05%      81.13%

Coefficients

Term            Coef  SE Coef  T-Value  P-Value   VIF
Constant       182.3     62.4     2.92    0.010
Size(100sqft)  44.95     4.37    10.29    0.000  1.00

Regression Equation

Price($1000) = 182.3 + 44.95 Size(100sqft)
```

   (a) Construct a 95% confidence interval for β_1.

   (b) What is the meaning of the confidence interval for β_1?

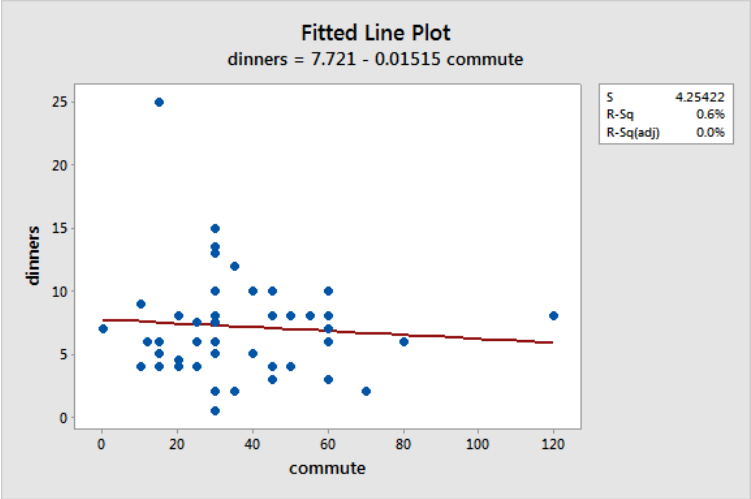
   (c) What is the meaning of a 95% confidence interval for β_0? In the context of the housing data, is this useful?

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

2. 51 students reported their commute times (in minutes) and the number of times they go out to dinner in a typical month. We will use this data to examine the relationship between these two variables. We fit the model

$$\text{Dinners} = \beta_0 + \beta_1 \text{Commute} + \varepsilon$$

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




```
Model Summary

      S   R-sq  R-sq(adj)  R-sq(pred)
4.25422  0.56%      0.00%       0.00%

Coefficients

Term        Coef  SE Coef  T-Value  P-Value   VIF
Constant    7.72     1.21     6.38    0.000
commute  -0.0152   0.0288    -0.53    0.601  1.00

Regression Equation

dinners = 7.72 - 0.0152 commute
```

(a) Quantify the relationship between dinners and commute time using a 95% confidence interval. (You will need the value $t_{.025,49} \approx 2.021$.)

(b) Perform a hypothesis test to determine if there is a significant linear relationship between dinners and commute time.

Forecasting

3. We used the regression model fit to the housing data to predict price at size 2000 ft^2:

```
Regression Equation

Price($1000) = 182.3 + 44.95 Size(100sqft)


Variable       Setting
Size(100sqft)       20


    Fit   SE Fit        95% CI              95% PI
1081.27  38.1287  (1000.44, 1162.10)  (851.667, 1310.88)
```

   (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?

4. We fit a regression model to the 294 restaurants from the 2003 Zagat data. Our predictor variable is food quality (1–30), and our response variable is price ($). Here is the result of using the fitted model to predict the price when the food quality is 25.

```
Model Summary

      S    R-sq  R-sq(adj)  R-sq(pred)
12.5559  27.93%     27.68%      26.86%

Coefficients

Term       Coef  SE Coef  T-Value  P-Value   VIF
Constant  -4.74     3.95    -1.20    0.232
Food      2.129    0.200    10.64    0.000  1.00

Regression Equation

Price = -4.74 + 2.129 Food

Variable  Setting
Food           25

    Fit   SE Fit        95% CI              95% PI
48.4832  1.33906  (45.8478, 51.1187)  (23.6315, 73.3349)
```

(a) What is the interpretation of the 95% confidence interval?

(b) What is the interpretation of the 95% prediction interval?

(c) Explain how the confidence interval is related to `Fit`, `SE Fit`, and `S`.

(d) Explain how the prediction interval is related to `Fit`, `SE Fit`, and `S`.