**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 $\beta_1$.

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

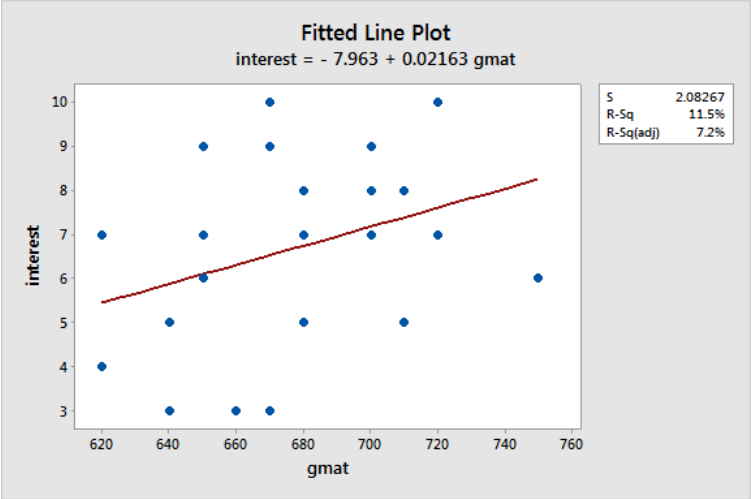
   (c) What is the meaning of a 95% confidence interval for $\beta_0$? In the context of the housing data, is this useful?

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

2. 23 students reported their GMAT scores and their interest levels in the course material (1–10). We will use this data to examine the relationship between these two variables. We fit the model

$$\text{Interest} = \beta_0 + \beta_1 \text{GMAT} + \varepsilon$$

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




Model Summary

| S | R-sq | R-sq(adj) | R-sq(pred) |
|---|---|---|---|
| 2.08267 | 11.45% | 7.24% | 0.00% |

Coefficients

| Term | Coef | SE Coef | T-Value | P-Value | VIF |
|---|---|---|---|---|---|
| Constant | -7.96 | 8.90 | -0.89 | 0.381 | |
| gmat | 0.0216 | 0.0131 | 1.65 | 0.114 | 1.00 |

Regression Equation

interest = -7.96 + 0.0216 gmat

(a) Quantify the relationship between GMAT score and interest using a 95% confidence interval. (You will need the value $t_{.025,21} = 2.080$.)

(b) Perform a hypothesis test to determine if there is a significant linear relationship between GMAT score and expected interest in the course.

## 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`.