**Introduction to Linear Regression**
COR1-GB.1305 – Statistics and Data Analysis

## Linear Regression

1. In the following scenarios, which would you consider to be predictor ($x$) and which would you consider to be response ($y$)?
   (a) Sales revenue; Advertising expenditures
   (b) Starting salary after college; Undergraduate GPA
   (c) The current month's sales; the previous month's sales
   (d) The size of an apartment; the sale price of an apartment.
   (e) A restaurant's Zagat Price rating; a restaurant's Zagat Food rating.

2. Let $y$ be the payment (in dollars) for a repair which takes $x$ hours. Suppose that

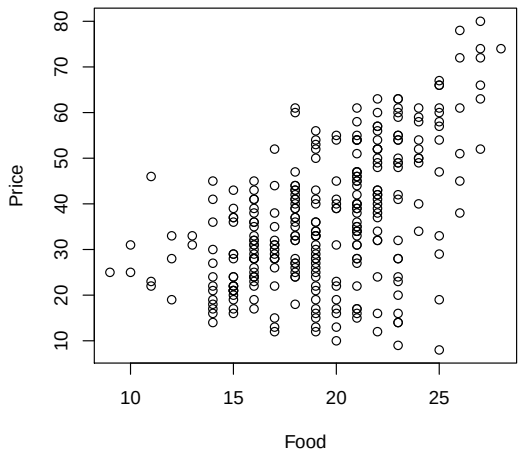
$$y = 25 + 30x.$$

   What is the interpretation of this model?

3. Consider two variables measured on 294 restaurants in the 2003 Zagat guide:

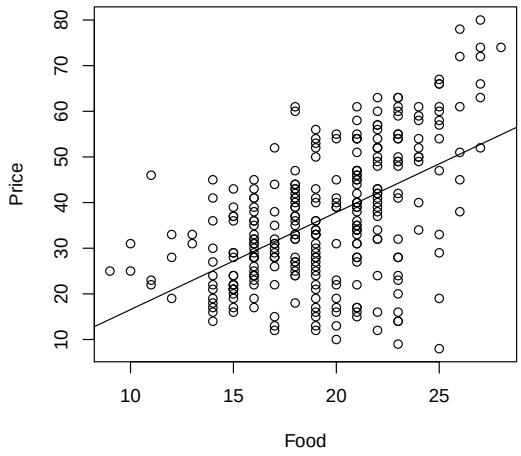
$y$ = typical dinner price, including one drink and tip ($)
$x$ = Zagat quality rating (0–30).

   Here is a scatterplot of $y$ on $x$:




   Why is an exact linear relationship inappropriate to describe the relationship between $y$ and $x$?

4. Here is the least squares regression fit to the Zagat restaurant data:




Here is the Minitab output from the fit:

```
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
```

(a) What are the estimated intercept and slope?

(b) Use the estimated regression model to estimate the average dinner price of all restaurants with a quality rating of 20.

(c) In the estimated regression model, what is the interpretation of the slope?

(d) In the estimated regression model, why doesn't the intercept have a direct interpretation?

5. Refer to the Minitab output from the previous problem, the regression analysis of the Zagat data.

(a) What is the estimated standard deviation or the error? What is the interpretation of this value?

(b) According to the estimated regression model, what is the range of typical prices for restaurants with quality ratings of 20?

(c) According to the estimated regression model, what is the range of typical prices for restaurants with quality ratings of 10?

## The Analysis of Variance Table

6. When we fit the regression model to the Zagat data with response "Price" and predictor "Food", we get the following "Analysis of Variance" table:

```
Analysis of Variance

Source          DF  Adj SS   Adj MS  F-Value  P-Value
Regression       1   17838  17838.4   113.15    0.000
  Food           1   17838  17838.4   113.15    0.000
Error          292   46034    157.7
  Lack-of-Fit   18    5394    299.7     2.02    0.009
  Pure Error   274   40640    148.3
Total          293   63873
```

   (a) Find the SSE. Explain how this value is computed.

   (b) Find the SSR. Explain how this value is computed.

   (c) Find the SST. Explain how this value is computed.

   (d) Explain how to compute $R^2$ from the ANOVA table.

   (e) Explain how to compute $s$ from the ANOVA table.