**Stat 204**
**Quiz 2 1/28/20 Name:** _____

1. 605 out of 1060 randomly selected teens aged 13 to 17 said that they had made a new friend online. Provide the relevant summary statistic using correct notation.

   The sample proportion is $\hat{p} = \frac{605}{1060} = 0.57 = 57\%$.

2. A Canadian longitudinal study examined whether giving antibiotics in infancy increased the likelihood of obesity later in life. Children were classified as having received antibiotics or not during the first year of life and as being overweight or not at 9 years old; the following table summarizes their findings:

| | Overweight | Not overweight | Total |
|---|---|---|---|
| Antibiotics | 144 | 294 | **438** |
| No antibiotics | 37 | 141 | **178** |
| Total | **181** | **435** | **616** |

   (a) Fill in the missing totals; you should add 5 numbers to the table above.

   (b) What proportion of all children in the study received antibiotics?

   The proportion is $\frac{438}{616} = 0.711 = 71.1\%$.

   (c) What proportion of all children in the study were overweight at age 9?

   The proportion is $\frac{181}{616} = 0.294 = 29.4\%$.

   (d) What proportion of those receiving antibiotics were classified as overweight at age 9? Call this $\hat{p}_A$.

   The proportion is $\hat{p}_A = \frac{144}{438} = 0.329 = 32.9\%$.

   (e) What proportion of those not receiving antibiotics were classified as overweight at age 9? Call this $\hat{p}_N$.

   The proportion is $\hat{p}_N = \frac{37}{178} = 0.208 = 20.8\%$.

   (f) What is $\hat{p}_A - \hat{p}_N$ ? What does this suggest?

   The difference is $\hat{p}_A - \hat{p}_N = 0.121 = 12.1\%$ More on interpreting this later.

3. The `Calories` column in the book's `NutritionStudy` dataset provides the daily calorie consumption for each subject in the study. Import this data into RStudio to answer the following questions.

   (a) What is the five-number summary of `Calories`?

   - Minimum = 445.2
   - First quartile ($Q_1$) = 1338.0
   - Median = 1666.8
   - Third quartile ($Q_3$) = 2100.4
   - Maximum = 6662.2

   (b) What is the IQR of `Calories`?

   The IQR for this sample is $Q_3 - Q_1 = 2100.4 - 1338.0 = 762.4$ .

   (c) What are the mean and standard deviation of `Calories`?

   The sample mean is $\bar{x} = 1796.7$ ; the sample standard deviation is $s = 680.3474$ .

   (d) Which of the following is the most likely shape of the histogram of `Calories`? why?

   Since the mean is larger than the median, the distribution is likely skewed to the right, i.e., right-tailed.