Worksheet 1

1. Often we are interested in estimating the center of a distribution. The sample mean ($\bar{x} = \sum x/n$) is sensitive to outliers and therefore we sometimes use other measures of center to estimate the center of a distribution. One such measure is the **midhinge** which is the average of Q_1 and Q_3. The quartiles break up the data into four equal quarters. Q_1 is the 25th percentile and Q_3 is the 75th percentile.
    a. Create a function that computes the midhinge. You will need to use the function `quantile()` to do this.
    b. Use your function to compute the midhinge of the numbers: 3,100,40,7,29,2,230,44,100,1200,8,15,900.
2. Suppose I wanted to use the midhinge to estimate the mean of a Poisson distribution, denoted as λ. The function `rpois(500,2)` will create a vector of 500 values that follows the Poisson distribution with $\lambda = 2$.

a. Use your function to calculate the midhinge of a sample of 500 values. Is the midhinge a good estimate of λ? Explain your reasoning.

3. The unknown parameters in simple linear regression are the slope, β_1, and the intercept, β_0. The common formula for these estimates are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})\, y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Write a function in R that computes both the intercept and the slope estimate. Your function will need to return two values. There are a few ways to do this but using the `list()` command will do this (i.e. your last line would look something like `list($\hat{\beta}_{0},\hat{\beta}_{1}$)`

4. In Denali National Park, Alaska, the wolf population is dependent on a large, strong caribou population. In this wild setting, caribou are found in very large herds. It is thought that wolves keep caribou herds strong by helping prevent over-population. Let x represent the number of fall caribou herds and y represent the late winter wolf population in the park. A random sample of recent years gives the following results:

- `x = 31, 34, 27, 25, 17, 23, 20`
- `y = 75, 85, 75, 60, 48, 60, 60`

Use your function to estimate the slope and the intercept. Then use the `lm()` function in R to check your code.