MATH 458 - Worksheet #1 Names: ____________________

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. The trimmed mean calculates the sample mean, however, deletes some of the largest or smallest outliers. The median is also a measure we use when there are outliers. Another measure of center 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. A useful function to use is *quantile*. To learn more about the quantile function you can type *?quantile* and information will appear on the right bottom window.**

   (b) **Use your function to compute the midhinge of the numbers 3,100,40,7,29,2,230,44,100,1200,8,15,900.**

   (c) **The function rpois(500,2) will create a variable that follows the Poisson distribution of length 500 with $\lambda = 2$. Suppose I wanted to use the midhinge to estimate λ. Create the variable using the function $x < -rpois(500, 2)$ and try estimating λ (we know the true value of $\lambda = 2$). Is the midhinge a good estimate of λ? Why or why not?**

   (d) **To really address the question *Is the midhinge a good estimate of λ?* we would need to repeat the random sample from the Poisson distribution with $\lambda = 2$ numerous times. Do this at least 10 times and estimate λ each time with the midhinge and also the sample mean. Comment on your simulations. Is the sample mean or the midhinge a better estimator when $\lambda = 2$. Explain.**

   (e) **Let's try a different value for the sample size. Rather than 500, let's use 25. Repeat d) but use a sample size of 25 (rpois(25,2)). Comment on the results.**

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

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

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

(a) **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($\hat{\beta}_0, \hat{\beta}_1$))**

(b) **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. You can run the following code to get the model *lm(y~x)*. Report the model below.