# Problem Set #3

*Name(s):* ___________________________________

## Q1

Use the "cars" dataset and the variable of interest, 'weight' in the openintro package (remember… in order to access such dataset you will need to use the R function, library(openintro)). Assume this dataset represents one sample from a population where the population standard deviation is *unknown*.

a) Calculate the sample statistics, $\bar{x}$ and $s$ for the sample.

b) Determine the 95% confidence interval for the true population mean using the parameter estimate $\bar{x}$. Remember all the necessary step and show all your work! (**Hint**: Find $\alpha$, $t_{\frac{\alpha}{2}}$ (use R function qt(p,df)), standard error, and MOE)

## Q2

a) Assume the population standard deviation is *known*. Determine a theoretical function that will represent the sample size ($n$) required to receive a certain margin of error. (**Hint**: Start by writing down the function for MOE)

b) If the population standard deviation is $\sigma = 5$, the level of significance is $\alpha = 0.05$, and the necessary $MOE = 2$, what is the required sample size?

## Q3

The Timken Company manufacturers ball bearings for industrial products. Use the "ballBearing" dataset and the variable of interest, 'lifeSpan' in the openintro package (remember… in order to access such dataset you will need to use the R function, library(openintro)). The chief industrial engineer claims that the average lifespan of a ball bearing produced on any given day is 7 years with *known* standard deviation, 1.2 years. A quality control specialist regularly checks this claim to ensure proper specification limits are maintained and proper information is provided to customers. On one production run (the "ballBearing" dataset), 75 ball bearings were tested until failure and the lifespace recorded. The production process wants to ensure that the average ball bearing lasts at least 7 years.

a) Determine the appropriate hypothesis test (Use the level of significane, $\alpha = 0.05$)

b) Use whatever hypothesis testing procedure (p-value, classic test statistic, confidence interval, or computational) to the test the hypothesis constructed in part a. Show all work and state conclusion.

## Q4

The 'MLB' dataset in the openintro package refers to the salaries ("salary") of each major league baseball player in 2010 (assume population). Follow the directions below:

a) Calculate the population parameters, $\mu$ and $\sigma$

b) Take a random sample of 100 players and assign it the name of "samp" using the following code:

```r
library(openintro)
```

```
## Warning: package 'openintro' was built under R version 3.2.3

## Please visit openintro.org for free statistics materials
```

```
##
## Attaching package: 'openintro'

## The following object is masked from 'package:datasets':
##
##     cars
```

```r
set.seed(1986)
samp <- sample(MLB$salary, 100)
```

c) Calculate the sample statistics, $\bar{x}$ and $s$ from the sample taken

d) Contrust the 95% confidence interval assuming the population standard deviation is *known*; Use $\sigma$

e) Construct the 95% confidence interval assuming the population standard deviation is *unknown*; Use $s$

f) Compare the two confidence intervals in terms of their range. Do both intervals "capture" the true population parameter, $\mu$?

## Q5

In statistical inference such test statistics $z$ and $t$ are used when certain assumptions about the population parameter are understood. Recall, the $z$ statistic/value is used when the population standard deviation is known and the $t$ statistic/value is used when the population standard deviation is unknown (we use the sample standard deviation, $s$ in this case). Explain the case where it would not matter what test statistic, $z$ or $t$ could be used and the resulting hypothesis test or confidence interval would provide approximately the same result. Provide an example to help argue such.