STAT 509 Section E01 Homework 4
Due in class on March 17th

---

1. A textile fiber manufacturer is investigating a new drapery yarn, which the company claims has a mean thread elongation of 12 kilograms with a standard deviation of 0.5 kilograms. The company wishes to test the hypothesis $H_0 : \mu = 12$ against $H_a : \mu \neq 12$ using a random sample of four specimens. Suppose the random sample is from a normal population.
   (a) Follow the 4 steps of conducting a hypothesis test, what is your conclusion if the sample mean $\overline{y} = 11.3$ and we use $\alpha = 0.05$?
   (b) Calculate a 95% two-sided confidence interval for μ. Does the confidence interval cover 12?

2. A manufacturing firm claims that the batteries used in their electronic games will last an average of at least 30 hours. To maintain this average, 16 randomly selected batteries are tested each month. Suppose the battery lifetime follows a normal distribution.

   ```
   20,25,21,28,21,30,23,27,26,26,28,31,26,32,33,35
   ```

   (a) State the null and alternative hypotheses.
   (b) Calculate the appropriate test statistic.
   (c) What is the p-value of the test? What is your conclusion based on p-value ($\alpha = 0.01$).
   (d) If we want to draw the conclusion based on a 99% confidence interval, what is the proper form of the confidence interval? What is your conclusion based on the confidence interval approach?
   (e) Redo the above analysis using t.test in R, does your calculation by hand match the R output? Attach your R output.

3. An industrial engineer has randomly chosen 20 experienced operators and timed them as they assembled a given piece of equipment. The times are in minutes:

   ```
   30 33 31 24 28 27 33 28 32 30 27 30 28 32 36 30 29 26 31 23
   ```

   The industrial engineer are interested in whether the mean assembly time equals to 28 or not. Conduct a complete data analysis in R. Attach your R code and output
   Use this code to read in data:

   ```
   time<-c(30,33,31,24,28,27,33,28,32,30,27,30,28,32,36,30,29,26,31,23)
   ```

4. Inexperienced data analysts often erroneously place too much faith in qq plots when assessing whether a distribution adequately represents a data set (especially when the sample size is small).The purpose of this problem is to illustrate to you the dangers that can arise. In this problem, you will use R to simulate the process of drawing repeated random samples from a given population distribution and then creating normal probability plots (Q-Q plots). Follow the code provided
   (a) Generate your own data and create a qq plot for each sample using this R code:

```
# create 2 by 2 figure
par(mfrow = c(2,2))
B = 4
n = 10
# create matrix to hold all data
data = matrix(round(rnorm(n*B,0,1),4), nrow = B, ncol = n)
# this creates a qq plot for each sample of data
for (i in 1:B){
   qqnorm(data[i,],pch=16,main="")
   qqline(data[i,])
}
```

mark the qq plot that appears to violate the normal assumption the most. Note: In theory, all of these plots should display perfect linearity! Why? Because we are generating the data from a normal distribution! **Therefore, even when we create normal qq plots with normally distributed data, we can get plots that don't look perfectly linear.** This is a byproduct of sampling variability. This is why you don't want to rush to discount a distribution as being plausible based on a single plot, especially when the sample size n is small (like $n = 10$).

(b) Increase your sample size to $n = 100$ and repeat. What happens? What if $n = 1000$? Just change n in the R code and re-run.

(c) Take $n = 100$, replace

```
data = matrix(round(rnorm(n*B,0,1),4), nrow = B, ncol = n)
```

with

```
data = matrix(round(rexp(n*B,1),4), nrow = B, ncol = n)
```

and re-run. By doing this, you are changing the underlying population distribution from $\mathcal{N}(0, 1)$ to exponential(1). What do these normal qq plots look like? Are you surprised?

5. A study to characterize the physical behavior of steel concrete beams under periodic load was carried out at the National Institute of Standards and Technology. The response variable is deflection (from rest point) of the steel concrete beam. The variance should be less than 70. A sample of fifteen beams gave the following summary statistics:

Variable	n	mean	variance	std. Dev.
size	15	569.4	8.829	7.670

(a) Estimate the variance with a 95% confidence interval.

(b) Perform a hypothesis test that tests the alternative hypothesis that the variance is less than 70. Be sure to include the hypotheses, calculation of the test statistic, calculation of the p-value, and statement of conclusion.