# Stat 509: Statistics for Engineers
## Homework Assignment 7

Make sure to include the R code and graphs used in your analyses and provide support for the validity of all underlying assumptions.

1. You are in charge of monitoring errors in the production of a certain part. Let $X$ represent the absolute difference in mm between the length of a part and its desired length. You have collected data on these differences in 50 randomly selected parts. You wish to use this data to make inferences about the average population-level error in the length of a part. Read the *part* data into R in order to answer the following questions:

   (a) Construct <u>and</u> interpret a 99% confidence interval for the true mean error in the length of a part.

   ```
   part=read.csv("part.csv")[,-1]
   t.test(part,conf.level=.99)
   qqnorm(part)
   lines(c(-1000^2,1000^2),c(-1000^2,1000^2),col="red")
   ```

   We are 99% confident that the average error in the length of a part is between 0.58 and 0.9 mm.

   (b) Based on your interval, are you confident that the mean error does not exceed 1 mm?

Yes. The entire confidence interval is below 1. Thus, we are 99% confident that the mean error is less than 1 mm.

We have a random sample and the mostly linear-trend of the qqplot provides evidence that the populationlevel error in the length of a part is not grossly non-normal. In addition, we have a sufficiently large sample ($\approx n > 30$). Thus, the assumptions of this procedure are satisfied.

2. You are in charge of monitoring the production of nuts and bolts. Let $X$ represent the diameter in mm of a nut. You have collected data on these diameters in 2300 randomly selected nuts. You wish to use this data to understand the variability in deez nuts. Read the *diameter* data into R in order to answer the following questions:

   (a) Construct <u>and</u> interpret a 90% confidence interval for the variance in the diameter of a nut.

   ```
   diameter=read.csv("diameter.csv")[,-1]
   one.sample.var(diameter,.9)
   qqnorm(diameter)
   ```

   We are 90% confident that the variance of the diameter in a nut is between 0.61 and 0.67 mm2 .

   (b) If the standard deviation in the diameter is more than 0.75 mm the machine must be re-calibrated. Based on your interval, does the machine need recalibration?

A confidence for the standard deviation can be computed as the square-root of the confidence interval for the variance.

(0.78, 0.82)

Yes. We have evidence that the machine needs recalibration because the entire confidence interval for the standard deviation of the diameter of a part is above 0.75 mm.

We have a random sample and the linearity of the qqplot provides support for the normality of the nut diameters. Thus, the assumptions of this procedure are satisfied.

3. You have collected data on 40 earthquakes. You have found that 8 of the 40 earthquakes are above a 3.0. Construct <u>and</u> interpret a 98% confidence interval for the proportion of earthquakes that are above a 3.0 and use it to determine if we have evidence that at least 10% of earthquakes are above a 3.0.

```
prop.test(8,40,conf.level=.98)
```

We are 98% confident that the true proportion of earthquakes that are above a 3.0 is between 0.08 and 0.39.

Because 0.1 is within the interval we do not have evidence that at least 10% of earthquakes are above a 3.0. Inconclusive.

We have at least 5 successes and failures (8 and 32 respectively), so assuming we have a random sample, our assumptions are satisfied.

4. You have been tasked with collecting random soil samples at a prospective build site to ensure the location is suitable. Specifically, you will be required to compute a 99.9% confidence interval for the average soil density with a margin of error no greater than 0.3 g/cm$^3$. It is estimated that the standard deviation of the soil density is 0.38 g/cm$^3$. What is the minimum number of samples that you must collect in order to ensure that the

confidence interval is constructed to the desired specifications?

First we must realize that we are computing a confidence interval for a single population mean $\mu$. Also, as we have a 99.9% confidence level, $\alpha = 0.001$. We seek a sample size n such that t1−0.001/2,n−1 0.38 p n ≤ 0.3. Simple trial-and-error yields:

```
n=23

qt(.9995,n-1)*.38/sqrt(n)

[1] 0.3004713

n=24

qt(.9995,n-1)*.38/sqrt(n)

[1] 0.2922442
```

Thus, a sample size of n=24 is the smallest such that the margin of error should be no greater than 0.3.

5. You are in charge of selecting a new battery system to be implemented. You must choose between class A batteries and class B batteries. In order to make a decision, you collect data on the lifetimes (years) of 50 batteries of each class. Read the *battery* data into R in order to answer the following questions:

    (a) Construct <u>and</u> interpret a 96% confidence interval for the difference in the mean lifetimes of the two battery classes. Provide support for your choice of interval.

```
battery=read.csv("battery.csv")[,-1]

attach(battery)

sd(A)

sd(B)

t.test(A,B,conf.level=.96)

qqnorm(A)

qqnorm(B)
```

First we must realize that we are computing a confidence interval for $\mu 1 - \mu 2$. We then need to realize that we have independent samples (which rules out method 3). Next, we must determine if we can assume the population variances are equal. With standard deviations of 2.12 and 1.13 respectively, it does not seem safe to assume the variances are equal. Thus, we proceed with method 2.

We are 96% confident that the difference between the mean lifetime of class A batteries and the mean lifetime of class B batteries is between 1.27 and 2.69 years.

Assuming we have random samples, it remains only to verify the assumption of normality, which seems reasonable based on the linearity of the qqplots.

    (b) Construct <u>and</u> interpret a 96% confidence interval for the ratio of the variances in the lifetimes of the two battery classes.

```
var.test(A,B,conf.level=.96)
```

We are 96% confident that the ratio of the variance in the lifetimes of A batteries to the variance in the lifetimes of B batteries is between 1.93 and 6.34.

We have already verified independent samples and normality of the lifetimes for each class of battery. Again, we assume that these are random samples.

    (c) Based on all of your analyses, which class of batteries do you choose? Provide support for your decision.

The decision in this case is not straight forward. Class A batteries last 1.3 to 2.7 years longer on average than class B batteries, but also have a variance which is 1.9 to 6.3 times greater than the variance of class B batteries. Thus, class A batteries last longer on average, but are less consistent. Depending on the application, it may be that this increased variability in the lifetime is unacceptable.

6. You have collected data on the processing speed (GHz) for 20 computers, both before and after a newly designed chip was implemented. You then take the difference in processing speed before the chip and after chip for each computer. Read the *computer* data into R in order to answer the following questions:

    (a) Construct <u>and</u> interpret an 80% confidence interval for the difference between the average processing speed before and after the implementation of the new chip. Provide support for your choice of interval.

```
computer=read.csv("computer.csv")[,-1]

t.test(computer,conf.level=.8)

qqnorm(computer)

lines(c(-1000^2,1000^2),c(-1000^2,1000^2),col="red")
```

We must realize that we are comparing two means using a matched-pairs design. Also, a one-sample ttest on differences is the same as a two-sample t-test on matched pairs data.

We are 80% confident that the difference between the average processing speed before and after the implementation of the new chip is between -0.57 and -0.31 GHz.

(b) Do you have evidence that the chip improved processing speed?

Yes. The interval given shows that the average processing speed increases by about 0.2 GHz after implementing the chip.

Because the entire confidence interval is negative, we have evidence that the average processing speed is higher after the implementation of the chip (second group) and thus the chip does help improve processing speed.

It is obvious that normality is grossly violated. This, combined with a small sample size, make any inferences drawn from this interval questionable. (This is an example of what a sample from an exponential distribution looks like.)

7. You are comparing the quality of a certain part that is produced by two different manufacturers. You sample 1000 parts from each manufacturer. You find 4 defective parts from the first manufacturer and 20 defective parts from the second manufacturer.

(a) Construct <u>and</u> interpret a 98% confidence interval for the difference between the proportions of defective parts for the two manufacturers.

```
prop.test(c(4,20),c(1000,1000),conf.level=.98)
```

We are 98% confident that the difference between the proportion of defective parts at the first manufacturer and the proportion of defective parts at the second manufacturer is between -0.028297527 and -0.003702473.

(b) Do we have evidence that the quality of one manufacturer’s parts is superior?

Because the interval is completely negative, we have evidence that the second manufacturer has a higher defect rate and thus we have evidence that the quality of the first manufacturer’s parts is superior.

Contextually, it appears we have independent samples. We also rely on having random samples. We do not have at least 5 successes and failures for each group; however, the prop.test() function will actually alert us when this assumption is violated in such a way as to grossly affect the reliability of the results.