# Stat 321 - Day 38
## Introduction to Interval Estimation

Point estimates of a parameter value are of limited usefulness without information about the uncertainty inherent in the estimate. We can correct this limitation by producing an interval of values that we believe is likely to contain the unknown parameter value. The key to this procedure is knowledge of the sampling distribution of the statistic.

**Example: Potato Chip Assembly Line (cont.)**
Suppose that the weights of bags of potato chips coming off an assembly line have some unknown mean $\mu$ and standard deviation $\sigma$=0.4 ounces. Suppose further that you take a random sample of $n$=100 bags. The sample mean $\bar{X}$ is a reasonable point estimator of $\mu$. We know that $E(\bar{X}) = \mu$, so $\bar{X}$ is unbiased, and we know that SD($\bar{X}$)=$\sigma/\sqrt{n}$ = 0.04. We also know from the Central Limit Theorem that $\bar{X}$ follows a normal distribution.

(a) Determine the probability that $\bar{X}$ falls within .08 ounces of the population mean $\mu$.

(b) Determine the probability that the interval from $\bar{X}$ -.08 to $\bar{X}$ +.08 captures the population mean $\mu$. [*Hint*: How does this calculation differ from the previous one, if at all?]

(c) If we were to take a very large number of random samples from this population, and if we were to construct the interval $\bar{X} \pm .08$ from each sample, then in the long run what proportion of those intervals would succeed in capturing the population mean $\mu$?

(d) Answer c) for the interval $\bar{X} \pm .066$.

These questions reveal the fundamental idea behind the idea of statistical *confidence*: We refer to the $\bar{X} \pm .08$ as an approximate *95% confidence interval* because if we were to repeatedly take random samples from the population and construct intervals in this manner, then in the long run 95% of those intervals would succeed at capturing the value of the population mean $\mu$.

This multiplier value of "2" is only an approximation for producing 95% confidence. The exact value is 1.96, because P(-1.96<$Z$<1.96)=.95, where $Z$ denotes a standard normal distribution. You can choose the *confidence level* to be whatever you want and find the corresponding multiplier. In general, a 100(1-$\alpha$)% confidence interval for a population mean $\mu$ is given by: $\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$, where $z_{\alpha/2}$ the *critical value* from the standard normal distribution such that P(Z>$z_{\alpha/2}$)=$\alpha$/2. In order for this procedure to be valid, we need to have a random sample and we need either a large sample size ($n \geq 30$ as a rule-of-thumb) or a normally distributed population.

(e) Determine the critical value for a 99% confidence interval, and then indicate what that interval would turn out to be in the potato chip example.

(f) Suppose that you take a random sample of 100 potato chip bags and find the sample mean weight to be 12.14 ounces. Determine a 95% confidence interval for the population mean $\mu$.

(g) Can you be certain that this interval contains the actual value of the population mean $\mu$?

(h) Explain the basis for your "confidence" that this interval contains the population mean $\mu$.

(i) What happens to the width of a confidence interval as you increase the confidence level? Explain why this makes sense.

(j) What happens to the width of a confidence interval as you increase the sample size? Explain why this makes sense.

(k) Suppose that you want a 95% confidence interval to estimate $\mu$ (the population mean of the potato chip bag weights) to within $\pm$.04 ounces. Determine the sample size necessary. B y what factor is this sample size larger than the original one?

**Example: Mothers' Ages**

The file `agemom.mtw` contains data gathered as part of the 1998 General Social Survey. Each mother in the survey reported their age when they gave birth to their first child. Consider the data on 1199 mothers in C1 as the population of interest.

(l) Create a histogram of the population data. Describe the distribution. Does it look roughly normal? Calculate the mean and standard deviation. Record these values with their appropriate symbols.

(m) Use Minitab to take a sample of 35 mothers from the population in C1:

```
MTB> sample 35 c1 c3
```

Then use Minitab to construct a 90% confidence interval for the population mean:

```
MTB> let c4(1) = mean(c3)-1.645*4.885/sqrt(35)
MTB> let c4(2) = mean(c3)+1.645*4.885/sqrt(35)
```

[*Note*: The two values in C4 provide the interval's lower and upper endpoints.]

Does this interval succeed in capturing the population mean $\mu$ (which you recorded in a)?

(n) Copy and paste these three commands until you have taken 25 samples. Keep a tally of whether or not each interval contains the population mean.

(o) Combine your results with the rest of the class. What proportion of these samples led to an interval that did capture the population mean? Is this value close to what you expected? Explain.