**Introduction:**
I hope you feel comfortable with chapter 5. The knowledge of chapter 5 is essential to this and the chapters after this one.
In this chapter you will learn how to find confidence intervals for different parameters (μ, p, σ).
First, you should know what we mean by confidence interval. I will explain the situation:
Suppose you take a random sample of size n from RCC students and you want to gather some information about the height of all the students at RCC.
One piece of information you can obtain from this sample is the sample mean (x̄). Having the x̄ will give a very good idea about μ. We say that x̄ estimates μ. For that reason we call x̄ a point estimator (for μ). In general, x̄ is either bigger or smaller than μ, by how much we don’t really know (the difference between x̄ and μ is called the error of approximation, E). To overcome this uncertainty, we will find an interval for μ with some certainty (confidence). We will also find confidence interval for other parameters. The concept is the same.

**6.1-Confidence Interval for μ:**
In section one we will find confidence interval for the population mean when one of the following conditions is satisfied: Either n ≥ 30 OR σ is given (known).
The formula is: x̄ − E < μ < x̄ + E, this should be obvious from what we talked above.
The formula for E, the maximum error of estimate, is: E = z_{α/2} \frac{σ}{\sqrt{n}}. In the text instead of z_{α/2}2 they have cz_{α/2}. I prefer z_{α/2}, because it is a more meaningful notation. You can choose the notation that you like.
Note:
1. When n ≥ 30 and s is given, use s for σ in the formula.
2. Level of confidence (1 − α) tells you the chance (probability) that μ is actually in the interval you have found.
3. α Is the probability that μ is outside of the interval you have found.
Q. What is the effect on the confidence interval if we make α very small?
I am going to give you one example for confidence interval. The solution to other questions is similar.

**Example 1:** From a random sample of size 64 from RCC students we found that the mean height was 180 cm with a standard deviation of 10 cm. Find a 95% CI for the height of all students at RCC.
**Solution:**
Step 1: I like to summarize the information symbolically.
\[ x̄ = 180 \]
\[ s = 10 \]
\[ 1 − α = .95 \]
Step 2: Write down the formulas: \[ E = z_{α/2} \frac{σ}{\sqrt{n}} \]
\[ x̄ − E < μ < x̄ + E. \]
Everything is directly given except \( z_{α/2} \). I will explain what it is and how to find it.
\( z_{α/2} \) or \( z_{} \) is a point on the z axis whose area to the right is \( α/2 \).
For this example, $\alpha = .05$, and $\frac{\alpha}{2} = .025$. In order to find $z_c$, you must go to the z-table and hunt for .025 and then find the corresponding $z$ value. The corresponding $z$ value is 1.96 (note that this number is always positive).

Step 3: Find E: $E = 1.96 \cdot \frac{10}{\sqrt{64}} = 2.45$ (Since n is large, we could use $s$ for $\sigma$.)

Step 4: $180 - 2.45 < \mu < 180 + 2.45 \Rightarrow 177.55 < \mu < 182.45$.

So, we are 95% sure that the mean height students at RCC is between the above range.

**Required Sample Size** - Taking a sample is expensive and time consuming, therefore before we take a sample we like to know the smallest sample size that we can take with a given confidence and error.

Example: Suppose I want to have a 90% confidence interval for the mean weight of RCC students. From a previous study, I know that the standard deviation for the weight of RCC students is 10 pounds. What is the required sample size? Let the maximum error of estimate be 2 pounds.

Solution: $\alpha = .1$, $z_{.05} = 1.645 = z_c$, $E = \sigma = 10$

The formula is: $n = \left(\frac{z_c \sigma}{E}\right)^2 = \left(\frac{1.645 \cdot 10}{2}\right)^2 = 67.65$, so $n = 68$. Note: You must always round up the value you get from the formula.

**6.2 - Confidence Interval for $\mu$ ($\sigma$ is unknown and $n < 30$)**

In this section we like to have a confidence interval for the population mean when the sample is small and the population standard deviation is not known.

You have seen the z-distribution, now you will become familiar with a distribution, which is called the t-distribution and in many ways it is similar to the z-distribution.

Characteristics of the t-distribution, which are similar to the z-distribution:
1. The distribution looks like a bell.
2. The area under the curve is exactly 1.
3. $\mu = 0$.
4. The distribution is symmetric with respect to the mean.

Characteristics of the t-distribution, which are different from the z-distribution:
1. $\sigma > 1$, But as $n$ approaches 30, then $\sigma$ will approach 1.
2. The shape of the distribution depends of the degrees of freedom ($df = n - 1$). (The smaller the sample size the wider the distribution will be.)

Example: Suppose we want a 95% confidence interval for the population mean weight of RCC students. We took a random sample 16 RCC students and found the sample mean and standard deviation of 140 and 5 pounds, respectively.

Solution: Summarize the problem: $df = n - 1 = 16 - 1 = 15$, $\bar{x} = 140$, $s = 5$, $1 - \alpha = .95$.

The formula is identical to the previous section: $\bar{x} - E < \mu < \bar{x} + E$. So we have to find E first:

$$E = t_c \cdot \frac{s}{\sqrt{n}}$$

which is very similar to section 6.1. In order to find $t_c$, you must go to the t-table (it is right after the z-table). The first column gives you the degrees of freedom ($df = 15$), and from the first row you can select the confidence level (.95). Using these two numbers, you will find
\[ t_c = 2.131. \] Now we can find the error: \[ E = t_c \frac{s}{\sqrt{n}} = 2.131 \left( \frac{5}{\sqrt{16}} \right) = 2.66 \quad \text{and} \]

\[ 140 - 2.66 < \mu < 140 + 2.66 \]

\[ 137.34 < \mu < 142.66 \]

Note that there are situations that we cannot have large sample sizes. In those cases, given that the population is normally distributed, the t-distribution is very useful.

\[ 6.3 \text{ -- Confidence Intervals for Population Proportions (p)} \]

So far you have seen confidence intervals for \( \mu \). In this section you will learn how to create confidence intervals for \( p \). This \( p \) is the same \( p \) that you worked with in binomial distributions. In this section we will use the z-table.

**Example:** Suppose you like to have some ideas about the proportion of RCC students who smoke cigarettes. You randomly select a sample of size 50 from RCC students. In your data you noticed that 5 of these 50 students are smokers. First determine the point estimator for the population proportion of RCC students who smoke cigarettes, then find a 90% confidence interval for \( p \).

In order to find a point estimator for \( p \) we must divide the number smokers by the sample size:

\[ \hat{p} = \frac{x}{n} = \frac{5}{50} = .1. \]

The formula for the confidence interval is: \( \hat{p} - E < p < \hat{p} + E \), where \( E = z_c \sqrt{\frac{pq}{n}} \).

\( E \) and \( z_c \) are the same as before, and \( \hat{q} = 1 - \hat{p} = 1 - .1 = .9 \). Looking at the z-table you find \( z_c = 1.645 \). We can find \( E = 1.645 \sqrt{\frac{(1)(.9)}{50}} = .07 \). \( .1 - .07 < p < .1 + .07 \Rightarrow .03 < p < .17 \).

With this experiment we are 90% sure that the proportion of RCC students who smoke cigarettes is between .03-.17.

Sometimes before we take a sample to investigate a population proportion we like to know what the smallest sample would meet the maximum error. You need to know 2 things: The level of confidence and the maximum error you can tolerate.

**Example:** Suppose you want to have 95% confidence and the maximum error you can have is 4%.

Formula: \( n = \hat{p} \hat{q} \left( \frac{z_c}{E} \right)^2 \), \( E = .04 \), \( z_c = 1.96 \) and since we don’t have any idea what \( \hat{p} \) is we will use .5 for it. \( n = (.5)(.5) \left( \frac{1.96}{.04} \right)^2 = 600.25 \). Since we can’t have a decimal number for \( n \), we choose \( n = 601 \).

Note 1: Do not round off for \( n \), always round up.

Note 2: The selection of .5 for \( \hat{p} \) will give the largest possible \( n \) for the prescribed level of confidence.
6.4-Confidence Intervals for Variance ($\sigma^2$) and Standard Deviation ($\sigma$)

In this section we would like to find confidence interval for a population variance. First we find a point estimator for $\sigma^2$. By now you should know that the best estimator for the population variance ($\sigma^2$) is the variance of a sample ($s^2$) taken from that population.

So far you have seen two distributions, the t and z distributions. In this section you will learn about a new distribution called the $\chi^2$ (pronounced chi or ky squared) distribution.

Let me explain what this new random variable is: Suppose you randomly select samples of size $n$ from a population whose random variable is normally distributed with a variance of $\sigma^2$. If you compute $s^2$ for each sample, then the numbers (random variable) you obtain from the following fraction will make up a $\chi^2$ distribution: $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$.

Note: You should remember this formula for the upcoming sections.

From this fraction you can see that $\chi^2$ is never negative. Other characteristics of $\chi^2$ distributions are:

- They are all skewed to the right.
- The shape of the distribution depends on the sample size.
- The degrees of freedom is $df = n - 1$.
- The larger the sample size is the less skewed the distribution will be.
- Like any other probability function, with a continuous random variable, the area under the curve is exactly one.

The way we use the $\chi^2$ table is very much like the t-table. Since the distribution is not symmetric, we have to find two critical values. They are: $\chi^2_L$ and $\chi^2_R$.

Example: Suppose $n=20$ and $\alpha=.05$. Find the left and right critical values.

The table is on page A21. Look at the figure that has both tails shaded. $\alpha = .05$, $\frac{\alpha}{2} = .025$ and $df = 19$. For the $\chi^2_L$, you must look under the column that has .975 and $df = 19$. You should see 8.907. For the $\chi^2_R$, you must look under the column that has .025 with the same degrees of freedom. You will find 32.8532.

**Example:** This is question 12 on page 307.

Summarize the problem:

1. $n=26$, $s=150$, $1-\alpha = .95$, $df = 25$ .
2. Find the critical values: If you follow the guidelines above you will find that $\chi^2_L = 13.120$, $\chi^2_R = 40.646$.

Note: When I was a student, we had better notations for these critical values. For the above critical values we would write $\chi^2(25,.975)=13.120$, $\chi^2(25,.025)=40.646$. What you see inside the parenthesis are degrees of freedom and then the area to the right of the critical value. Use the one you like.

3. The formula is: $\frac{(n-1)s^2}{\chi^2_R} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_L} \quad \Rightarrow \quad \frac{(25)150^2}{40.646} < \sigma^2 < \frac{(25)150^2}{13.120} \quad \Rightarrow$

   $13839 < \sigma^2 < 42873.48$. To find a confidence interval for $\sigma$, all you have to do is to take the square root of these numbers: $117.64 < \sigma < 207.06$