**Chapter 8 – Confidence Interval – One Population**

---

## Section 8.1 - Estimating a Population Mean: σ Known

**Assumptions:**

- **The sample is a simple random sample**
- **The value of the population standard deviation σ is known.**
- **Either or both of these conditions are satisfied:**
    **i) The population is normally distributed, or**
    **ii) n ≥ 30 (The sample has 30 or more values)**

**Procedure for Constructing a Confidence Interval for μ (with Known σ)**

1. **Verify that the required assumptions are satisfied.**
2. **Find the critical value $z_c$ (from table 5).**

3. **Evaluate the margin of error E.     (E = $z_c \dfrac{\sigma}{\sqrt{n}}$ )**

4. **Then using E and the sample mean the confidence interval is:**

$$\overline{x} - E < \mu < \overline{x} + E \text{ or } (\overline{x} - E, \overline{x} + E)$$

**Using the TI-83 to Construct Confidence Intervals for μ:**
**STAT>>TESTS  select 7:ZInterval**

**Round-off Rule for Confidence Intervals used to Estimate μ:**

**a) If original data is given: use one more decimal place than original values.**
**b)  If you are given summary statistics from a data set, use the same number of**
     **decimal places used for the sample mean.**

## Section 8.1 -Estimating a Population Mean: σ Known

1) A simple random sample of size n is drawn from a population whose population standard deviation, σ, is known to be 3.8. The sample mean, x-bar, is determined to be 59.2.
a) Compute the 90% confidence interval about μ if the sample size, n, is 45.

$90\% \, CI \to z_c = 1.645$

$n = 45$

$\sigma = 3.8$

$\bar{x} = 59.2$

$\bar{x} \pm z_c * \dfrac{\sigma}{\sqrt{n}}$

$59.2 \pm 1.645 * \dfrac{3.8}{\sqrt{45}}$

**59.2 ± .9318** → **(58.27, 60.13)**

$58.27 < \mu < 60.13$

**For calculator feature use STAT, arrow to TESTS, and select 7:ZInterval, select Stats enter the required information, and CALCULATE**

b) Compute the 90% confidence interval about μ if the sample size, n, is 55. How does increasing the sample size affect the margin of error E? How does it affect the length of the interval?

$90\% \, CI \to z_c = 1.645$

$n = 55$

$\sigma = 3.8$

$\bar{x} = 59.2$

$\bar{x} \pm z_c * \dfrac{\sigma}{\sqrt{n}}$

$59.2 \pm 1.645 * \dfrac{3.8}{\sqrt{55}}$

**59.2 ± .8429** → **(58.357, 60.043)**

**This is a shorter interval than the one in part (a). It's more precise. At the same confidence level, a larger sample size produces a more precise (shorter) interval**

**For calculator feature use STAT, arrow to TESTS, and select 7:ZInterval, select Stats enter the required information, and CALCULATE**

c) Compute the 98% confidence interval about μ if the sample size, n, is 45. Compare the results to those obtained in part (a). How does increasing the level of confidence affect the size of the margin of error, E? How does it affect the length of the interval?

$98\% \, CI \to z_c = 2.33$ (Higher confidence level)

$n = 45$

$\sigma = 3.8$

$\bar{x} = 59.2$

$\bar{x} \pm z_c * \dfrac{\sigma}{\sqrt{n}}$

$59.2 \pm 2.33 * \dfrac{3.8}{\sqrt{45}}$

**59.2 ± 1.3199** → **(57.88, 60.52)**

**This is longer (less precise) than the interval produced in part (a). A higher confidence level produces a longer interval. With a longer interval we are now more confident that the interval contains the population mean.**

Use calculator feature to check answer: **Z-Interval**

d) Can we compute a confidence interval about μ based on the information given if the sample size is n = 15? **No.** Why? **The sample size is small.** If the sample size is 15, what must be true regarding the population from which the sample was drawn? **When we have a small sample size, the population should be normally distributed in order to be able to use this method to construct a confidence interval.**

**<u>Confidence Intervals Interpretations</u>**

Refer to the first problem from the previous page.

Let's assume thousands of students took this test and we are trying to estimate the mean score (mu) of all students who took the test. In order to estimate mu we select a simple random sample of 45 students who took the the test and find that the mean score of those 45 students is x-bar = 59.2. Let's also assume that the standard deviation of the population (all thousands of test scores) is known to be σ = 3.8.

You must be careful about what you say about confidence intervals.

Remember that a confidence interval is an ESTIMATE for the UNKNOWN parameter mu, which is the mean score of all students who took the test.

We are trying to estimate mu by selecting a sample. Our sample yields an x-bar of 59.2. This is OUR POINT ESTIMATE, and we provide a range of PLAUSIBLE values for mu by constructing a 90% confidence interval around x-bar = 59.2. Our interval estimate happens to be **(58.27, 60.13)**

**Remember that different samples will produce different x bars, and hence, different intervals. The <u>confidence level (90%) refers to the accuracy of the method used in constructing the interval</u>.** <u>We are expecting about 90% of random samples to yield an interval that captures the true mean, which means that about 10% of the intervals constructed will fail to capture the true value of the parameter mu.</u> We will never know whether the interval constructed contains the mean mu or not, but we are in this case 90% confident that it does.

**<u>We should say:</u>**
- I am 90% confident that the interval from 58.27 to 60.13 contains the mean score of all the students who took the test.
- It's also OK to say something less formal: "<u>I'm 90% confident</u> that the mean score of all students who took the test is between 58.27 and 60.13.
- About 90% of the intervals constructed with this method will capture the mean score of all students who took the test.
- If we were to select 100 random samples of size 45 and construct a confidence interval about each of the sample means, approximately 90 of the 100 intervals would contain the population mean mu.

**1) Don't suggest that the parameter varies.**
A statement like "There is a 90% chance that the true mean score of all students who took the test is between 58.27 and 60.13" is just wrong. This sounds as though you think the population mean wanders around and sometimes happens to fall between 58.27 and 60.13.
When you interpret a confidence interval, make it clear that you know **that the population parameter mu is fixed and that it is the interval that varies from sample to sample.**

**2) Don't say** "90% of the test scores are between 58.27 and 60.13". The confidence interval is about the MEAN scores not about the INDIVIDUAL scores.

**3) Don's say**, "The mean of all test scores is between 58.27 and 60.13,   90% of the time" That implies that the true mean varies, when in fact it is the confidence interval that would have been different had we gotten a different sample.

**4) Don't be certain about the parameter.**
Saying "The mean score of all students who took the test is between 58.27 and 60.13 asserts that the population mean cannot be outside that interval. Of course, we can't be absolutely certain of that. (Just 90% confident on this case)
It's necessary to add the statement "With 90% confidence, we can say that....."

**Chapter 8 – Confidence Intervals – One Population**

---

### Section 8.2 - Estimating a Population Mean:  σ Not Known

**Conditions for Using the Student t Distribution**
- **σ is unknown, (if σ is known we use z);**
- **The sample is a simple random sample**
- **Either or both of these conditions is satisfied:**
  - **i) The population is normally distributed, or**
  - **ii) n ≥ 30 (The sample has 30 or more values)**

**Confidence Interval for the Estimate of μ (With σ Not Known)**

1. **Verify that the required assumptions are satisfied.**
2. **Find the critical value $t_c$ from table 6 (n -1 degrees of freedom).**

3. **Evaluate the margin of error E.      (E = $t_c \dfrac{s}{\sqrt{n}}$  )**

4. **Then using E and the sample mean the confidence interval is:**
$$\overline{x} - E < \mu < \overline{x} + E \qquad \textbf{or} \qquad (\overline{x} - E, \overline{x} + E)$$

### Using the TI-83 to Construct a Confidence Interval for Estimating μ, (σ Not Known)
**STAT>>TESTS  select 8:TInterval**

**Round-off Rule for Confidence Intervals used to Estimate μ:**

**a) If original data is given: use one more decimal place than original values.**
**b)  If you are given summary statistics from a data set, use the same number of**
     **decimal places used for the sample mean.**

**Section 8.2 -Estimating a Population Mean: σ Not Known**

2) A simple random sample of size n is drawn from a population. If that sample has a mean of 59.2 and a standard deviation of 3.8,
a) Compute the 90% confidence interval about μ if the sample size, n, is 45.
$90\% CI \rightarrow t_c = 1.684$  (**Use the closest degrees of freedom that is smaller than 45**)

$$\bar{x} \pm t_c * \frac{s}{\sqrt{n}}$$

$$59.2 \pm 1.684 * \frac{3.8}{\sqrt{45}}$$

$59.2 \pm .9539 \quad \rightarrow \quad$ **(58.25, 60.15)**

**For calculator feature use STAT, arrow to TESTS, and select 8:TInterval, select Stats enter the required information, and CALCULATE**

b) Compare with your answer to part 1-a. Which interval is longer? Which interval is more precise?

**Results from part (c):**        $59.2 \pm .9539 \quad \rightarrow \quad$ **(58.25, 60.15)**

**Results from part (a):**        $59.2 \pm .9318 \quad \rightarrow \quad$ **(58.27, 60.13)**

**The z interval is smaller, more precise. When we don't have the standard deviation of the population, σ, we use the standard deviation of the sample, s. In this case we use the t-distribution instead of the z-distribution. Corresponding t-intervals are a little longer than z-intervals.**

Finding the Point Estimate and the Margin of Error from a Confidence Interval

3) We are 95% confident that the interval from 98.08°F to 98.32°F actually contains the true mean value of the body temperatures of all healthy adults.
What is the point estimate in this case? What is the margin of error?

$$\bar{x} = \frac{98.32 + 98.08}{2} = 98.2$$

$$E = \frac{98.32 - 98.08}{2} = .12$$

## Section 8.2 -Estimating a Population Mean: σ Not Known

4) In order to correctly diagnose the disorder of hydrocephalus, a pediatrician investigates head circumferences of two month old babies. Use the sample MHED, augment it with the sample FHED. This will give you a new sample of size 100. Use this sample to estimate the head circumference of all two months old babies.

To augment a List we'll do the following:
First: STAT, Edit, go UP and to the RIGHT until you reach a list with no name. Name it BABY.
Second: Stay in the editor, press 2$^{nd}$ STAT [LIST], arrow right to OPS, go down and select 9:augment( by pressing enter, select FHED, press "comma", then select MHED, and press ENTER. The BABY list should fill up with 100 numbers  The statement should read augment(FHED, MHED))

   a)  What is the point estimate?

$$\overline{x} = 40.573$$

   b)  Verify that the requirements for constructing a confidence interval about x-bar are satisfied.

**Since the sample size is large, the distribution of sample means is approximately normally distributed. (Central Limit theorem, section 7.2)**

   c)  Construct a 99% confidence interval estimate for the head circumference of all two months old babies.  (Are you using z or t? Why?)

$n = 100$

$s = 1.649$

$\overline{x} = 40.573$

$$\overline{x} \pm t_c * \frac{s}{\sqrt{n}}$$

$$40.573 \pm 2.639 * \frac{1.649}{\sqrt{100}}$$

**40.573 ± 0.43517 → (40.14, 41.006)**

Use calculator feature to check answer: **T-Interval (Use DATA option)**

   d)  The statement "99% confident" means that, if 100 samples of size __**100**___ were taken, about __**99**___ intervals will contain the parameter μ and about __**1**__ will not.

   e)  Complete the following: We are __**99**_% confident that the mean head circumference of all two months old babies is between __**40.14 cm**___ and __**41.01 cm**____

   f)  With __**99**_% confidence we can say that the mean head circumference of all two months old babies is __**40.573 cm**____ with a margin of error of __**0.435 cm**_____

   g)  How can you produce a more precise confidence interval?
**If we maintain the same degree of confidence, (99%) we'll get a shorter interval by selecting a larger sample**

**Chapter 8 – Confidence Intervals –One Population**

---

### Section 8.3 - Estimating a Population Proportion

**Assumptions**
- **The sample is a simple random sample**
- **The conditions for a binomial distribution are satisfied by the sample. That is: there are a fixed number of trials, the trials are independent, there are two categories of outcomes, and the probabilities remain constant for each trial. A "trial" would be the examination of each sample element to see which of the two possibilities it is.**
- **The normal distribution can be used to approximate the distribution of sample proportions because np > 5 and nq > 5 are both satisfied. (q = 1 – p)**


**Procedure for Constructing a Confidence Interval for p**

1) **Verify that the assumptions are satisfied.**

2) **Find the critical value $z_c$ (from table 5).**

3) **Evaluate the margin of error E.** $E = z_c \sqrt{\dfrac{p \bullet q}{n}}$

4) **Find the interval and write it in one of the following forms**

$$p \pm E; \qquad p - E < p < p + E; \qquad (p - E, p + E)$$

**Round-Off Rule for Confidence Interval Estimates of p**
**3 significant digits**


**Using the TI-83 to Construct Confidence Intervals for p:**
**STAT>>TESTS choose A:1-propZInt.**
**Notice: x must be a whole number**
**If you have to calculate it,**
**round it to the nearest whole number**

## 8.3 – Estimating a Population Proportion – Confidence Intervals

5) In an October 2003 poll conducted by the Gallup Organization, 684 of 1006 randomly selected adults aged 18 years old or older stated they think the government should make partial birth abortions illegal, except in cases necessary to save the life of the mother.

**We survey adults 18-year-old or older**
**We ask the question: Do you think the government should make partial birth abortions illegal, except in cases necessary to save the life of the mother?**
**There will be many YES and NO answers to this question.**
**This is a problem in which the variable is <u>qualitative</u>. In order to be able to handle this problem numerically we consider our success attribute the YES answer, and concentrate in the percentage of YES answers.**

**Objective of the problem: We want to know what percentage of adults 18-year-old or older think the government should make partial birth abortions illegal, except in cases necessary to save the life of the mother? It's impossible to survey ALL adults, so we work with a sample of size 1006.**

a) Obtain a point estimate for the proportion of adults aged 18 or older who think the government should make partial birth abortions illegal, except in cases necessary to save the life of the mother.

**Our point estimate is** $p = \dfrac{684}{1006} = .68$

b) Verify that the requirements for constructing a confidence interval about p-hat are satisfied.
**See assumptions on the prior page.**
**Also note that np and nq are both greater than 5, so the normal approximation is appropriate.**

c) Construct a 98% confidence interval for the proportion of adults aged 18 or older who think the government should make partial birth abortions illegal, except in cases necessary to save the life of the mother. Interpret the interval.

$$p \pm z_c \sqrt{\dfrac{p \bullet q}{n}} = .68 \pm 2.33 \sqrt{\dfrac{.68 \bullet .32}{1006}} = .68 \pm .0343 \rightarrow \textbf{(.6457, .7143)}$$

Use calculator feature to check answer: **1-PropZInterval**

d) We are ___**98**__% confident that the true proportion of adults aged 18 or older who think the government should make partial birth abortions illegal, except in cases necessary to save the life of the mother is between ____**64.6**_____% and ____**71.4**_____%

e) With __**98**____% confidence we can say that ___**68**__%  adults aged 18 or older think the government should make partial birth abortions illegal, except in cases necessary to save the life of the mother with a margin of error of __**3.4**_%

f) The statement "98% confident" means that, if 100 samples of size _**1006**____ were taken, about ____**98**__ of the intervals will contain the parameter p and about _**2**___ will not.

g) **Section 8.4 -** <u>How many more adults</u> should be included in the sample to be 98% confident that a point estimate p-hat will be within 1% of p? Use a p-hat of 0.68.

$$n = pq(\frac{z}{E})^2 = .68 * .32 * (\frac{2.33}{.01})^2 = 11814$$

**11814 – 1006 = 10808**

**Since we have the results of the 1006 adults surveyed, we need to survey 10808 more people.**
**We then have to use the p-hat of the sample of 11,814 adults to construct a new confidence interval. This new interval will have a margin of error of at most 1%, so our estimate will be more precise than the one obtained in part (c).**

h) **Section 8.4 -** If no preliminary sample is taken to estimate p, how large a sample is necessary to be 98% confident that the point estimate p-hat will be within a distance of 0.01 from p?

$$n = .25(\frac{z}{E})^2 = .25 * (\frac{2.33}{.01})^2 = 13573$$

**We have to survey 13573 adults and use the p-hat of the sample to construct a new confidence interval. This new interval will have a margin of error of at most 1%, so our estimate will be more precise than the one obtained in part (c).**

<u>**Notice**</u> **that when we don't have a prior point estimate p-hat, the sample size is much larger. It results in a more expensive study**

Section 8.5 – Confidence Intervals – Two Populations Means

---

### Inferences about Two Means with Unknown Population Standard Deviations
### Independent Samples
### Population Standard Deviations not Assumed Equal (Non-Pooled t-Interval)

**Assumptions**
- **The samples are obtained using simple random sampling**
- **The samples are independent**
- **The populations from which the samples are drawn are normally distributed or the sample sizes are large ($n_1 > 30, n_2 > 30$)**

**The procedure is robust, so minor departures from normality will not adversely affect the results. If the data have outliers, the procedure should not be used.**

**Use normal probability plots to assess normality and box plots to check for outliers.**

**A normal probability plot plots observed data versus normal scores. If the normal probability plot is roughly linear and all the data lie within the bounds provided by the software (our calculator does not show the bounds), then we have reason to believe the data come from a population that is approximately normal.**

### Using the TI-83/84 to Construct Confidence Intervals for µ1- µ2:
### STAT>>TESTS  select 0:2SampT Interval
### Use the DATA option if you have been given data

6) – Schizophrenia and Dopamine (with 2 populations we'll use the calculator only)
Previous research has suggested that changes in the activity of dopamine, a neurotransmitter in the brain, may be a causative factor for schizophrenia. In the paper "Schizophrenia: Dopamine b-Hydroxylase Activity and Treatment Response", Sternberg et al. published the results of their study in which they examined 25 schizophrenic patients who had been classified as either psychotic or not psychotic by hospital staff. The activity of dopamine was measured in each patient by using the enzyme dopamine b-hydroxylase to assess differences in dopamine activity between the two groups.
The following are the data, in nanomoles per milliliter-hour per milligram (nmol/ml-h/mg).

**Psychotic**

| | | | | |
|---|---|---|---|---|
| 0.0150 | 0.0222 | 0.0204 | 0.0275 | 0.0306 |
| 0.0270 | 0.0320 | 0.0226 | 0.0208 | 0.0245 |

**Non-psychotic**

| | | | | |
|---|---|---|---|---|
| 0.0104 | 0.020 | 0.0210 | 0.0105 | 0.0112 |
| 0.0230 | 0.0116 | 0.0252 | 0.0130 | 0.0200 |
| 0.0145 | 0.0180 | 0.0154 | 0.0170 | 0.0156 |

a) Because the sample sizes are small, for each population we must verify that the variable is normally distributed and the sample does not contain any outliers. Construct a normal probability plot to assess normality and a boxplot in order to check for outliers.
*For the firs data set, enter the data in L1 (press STAT, select Edit) of the calculator and open two plots, one with a modified box plot (the fourth icon) and another with the normal probability plot, which is the last icon type in the 2nd Y= [STAT PLOT] window. Do the same with the second data set after you enter it into L2.*

Use the calculator to obtain the following statistics (1-Var-Stats for each list)
Dopamine activity (nmol/ml-h/mg)

| Psychotic | Non-Psychotic |
|---|---|
| x1bar = .02426 | x2bar = .01643 |
| s1 = .00514 | s2 = .0047 |
| n1 = 10 | n2 = 15 |

b) Give the point estimate for the psychotic patients

x1bar = .02426

c) Give the point estimate for the non-psychotic patients

x2bar = .01643

d) We observe that x1bar is ……………*higher than / lower than* x2bar.
Is it *higher* by chance; or is it significantly *higher*?
In order to discover this, we'll do the following:

e) Give the point estimate for the difference between dopamine-activity in the two groups:
x1bar – x2bar = .02426 - .01643 = .00783

f) Use a calculator feature to obtain a 98% confidence interval for the difference between the two population means. (Are you using z or t? Why?)
We are using t because the standard deviations of the population are not known. (We know the standard deviations of the samples)
Use a 2-Samp T Interval (non-pooled because we can't assume $\sigma_1 = \sigma_2$)

….00266………. <   μ1- μ2  < …… .013…………..

g) What does the interval suggest about μ1 and μ2?

μ1 < μ2                    μ1 > μ2                    μ1 = μ2

Explain your choice; be very specific in your explanation

The interval gives plausible values for μ1- μ2 . Since the interval is completely ABOVE ZERO, all possible values for μ1- μ2  are positive, which implies than μ1 > μ2

h) We are __98_____% confident that dopamine activity in psychotic patients, on average

*(choose from the following choices)*
*is higher than,*          *is lower than,*          *may be equal to*

the dopamine activity in non-psychotic patients.

**Section 8.5 – Confidence Intervals - Two Population Proportions**

---

**Assumptions**
- **The samples are independently obtained using simple random sampling.**
- **For both samples, the conditions np ≥ 5 and n(1 – p) ≥ 5 are both satisfied.**
  **For both samples, the sample size, is no more than 5% of the population size**

7) Eating Out Vegetarian
A Zogby International poll of 1181 US adults was conducted in March 1999, to gauge the demand for vegetarian meals in restaurants. The study was commissioned by the Vegetarian Resource Group and was published in the September/October 1999 issue of the Vegetarian Journal.
In the survey, independent random samples of 747 US men and 434 US women were taken. Of those sampled, 276 men and 195 women said that they sometimes order a dish without meat, fish, or fowl when they eat out.

a) Find the point estimate p1hat – p2hat

| men | women |
|---|---|
| n1= 747 | n2 = 434 |
| x1 = 276 | x2 = 195 |
| p1-hat = .3695 | p2-hat = .4493 |

$$p_1 - p_2 = .3695 - .4493 = -.0793$$

b) Construct a 90% confidence interval for the difference, p1 – p2, between the proportions of US men and US women who sometimes order a dish without meat, fish, or fowl.

Use 2-Prop-Z-Interval to obtain the interval

**-.1287 < p1 – p2 < -.031**

c) Do the data provide sufficient evidence to conclude that, in the United States, the percentage of men who sometimes order a dish without meat, fish, or fowl is smaller than the percentage of women who do the same? Explain why or why not.
**The interval is completely below zero. This means that any possible value for p1 – p2 is negative, which implies than p1 is smaller than p2.**

**With 90% confidence we can conclude that in the United States, the percentage of men who sometimes order a dish without meat, fish, or fowl is smaller than the percentage of women who do the same.**

d) We are _**90**___% confident that, in the United States, the percentage of women who sometimes order a vegetarian meal is___ **larger than**_____ **(larger than, smaller than, may be equal to)** the percentage of men who order a vegetarian meal by somewhere between ____**3.1**_____ and ___**12.9**_____ percentage points