Section 8.3 - **Estimating a Population Mean** (pp. 507-530)

1. When σ Is Known: The One-Sample z Interval for a Population Mean

Returning to the "mystery mean" activity we conducted in Section 8.1, our point estimate was $\bar{x}$ = 249.912. Our task was to build a reasonable interval for the population mean μ. We kind of took a "seat of the pants" approach but what we did was conceptually correct. When constructing a confidence interval for the population mean when the population standard deviation is known, the **one-sample z interval for a population mean** is used.

> **One-Sample z Interval for a Population Mean**
>
> Draw an SRS of size *n* from a population having unknown mean μ and known standard deviation σ. As long as the Normal and Independent conditions are met, a level C confidence interval is
>
> $$\bar{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}}$$
>
> NEVER USE THIS!!!
>
> The critical value z^* is found from the standard Normal distribution.

This method is not very useful in practice because we do not know the population standard deviation but we can use the one-sample z interval for a population mean in order to *estimate the sample size* need to achieve a specified margin of error.

> **Choosing Sample Size for a Desired Margin of Error When Estimating μ**
>
> To determine the sample size *n* that will yield a level C confidence interval for a population mean with a specified margin of error ME:
>
> - Get a reasonable value for the population standard deviation σ from an earlier pilot study.
> - Find the critical value z^* from a standard Normal curve for confidence level C.
> - Set the expression for the margin of error to be less than or equal to ME and solve for *n*:

Example - Administrators at WCHS want to estimate how much time students spend on homework, on average, during a typical week. They want to estimate μ at the 90% confidence level with a margin of error of at most 15 minutes. A pilot study indicated that the standard deviation of time spent on homework per week is about 154 minutes. How many students must be surveyed to meet the conditions of this task?

$$z^* \frac{\sigma}{\sqrt{n}} \le ME$$

$$1.645\left(\frac{154}{\sqrt{n}}\right) \le 15$$

$$1.645\left(\frac{154}{15}\right) \le \sqrt{n}$$

$$\left(1.645\left(\frac{154}{15}\right)\right)^2 \le n$$

$\rightarrow n \approx 285.2 \Rightarrow$ $\boxed{n = 286}$

Section 8.3 - **Estimating a Population Mean** (pp. 507-530)

1. When σ Is Known: The One-Sample z Interval for a Population Mean

Returning to the "mystery mean" activity we conducted in Section 8.1, our point estimate was $\bar{x}$ = 249.912. Our task was to build a reasonable interval for the population mean μ. We kind of took a "seat of the pants" approach but what we did was conceptually correct. When constructing a confidence interval for the population mean when the population standard deviation is known, the **one-sample z interval for a population mean** is used.

> **One-Sample z Interval for a Population Mean**
>
> Draw an SRS of size *n* from a population having unknown mean μ and known standard deviation σ. As long as the Normal and Independent conditions are met, a level C confidence interval is
>
> $$\boxed{\bar{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}}}$$
>
> NEVER USE THIS !!!
>
> The critical value z^* is found from the standard Normal distribution.

This method is not very useful in practice because we do not know the population standard deviation but we can use the one-sample z interval for a population mean in order to *estimate the sample size* need to achieve a specified margin of error.

> **Choosing Sample Size for a Desired Margin of Error When Estimating μ**
>
> To determine the sample size *n* that will yield a level C confidence interval for a population mean with a specified margin of error ME:
>
> - Get a reasonable value for the population standard deviation σ from an earlier pilot study.
> - Find the critical value z^* from a standard Normal curve for confidence level C.
> - Set the expression for the margin of error to be less than or equal to ME and solve for *n*:

Example - Administrators at WCHS want to estimate how much time students spend on homework, on average, during a typical week. They want to estimate μ at the 90% confidence level with a margin of error of at most 15 minutes. A pilot study indicated that the standard deviation of time spent on homework per week is about 154 minutes. How many students must be surveyed to meet the conditions of this task?

$$z^* \frac{\sigma}{\sqrt{n}} \le ME$$

$$1.645\left(\frac{154}{\sqrt{n}}\right) \le 15$$

$$1.645\left(\frac{154}{15}\right) \le \sqrt{n}$$

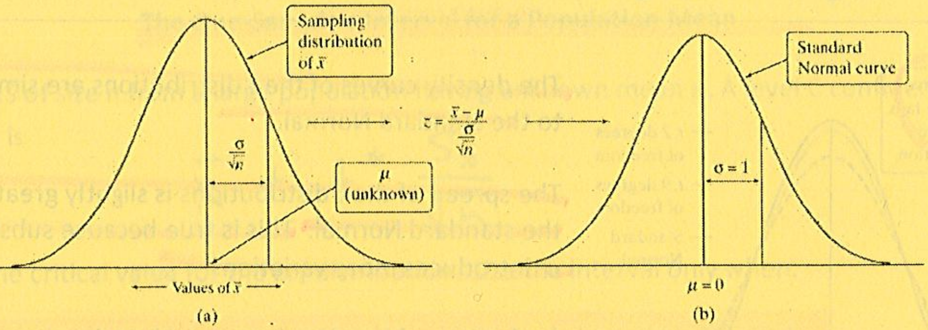
$$\left(1.645\left(\frac{154}{15}\right)\right)^2 \le n$$

$\rightarrow n \approx 285.2 \Rightarrow$ $\boxed{n = 286}$

2. When σ Is Unknown: The t Distributions

When the sampling distribution of $\bar{x}$ is close to Normal, we can find probabilities involving $\bar{x}$ by standardizing:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$



The figures above show the sampling distribution of $\bar{x}$ and the standardized values of $\bar{x}$.

We do not know σ so we will estimate it using the sample standard deviation s_X. What happens when we now standardize?

$$?? = \frac{\bar{x} - \mu}{s_X / \sqrt{n}}$$

A simulation using the calculator can be used to compare the shape of the distribution of standardized $\bar{x}$ values when σ is used with the shape of the distribution when s_X is used. We will simulate taking repeated SRSs of size n = 4 from a Normal population with mean μ = 100 and standard deviation σ = 5. We will graph the results in order to compare the distributions.

ROSSMAN CHANCE CI APPLET.

① MEANS Z WITH σ

$\mu = 100$ $\sigma = 5$ CL = 99

$n = 4$ INTERVALS = 100

② MEANS Z WITH S

SAME

③ MEANS t

SAME

How did the distributions compare?

t-DIST MATCHED BEST.

The distribution using s_X is known as the **t distribution**. It has a different shape than the standard Normal curve: still symmetric with a single peak at 0 but with much more area in the tails.

The statistic t has the same interpretation as any standardize statistic: it tells us how far $\bar{x}$ is from its mean μ in standard deviation units.

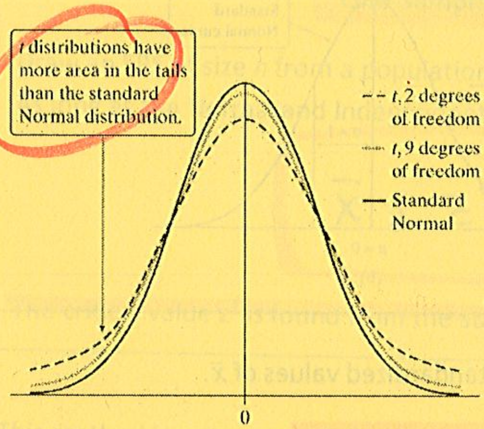
There is a different t distribution for each sample size. We specify a particular t distribution by giving its **degrees of freedom (df)**. The appropriate degrees of freedom are found by subtracting 1 from sample size n, making df = n -1. The t distribution with n - 1 degrees of freedom is denoted by t_{n-1}.

The t Distributions: Degrees of Freedom

Draw an SRS of size n from a large population that has a Normal distribution with mean μ and standard deviation σ. The statistic

$$t = \frac{\bar{x} - \mu}{s_x / \sqrt{n}}$$

has the ***t* distribution** with **degrees of freedom** df = n-1. The statistic will have approximately a t_{n-1} distribution as long as the sampling distribution of $\bar{x}$ is close to Normal.



The density curves of the t distributions are similar in shape to the standard Normal.

The spread of the t distributions is slightly greater than that of the standard Normal. This is true because substituting s_x for σ introduces more variation.

As degrees of freedom increase, the t density curve approaches the standard Normal more closely. This happens because s_x approaches σ as the sample size gets larger.

Table B gives critical values for t^* for the t distributions.

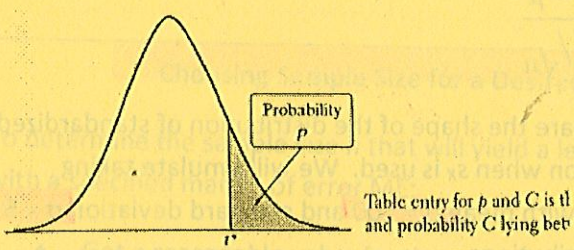


Table B t distribution critic

Upper tail prob

df	.25	.20	.15	.10	.05	.025
1	1.000	1.376	1.963	3.078	6.314	12.71
2	0.816	1.061	1.386	1.886	2.920	4.303
3	0.765	0.978	1.250	1.638	2.353	3.182
4	0.741	0.941	1.190	1.533	2.132	2.776
5	0.727	0.920	1.156	1.476	2.015	2.571

Suppose you want to construct a 95% confidence interval for the mean μ of a Normal population based on an SRS of size n = 6. What critical value t^* should you use?

.95 ⇒ UPPER TAIL = $\frac{0.5}{2}$ = 0.25

df = n - 1 = 6 - 1 = 5

$t^* = 2.571$

Technology can also be used to find critical values of t^*. Use [2nd] [VARS] 4:invT(and enter the ***area of the lower tail*** and the ***degrees of freedom (n - 1)***.

Applications - Use Table B to find the critical value t^* that you would use for a confidence interval for a population mean μ in each situation. Then check your answer with your calculator.

1.323

(a) A 98% confidence interval based on 22 observations. TAIL = $\frac{0.2}{2}$ = 0.1 df = 22 - 1 = 21

1.833

(b) A 90% confidence interval based on 10 observations. TAIL = $\frac{0.1}{2}$ = 0.05 df = 10 - 1 = 9

(c) A 95% confidence interval from a sample of size 7. TAIL = $\frac{0.5}{2}$ = 0.25 df = 7 - 1 = 6

0.718

CALC ✓

3. Constructing a Confidence Interval for μ

Statistic ± (critical value)•(standard deviation of the statistic)

The **standard error of the sample mean** $\bar{x}$ is $s_x/\sqrt{n}$ where s_x is the sample standard deviation. It describes how far $\bar{x}$ will be from μ, on average, in repeated SRSs of size *n*.

The One-Sample *t* Interval for a Population Mean

Choose an SRS of size *n* from a large population having unknown mean μ. A level C confidence interval for μ is

$$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$$

Where t^* is the critical value for the t_{n-1} distribution. Use the interval only when:

(1) the population distribution is Normal *or* the sample size is large ($n \geq 30$), and

(2) the population is at least 10 times as large as the sample.

As with confidence intervals for population proportions, we have to verify the Random, Normal, and Independent conditions. However in practice, this may be more complicated when we do not know the population standard deviation σ.

Example - As part of their final project in AP Statistics, Christina and Rachel randomly selected 18 rolls of a generic brand of toilet paper to measure how well this brand could absorb water. To do this, they poured ¼ cup of water onto a hard surface and counted how many squares it took to completely absorb the water. Here are the results:

29 20 25 29 21 24 27 25 24 29 24 27 28 21 25 26 22 23

Construct and interpret a 99% confidence interval for μ = the mean number of squares of generic toilet paper needed to absorb ¼ cup of water.

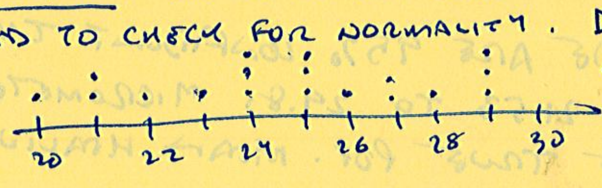
State: We want to estimate μ = the mean number of squares of generic toilet paper needed to absorb ¼ cup of water with 99% confidence:

Plan: WE WILL CONSTRUCT A ONE-SAMPLE t-INTERVAL FOR POP. MEAN PROVIDING CONDITIONS ARE MET.

(DISCUSS PLOT)

RANDOM: RANDOM SAMPLE ✓

NORMAL: SINCE n IS SMALL AND DIST. IS NOT SPECIFIED, WE NEED TO CHECK FOR NORMALITY. DOT PLOT DOES NOT SHOW ANY OUTLIERS OR STRONG SKEWNESS SO IT IS ASSUMED POP IS NORMAL.



IND: IT IS REASONABLE TO ASSUME MORE THAN 10(18) = 180 BRANDS OF GENERIC TOILET PAPER.

CONDITIONS MET ✓

Do: $\bar{x} = 24.94 \quad s_x = 2.86 \quad t^* = 2.898 \quad df = 18-1=17$

$$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}} = 24.94 \pm 2.898 \frac{2.86}{\sqrt{18}} = [22.99, 26.89]$$

Conclude: WE ARE 99% CONFIDENT THAT THE INTERVAL FROM 22.99 TO 26.89 CAPTURES THE TRUE POP. MEAN NUMBER OF SQUARES OF T.P. NEEDED TO ABSORB 1/4 CUP OF WATER.

Note: Since Table B does not include every possible sample size *n*, when the actual df does not appear in the table, use the greatest df available that is less than your desired df.

Application - Complete CYU on p. ~~511~~ 522.

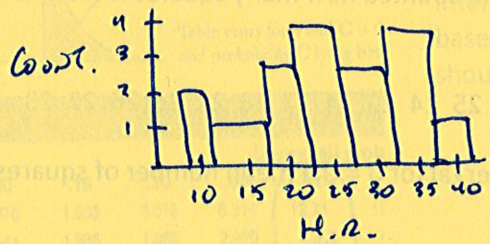
① ~~READ~~ STATE: WE WANT TO BUILD A 95% CI FOR THE MEAN HEALING RATE

② ~~WE~~ PLAN: WE WILL USE A ONE-SAMPLE t-INTERVAL FOR POP. μ.

CONDS: RANDOM: RANDOMLY CHOSEN ✓

NORMAL: $n < 30$ AND DO NOT KNOW DIST SO GRAPH.

GRAPH IS REASONABLY SYMMETRIC W/NO OUTLIERS ✓



IND: SAFE TO ASSUME MORE THAN 10(18) = 180 NEWTS IN POPULATION ✓

CONDITIONS ARE MET

③ DO. $\bar{x} = 25.67 \quad s_x = 8.32 \quad t^* = 2.110 \quad df = 18-1=17$

$$25.67 \pm 2.110 \frac{8.32}{\sqrt{18}} = [21.53, 29.81]$$

④ CONCLUDE: WE ARE 95% CONFIDENT THAT THE INTERVAL FROM 21.53 TO 29.81 MICROMETERS/HR CAPTURES THE TRUE POP. MEAN HEALING RATE FOR NEWTS.

4. Using *t* Procedures Wisely

The stated confidence level of a one-sample *t* interval for μ is exactly correct when the population distribution is exactly Normal. No population of real data is exactly Normal. It turns out that *t* procedures are not strongly affected when this condition is violated. Procedures that are not strongly affected when a condition for using them is violated are called **robust**.

Definition: An inference procedure is called **robust** if the probability calculations involved in that procedure remain fairly accurate when a condition for using the procedure is violated.

It should be noted that *t* procedures are not robust against **outliers**.

Using One-Sample *t* Procedures: The Normal Condition

- *Sample size less than 15*: Use *t* procedures if the data appear closely Normal (roughly symmetric, unimodal, no outliers). If the data are clearly skewed or if outliers are present, do not use *t*.
- *Sample size at least 15*: The *t* procedures can be used except in the presence of outliers or strong skewness.
- *Large samples*: The *t* procedures can be used even for clearly skewed distributions when the sample size is large, roughly $n \geq 30$.

Technology - The calculator can be used to construct a confidence interval for an unknown population mean. Refer to p. 64 of NTA or p. 514 of the text.

[STAT] <TESTS> 8: TINTERVAL

It should be noted that if you use the calculator, it is recommended that you check your answer with the calculations of the formula.

HW: Read Sec 8.3; pp. 527-530 problems 57, 59, 63, 69, 71-73, 75-78, 80*