Chapter 6 Confidence Intervals

CI

STA2023 SECTION 6.1 Confidence Intervals for the Mean (σ Known) NOTES

$\boxed{(LB, UB)}$ 90%

μ

Learning Outcomes:

1) Find a point estimate and a margin of error
2) Construct and interpret confidence intervals for the population mean when σ is known
3) Determine the minimum sample size required when estimating a population mean

Inferential statistics are then used to generalize and predictions about the parameter based on the statistic gained from the sample.

$\bar{x}$, Sx

Point Estimate – a point or single value used to approximate a population parameter (population mean μ or population proportion p). The sample mean $\bar{x}$ is the best point estimate for the population mean μ and the sample proportion $\hat{p}$ is the best point estimate for the population proportion p.

*Point estimates (sample values) are a good approximation to population values, but are not exact (sampling error), so we give a range or interval of values instead.

Confidence Interval or Interval Estimate – a range or interval of values used to estimate the true value of a population parameter.

Confidence Level – a percentage (0%-100%) that measures the success rate of the method used to construct the confidence interval. If we draw many samples and use each one to construct a confidence interval, in the long run, the percentage of confidence intervals that cover the true value of the population would be equal to the confidence level.

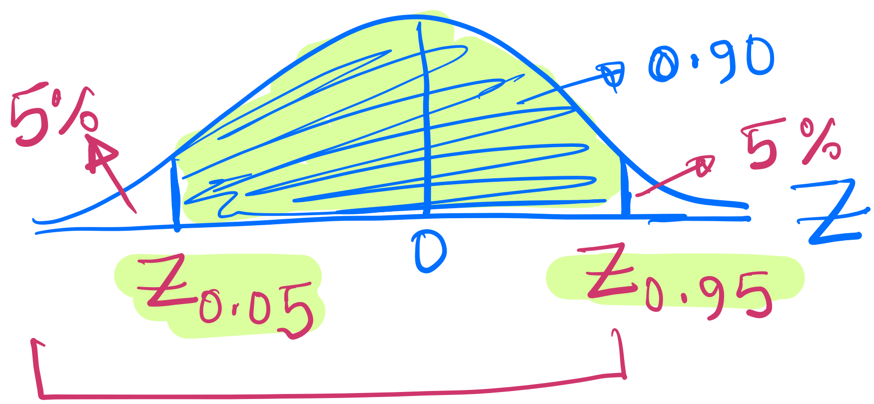
$c = 90\%$, $\alpha = \text{error} = 1 - c = 10\%$

Let $\boldsymbol{c = confidence\ level}$ and let $\boldsymbol{\alpha = 1 - c}$. Therefore c *and* α *are complements*. Confidence levels that are most commonly used are $\boldsymbol{c = 90\%}$ ***or*** $\boldsymbol{c = 95\%}$ ***or*** $\boldsymbol{c = 99\%}$.

$\alpha = 1\%$

$\alpha = 5\%$

$\boxed{\text{Let } c = 90\%.}$



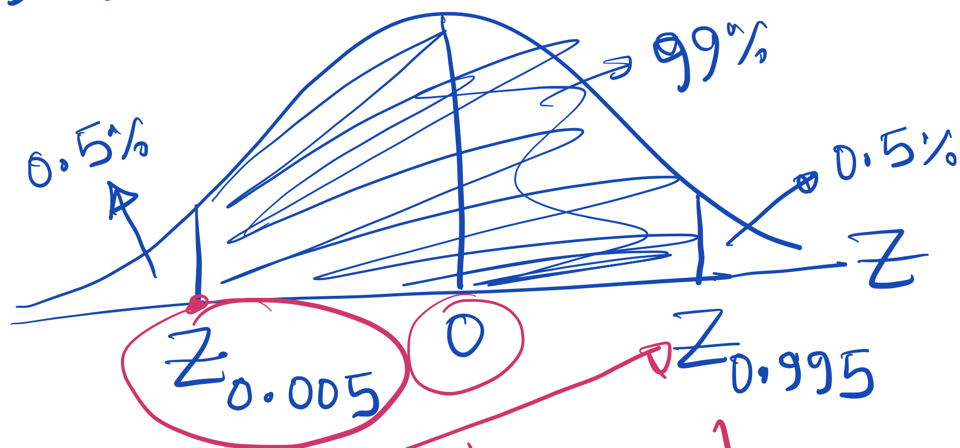
Let $c = 95\%$.

Let $c = 99\%$.

$Z_{0.05} = \text{invNorm}(0.05, 0, 1)$
$= -1.6449$

$Z_{0.95} = 1.6449$

$c = 99\%$



2.5758

$Z_{0.995} = \text{invNorm}(\qquad) = \boxed{2.5758}$

$Z_{0.005} = -2.5758$

$Z_{\frac{\alpha}{2}}$ $\qquad$ $Z_{1-\frac{\alpha}{2}}$

critical value

CI → 90% , 95% , 99% → C

for pop. parameter

$\alpha = 1 - C$

6.1	6.2	6.3
CI for pop mean	CI for pop mean	CI for pop proportion
Z interval (σ is known)	T- interval (σ is unknown)	1- prop. Z interval [Sample prop. will be given as % or x and n formate.]
point estimate (PE) $= \bar{x}$	PE is $\bar{x}$	PE is $\hat{p} = \frac{x}{n}$
Critical value is $Z_{\frac{\alpha}{2}}$ (lower critical value) and $Z_{1-\frac{\alpha}{2}}$ (upper) $Z_{\frac{\alpha}{2}} = \text{invNorm}\left(\frac{\alpha}{2}, 0, 1\right)$	Critical value is $t_{\frac{\alpha}{2}}$ (lower), $t_{1-\frac{\alpha}{2}}$ (upper) $t_{\frac{\alpha}{2}} = \text{invT}\left(\frac{\alpha}{2}, df\right)$ $\boxed{df = n-1}$	Crical value $Z_{\frac{\alpha}{2}}$ (lower) and $Z_{1-\frac{\alpha}{2}}$ (upper)

6.1	6.2	6.3
Standard error (SE) $SE = \dfrac{\sigma}{\sqrt{n}}$	SE is $SE = \dfrac{s}{\sqrt{n}}$	SE is $SE = \sqrt{\dfrac{\hat{p}\,\hat{q}}{n}}$ $\hat{p} = \dfrac{x}{n}$ $\hat{q} = 1 - \hat{p}$ n = sample size
Margin of error (ME) $ME = SE \times$ critical value $= \dfrac{\sigma}{\sqrt{n}} \times Z_{1-\frac{\alpha}{2}}$	$ME = SE \times$ critical value $ME = \dfrac{s}{\sqrt{n}} \times t_{1-\frac{\alpha}{2}}$	$ME = SE \times$ critical value $ME = \sqrt{\dfrac{\hat{p} \cdot \hat{q}}{n}} \times Z_{1-\frac{\alpha}{2}}$
CI is $(PE - ME,\ PE + ME)$	CI is $(PE - ME,\ PE + ME)$	CI is $(PE - ME,\ PE + ME)$

$z_{\frac{\alpha}{2}}$ is called the critical value of z. It represents –

Margin of Error – the maximum likely difference between the observed sample value and the true population parameter. (maximum likely difference between $\bar{x}$ *and* μ *or between* $\hat{p}$ *and* p).

CONSTRUCTING A CONFIDENCE INTERVAL ESTIMATE FOR THE POPULATION MEAN μ

For means, we are using sample means from a sample of size n – The Central Limit Theorem says we can use the normal distribution if ___________________ and _________________________. The standard error of the sampling distribution will be :

And therefore, the Margin of Error will be:

To construct a confidence interval estimate: $point\ estimate\ \pm margin\ of\ error$

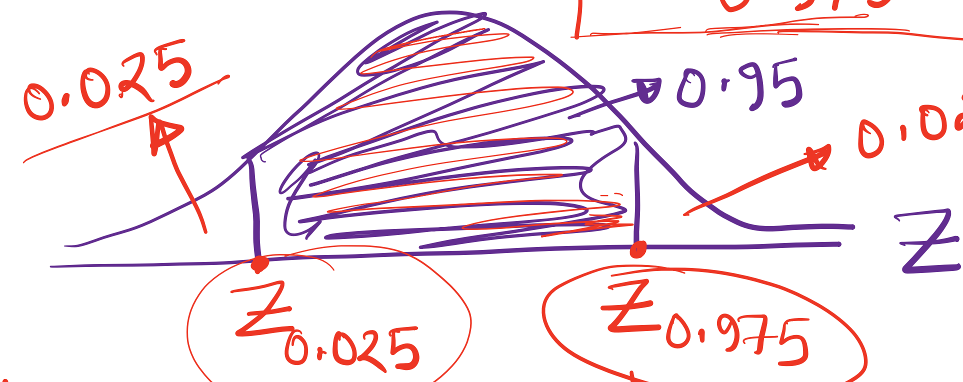
Example #1: In a simple random sample of 150 Fort Myers households, the sample mean number of personal computers was $\bar{x} = 1.32$. Assume that the population standard deviation is $\sigma = 0.41$.

a. Construct a 95% confidence interval estimate for the mean number of personal computers for all households in Fort Myers.

$n = 150$, $\sigma = 0.41$, $\bar{x} = 1.32$

$c = 95\% = 0.95$, $\alpha = 1 - c = 0.05$

critical value, $\boxed{Z_{0.975} = \text{invNorm}(.975, 0, 1) = 1.9599}$



$\boxed{.95 + .025 = .975}$

$SE = \frac{\sigma}{\sqrt{n}} = \frac{.41}{\sqrt{150}} = .0335$

$ME = SE \times \text{critical value} = 0.0335 \times 1.9599$

$= 0.0657$

95% CI is

$(PE - ME,\ PE + ME)$

$= (1.32 - 0.0657,\ 1.32 + 0.0657)$

$\boxed{= (1.2543,\ 1.3857)}$

We are 95% confident that our mean # of PC in all households in Fort Myers is between $(1.2543,\ 1.3857)$

$\boxed{\sigma \text{ Known}}$ → Z-interval

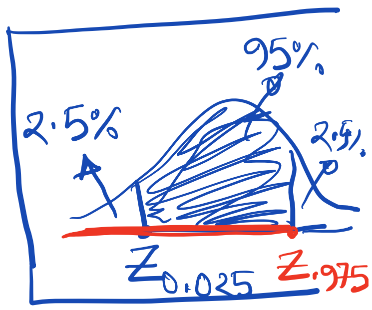
$c = 90\% = 0.90$, $\alpha = 1 - c = 10\% = 0.10$

Critical value $\boxed{Z_{1-\frac{\alpha}{2}}} > Z_{\frac{\alpha}{2}}$

$Z_{1-\frac{\alpha}{2}} = \text{invNorm}\left(1-\frac{\alpha}{2}, 0, 1\right)$

→ $SE = \frac{\sigma}{\sqrt{n}}$

$$\boxed{ME = Z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}}}$$



→ ?

$\boxed{Z_{0.975}}$

$= \text{invNorm}(0.975, 0, 1)$

$= 1.9599$

→ $25 = 1.9599 \times \frac{210}{\sqrt{n}}$

→ $\boxed{\frac{25}{1.9599 \times 210}} = \frac{1}{\sqrt{n}}$

→ $0.0607 = \frac{1}{\sqrt{n}}$

→ $(\sqrt{n})^2 = \frac{1}{0.0607} = (16.46316)^2$

→ $n = 271.0356 \approx \boxed{272}$

Example 2: You wish to estimate the mean amount of time that internet users spend on *Facebook* each month. How many internet users must be surveyed in order to be 95% confident that your sample mean is within 25 minutes of the true population mean? From a prior *Nielson* survey, assume the standard deviation of the population is ~210 minutes.

$\sigma = 210, \quad c = 0.95, \quad ME = 25$

$$\boxed{n = 272}$$

Example 3: An admissions director wants to estimate the mean age of all students enrolled at a college. The estimate must be with 1.5 years of the population mean. Assume the population of ages is normally distributed and $\sigma = 2.7$ years. Determine the minimum sample size required to estimate the population mean with 95% confidence.

$\boxed{c = 0.95}, \quad \sigma = 2.7, \quad ME = 1.5$

$$ME = z_{0.975} \cdot \frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow 1.5 = 1.9599 \cdot \frac{2.7}{\sqrt{n}}$$

$$\Rightarrow \frac{1.5}{(1.9599)\cdot(2.7)} = \frac{1}{\sqrt{n}}$$

$$\Rightarrow 0.2835 = \frac{1}{\sqrt{n}}$$

$$\Rightarrow \sqrt{n} = \frac{1}{0.2835} = 3.5278$$

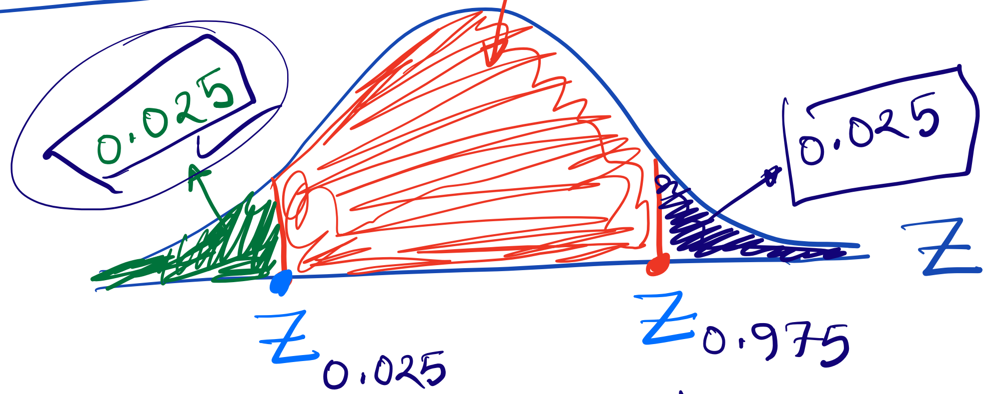
$$n = (3.5278)^2 = 12.45 \approx \boxed{13}$$

Example 4: An admissions director wants to estimate the mean age of all students enrolled at a college. The estimate must be with 2 years of the population mean. Assume the population of ages is normally distributed and $\sigma = 1.3$ years. Determine the minimum sample size required to estimate the population mean with 90% confidence.

$\boxed{\alpha = 5\%}$

Normal dirt

$C = 90\%,\ 95\%,\ 99\%$

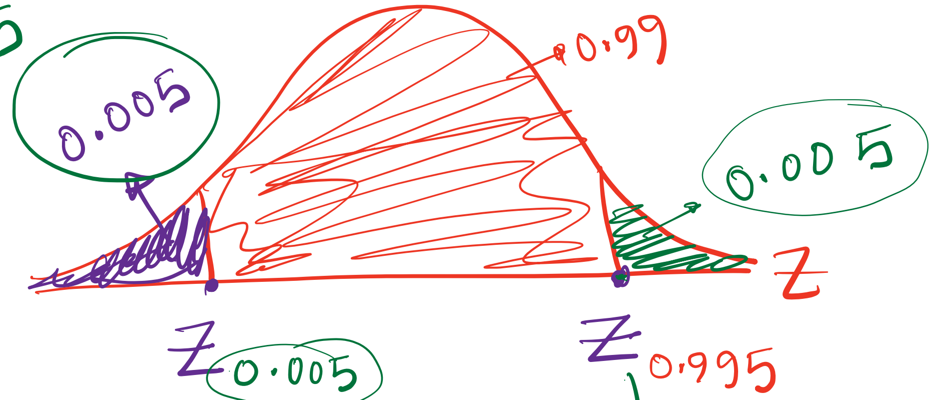


$$0.95 + 0.025 = 0.975$$

$$Z_{0.975} = \text{invNorm}(0.975, 0, 1)$$
$$= 1.9599$$

$C = 99\%$

$0.99 + 0.005$
$=$



$\boxed{AL = \text{subscript}}$

$1 - 0.005$
$=$

STA 2023 SECTION 6.2 Confidence Intervals for the Mean (σ Unknown) NOTES

T- interval

Learning Outcomes:

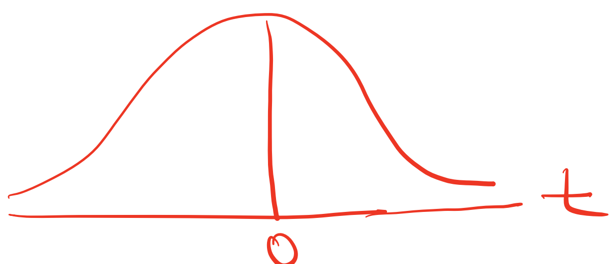
- Interpret the t-distribution
- Construct and interpret confidence intervals for a population mean when σ is not known

When trying to estimate the value of the population mean μ, if the value of the population standard deviation σ is known and $n > 30$, then we can use the standard normal distribution to build confidence intervals.

If a value of σ is not known (which is usually the case in research situations), then we must use the **sample standard deviation (s)**, which does not always follow a normal distribution. So, instead of the normal distribution, we use another distribution called the **t-distribution.**

Properties of the t-distribution:

- The graph of the distribution **changes for different sample sizes**. As n gets larger, the shape of the graph approaches that of the normal distribution.
- The graph of the distribution is a somewhat bell-shaped symmetric curve with a mean of $\boldsymbol{t = 0}$.
- The graph reflects greater variability that is expected with smaller samples, therefore the standard deviation is greater than 1 and **changes depending on sample size.**



$\boxed{df = n-1}$

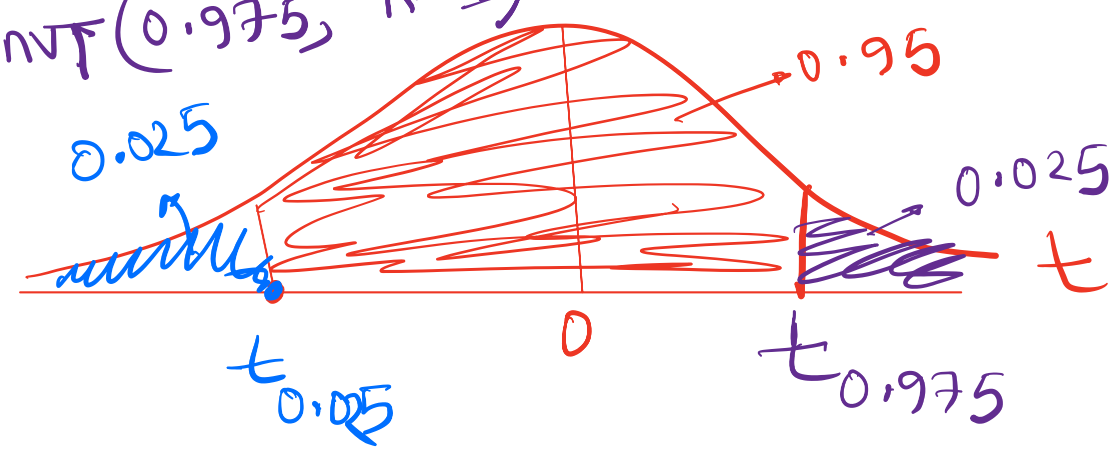
Degrees of Freedom (df) – For a collection of sample data, the degrees of freedom is the number of sample values that can vary after a certain restriction has been imposed on all data values. The degrees of freedom is always one less than the sample size. Therefore $\boldsymbol{df = n - 1.}$

Finding Critical Values of t: *Note:

Suppose $n = 15$ and we want 95% confidence:

$c = 0.95, \ \alpha = 0.05$

$t_{0.975} = \text{invT}(0.975, n-1)$



Constructing a Confidence Interval Estimate for a Population Mean μ; σ *is NOT known*

$c = .95, .90, .99$, $\alpha = 1 - c$

$t_{1-\frac{\alpha}{2}} =$, $\boxed{SE = \frac{s}{\sqrt{n}}}$

$ME = t_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$

Example 1: The *University of Arizona* (Phoenix) conducted *The Garbage Project.* For 16 randomly chosen families, the mean weight of plastics discarded in the trash (that could be recycled) was $\bar{x} = 1.911\ pounds$ with standard deviation of $s = 0.892\ pounds$. Construct a 90% confidence interval estimate for the true mean amount of plastics discarded in the trash for **all** families in Phoenix.

$c = 0.90$, $s = 0.892$, $\bar{x} = 1.911$

90% CI $\Rightarrow$ (1.5201, 2.3019)

90% confident that μ is bet^n (1.5201, 2.3019)

Example 2: In a random sample of 10 people, the mean commute time to work was 40 minutes with a standard deviation of 6.5 minutes. Find a point estimate for the population mean. Assume the population is normally distributed.

a) Find a point estimate for the population mean commute time.

b) Construct and interpret a 95% confidence interval to estimate the population mean. Do not round.