Calculator
10.2 NB Slides

10.2 Estimating a Population Mean

Remember: Conditions for Inference about μ

- SRS
- Normality
- Independence $N \geq 10n$

If the population standard deviation, σ, is unknown you must use the sample standard deviation, s.

standard deviation of sampling distribution $= \boxed{\frac{s}{\sqrt{n}}}$ = Standard error S.E

Sampling Distribution

	σ known	σ unknown
Standard error:	$\frac{\sigma}{\sqrt{n}}$	$\frac{s}{\sqrt{n}}$
distribution:	normal z-distribution	degrees of freedom t-distribution $df = n-1$
Confidence Interval:	$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$	$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$

(slide 8) Density Curves

Similar to N: symmetric to 0, bell shaped, single peak
Contrast to N: larger spread, more area in tails,
As df increases, curve approaches N

When df does not show up in Table C, use the greatest value less than df

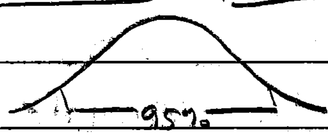
Finding t^* Table C

95% CL with $n=10 \Rightarrow df=9 \quad t^*=2.262$

80% CL with $n=7 \Rightarrow df=6 \quad t^*=1.440$

99% CL with $n=51 \Rightarrow df=50 \quad t^*=2.678$

Calculator: $\boxed{\text{2nd}}$ $\boxed{\text{VARS}}$ invT(area, df)



$n=10$

$z=1.96$

t^*

area to the left = .975

df = 9

$\boxed{t = 2.262}$

$$\text{area} = \frac{1+C}{2}$$

Steps for Confidence Interval with σ unknown

① one sample t^* Confidence interval for mean ______ (context)

② SIN – SRS, independent, ~N

③ Math $\bar{x} \pm t^*\left(\frac{s}{\sqrt{n}}\right)$; df: $n-1$

④ Interpretation in Context

Example 1 (Slide 19 - Day 1)

① one-sample t confidence interval for mean mileage (one way) for all college students to school

② (Simple random sample is stated) $N \geq 10n = 10(30) = 300$. Since there are ≥ 300 college students that commute we can assume independence.

10.2 cont'd

Example 1 (cont'd)

Data: L_1 — 2 A graph of the distances is $\sim N$, so we can assume the population of distances is $\sim N$.

Graph Box-whisker

1VAR Stat — 3 $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$ df $n-1$ — 95% CI df = 29

$t^* = 2.045$

STAT-TESTS — $18.867 \pm 2.045\left(\frac{9.989}{\sqrt{30}}\right)$ — $\bar{x} = 18.867$

TInterval: $\boxed{\text{Data}}$ — $(15.137, 22.597)$ — $S_x = 9.989$

df = 29

4 Based on this sample, we are 95% confident that the true mean distance driven one-way to college by all students is between 15.137 and 22.597 miles.

HW pg. 648 27-28, 31

Review Ch. 5

Matched pairs block design - two subjects are paired and either:

1. each subject receives both treatments
2. one subject of the pair receives the treatment

Paired t procedures - to compare the responses to the two treatments in a matched pairs design or before-and-after measurements on the same subjects, apply one sample t procedures to the observed differences.

The parameter μ in a paired t procedure, μ_D

- The mean difference in the responses to the two treatments within matched pairs of subjects in the entire population (subjects are matched in pairs)
- pg 651 The mean difference in response to the two treatments for individuals in the population (when the same subject receives both treatments)
- the mean difference between before-and-after measurements for all individuals in the population (for before-and-after on the same individuals)

Example 2 (slide B-19) Is Caffeine Dependence Real

① Parameter: One-sample t confidence interval for the mean difference in depression scores for this population. $\mu_D = \mu_{\text{placebo}-\text{caffeine}}$

② Conditions (Simple random sample of differences in depression scores is stated) $N \geq 10n = 10(11) = 110$.

SIN pg 652 It is safe to assume there are more than 110 people addicted to caffeine, so the condition of indepence is satisfied. A graph of the differences is approximately normal. 

We will use one-sample t-procedures to construct the CI for μ_{Diff} since σ is unknown.

10.2 cont'd

(3) Calculations: $\bar{x} = t^*\left(\frac{s}{\sqrt{n}}\right)$

$7.364 \pm 1.812\left(\frac{6.918}{\sqrt{11}}\right)$

$t = 1.812$
$\bar{x} = 7.364$
$df = 11-1=10$

$(3.583, 11.144)$

(4) Interpretation: Based on this sample we are 90% confident that the true mean difference in depression scores (placebo - caffeine) for the population is between 3.583 and 11.144 units.

Random selection of individuals allows us to generalize the results of the study to a larger population

Random assignment of treatments to subjects in an experiment lets us investigate whether there is evidence of a treatment effect, which might suggest the treatment caused the difference

pg. 654
Definition

Procedures not strongly affected by lack of normality are called robust.

The t procedures are not robust against outliers, because $\bar{x}$ and s are not resistant to outliers

The t procedures are robust against non-Normality when there are no outliers ; CLT

Using t-Procedures

- SRS is more important than ~N. Except in the case of small samples:

  $n < 15$: Use t-procedures if ~N.

  If not N or outliers are present, do not use t

  $n \geq 15$: t-procedures may be used except in the presence of outliers or strong skewness

  large samples: t-procedures can be used even for clearly skewed when $n \geq 30$

Example 3 Can we use t-procedures? (Slides 24-

24 yes $n \geq 30$ Population

25 yes

26 yes, if $n \geq 30$ maybe ≥ 15?

vertical axis: count, %