Confidence Intervals and the t Distribution

Cohen Chapter 6

EDUC/PSY 6600

“It is common sense to take a method and try it.
If it fails, admit it frankly and try another.
But above all, try something.”

"

-- Franklin D. Roosevelt

Problems with z-tests

Often don't know σ^2, so we cannot compute SE_M, *Standard Error for the Mean* or $\sigma_{\bar{x}}$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

Can you use s in place of σ in $SE_{\bar{x}}$ and do z test?

- Small samples – No, **inaccurate** results
- Large samples – Yes (> 300 participants)

$$z = \frac{\bar{x} - \mu_x}{\frac{s}{\sqrt{n}}}$$

Small samples

As samples get smaller: $N \downarrow$

- the skewness of the sampling distribution of $s^2 \uparrow$
- s^2 **under**estimates σ^2
- z will $\uparrow$
- an overestimate $\uparrow$ risk of **Type I error**

Small samples

As samples get smaller: $N \downarrow$

- the skewness of the sampling distribution of $s^2 \uparrow$
- s^2 **under**estimates σ^2
- z will $\uparrow$
- an overestimate $\uparrow$ risk of **Type I error**

Comparatively... in LARGE samples

- s^2 **un**biased estimate of σ^2
- σ is a constant, *unknown truth*
- s is NOT a constant, since it varies from sample to sample
- As N increases, $s \to \sigma$

The t Distribution, “student’s t”


1908, William Gosset

- Guinness Brewing Company, England
- Invented `t-test` for **small** samples for brewing quality control

Wrote paper using moniker “a student” discussing nature of SE_M when using s^2 instead of σ^2

- Worked with Fisher, Neyman, Pearson, and Galton

Student's t & Normal Distributions

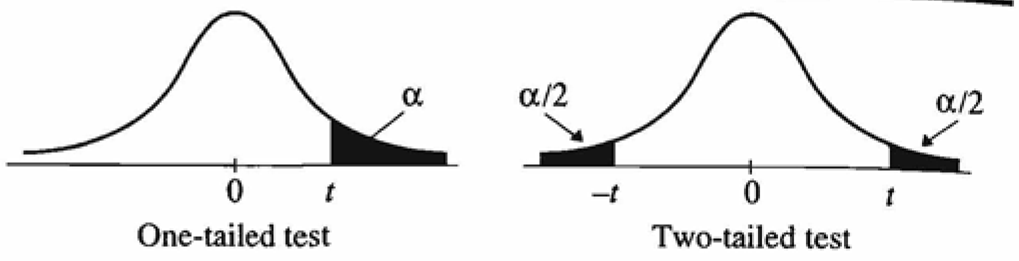
Similarities

- Follows mathematical function
- Symmetrical, continuous, bell-shaped
- Continues to $\pm$ infinity
- Mean: $M = 0$
- Area under curve = $p(event[s])$
- When N is **large** --- ≈ 300 --- $t = z$

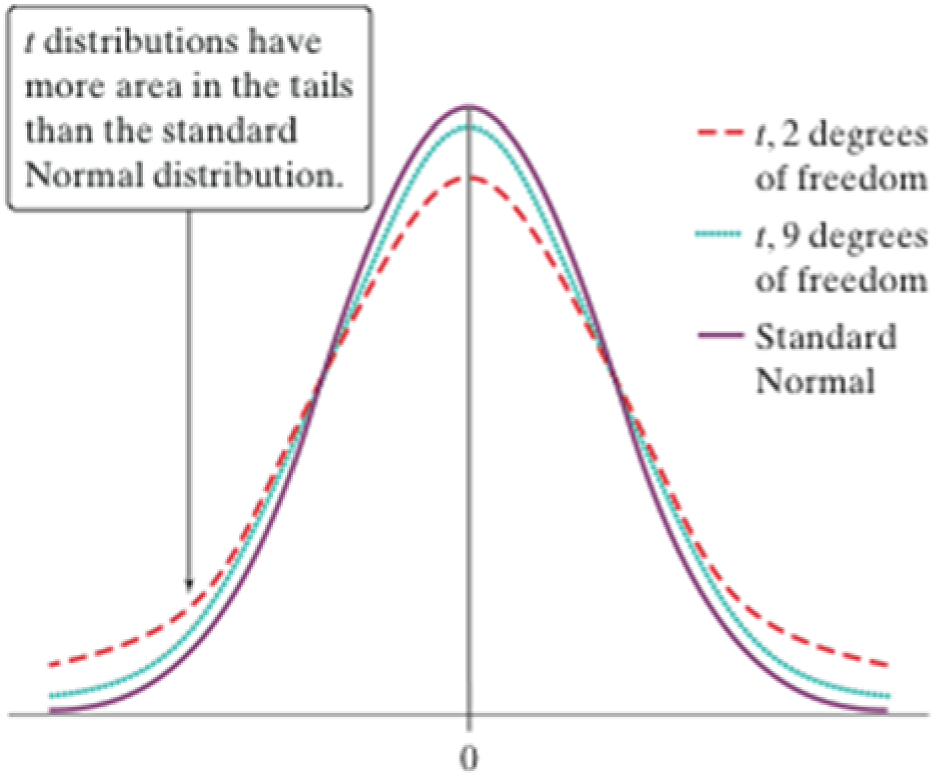
Differences

- Family of distributions
- Different distribution for each N (or df)
- Larger area in **tails** (%) for any value of t corresponding to z
  - $t_{cv} > z_{cv}$, for a given α
- More difficult to reject H_0 w/ t-distribution
  - $df = N - 1$
- As $df \uparrow$, the critical value of $t \rightarrow z$

The t Table



	Level of Significance for One-Tailed Test					
	.10	.05	.025	.01	.005	.0005
	Level of Significance for Two-Tailed Test					
df	**.20**	**.10**	**.05**	**.02**	**.01**	**.001**
1	3.078	6.314	12.706	31.821	63.657	636.620
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	[illegible]



Calculating the t-Statistic

x is interval/ratio data (ordinal okay: $\geq 10 - 16$ levels or values)

Like z, t-statistic represents a **SD** score (the # of SE's that $\bar{x}$ deviates from μ)

$$t = \frac{\bar{x} - \mu_x}{\frac{s_x}{\sqrt{N}}}$$

$$df = N - 1$$

When σ is known, t-statistic is sometimes computed (rather than z-statistic) if N is small

Estimate the population SE_M with sample data:

Estimated SE_M is the amount a sample's observed **mean**
may have deviated from
the true or population value
just due to random chance variation due to sampling.

Assumptions (same as z tests)

Sample was drawn at random (at least as representative as possible)

- Nothing can be done to fix NON-representative samples!
- Can not statistically test

Assumptions (same as z tests)

Sample was drawn at random (at least as representative as possible)

- Nothing can be done to fix NON-representative samples!
- Can not statistically test

SD of the sampled population = SD of the comparison population

- Very hard to check
- Can not statistically test

Assumptions (same as z tests)

Sample was drawn at random (at least as representative as possible)

- Nothing can be done to fix NON-representative samples!
- Can not statistically test

SD of the sampled population = SD of the comparison population

- Very hard to check
- Can not statistically test

Variables have a normal distribution

- Not as important if the sample is large (Central Limit Theorem)
- IF the sample is far from normal &/or small n, might want to transform variables
  - Look at plots: histogram, boxplot, & QQ plot (straight 45\degree line)
  - Skewness & Kurtosis: Divided value by its SE & $> \pm 2$ indicates issues
  - Shapiro-Wilks test (small N): p < .05 ??? not normal
  - Kolmogorov-Smirnov test (large N)

EX) 1 sample t Test: mean vs. *historic control*

A physician states that, in the past, the average number of times he saw each of his patients during the year was 5. However, he believes that his patients have visited him significantly **more frequently** during the past year. In order to validate this statement, he randomly selects 10 of his patients and determines the number of office visits during the past year. He obtains the values presented to the below.

9, 10, 8, 4, 8, 3, 0, 10, 15, 9

Do the data support his contention that the average number of times he has seen a patient in the last year is different that 5?

EX) 1 sample t Test: mean vs. *historic control*

```r
x = c(9, 10, 8, 4, 8, 3, 0, 10, 15, 9)

length(x)
```

```
[1] 10
```

```r
sum(x)
```

```
[1] 76
```

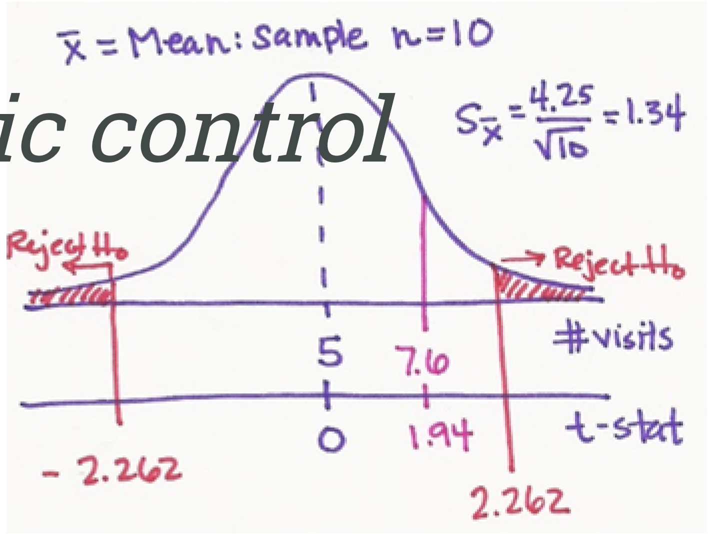
```r
mean(x)
```

```
[1] 7.6
```

```r
sd(x)
```

```
[1] 4.247875
```

EX) 1 sample t Test: mean vs. *historic control*



1. Null/Alt Hypotheses

$H_0: \mu = 5$

$H_1: \mu \neq 5$

2. Choose Test Stat, α, & # tails

CLT: mean of repeated SRS → normally dist. w/o pop SD known

→ So use the t-stat

$\alpha = .05$ & 2 tails (default)

3. SRS data → Sample Mean

$$\bar{X} = \frac{\sum X}{n} = \frac{76}{10} = \mathbf{7.6}$$

S = 4.25

4. Rejection Region?

df = n – 1 = 10 – 1 = 9

→ Critical t = +/- 2.262 ...

→ Reject if t-score is >2.262 or < -2.262

5. Calculate the Test Stat

Distribution of all sample means:

$$Mean_{mean} = \mu_{\bar{X}} = \mu_{pop} = 5$$

$$SE_{mean} = s_{\bar{X}} = \frac{s_{sample}}{\sqrt{n}} = \frac{4.25}{\sqrt{10}} = 1.34$$

$$t = \frac{\bar{X} - \mu_{\bar{X}}}{s_{\bar{X}}} = \frac{7.6-5}{1.34} = \mathbf{1.94}$$

6. Conclusion

t-stat does NOT falls in the rejection region

No evidence the population's mean is not 5

"FAIL to reject the Null"

"Even though this sample's mean of 7.6 was more than 5, this could be due to random chance and another sample may find the opposite effect."

Confidence Intervals

Statistics are point estimates, or *population parameters*, **with error**

How **close** is estimate to population parameter?

- Confidence interval (CI) around point estimate *(Range of values)*
  - Upper limit: UL or UCL
  - Lower limit: LL or LCL

CI expresses our **confidence** in a statistic & the *width* depends on SE_M and t_{cv}

- Both are function of N
  - Larger $N \rightarrow$ Smaller CI

- More confident that sample point estimate (statistic) approximates population parameter
  - Narrow CI: Less confidence, more precision *(less error)*
  - Wide CI: More confidence, less precision *(more error)*

Steps to Construct a Confidence interval

1. Select your random sample size
2. Select the **Level of Confidence**
    - Generally 95% *(can by 80, 90, or even 99%)*
3. Select random sample and collect data
4. Find the **Region of Rejection**
    - Based on $\alpha = 1 - Conf$ & # of tails = 2
5. Calculate the Interval **End Points**

$$Est \pm CV_{Est} \times SE_{Est}$$

Steps to Construct a Confidence interval

1. Select your random sample size
2. Select the **Level of Confidence**
   - Generally 95% *(can by 80, 90, or even 99%)*
3. Select random sample and collect data
4. Find the **Region of Rejection**
   - Based on $\alpha = 1 - Conf$ & # of tails = 2
5. Calculate the Interval **End Points**

$$Est \pm CV_{Est} \times SE_{Est}$$

Narrow CI:

- large smaple
- Lower %

Wider CI:

- smaller sample
- Higher %

95% CI with z score

$$\bar{x} \pm 1.96 \times \frac{\sigma}{\sqrt{N}}$$

99% CI with z score

$$\bar{x} \pm 2.58 \times \frac{\sigma}{\sqrt{N}}$$

EX) Confidence Interval: for a Mean

A physician states that, in the past, the average number of times he saw each of his patients during the year was 5. However, he believes that his patients have visited him significantly **more frequently** during the past year. In order to validate this statement, he randomly selects 10 of his patients and determines the number of office visits during the past year. He obtains the values presented to the below.

9, 10, 8, 4, 8, 3, 0, 10, 15, 9

Construct a 95% confidence interval for the mean number of visits per patient.

EX) Confidence Interval: for a Mean

A physician states that, in the past, the average number of times he saw each of his patients during the year was 5. However, he believes that his patients have visited him significantly **more frequently** during the past year. In order to validate this statement, he randomly selects 10 of his patients and determines the number of office visits during the past year. He obtains the values presented to the below.

9, 10, 8, 4, 8, 3, 0, 10, 15, 9

Construct a 95% confidence interval for the mean number of visits per patient.

1. Best estimate

$$\bar{X} = \frac{\sum X}{n} = \frac{76}{10} = \mathbf{7.6}$$

2. Critical Value

df = n – 1 = 10 – 1 = 9

Always use TWO tails

→ Critical t = 2.262

3. Standard Error for the Estimate

Sample standard deviation, S = 4.25

$$SE_{mean} = s_{\bar{X}} = \frac{s_{sample}}{\sqrt{n}} = \frac{4.25}{\sqrt{10}} = 1.34$$

4. Put it together

$$Est \pm CV \times SE_{est} \rightarrow$$

$$7.60 \pm 2.262 \times 1.34$$

$$7.60 \pm 3.03$$

$$\boxed{\mathbf{4.57,\ 10.63}}$$

Estimating the Population Mean

Point estimate (M) is in the center of CI

Degree of confidence determined by α and corresponding critical value (CV)

- Commonly use 95% CI, so $\alpha = .05$
- Can also compute a .90, .99, or any size CI

z-distribution:
Known population variance or N is large (about 300)

$$\bar{x} \pm z_{cv} \times \frac{\sigma}{\sqrt{N}}$$

t -distribution:
Do not know population variance or N is small

$$\bar{x} \pm t_{cv} \times \frac{s}{\sqrt{N}}$$

Estimating the Population Mean

Point estimate (M) is in the center of CI

Degree of confidence determined by α and corresponding critical value (CV)

- Commonly use 95% CI, so $\alpha = .05$
- Can also compute a .90, .99, or any size CI

z-distribution:
Known population variance or N is large (about 300)

$$\bar{x} \pm z_{cv} \times \frac{\sigma}{\sqrt{N}}$$

t -distribution:
Do not know population variance or N is small

$$\bar{x} \pm t_{cv} \times \frac{s}{\sqrt{N}}$$

NOT the meaning of a 95% CI

There is **NOT** a 95% chance that the population M lies between the 2 CLs from your sample's CI !!!

Each random sample will have a different CI with different CLs and a different M value

Meaning of a 95% CI

95% of the CIs that could be constructed over repeated sampling will contain M Yours **MAY** be one of them

5% chance our sample's 95% CI does not contain μ
Related to **Type I Error**

APA Style Writeup

Z-test

(happens to be a statistically significant difference)

The hourly fee (M = \$72) for our sample of current psychotherapists is significantly greater, $z = 4.0$, $p < .001$, than the 1960 hourly rate (M = \$63, in current dollars).

APA Style Writeup

Z-test

(happens to be a statistically significant difference)

The hourly fee (M = \$72) for our sample of current psychotherapists is significantly greater, $z = 4.0$, $p < .001$, than the 1960 hourly rate (M = \$63, in current dollars).

T-test

(happens to not quite reach .05 significance level)

Although the mean hourly fee for our sample of current psychotherapists was considerably higher (M = \$72, SD = 22.5) than the 1960 population mean (M = \$63, in current dollars), this difference only approached statistical significance, $t(24) = 2.00$, $p = .06$.

Let's Apply This to the Cancer Dataset

Read in the Data

```
library(tidyverse)     # Loads several very helpful 'tidy' packages
library(haven)         # Read in SPSS datasets
library(psych)         # Lots of nice tid-bits
library(car)           # Companion to "Applied Regression"
```

```
cancer_raw <- haven::read_spss("cancer.sav")
```

And Clean It

```
cancer_clean <- cancer_raw %>%
  dplyr::rename_all(tolower) %>%
  dplyr::mutate(id = factor(id)) %>%
  dplyr::mutate(trt = factor(trt,
                             labels = c("Placebo",
                                        "Aloe Juice"))) %>%
  dplyr::mutate(stage = factor(stage))
```

1 sample t Test vs. Historic Control

Do the patients weigh more than 165 pounds at intake, on average?

```
cancer_clean %>%
  dplyr::pull(weighin) %>%
  t.test(mu = 165)
```

```

    One Sample t-test

data:  .
t = 2.0765, df = 24, p-value = 0.04872
alternative hypothesis: true mean is not equal to 165
95 percent confidence interval:
 165.0807 191.4793
sample estimates:
mean of x
   178.28
```

...Change the Confidence Level

Find a 99% confience level for the population mean weight.

```
cancer_clean %>%
  dplyr::pull(weighin) %>%
  t.test(mu = 165,
         conf.level = 0.99)
```

```
    One Sample t-test

data:  .
t = 2.0765, df = 24, p-value = 0.04872
alternative hypothesis: true mean is not equal to 165
99 percent confidence interval:
 160.3927 196.1673
sample estimates:
mean of x
   178.28
```

...Restrict to a Subsample

Do the patients with stage 3 & 4 cancer weigh more than 165 pounds at intake, on average?

```
cancer_clean %>%
  dplyr::filter(stage %in% c("3", "4")) %>%
  dplyr::pull(weighin) %>%
  t.test(mu = 165)
```

```

	One Sample t-test

data:  .
t = 0.82627, df = 5, p-value = 0.4463
alternative hypothesis: true mean is not equal to 165
95 percent confidence interval:
 137.0283 219.4717
sample estimates:
mean of x 
   178.25 
```

Questions?

Next Topic

Independent Samples t Tests for Means