Outline for Ch. 7

1. Underlying logic of null hypothesis significance testing (NHST)
2. Steps of NHST
3. Common misunderstandings related to NHST
4. **Effect sizes (p. 140)**
5. Using confidence intervals for hypothesis testing (Ch. 8 pp. 154-158)
6. Types of Errors when engaging in NHST

Effect Sizes (p. 140)

- Standardized measures of the magnitude of an *effect* (difference/change/relationship)

- Example: If blueberry supplements have a statistically significant effect on cognitive functioning . . .
  - they could have a large effect (e.g., improve functioning *a lot* $\bar{\mathbf{X}}$ = 84)
  - they could have a small effect (e.g., improve it, but *just by a little bit* $\bar{\mathbf{X}}$=81)

Effect Sizes

- Standardized measures of the magnitude of an effect
  - *Standardized* implies that effect sizes can be compared across studies

- Why is it important to be able to compare across studies?
  - EX 1: Effect of blueberry supplements on cognitive functioning . . .
    - when cog functioning is measured by Test X, which has a mean of 80 (Study 1)
    - when cog functioning is measured by Test Y, which has a mean of 15 (Study 2)
  - EX 2: Effect of blueberry supplements on physical diseases
    - progression of cancer
    - progression of heart disease

Effect Sizes

- Standardized measures of the magnitude of an effect
  - *Standardized* implies that effect sizes are comparable across studies
- **Take home point: When your data reveal a statistically significant result, you will also calculate & report the effect size!**
- There are several effect size measures that can be used, for example
  - Cohen's *d*, Pearson's *r*
    - We'll return to these later in the course.

Outline for Ch. 7

1. Underlying logic of null hypothesis significance testing (NHST)
2. Steps of NHST
3. Common misunderstandings related to NHST
4. Effect sizes
5. **Using confidence intervals for hypothesis testing (Ch. 8 pp. 154-158)**
6. Types of Errors when engaging in NHST

Some Important Ideas That You Already Know

We can use a sample statistic (e.g., x-bar) as an estimate of the population parameter (μ).

Even a randomly-drawn sample's mean won't perfectly match the population mean.

We can calculate the *SE* from one sample of data.

SE tells us how much our sample deviates from other samples drawn from the same population, and deviates from the population mean.

NEW: We can use the *SE* to calculate boundaries or an interval that we think contains the true population mean.

confidence intervals

Confidence intervals (CI) around the mean – first, some definitions

- CIs around the mean: **boundaries** constructed such that for a certain percentage of samples (e.g., 95% of samples), the boundaries constructed will capture the true value of the population mean

- In easier terms, "**a 95% confidence interval around the mean**" can be interpreted as . . .

  we are 95% confident that the interval (the boundaries) contains the true population mean.

To determine these boundaries we need to know: the **point estimate** and the **margin of error**

See handout

What is meant by the point estimate?

- in this case, the *sample mean* (i.e., our best estimate of the true pop. mean)
- center of interval

What is meant by the margin of error?

- distance between point estimate and boundary line
- calculated as: **the critical value x *SE***
- and how do we know the critical value?
    - For **95**% CIs, the critical value for z is **1.96**
    - For **99**% CIs, the critical value for z is **2.58**

Calculating CIs

APA style: 95% CI [9.61, 10.39]

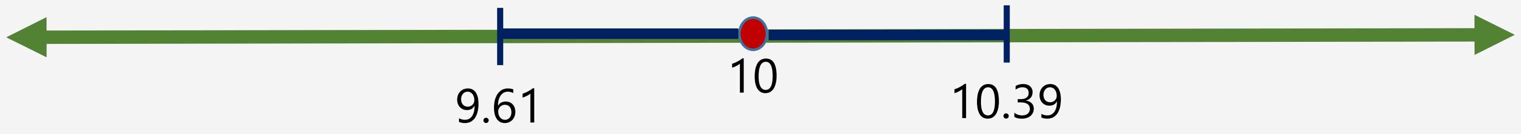
- lower bound of CI = point estimate *minus* margin of error
- upper bound of CI = point estimate *plus* margin of error

If our sample mean = 10, & *SE* = 0.20, & we wanted to find the 95% CI

- first calculate the **margin of error:** crit value x *SE* = 1.96 x 0.20 = **0.392**
- then subtract/add the margin of error from/to our point estimate of the mean

Lower bound
10 – 0.392 = **9.61**

Upper bound
10 + 0.392 = **10.39**



As the *SE* gets larger, our **interval** . . . also gets larger

If our sample mean = 10 & ***SE* = 1.10**, and we wanted to find the 95% CI

Lower bound
10 – 2.156= **7.84**

Upper bound
10 + 2.156 = **12.16**

If our sample mean = 10 & ***SE* = 1.80**, and we wanted to find the 95% CI

Lower bound
10 – 3.528= **6.47**

Upper bound
10 + 3.528 = **13.53**

If our sample mean = 10 & SE = 0.20, and we wanted to find the 95% CI

- first calculate the margin of error (1.96 x 0.20= **0.392**)
- then subtract and add the margin of error to our point estimate of the mean

Lower bound
10 – 0.392 = **9.61**

Upper bound
10 + 0.392 = **10.39**

Interpreting 95% vs. 99% CIs

95% CI: We are **95% confident** that the interval contains the true population mean.

We saw that if our sample mean = 10 & *SE* = 1.10, the 95% CI ranges from **7.84 to 12.16**

99% CI: We are **99% confident** that the interval contains the true population mean.

If our sample mean = 10 & *SE* = 1.10, what is the ***99% CI***?

2.58 x 1.10 = 2.838 (margin of error)

Lower bound
10 – 2.838= **7.16**

Upper bound
10 + 2.838 = **12.84**

99% CIs will be larger than 95% CIs.

Practice with Confidence Intervals

1) 0.474
2) 1.222
3) Lower = 6.28 Upper = 8.72

A sample of n = 10 students has seen, on average, 7.5 Harry Potter movies. The standard deviation (s) for the sample is 1.50. Determine:

1. The standard error (round to 3 decimal places)
2. The margin of error for the 99% CI (round to 3 dec places; crit val = 2.58)
3. The lower and upper bound of the 99% CI (round to 2 dec places)
4. Finally, *interpret* the meaning of the 99% CI in this example. (Translate the #s into words.)

4) We are 99% certain that 6.28 and 8.72 captures the true mean number of HP movies seen by students in the population.

True or False – correct any false statements.

1. If your standard error is larger, it means your point estimate is less accurate.
2. If your standard error is larger, it means that your confidence intervals will be smaller.

1 True
2. False – your CIs will be *wider* (larger).

How do you use CIs for hypothesis testing?

- Assess whether or not the null hypothesis is credible, given the CI.
- EX. Blueberry-supplements
  - Null hyp: The mean of the b-s population is 80.
- Use your data to calculate the CI and ask, *does the CI contain 80?*
  - **Yes**, the CI contains 80 → retain the null.
    - Null is credible. B-s population mean *might be* 80
  - **No**, the CI does **not contain** 80 → reject the null.
    - Conclude it’s unlikely that the mean test score in b-s population is 80.

Example of all four steps using CIs

Dr. Valenti wants to compare her overall student evaluation scores from fall 2020, to the population of all her evaluations from prior semesters ($\mu = 1.50$)

1. State hypotheses H_0: $\mu_{fall2020} = 1.50$ / H_1: $\mu_{fall2020} \neq 1.50$
2. Set significance/alpha level. $\alpha = .05$ or $\alpha = .01$
3. Use statistical program to calculate test statistic from fall 2020 data, and determine corresponding p-value. Let's assume test statistic = 1.09, and $p = .69$
   - Also use fall 2020 data to calculate CI. Assume I calculated… 95% CI [1.10, 1.90]

4. Make decision to retain or reject null using p-values and CIs.

- $p = .69$, which is **greater than α** = .05, meaning I should **retain** the null.
- Also, 95% CI [1.10, 1.90] **contains the value associated w/the null** hypothesis, 1.50, within the boundaries. The true, fall 2020 mean *could be* 1.50, which also suggests that I should retain the null.

I conclude that there is no evidence that Dr. V's fall 2020 evaluations are significantly different from her typical scores.

Example of all four steps using CIs

Dr. Valenti wants to compare her overall student evaluation score from fall 2020, to the population of all her evaluations from prior semesters (μ =1.50)

1. State hypotheses H_0: $\mu_{fall2020}$= 1.50 / H_1: $\mu_{fall2020} \neq$ 1.50
2. Set significance/alpha level. α = .05 or α = .01

3. Use statistical program to calculate test statistic from your fall 2020 data, and determine corresponding p-value. Let's assume test statistic = 2.80, and p = .005

   - Also use fall 2020 data to calculate CI. Assume I calculated… 95% CI [1.60, 1.90]

4. Make decision to retain or reject null using p-values & CIs. Describe thought process:

- p = .03, which is less than α = .05 , meaning I should reject the null.
- Also, 95% CI [1.60, 1.90] *does not contain* the value associated w/the null hypothesis, 1.50, within the boundaries. This also means I should reject the null.
- **There is evidence that Dr. V's fall 2020 evaluations are significantly different from her typical scores.**

Outline for Ch. 7

1. Underlying logic of null hypothesis significance testing (NHST)
2. Steps of NHST
3. Common misunderstandings related to NHST
4. Effect sizes
5. Using confidence intervals for hypothesis testing (Ch. 8 pp. 154-158)
6. **Types of Errors when engaging in NHST**

Reality (we'll never know this)

	H_0 is actually true (no differences, no effect)	H_0 is actually false (there *are* differences, H_1 true)

Reality (we'll never know this)

		H_0 is actually true (no differences, no effect)	H_0 is actually false (there *are* differences, H_1 true)
Researcher's decision *(based on test statistic & p-value)*	I conclude H_0 is credible (b/c $p \geq \alpha$)		
	I reject H_0 (conclude H_0 NOT credible, b/c $p < \alpha$)		

		Reality (we'll never know this)	
		H_0 is actually true (no differences, no effect)	H_0 is actually false (there *are* differences, H_1 true)
Researcher's decision *(based on test statistic & p-value)*	I conclude H_0 is credible (b/c $p \geq \alpha$)	*correct decision*	
	I reject H_0 (conclude H_0 NOT credible, b/c $p<\alpha$)		*correct decision*

Reality (we'll never know this)

		H_0 is actually true (no differences, no effect)	H_0 is actually false (there *are* differences, H_1 true)
Researcher's decision *(based on test statistic, p-value)*	I conclude H_0 is credible (b/c $p \geq \alpha$)	*correct decision*	
	I reject H_0 (conclude H_0 NOT credible, b/c $p<\alpha$)	**Type I Error** (false positive)	*correct decision*

Reality (we'll never know this)

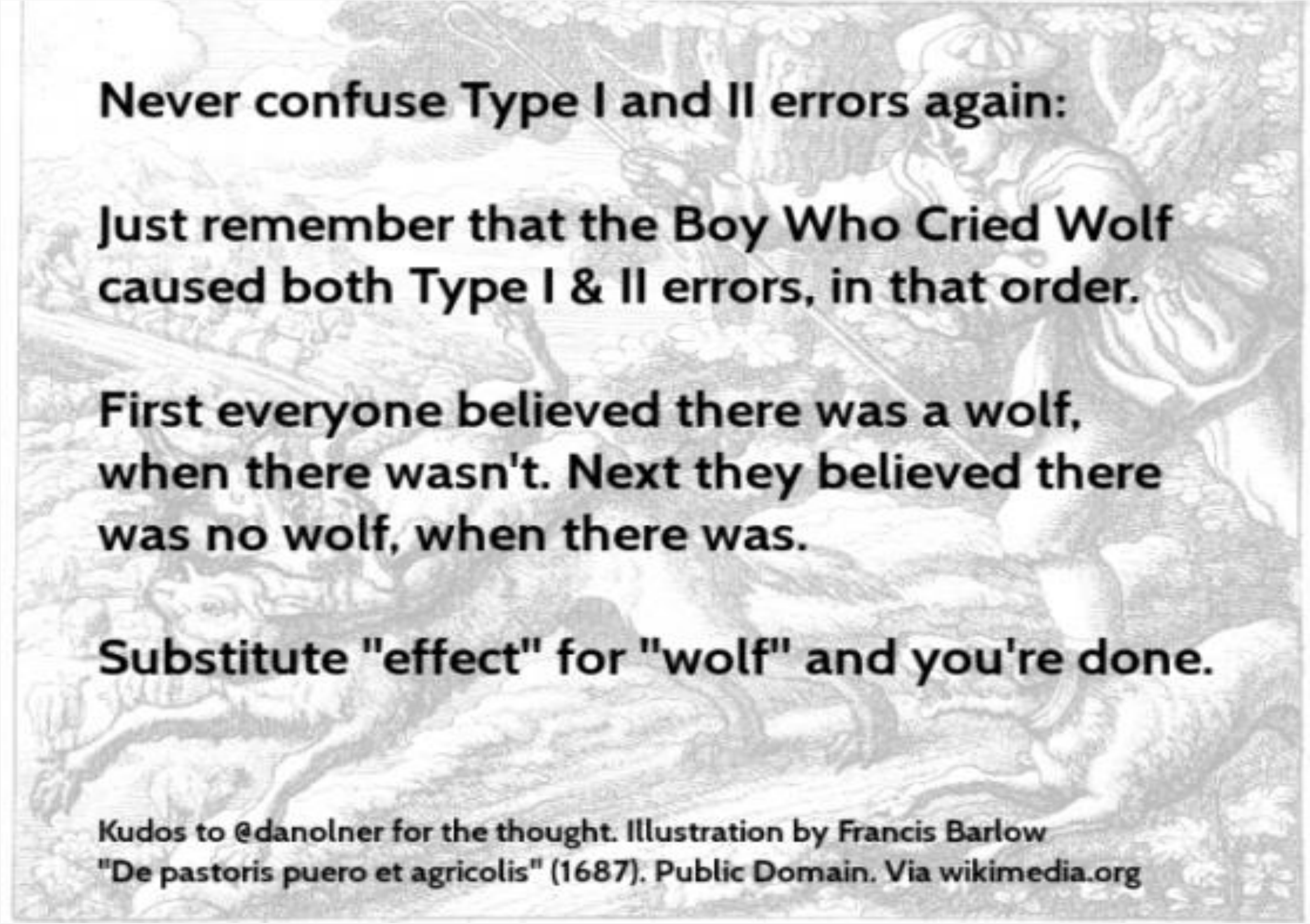
		H_0 is actually true (no differences, no effect)	H_0 is actually false (there *are* differences, H_1 true)
Researcher's decision *(based on test statistic, p-value)*	I conclude H_0 is credible (b/c $p \geq \alpha$)	*correct decision*	**Type II Error** (false negative)
	I reject H_0 (conclude H_0 NOT credible, b/c $p < \alpha$)	**Type I Error** (false positive)	*correct decision*

Possible Errors in Hypothesis Testing

After computing a test statistic, and examining its p-value and CIs, researchers may make one of two errors:

- **Type I error**: The researcher concludes there **is** a significant effect (difference, relationship) when, in fact, there is no true effect.
  - E.g., Conclude that blueberries had an effect on test scores, when they did not.
- **Type II error**: The researcher concludes there is **no evidence** for a significant effect (difference, relationship) when, in fact, there is a true effect.
  - E.g., Conclude that blueberries had no effect on test scores, when they did.

One way to remember the difference btwn Type I & Type II errors



Class Business

- We'll finish Ch. 7 today & introduce Ch. 9-10
- By Tuesday, make sure to be up to date on the reading for all the material since exam 1 (Ch. 6, 7, portions of Ch. 8, and Ch. 9-10).
- There are self-graded HWs on Moodle for Ch. 6 & 7.
  - Consider completing these over the weekend & coming to office hours Monday (2-3 pm) to ask questions!
- Academic Resource Center PY hours are here: https://www.bsc.edu/academics/arc/index.html
- Next week we have class Tuesday and Wednesday. Use Thursday as a mental health day for our class.