Wrapping up and review of Parts I and II

Objectives

- revisit the original goals of the course and check in
- suggest some strategies for final exam
- broad overview of part III

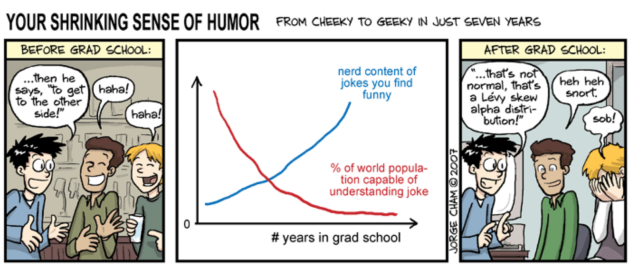
General wrap up

Goals for the semester

In addition the the learning objectives listed in your syllabus our overarching goals for the semester are to develop:

- your ability to critically assess statistical information presented to you in scientific and non-scientific fora
- your sense of how to approach answering real world questions with data
- your ability to concisely and accurately describe statistical methods and results

Goals for the semester



***footnote:** Thanks to Daragh at George Mason U. for this comic idea.*

Day 1 argument: This is a relevant class

I hoped to convince everyone here that statistics is relevant to everyone

You make many decisions during your day that are influenced by statistics

Statistics is not just relevant for public health, but also for other professions, including: policy, journalism and law

As we have tried to illustrate via the recurring "statistics is everywhere" segments, statistics is useful for understanding the news and the world around us - certainly during this pandemic we have seen a lot of public health and statistics in the news.

Non-statistical considerations

Statistics tells us about the role of chance - what would happen over many repeated samples

We want to think about the quality of the study design, what population was studied and whether the results are generalizable, sources of potential biase...

And remember that in our interpretation process we want to think about not just the statistical results.... - how meaningful is the effect - if the results suggest a relationship what are the risks and benefits of the exposure/treatment - what are the costs or implications of changing recommendations or guidelines?

Interpreting evidence: Mask use study published February 4, 2022

Effectiveness of Face Mask or Respirator Use in Indoor Public Settings for Prevention of SARS-CoV-2 Infection - California, February-December 2021
Kristin L. Andrejko, Jake M. Pry, Jennifer F. Myers, Nozomi Fukui, Jennifer L. DeGuzman, John Openshaw, James P. Watt, Joseph A. Lewnard, Seema Jain, California COVID-19 Case-Control Study Team

Test negative design case control study - matched observations.

Comparing self reported mask use between cases and controls.

Any face mask or respirator use in indoor public settings was associated with significantly lower odds of a positive test result compared with never using a face mask or respirator (aOR = 0.51; 95% CI = 0.29–0.93). Always using a face mask or respirator in indoor public settings was associated with lower adjusted odds of a positive test result compared with never wearing a face mask or respirator (aOR = 0.44; 95% CI = 0.24–0.82);

Interpreting evidence: Mask use study published February 4, 2022

TABLE 2. Face mask or respirator use in indoor public settings among persons with positive and negative SARS-CoV-2 test results — California, February–December 2021

	SARS-CoV-2 infection status, no. (%)		Odds ratio (95% CI)	
Mask type and use*	Positive (case-participant) N = 652	Negative (control-participant) N = 1,176	Unadjusted† [p-value]	Adjusted§ [p-value]
None (Ref)	44 (6.7)	42 (3.6)	—	—
Any use†	608 (93.3)	1,134 (96.4)	0.57 (0.37–0.90) [0.02]	0.51 (0.29–0.93) [0.03]
Some of the time	62 (9.5)	76 (6.5)	0.81 (0.47–1.41) [0.49]	0.71 (0.35–1.46) [0.36]
Most of the time	153 (23.5)	239 (20.3)	0.64 (0.40–1.05) [0.08]	0.55 (0.29–1.05) [0.07]
All of the time	393 (60.3)	819 (69.6)	0.49 (0.31–0.78) [<0.01]	0.44 (0.24–0.82) [<0.01]

Abbreviation: Ref = referent group.

* Trained interviewers administered a structured telephone-based questionnaire and asked participants to indicate whether they attended indoor public spaces during the 2 weeks before seeking a SARS-CoV-2 test. Participants who indicated attending these settings were further asked to specify whether they typically wore a face mask or respirator all, most, some, or none of the time while in these settings.

† Conditional logistic regression models were used to estimate the unadjusted odds of mask use by type of face mask or respirator worn in indoor public settings during the 2 weeks before testing. Models included matching strata defined by (for the period before June 15, 2021) the reopening tier of California in the county of residence and the week of SARS-CoV-2 testing.

§ Conditional logistic regression models were used to estimate the odds of face mask or respirator use in indoor public settings during the 2 weeks before testing, adjusting for COVID-19 vaccination status, household income, race/ethnicity, age group, sex, state region, and county population density. All models included matching strata defined by (for the period before June 15, 2021) the reopening tier of California in the county of residence, and the week of SARS-CoV-2 testing. To understand the effects of masking in community settings, this analysis was restricted to a subset of persons who did not indicate a known or suspected exposure to a SARS-CoV-2 case within 14 days of seeking a SARS-CoV-2 test. Adjusted models used a complete case analysis (454 case-participants and 789 control-participants). A sensitivity analysis using multiple imputation of missing covariate values obtained results similar to those reported in the table: adjusted odds ratios were 0.54 (95% CI = 0.33–0.89) for any mask use, 0.44 (95% CI = 0.27–0.73) for mask use all of the time, 0.62 (95% CI = 0.37–1.04) for mask use most of the time, and 0.77 (95% CI = 0.43–1.40) for mask use some of the time. An additional sensitivity analysis was conducted with additional adjustment for the reasons for SARS-CoV-2 testing as listed in Table 1 (experiencing symptoms, testing required for medical procedure, routine screening through work or school, pre-travel test, just wanted to see if I was infected, test required for admission to an event or gathering). The adjusted odds ratio was 0.42 (95% CI = 0.20–0.89) for any mask use as compared to no mask use upon additional adjustment for testing indications.

Recomendations/reminders

Extra credit

Please submit your course evaluations - only the screenshot that shows your name and the class - for an extra point on your final exam.

If we get to a 90% response rate for the evaluations, the whole class will get an extra point on the final.

Extra credit assignments: only one per student, due April 26th

Worth up to 2 points on your overall grade

Study recommendations

Focus on part III but understand parts I and II

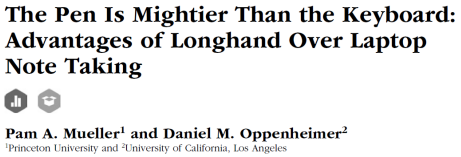
Make sure you understand where and why you lost points on previous exams/quizzes

Try explaining your answers to another student

Do what works for you - take suggestions in context of your experience

Evidence based suggestions: longhand notes

Pam A. Mueller and Dan Oppenheimer Psychological Science 2014, Vol. 25(6) 1159 -1168



The Pen Is Mightier Than the Keyboard: Advantages of Longhand Over Laptop Note Taking

Pam A. Mueller[1] and Daniel M. Oppenheimer[2]
[1]Princeton University and [2]University of California, Los Angeles

Evidence based suggestions: longhand notes

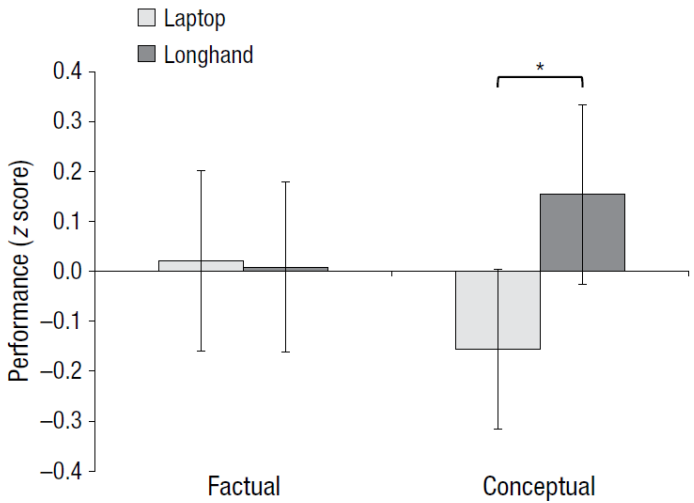
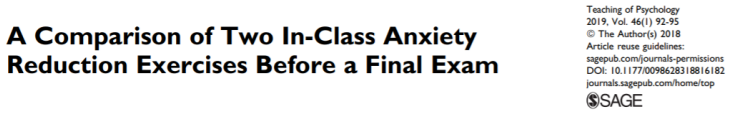


Fig. 1. Mean z-scored performance on factual-recall and conceptual-application questions as a function of note-taking condition (Study 1).

Evidence based suggestions: anxiety reduction

Virginia Clinton and Stacy Meester Teaching of Psychology 2019. vol 26(1)92-95



A Comparison of Two In-Class Anxiety Reduction Exercises Before a Final Exam

Teaching of Psychology
2019, Vol. 46(1) 92-95

DOI: 10.1177/0098628318816182
journals.sagepub.com/home/top
SAGE

Virginia Clinton[1] and Stacy Meester[2]

Evidence based suggestions: anxiety reduction

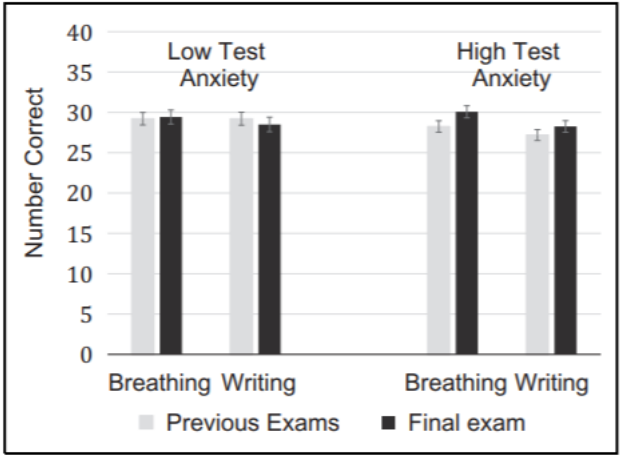


Figure 1. Previous exam and final exam performance by condition and level of trait test anxiety (means and $\pm$ 1 *SE*).

Selected review of parts I and II

Conditional and marginal proportions/probabilities: table

exposure	Concussion	No Concussion	Total
Helmet	10	90	100
No Helmet	140	160	300
Total	150	250	400

Conditional and marginal proportions/probabilities: ven diagram

Conditional and marginal proportions/probabilities: screening tests

Conditional and marginal proportions/probabilities: power

Study desing from an abstract

Background Socioeconomic status in the risk of developing type 1 diabetes seems inconsistent. We investigated whether risk of childhood-onset type 1 diabetes differed by parental education or occupation in a nationwide cohort.

Methods This cohort study included all children born in Norway from 1974 to 2013. In individually linked data from nationwide population registries following children born in Norway up to 15 years of age, we identified 4647 with newly diagnosed type 1 diabetes during 15,381,923 person-years of follow-up.

Study design? Exposure and outcome variables?

Study design

We randomly assigned patients who had heart failure with preserved ejection fraction, a body-mass index (the weight in kilograms divided by the square of the height in meters) of 30 or more, and type 2 diabetes to receive once-weekly semaglutide (2.4 mg) or placebo for 52 weeks. The primary end points were the change from baseline in the Kansas City Cardiomyopathy Questionnaire clinical summary score (KCCQ-CSS; scores range from 0 to 100, with higher scores indicating fewer symptoms and physical limitations) and the change in body weight. Confirmatory secondary end points included the change in 6-minute walk distance; a hierarchical composite end point that included death, heart failure events, and differences in the change in the KCCQ-CSS and 6-minute walk distance; and the change in the C-reactive protein (CRP) level.

Confounders

Sampling distribution vs data distribution vs test distribution

Recap - confidence intervals and testing

How confidence intervals behave

Recall the form of a CI:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

Where $z^* \frac{\sigma}{\sqrt{n}}$ is the margin of error.

How confidence intervals behave

The margin of error gets smaller when:

- z* is smaller (i.e., you change to a smaller confidence level). Thus, there is a trade-off between the confidence level and the margin of error.
- σ is smaller. You might be able to reduce σ if there is measurement error. Often times, the σ can't be reduced, it is just a characteristic of the population
- n is larger.

How hypothesis tests behave

- Statistical significance depends on sample size (since sample size determines the standard error of the sampling mean)
- Recall the form of the z-test:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\text{magnitude of observed effect}}{\text{size of chance variation}} = \frac{\textit{signal}}{\textit{noise}}$$

- The numerator quantifies the distance between what you observe in the sample and the null hypothesized parameter.
- The denominator represents the size of chance variations from sample to sample

How hypothesis tests behave

- Statistical significance depends on:
  - The size of the observed effect ($\bar{x} - \mu$)
  - The variability of individuals in the population (σ)
  - The sample size (n)
  - Your criteria for rejection the null (α)

If you obtain a small p-value it is not necessarily because the effect size is large.

Type I error, and Type II error in hypothesis tests

	H_a is true	H_0 is true
Reject H_0	Correct decision	Type I error (α)
Fail to reject H_0	Type II error (β)	Correct decision

This table should remind you of something we have seen before....

Power

- The power is the chance of making the correct decision when the alternative hypothesis is true.
- Thus, it is the complement of β
- Power $= 1 - \beta$

	H_a is true	H_0 is true
Reject H_0	Correct decision	Type I error (α)
Fail to reject H_0	Type II error (β)	Correct decision

However, there are an infinite number of possible values that μ could assume that are not $= \mu_0$

Thus we must choose a value at which to evaluate the β and power for an alternative hypothesis...

When we evaluate β we do so at a single such value μ_1

Calculating power in R

Example of calculating power

Suppose you have a known standard deviation $\sigma = 1$. $H_0 : \mu = 0$ vs. $H_a : \mu > 0.8$ and choose $\alpha = 0.05$. Calculate the power when $n = 10$.

You can calculate the minimum z-value required to reject H_0:

```
qnorm(p = 0.05, mean = 0, sd = 1/sqrt(10), lower.tail = F)
```

```
## [1] 0.5201484
```

So for any z-test with this value or higher, you will reject H_0 in favor of H_a.

This is often called Z_α

Example of calculating power

Now suppose that H_a is true. The test will reject H_0 about what percent of the time when H_a is true? To calculate this probability, we take the value from the previous calculation and calculate the *probability* above its value under H_a:

```
pnorm(q = 0.5201484, mean = 0.8, 1/sqrt(10), lower.tail = F)
```

```
## [1] 0.8119132
```

Example of calculating power, illustrated

Size of the difference

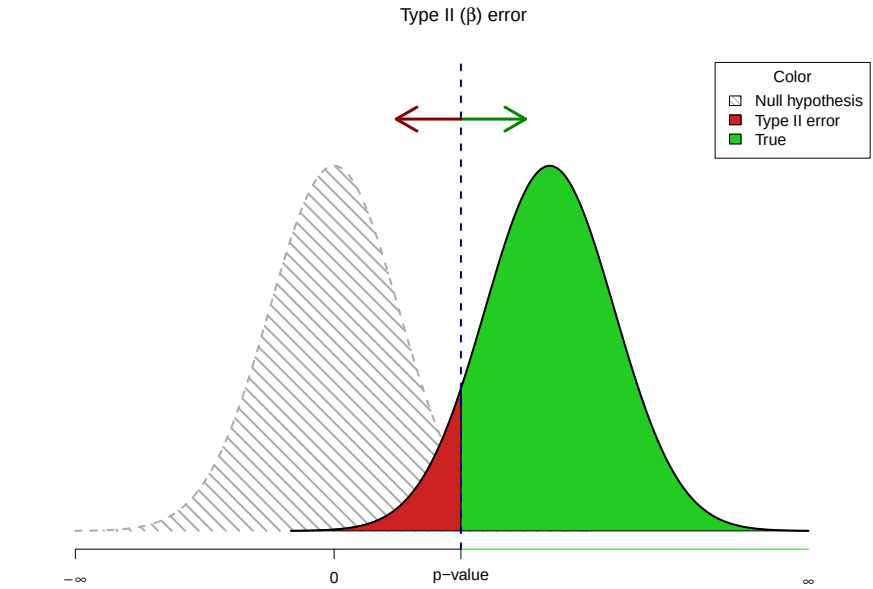
Effect of changing the μ_1

Imagine we our H_0 is a standard normal (mean=0, SD=1) and we set our α at 0.05.

If the true mean of our sampled population is 1.7 standard deviations above the μ_0,

what does our β look like?

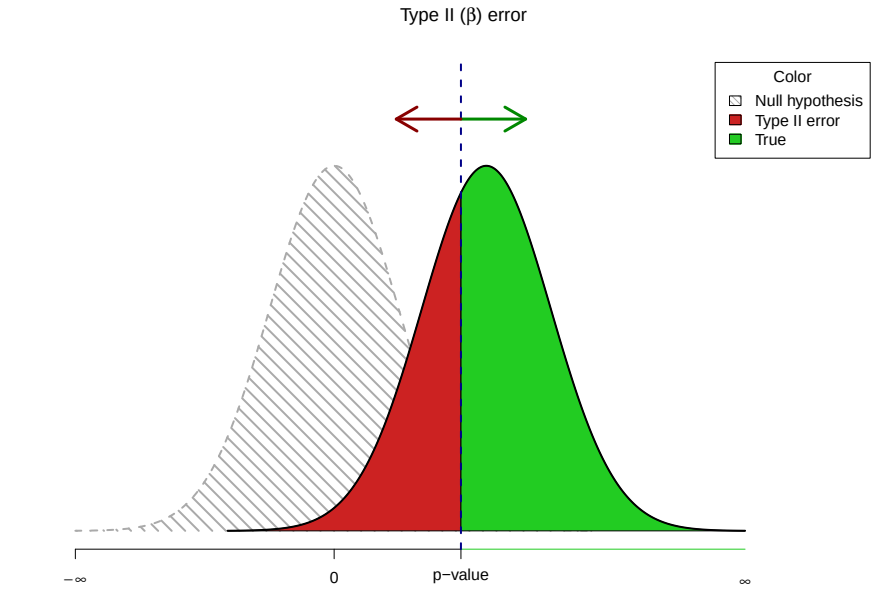
Effect of changing the μ_1



Effect of changing the μ_1

What happens if the “true” mean is closer to the Null?

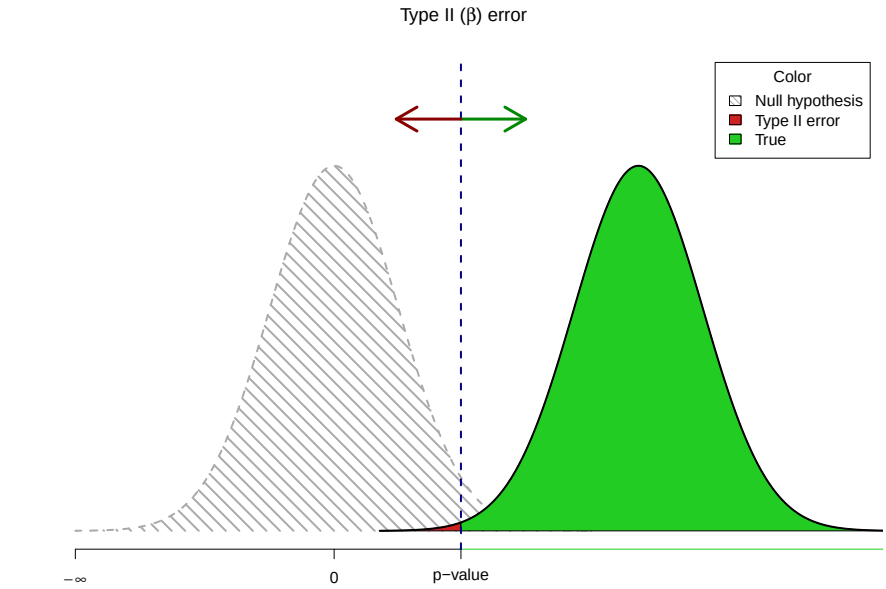
Effect of changing the μ_1



Effect of changing the μ_1

What happens if the “true” mean is further from the null?

Effect of changing the μ_1



Example

Lecture 21 has a worked example from the Pagano text that is worth reviewing Here we will go through another example from the Baldi and Moore textbook. This example assumes you are planning a quality control study to look at whether storage impacts the percieved sweetness of a beverage. Ten professional tasters will rate the sweetness on a 10 point scale before and after storage. We know that the standard deviation of sweetness ratings is $= 1$. We also know that a mean sweetness change of 0.8 on this scale is noticed by consumers. We want 90% power and an alpha of 0.05 for our study. We have a set of 10 values representing the difference in sweetness caused by storage.

What is the null hypothesis here?

What is our alternative?

Is our hypothesis one or two sided?

Example

We will start by finding the Z alpha:

$$Z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

```
qnorm(.05)
```

```
## [1] -1.644854
```

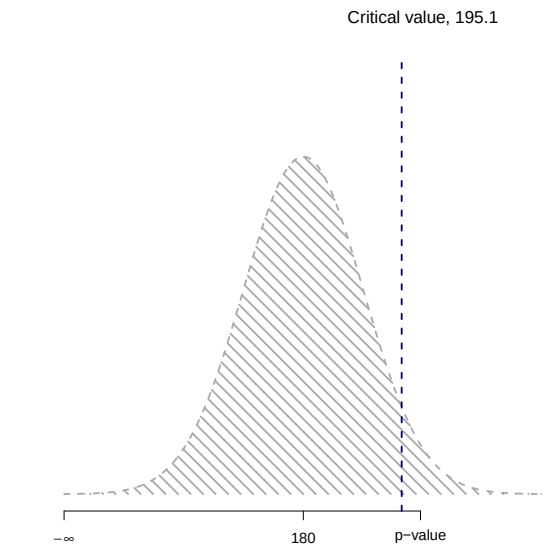
$$-1.645 = \frac{\overline{x} - 0}{\frac{1}{\sqrt{10}}}$$

Solve this for $\overline{x}$

$$\bar{x} = -1.645 \times \frac{1}{\sqrt{10}} = -0.522$$

null distribution

So here we have our null distribution with the value at which we reject the null



Example

We must choose a value at which to evaluate β. Here we will choose an alternate hypothesis that the mean sweetness difference is -0.8. Since we know a sample mean greater than -0.522 causes us to fail to reject H_0 we need to calculate the proportion of a distribution centered at 0.8 that would be below this value.

$$Z = \frac{-0.8}{\frac{1}{\sqrt{10}}}$$

$$Z = -0.253$$

Example:

Using R to calculate the probability,

```
pnorm(-0.253, mean=0.8)
```

```
## [1] 0.1461705
```

Thus β P(do not reject null(0)|Null is false (true sweetness change is -0.8)) is $\sim$ 0.146

Remember that Power is 1-β = P(reject null | null is false)

In this example, Power is 1-0.146 or $\sim$ 0.854

Calculating sample size

Calculating sample size

To think about sample size for a z-test (or more generally for any test), four things matter:

- Significance level α : How much protection do we want against getting a statistical significant results from our sample when there really is no effect in the population?
- Effect size: How large an effect in the population is important in practice?
- Power $(1 - \beta)$: How confident do we want to be that our study will detect an effect of the size we think is important? I.e., what is the probability of rejecting H_0 when the alternative hypothesis is true?
- variability in the population: Remember that the underlying variability in our population affects the variability of our sample mean

sample size

When we are calculating sample size the steps we follow are: - Find the Z alpha and use this to calculate the value of our variable at which we would reject the null. - Find Z beta and use that to calculate what value this would be on the curve of the alternative hypothesis - Set these values equal to each other and solve for n

Example

For example, using our previous study of mean serum cholesterol levels, if we remember that we assumed the following:

$H_0 : \mu \leq 180mg/100ml$

α: 0.01

σ: 46

If the true population mean is as large as 211 and we want to risk only a 5% chance of failing to reject the null, so $\beta = 0.05$ and power would be $= 1 - \beta = 0.95$

###Calculating Sample Size We start by finding the Z value at which we would reject H_0 at $\alpha = 0.01$

We call this value Z_α

```
qnorm(0.01, lower=FALSE)
```

```
## [1] 2.326348
```

Determining sample size for a margin of error

Determining sample size for a margin of error

You can also determine the sample size you will need to have a given margin of error. Suppose you want your margin of error to equal m. What sample size do you need to obtain a margin of error of m?

You can re-frame the sample size formula for a two sided hypothesis test

$$n = \left(\frac{(Z_{(\frac{\alpha}{2})} + Z_{\beta}) * (\sigma)}{(\mu_1 - \mu_0)} \right)^2$$

to a margin of error

$$n = \left(\frac{Z^* \sigma}{m} \right)^2$$

Example of calculation sample size

Body temperature has a known $\sigma = 0.6$ degrees F. We want to estimate the mean body temperature μ for healthy adults within ± 0.05 F with 95% confidence. How many healthy adults must we measure?

$$n = \left(\frac{z^* \sigma}{m}\right)^2$$

$$n = \left(\frac{1.96 \times 0.6}{0.05}\right)^2 = 553.2$$

We must recruit 554 (round up!) healthy adults for this study.

Common Issues or questions

Parametric vs non-parametric

What does it mean when I say non-parametric?

Which tests have we covered that are non-parametric?

Common Issues

method used vs P-value and CI vs. conclusion

Appropriate interpretation of a P-value and CI

Relationship between a confidence interval and a p-value