PHP 2510 Principles of Biostatistics & Data Analysis

Week 8 & 9: Confidence Intervals & more Hypothesis Testing

This Week's Plan

1. **Motivation**
   - Literature
2. **Derivation**
   - General Form
   - Attributes
3. **Simulations**
   - Coverage
   - Other Discussion Topics
4. **Practice Problems**

THEN: (1) COLLECT CHALLENGE AREAS; (2) RECAP/SYNTHESIZE

OUTCOMES

Required readings

NOW
CHIHARA Chapter 7, 7.1-7.4

SOON
SPEEGLE Chapter 8, 8.1-8.7 & 10.2
Hypothesis Tests & CIs (together)
CHIHARA Chapter 8, 8.4.4-8.5
Interpretation & Common Issues

- Explain the purpose of a confidence interval
- Calculate a confidence interval for a single population mean
- Describe how changes in the sample size and confidence level affect a confidence interval
- Describe the duality between a two-sided hypothesis test and a confidence interval

Part 1

Abstract

Importance Firearms are the leading cause of death in US children and adolescents, but little is known about whether the overall legal landscape was associated with excess mortality after a landmark US Supreme Court decision in 2010.

Objective To measure excess mortality due to firearms among US children aged 0 to 17 years after the *McDonald v Chicago* US Supreme Court decision (2010).

Design, Setting, and Participants An excess mortality analysis was conducted using the US Centers for Disease Control and Prevention's Wide-Ranging Online Data for Epidemiologic Research (WONDER) database before and after *McDonald v Chicago*, the landmark 2010 US Supreme Court decision on firearms regulation. States were divided into 3 groups based on legal actions taken before and since 2010, most permissive, permissive, and strict. Firearm mortality trends before (1999-2010) and after (2011-2023) were determined and compared across the 3 groups for all intents and by intent (homicide and suicide). Subgroup analysis by observed race and ethnicity was conducted. For each US state, pre-and post-*McDonald v Chicago* all-intent pediatric firearm mortality incident rates were compared. These data were analyzed January 2011 through December 2023.

Exposure The pre- and post-*McDonald v Chicago* legal landscape.

Main Outcomes and Measures Excess mortality during the post-*McDonald v Chicago* period.

Results During the post-*McDonald v Chicago* period (2011-2023), there were 6029 excess firearm deaths (incidence rate [IR], 158.6 per million population; 95% CI, 154.8-162.5) in the most permissive group. In the permissive group, there were 1424 excess firearm deaths (IR, 107.5 per million person-years; 95% CI, 103.8-111.3). In the strict group, there were −55 excess firearm deaths (IR, −2.5 per million person-years; 95% CI, −5.8 to 0.8). Non-Hispanic Black populations had the largest increase in firearm mortality in the most permissive and permissive state groupings. Four states (California, Maryland, New York, and Rhode Island) had decreased pediatric firearm mortality after *McDonald v Chicago*, all of which were in the strict firearms law group.

Conclusion States in the most permissive and permissive firearm law categories experienced greater pediatric firearm mortality during the post-*McDonald v Chicago* era. Future work should focus on determining which types of laws conferred the most harm and which offered the most protection.

Interpretation notes

- Likely typo: should say "per million person-years" (not "population"), which is the appropriate unit for an incidence rate
- "Excess" here means "more than I would have otherwise predicted using a model from the historical trends"
- We should only compare the IRs because the death counts factor in the population size
- "Yearly and period-specific 95% CIs were derived directly from the Poisson model"

https://jamanetwork.com/journals/jamapediatrics/article-abstract/2834530

Part 1

The effect of high-fructose corn syrup vs. sucrose on anthropometric and metabolic parameters: A systematic review and meta-analysis

Xiang Li [1,†], Yunqi Luan [2,†], Yuejin Li [3], Shili Ye [4], Guihui Wang [5], Xinlun Cai [5], Yucai Liang [6], Hamed Kord Varkaneh [6], Yunpeng Luan [1,7,*]

▸ Author information ▸ Article notes ▸ Copyright and License information

PMCID: PMC9551185 PMID: 36238453

Abstract

High-fructose corn syrup (HFCS) has been speculated to have stronger negative metabolic effects than sucrose. However, given the current equivocality in the field, the aim of the present study was to determine the impact of HFCS use compared to sucrose on anthropometric and metabolic parameters. We searched PubMed, Scopus, Cochrane Central and web of sciences, from database inception to May 2022. A random effects model and the generic inverse variance method were applied to assess the overall effect size. Heterogeneity analysis was performed using the Cochran Q test and the I^2 index. Four articles, with 9 arms, containing 767 participants were included in this meta-analysis. Average HFCS and sucrose usage equated to 19% of daily caloric intake. Combined data from three studies indicated that HFCS intake does not significantly change the weight (weighted mean difference (WMD): −0.29 kg, 95% CI: −1.34, 0.77, $I^2 = 0\%$) when compared to the sucrose group. Concordant results were found for waist circumstance, body mass index, fat mass, total cholesterol (TC), high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglyceride (TG), systolic blood pressure (SBP), and diastolic blood pressure (DBP). Moreover, overall results from three studies indicated a significant increase in CRP levels (WMD: 0.27 mg/l, 95% CI: 0.02, 0.52, $I^2 = 23\%$) in the HFCS group compared to sucrose. In conclusion, analysis of data from the literature suggests that HFCS consumption was associated with a higher level of CRP compared to sucrose, whilst no significant changes between the two sweeteners were evident in other anthropometric and metabolic parameters.

Interpretation notes

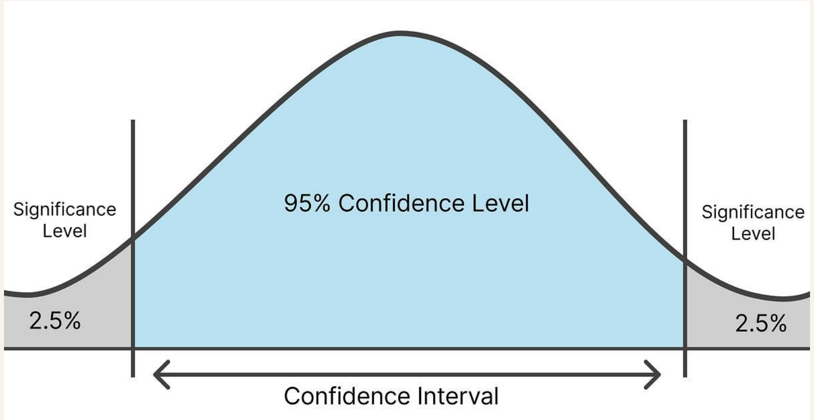
- Meta analysis == combine results across multiple (similar but independent) studies
  - Typically you don't have the subject-level data; only summary statistics
- Cochran Q & I^2 test: Checks for evidence that results within each sub-study are different from each other. Why?
- CRP = C-reactive protein (produced by the liver in response to inflammation)
- Weighted mean difference: uses the bigger sub-studies more. Why?

https://pmc-ncbi-nlm-nih-gov.revproxy.brown.edu/articles/PMC9551185/

Confidence Intervals - Motivation

- Testing for one null hypothesis doesn't tell us all the information we know
- P-values map back to the effect size (what we really care about), but not in a transparent way
- For a 95% confidence interval, it contains:
  - the true value 95% of the time *(in the long run)*
  - all the parameter values that we would NOT reject, if those parameter values were the null hypothesis
  - in other words, **the population values that seem aligned with the evidence** (but using an "innocent until proven guilty" framework, at the alpha-level, for "aligned")

General Form



typical distribution of test statistic (e.g. $\bar{X}$)

Most CIs share the same form:

Estimator ± Critical Value × SE of the estimator

- Estimator: An (unbiased and consistent) estimate of the population parameter of interest
- Critical Value: A look-up value based on the confidence level & the distribution of the estimator (test statistic). May rely on n (degrees of freedom for t distribution)
- Standard Error of the estimator: From standardizing the test statistic, using the CLT result. Will rely on assumptions (large n, pooled variance, binomial, etc.). Typically a function of n and the population standard error (as plug-in for standard deviation)

Let’s derive this ourselves for the case of a one-sample t-test

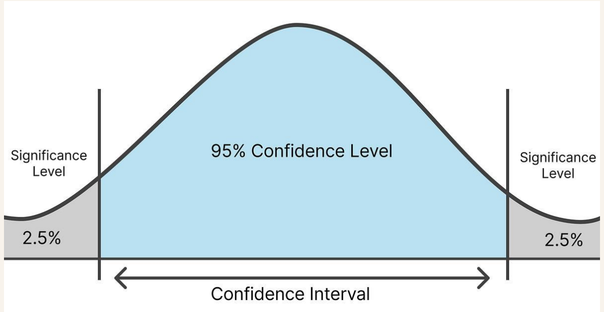
Confidence Intervals – Building Intuition

Do we want a wide or narrow CI? Why?

Does our CI get WIDER or NARROWER if:

- ❏ We get more data (larger n)
- ❏ We increase the confidence level (e.g.95% -> 99%)
- ❏ We have larger population variance?
- ❏ [CHALLENGE] We have a larger mean?

typical distribution of test statistic (e.g. $\bar{X}$)



Part 3

Coverage Simulations

Tests of Proportions

- For a fixed n, when is the confidence interval of a proportions test the widest? What is the intuition behind this?
- How does this influence your understanding and interpretation of political polling results?

Note: the "margin of error" is the (critical value) * (standard error) (the part after the +/- of the CI equation; the half the width of the CI)

This graphic says the margin of error in this poll is 2%

Discussion Points (chat with a neighbor)

- When might a confidence interval not have the correct coverage?

- How might you use confidence intervals to compare two parameters? Any concerns with your approach?

- [CHALLENGE] When might a confidence interval not be symmetric (the typical form)?

Practice Problems

CHIHARA 7.11

The data resampledata3::Olympics2012 has age, weight, and height on a random sample of 1000 London 2012 olympians. Compute a 95% confidence interval for the mean age of the 26 female athletes.

Do it "by hand", then use R's built-in function t.test
What assumptions did you make?

Practice Problems

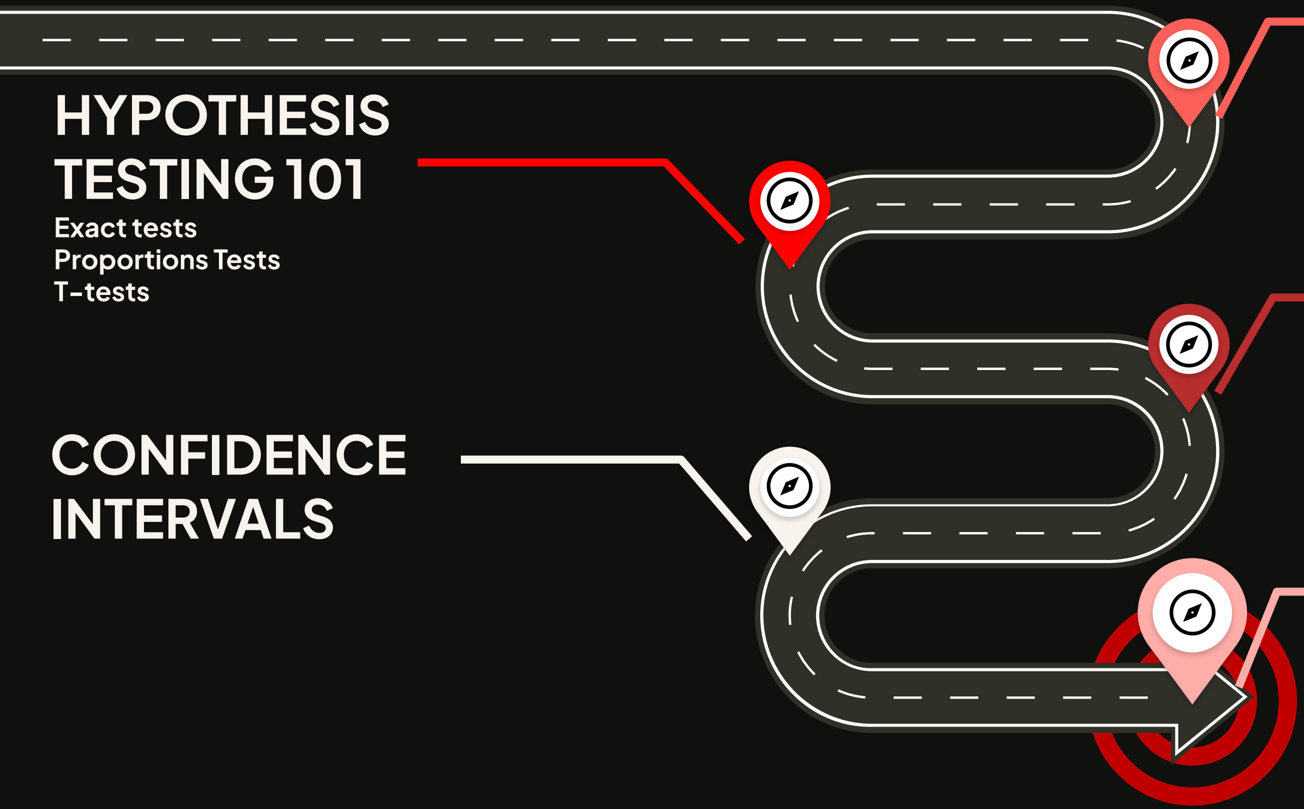
CHIHARA 7.3 (note: margin of error == half-width of CI)

20 years ago, mean cholesterol level of adult men in a certain town was 185 mg/dl, sigma = 50mg/dl

a. Suppose you now obtain a sample of 100 and find the sample mean = 210. Calculate your 90% CI, assuming sigma hasn't changed
b. If you start a new data collection process, how many people do you need to sample so that your margin of error on your 90% confidence interval is less than 10 mg/dl
c. Repeat (b) for 99% CI (before calculating, should this be bigger or smaller than (b)?)

ROADMAP

Free roadmap template



SAMPLING DISTRIBUTIONS & CLT

Exam Extra Credit

HYPOTHESIS TESTING 101

Exact tests
Proportions Tests
T-tests

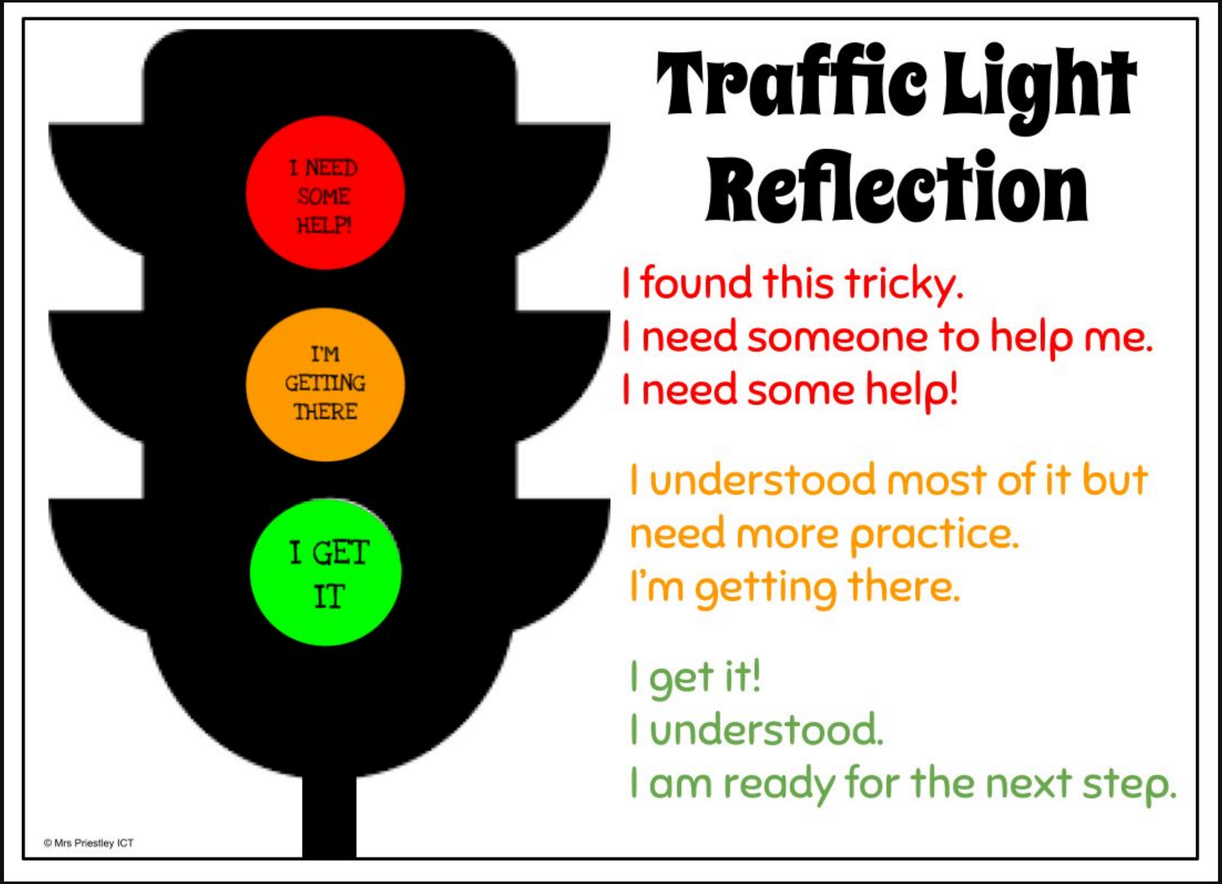
EXPERIMENTAL DESIGN

Handout

CONFIDENCE INTERVALS

SYNTHESIS: THE INFERENTIAL PROCESS

Week 9 Lab
Assignment #2
Extra Practice Problems



WHAT SPECIFIC CONCEPTS WITHIN THE FOLLOWING TOPICS ARE STILL CONFUSING?

Function of Random Variables

Sampling Distributions

Hypothesis Testing

Experimental Design

Confidence Intervals

WEEK 9

PLEASE READ:

- SPEEGLE Chapter 8, 8.1–8.7 & 10.2 - Hypothesis Tests & CIs (together)
- CHIHARA Chapter 8, 8.4.4–8.5 - Interpretation & Common Issues
- Suggested practice problems on Ed Discussion (Week 7 + 8 + 9; pinned to the top)

Feedback (link) & Discussion

Handouts

Practice Problems

CHIHARA 7.20

The data resampledata3::Groceries has a sample of grocery items and their advertised prices from one day. Create a 95% CI for the mean difference in prices between Target and Walmart. Check for outliers

Write out the notation for the hypothesis test
What is your test statistic's null distribution?
State your assumptions
Repeat calculation after checking and removing any outlier(s)

Practice Problems

CHIHARA 7.38

Let Xi ~ $N(\mu_1, \sigma^2)$, i = 1, ..., n
Let Yj ~ $N(\mu_2, 2*\sigma^2)$, j = 1, ..., m

Find the 95% CI for $3\mu_1 + \mu_2$
Assume X and Y are independent
Assume σ^2 is known;

TAKE HOME CHALLENGE: repeat for unknown σ^2

Practice Problems

- $Z^2 \sim \chi^2_1$.
- $\chi^2_\nu + \chi^2_\eta \sim \chi^2_{\nu+\eta}$ if independent.
- $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ when $X_1, \ldots, X_n$ iid normal.
- $\dfrac{Z}{\sqrt{\chi^2_\nu/\nu}} \sim t_\nu$ if independent.
- $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$ when $X_1, \ldots, X_n$ iid normal.

[CHALLENGE] CHIHARA 7.48

Robin buys 8 boxes of cereal and finds the following weights (in g). Compute a 90% CI for the **variance**. Why might one be interested in the variance?

560, 568, 580, 550, 581, 581, 562, 550

Hints / Notes:

- Use facts from SPEEGLE 5.5.4 pg 143
- Derive the CI structure by first principles (it’s not in the typical +/- form)
- We need a normality assumption for such a small n

OUTCOMES
Weeks 7–9

CHIHARA Chapter 3, 3.1–3.2
CHIHARA Chapter 8, 8.1–8.4.3

CHIHARA Chapter 7, 7.1–7.4

SPEEGLE Chapter 8, 8.1–8.7
SPEEGLE Chapter 10, 10.2
CHIHARA Chapter 8, 8.4.4 - 8.5

Q&A

Hypothesis Testing

- Define the core components and explain the logic of hypothesis testing
- Formulate appropriate null and alternative hypotheses for a given research question
- Evaluate the relationships between sample size, significance level, effect size, and power
- Identify the key assumptions underlying one- and two-sample t-tests
- Distinguish between designs that call for an independent t-test versus a paired t-test

Confidence Intervals

- Explain the purpose of a confidence interval
- Calculate a confidence interval for a single population mean
- Describe how changes in the sample size and confidence level affect a confidence interval
- Describe the duality between a two-sided hypothesis test and a confidence interval

OVERALL

- Carry out hypothesis tests in R and interpret results accurately; Synthesize the results to formulate a clear, jargon-free conclusion for a public health audience