# Confidence Intervals

## Part 1 - using Probability Models

Ryan Miller

Grinnell College
Statistics

# Introduction

- In the 1980s, AstraZeneca developed *Prilosec*, a highly successful heartburn medication
  - The FDA patent for Prilosec expired in 2001, prompting AstraZeneca to try to replace Prilosec with a new drug, *Nexium*
  - In 2005, the New York Times reported Nexium was over 5-times as expensive as Prilosec
- The clinical trial comparing the drugs was very large, involving over 6000 participants, and the difference in healing rate at 8-weeks was statistically significant ($p < 0.001$)
  - Based on this information, should doctors be using Nexium?

# Introduction (cont.)

- The $p$-value doesn't directly convey **effect size**, or how different the healing rates were in this study
  - 87.5% of patients assigned to receive Prilosec were healed, while 90% of patients assigned Nexium were healed
- Both healing rates are very high, so some patients might not want to pay the extra cost for Nexium
  - Furthermore, some of the observed difference of 2.5% could be due to sampling variability, and it is possible that the actual difference is smaller
  - This motivates the concept of *interval estimation*, which seeks to find a range of plausible values after accounting for sampling variability

# Interval Estimation

Consider an unknown population parameter, such as $p_1 - p_2$ in the Nexium vs. Prilosec example. Statisticians often consider two types of estimates:

1) **Point estimates** - a *single number* that is the *best guess* for what the population parameter is. For example, the sample mean $\overline{x}$ is a point estimate for the population's mean, $\mu$.

# Interval Estimation

Consider an unknown population parameter, such as $p_1 - p_2$ in the Nexium vs. Prilosec example. Statisticians often consider two types of estimates:

1) **Point estimates** - a *single number* that is the *best guess* for what the population parameter is. For example, the sample mean $\overline{x}$ is a point estimate for the population's mean, $\mu$.
2) **Interval estimates** - a *range of numbers* that represent *plausible values* of the population parameter. Interval estimates usually have the form: Point Estimate $\pm$ Margin of Error

For interval estimates, is it possible to use a margin of error such that the interval always contains the truth? What would the downside of this be?

# Interval Estimation

When developing an interval estimate there are two competing goals:

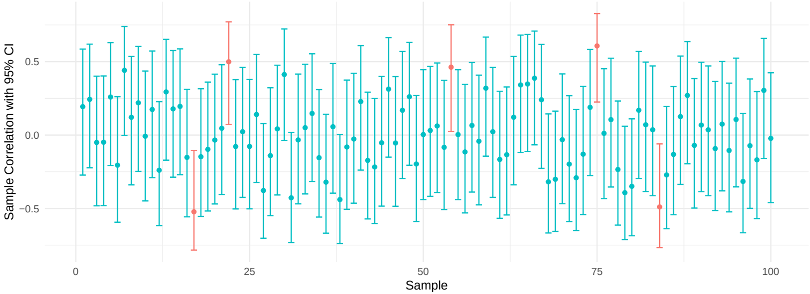
1. We want the interval to have a high chance of containing the true population parameter, something that is usually achieved by a wider interval
2. We want the interval to have a small enough margin of error to be useful - An interval suggesting a candidate could receive between 30% and 70% of the vote in an election isn't very useful.

# Confidence intervals

- **Confidence intervals** are a special type of interval estimate whose margin of error is calibrated to achieve a long-run "success rate" known as the *confidence level*
  - Here "success" is defined as the interval containing the true value of the parameter being estimated
- For example, the procedure used to construct the margin of error for a 95% confidence interval will make the interval wide enough to contain the parameter of interest in 95% of different random samples (or study replications)

# Confidence intervals

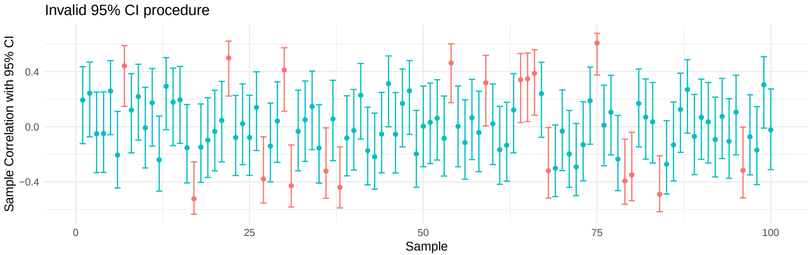
Below are 100 different 95% CI estimates that were made using data from a different random sample of $n = 20$ from a population with $\rho = 0$:




Notice how 5 of 100 samples produced a 95% CI that failed to contain the true population-level correlation!

# Confidence Interval Validity

- A *valid* confidence interval procedure will produce intervals where the stated confidence level matches the procedure's long-run success rate
  - The procedure used to make the intervals on the previous slide was valid
- An *invalid* procedure will produce intervals that fail to contain the true parameter more often than they should, for example the procedure shown below:




Grinnell College
Statistics

# Confidence Interval Validity

Achieving validity is the key challenge in confidence interval estimation, we'll begin with a framework that decomposes the margin of error into the *standard error* of the point estimate and a *calibration value*:

$$\text{Point Estimate} \pm c \cdot SE$$

- Central Limit theorem has given us the standard error of most point estimates (see table on Exam 1)
- The calibration value, $c$, can be found using an appropriate probability model

# One and Two-Sample $Z$-intervals

For a **single proportion** or **difference in proportions** the value of $c$ can be chosen from a Standard Normal distribution

- $c = 1.96$ corresponds to the middle 95% of the $N(0,1)$ distribution
  - This choice of $c$ will produce a valid 95% confidence interval (when the Normal distribution is an accurate probability model)
- $c = 1.28$ corresponds to the middle 80% of the $N(0,1)$ distribution, which can be used to calibrate an 80% confidence interval
  - Notice how this will lead to a smaller margin of error (relative to $c = 1.96$)

Grinnell College
Statistics

# Practice #1

We previously heard about a study by Johns Hopkins University Hospital where 31 of 39 babies born in their facilities at 25 weeks gestation (15 weeks early) went on to survive. Suppose our goal is to find an interval estimate the proportion of babies born under similar circumstances in similar hospitals that will survive.

1) In this application we are trying to estimate $p$, a population proportion, using $\hat{p} = 31/39$. What is the Standard Error of $\hat{p}$? *Hint*: This was on the cover page of Exam 1.
2) Use StatKey to determine the value of $c$ used to calibrate a 99% CI involving these data.
3) Find the endpoints of the 99% CI.
4) If the confidence level is decreased from 99% to 95% what will happen to the width of the interval?

# Practice #1 (solution)

1) $SE = \sqrt{\frac{(31/39)\cdot(1-31/39)}{39}} = 0.0646$
2) $c = 2.576$
3) $31/39 \pm 2.576 \cdot 0.0646 = (0.628, 0.961)$
4) The width would decrease, but we'd be less confident that the interval contains the truth.

# One and Two-Sample $T$-intervals

For a **single mean** or a **difference in means** the value of $c$ must be chosen from a $T$-distribution because the $SE$ of these statistics requires us to estimate population's standard deviation using the sample standard deviation, thereby introducing additional uncertainty

- If $df = 5$, $c = 2.571$ corresponds to 95% confidence
  - Notice how this is a much larger multiplier than $c = 1.96$, which corresponded with 95% confidence when the Normal distribution was an accurate probability model
- If $df = 30$, $c = 2.042$ corresponds to 95% confidence

# Practice #2

- We previously looked at a study of $n = 12$ professional swimmers who swam a 1500m time trial with and without a scientifically designed wetsuit.
  - Suppose we'd like to use these results to estimate the expected improvement in 1500m swim velocity due to the wetsuit.
- In this study, the average *paired difference* was 0.0775 m/s and the standard deviation of these differences was 0.022.
  - Find a 95% confidence interval estimate for the expected difference in velocity in the population.

# Practice #2 (solution)

To calculate the interval we need the two components of the margin of error, $c$ and $SE$:

- $c$ is given by the middle 95% of a T-distribution with $df = n-1 = 11$, which is 2.201 according to StatKey
- $SE = s/\sqrt{n}$ for a single mean, which is $0.022/\sqrt{12} = 0.00635$

So, the 95% CI is given by

$$0.0775 \pm 2.201 \cdot 0.00635 = (0.064, 0.091)$$

Thus, we conclude with 95% confidence that the expected improvement in swim velocity when wearing the wetsuit is between 0.064 m/s and 0.091 m/s. Since the entire interval is above zero, it's implausible that the wetsuit makes no difference.

# Confidence Intervals for Other Statistics

- Our examples in these slides focused on scenarios where *symmetric* confidence intervals of the form Point Estimate ± Margin of Error make sense
- Symmetric intervals *do not* make sense for quantities like odds ratios
  - $OR = 2.5$ indicates 2.5 times *higher odds* of the outcome relative to the reference group
  - $OR = 0.4$ indicates 2.5 times *lower odds* of the outcome relative to the reference group
- We will rely on software (`R`) for confidence interval estimates for descriptive statistics other than the four covered in these slides

# Conclusion

- A **point estimate** is our best guess at what an unknown population parameter might be based upon sample data
  - It will almost always be wrong by at least some amount due to sampling variability
- An **interval estimate** accounts for sampling variability by including a margin of error (MOE), which allows us to have a realistic shot at containing the unknown parameter
  - Interval estimation involves the conflicting goals of precision (small MOE) and accuracy (high chance of containing the truth)
- **Confidence intervals** are interval estimates made using a procedure that carries a *long-run success rate* known as the *confidence level*
  - We can find a $P\%$ confidence interval using the formula point estimate $\pm c \cdot SE$, where $c$ is selected to represent the middle $P\%$ of an appropriate probability model