Estimation Using Traditional and Bootstrap Methods
Recap

Last time, we discussed:

- Sampling distributions

- The traditional approach to frequentist inference: null hypothesis significance testing (NHST)

- The roots of NHST in the work of Fisher (null hypothesis, significance, p-values) and Neyman & Pearson (alternative hypotheses, $\alpha$, $\beta$, power, etc.)

- Some pros and cons of the traditional approach
Some Pros & Cons of Hypothesis Testing

Pros

- Objective method for making decisions regarding data
- Simple rules, do not require statistics expertise
- In the absence of auxiliary biases (and in scrupulous hands), guarantees correct decisions in the long run

Cons

- Rigid, 1-bit decision making
- Absolves scientists from thinking carefully about analysis
- Long-run guarantees rely on replication and unbiased reporting & publication
- p-values & significance level not useful for meta-analysis
Consequences: a Reckoning

The inadequacies of NHST within the context of current practices in research, reporting, and publication (especially within psychology) have recently garnered a lot of attention and hand-wringing.

- Kahneman, D. (October 2013) Trouble at the Lab, *The Economist*
- Simonsohn, U. (Too many relevant references to list here)

Some of these problems are structural, having to do with the way science is practiced (a topic outside the scope of a statistics course). However, researchers and scientific journals have made a number of recommendations regarding the use of statistics for inference.
The “New Statistics”: Alternatives to NHST

Among the most common recommendations are:

• **A greater emphasis on estimation, including the reporting of confidence intervals and effect sizes**
  
  – Allows us to assess the reliability and practical importance of the result
  
  – [e.g., Cumming, 2014 (Psychological Science)] This recommendation has near universal support and is quickly being adopted by major journals

• **A focus on explicit model comparison**
  
  – [e.g., Morey et al., 2014 (Psychological Science)]
  
  – This is slightly more controversial (and more complicated), but is quickly becoming standard in certain domains

• **A switch to Bayesian statistics**
  
  – [e.g., Wagenmakers et al., 2008] This is quite a bit more controversial and involves a fractious philosophical disagreement about the nature of probability
Risk of cardiovascular events and rofecoxib: cumulative meta-analysis

Peter, Jiin MD, Linda Naray DIPMD, Stephan Reichenbach MD, Rebekka Sterchi Prof, Paul A Dienke MD, Prof Matthias Egger MD
Estimation

Recall the general definition of a statistic (from first lecture):
• A function computed on a data set

Statistics can play three basic roles:
• As a summary description of data (descriptive statistic)
• As a value to test the adequacy of a hypothesis (test statistic)
• As an estimate of a population parameter (estimate)

Recall also that the same statistic (e.g., $\overline{X}$) can play any of the three roles, depending on the context.
Assessing the Quality of an Estimate

• **Bias**
  – The extent to which the expected value of an estimate $\hat{\theta}$ differs from the population value $\theta$
  
  \[
  \text{Bias}[\hat{\theta}] = E[\hat{\theta}] - \theta
  \]

• **Variance**
  – The square of the standard error, or the expected squared deviation of the estimate from its mean
  
  \[
  \text{Var}[\hat{\theta}] = E\left(\hat{\theta} - E[\hat{\theta}]\right)^2
  \]

• **Mean Squared Error (MSE)**
  – A measure of the overall estimation error
  
  \[
  \text{Var}[\hat{\theta}] + \left(\text{Bias}[\hat{\theta}]\right)^2
  \]
Confidence Intervals

Low bias

Low variance

High bias

High variance
Confidence Intervals: Traditional Approach

The traditional (Neyman-Pearson) approach to generating interval estimates are based on analytical results involving known parametric distributions.

• Assume either via special case or asymptotic properties, that the population and/or sampling distributions have a normal or related shape (e.g., t, F, $\chi^2$)

• Assume homoscedasticity, symmetry, etc.

    or

• Do some (difficult) math to derive results for some other special case(s)
Confidence Intervals: Traditional Approach

You know the form of the sampling distribution (e.g., \(N(\mu, \frac{\sigma}{\sqrt{n}})\)). Estimate the parameters of this distribution using sample statistics, then use the standard error or quantiles of this distribution to construct an interval of likely (RMS) errors. Generally

- Use a **pivotal statistic** (i.e., a statistic whose sampling distribution does not depend on unknown parameters)

- For a pivotal statistic, any interval constructed to include a \((1-\alpha)\) proportion of the sampling distribution will also **prospectively** contain the population parameter \((1-\alpha)\) proportion of the time when constructed around the statistic. This proportion is called the **confidence**

- Typically the interval is chosen to be as narrow (precise) as possible for the given confidence level
Confidence Intervals for the Population Mean: Pivot Method

Steps for computing a confidence interval (pivot method):

1. Compute the sample mean and the standard error
2. Choose the desired (1 - \( \alpha \)) level of confidence (e.g., 95%) and compute the corresponding \( z_{\alpha/2} \) values (e.g., \( \text{qnorm}(\alpha/2) \)).
3. Compute the confidence interval by inverting the z transformation

\[
CI_{1-\alpha}(\mu) = \bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}
\]

E.g., for a z-Statistic:

\[
= \bar{X} \pm \frac{z_{\alpha/2} \sigma}{\sqrt{n}}
\]
Confidence Intervals for the Population Mean: Percentile Method

Steps for computing a confidence interval (percentile method):

1. Compute the sample mean and the standard error
2. Choose the \((1 - \alpha)\) level of confidence (e.g., 95%) and compute the corresponding quantiles \(\bar{X}_{\alpha/2}\) and \(\bar{X}_{1-\alpha/2}\).
3. That’s it. The quantiles represent the endpoints of your confidence interval

\[
CI_{1-\alpha} (\mu) = [\bar{X}_{\alpha/2}, \bar{X}_{1-\alpha/2}]
\]
Example: Computing CI’s for the Mean (with known $\sigma$)

$\sigma = 7.0$
$n = 5$
$\bar{X} = 75.0$

**Pivot Method**

\[
\begin{align*}
z_{0.025} &\approx -1.96 \\
CI_{0.95}(\mu) &= \bar{X} \pm z_{0.025} \sigma_{\bar{X}} \\
&= 75 \pm 1.96 \frac{7.0}{\sqrt{5}} \\
&= 75 \pm 6.14 \\
CI_{0.95}(\mu) &= [68.86, 81.14]
\end{align*}
\]

**Percentile Method**

\[
\begin{align*}
\bar{X}_{0.025} &\approx qnorm(0.025,75,7/sqrt(5)) \approx 68.86 \\
\bar{X}_{0.975} &\approx qnorm(0.975,75,7/sqrt(5)) \approx 81.14 \\
CI_{0.95}(\mu) &= [\bar{X}_{0.025}, \bar{X}_{0.975}] = [68.86, 81.14]
\end{align*}
\]
Confidence Intervals

Computing CI’s for the Mean

\[ \bar{X} \] (point estimate of \( \mu \))

\[ CI_{95}(\mu) \] (interval estimate of \( \mu \))

95% CI

\[ z = 1.96 \]

\[ z = -1.96 \]

\[ z = 0 \]
Pros and Cons of Traditional Approach

Advantages of traditional approach

• If the assumptions are approximately satisfied
  – Analytical solutions, with known convergence characteristics
  – Easy and inexpensive to compute estimates
  – Very efficient (requires relatively little data to get a good estimate)

Disadvantages:

• If assumptions are not met, estimates can be terribly unreliable
• Difficult to derive estimates for unusual population distributions, non-algebraic statistics, or small sample sizes
Monte Carlo Simulation & Bootstrapping

What if you don’t know the form of the sampling distribution?

• If you know the underlying population distribution:
  – E.g., if you know the distribution, but it’s difficult to work with analytically, or if the statistic of interest is not algebraic (median, trimmed mean, etc.)
  – Use Monte-Carlo simulation to approximate the sampling distribution

• If you don’t know the underlying population distribution:
  – Use bootstrap resampling!
The Bootstrap

The central assumption of the bootstrap is that the sample is broadly representative of the population

• Though the sample rarely approximates the population precisely, it is often the best estimate that we have of the population distribution

• This is especially true if we don’t know anything \textit{a priori} about the form of the distribution

The bootstrap (in its simplest form) assumes that the population distribution is exactly the same as that of the sample, only bigger (i.e., with a much larger $n$)
The Bootstrap

Let’s say you’re interested in computing an interval estimate for a parameter $\theta$ from a sample $\{x_1, x_2, \ldots, x_n\}$.

From this sample, you can easily compute a point estimate $\hat{\theta} = t$ by applying the function $s(x_1, x_2, \ldots, x_n)$ to the sample data. But you want an interval estimate.

To do this using simple nonparametric bootstrapping:
Confidence Intervals

Real world

Real world parameter \( \theta \)

Population of N items

Observed sample \( x_1, x_2, \ldots, x_n \)

Estimate \( t_1 = s(x_1, \ldots, x_n) \)

Bootstrap world

Bootstrap world parameter \( \hat{\theta} = t_1 \)

Population of N items based on copying \( x_1, x_2, \ldots, x_n \)

Bootstrap sample \( x_1^*, x_2^*, \ldots, x_n^* \)

Bootstrap replicate

\( t^* = s(x_1^*, \ldots, x_n^*) \)
The Bootstrap

• In practice, you don’t actually need to make a bunch of copies of the \( \{x_1, x_2, \ldots, x_n\} \). Instead you can simply resample the original sample to obtain a distribution of values for the resampled statistic \( t^* \)

  – This is almost exactly the same as what you did in the sampling distribution lab, except that you will sample \textit{with} replacement

• After replicating this procedure many times to get \( t_1^*, \ldots, t_{10,000}^* \) (or so) you simply compute the confidence interval from the resulting sampling distribution

• The simplest and most general method (as with the traditional approach) is to use the \( \alpha/2 \) and \( (1 - \alpha/2) \) quantiles.
Sample R Code for Simple Bootstrap

```r
bootstrap_CI <- function(data, stat_func, conf=0.95, nr_samples=10000){
  bs_stats <- rep(NA, nr_samples)
  for(i in 1:nr_samples){
    bs_stats[i] <- stat_func(sample(data, replace=TRUE))
  }
  alpha <- 1-conf
  bs_CI <- c(quantile(data, alpha/2), quantile(data, 1-alpha/2))
  return(bs_CI)
}
```

For example, to compute the bootstrapped CI for the mean:

```r
mean_CI <- bootstrap_CI(sample_data, mean)
```
Female weights from cdc data:
Empirical distribution
Sample of size $n=10$ drawn from weight distribution
Empirical *sampling* distribution for n=10
Empirical \textit{sampling} distribution for \( n=10 \)
95% CIs computed using **bootstrap** and **normal approximation**
Simple Variants of the Bootstrap

• **Case-Resampling:**
  – This is just a more precise name for the simple non-parametric bootstrap that we just covered

• **Parametric Bootstrap:**
  – Instead of simply resampling the sample data, you fit a parametric model to the data and draw your bootstrap samples from that model
  – This is good for complex data models and/or complex statistical functions $s(.)$

• **Smooth Bootstrap:**
  – This is like case-resampling, except that you add some random (usually Gaussian) variability to the resampled values
  – The result is functionally equivalent to sampling from a kernel density estimate of the distribution
  – This is good when your sample size is small (and lumpy) but you know that the underlying density is smooth