BIOL 501: Boostrapping and Resampling

Peer-Feedback Survey

1\.

2\. **Jamboard anonymous feedback link Assignment #2 (Canvas announcements)**

https://jamboard.google.com/d/1p8FprE5AEnMYkG0RvxLycCxVcdUY4omA9_0274p_uZQ/edit?usp=sharing

20 March 2023 Dr. Beth Volpov, b.volpov@oceans.ubc.ca; UBC 2022 Term 2

Informal feedback and self-refelction on Assignment #2 on **Jamboard** (this is anonymous)

- **Self-reflection:**How do you feel overall about Assignment #2 on a scale from 1-5 (1=poor to 5=excellent)? Other comments?
- **Self-reflection:**How did you prepare? How far ahead did you start it?
- **Instructor feedback:**What could the instructor have done differently to help you prepare?

https://jamboard.google.com/d/1p8FprE5AEnMYkG0RvxLycCxVcdUY4omA9_0274p_uZQ/edit?usp=sharing

Outline

- What is bootstrapping and sampling with replacement
- Estimation and hypothesis testing
- Estimation
- Sampling distribution
- Bootstrap standard error
- Bootstrap confidence interval
- Comparing 2 groups
- Permutation test
- Summary
- **Workshop**

What is bootstrapping

Discussion (no jamboard link): what does bootstrapping mean?

- In pop culture? In life? In biology? Other?

Bootstrapping outside of statistics

- 18th and 19th century to "pull oneself up by one's bootstraps"
- Originally referred to an impossible task
- Now refers to the challenge of "making something out of nothing"

Bootstrapping in Business

- Bootstrapping: process of using **only existing resources** (savings, personal equipment, garage space) to start and grow a business
- **Stretch whatever you've got (no matter how small) to get the job done**
- Contrasts bringing in outside (new) funds from investors or taking on debt

Bootstrap, from Twitter

- A toolkit from Twitter to help kickstart development of web apps and sites
- Opensource on Github as of 2011
- Refers back to bootstrap idea of starting “with what you have” or from “nothing”

Bootstrapping in Statistics

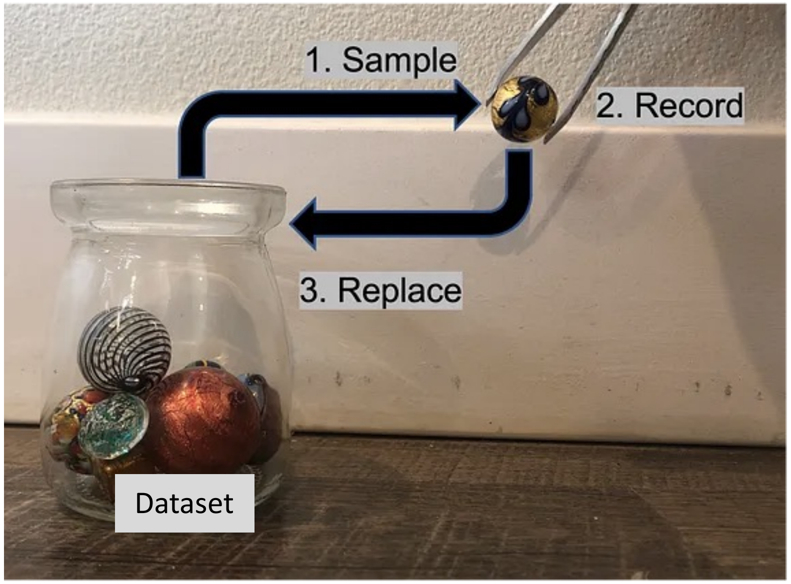
- Bootstrapping is a statistical process that **resamples with replacement** to create many samples (*without* collecting new data).
- Computer intensive,but has been used for a long time (since Efron 1979)
- Bootstrapping always includes sampling with replacement, but it is not the only type of sampling with replacement
- We will discuss using this **for 95% CI and estimation of S.E.**, but there are other uses for bootstrapping in stats that are not covered here

Sampling with replacement

1. Sample randomly from your original dataset
2. Record the sample (e.g. weight of ball)
3. Return sample back to dataset (***with replacement)***
4. Repeat #1-3 again (usually thousands of replicates)

Key things to remember

- The same ball can occur > once because replacement
- The probability of selecting any particular ball remains constant in future draws



Bootstrapping vs Simulation

- **Bootstrap:** procedure that uses resampling (sampling with replacement) to approximate the sampling distribution of an estimate.
  - **No new data collected**, data is not simulated
  - Unlike simulation, not directly intended for hypothesis testing
  - Used mostly for estimation (e.g. standard error) or boostrapping 95% CI on a parameter estimate
- **Simulation:** uses a computer to imitate the process of repeated sampling from a population to approximate the null distribution of a test stat

Estimation and hypothesis testing

Computer intensive methods

- **Question:** What to do if the assumptions of the best method available are violated?
  - We can’t collect more data (expensive, time limit)
  - We cannot turn to linear or generalized linear models (because their assumptions are also violated)?
  - It’s not desirable to transform the data
- **Optional answer:** Computer-intensive methods.
  - An approach in which the power of the computer is used to generate a sampling distribution.
    - For estimation: the bootstrap
    - For hypothesis testing: the permutation test

Estimation and Hypothesis Testing

- 2 types of statistical inference in conventional data analysis. Each is founded on a different sampling distribution.

Estimation (bootstrap)

- Uses the sampling distribution of an estimate:all the values for a parameter estimate we might obtain, when sampling from a population, and their probabilities.
- Used to obtain standard errors, **confidence intervals**.
- Most methods assume sampling distribution is approximately normal.

Hypothesis testing (permutation test)

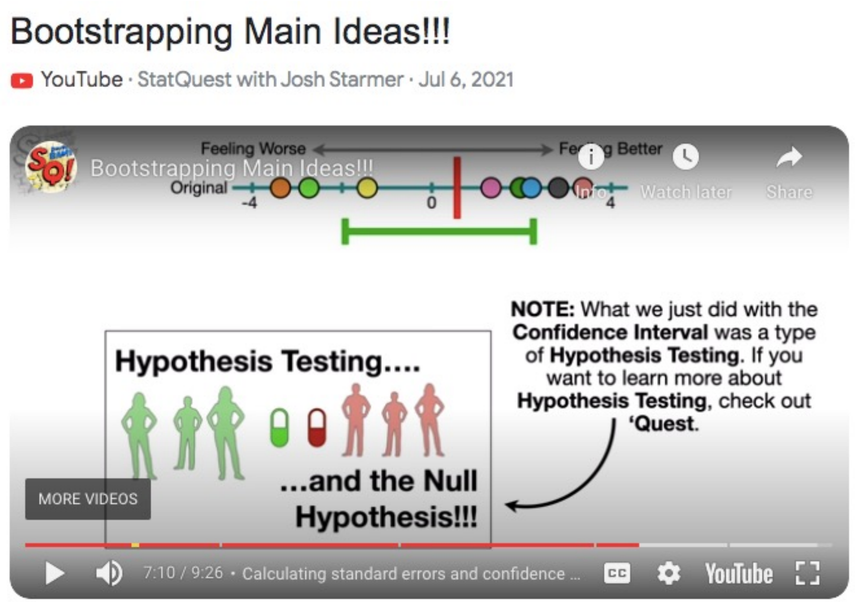
- Uses the null sampling distribution (or null distribution): the probability distribution of a test statistic if the null hypothesis is true.
- We frequently use the *t*, *F*, χ^2, and normal distributions to approximate null distributions, from which *P*-values are calculated.

Why use bootstrapping for estimation?

- Normally estimating something like the variance of the mean requires multiple independent samples, but you can't collect more data
- Bootstrapping allows you to perform estimates from a single population without collecting more data
- This is essentially allowing the estimate to "pull itself up by its bootstraps"
  - Do a seemingly impossible feat of physically pulling yourself up by your own bootstraps

Optional additional video resource

- StatQuest with Josh Starmer using Bootstrapping for hypothesis testing



https://www.google.com/search?q=statistics+boostrapping&source=lmns&bih=653&biw=1413&hl=en&sa=X&ved=2ahUKEwjd97H2pOz9AhWhJn0KHdR9CiYQ_AUoAHoECAEQAA#fpstate=ive&vld=cid:b27f5bd4,vid:Xz0x-8-cgaQ

Benefits of Bootstrapping

- Used for **estimation**, mainly*
- Provides standard errors and confidence intervals of useful parameters (magnitudes).
- **Nonparametric**→doesn't require normally-distributed data. It makes no assumptions about the distribution of the data.
- It can be applied to virtually any population parameter, including means, proportions, and linear model coefficients.
- **When to use it?** It is most handy when there is no ready formula for a standard error or confidence interval (e.g., median, trimmed mean, eigenvalue).
- It even works also for estimates based on complicated sampling procedures or calculations

*Bootstraping can also be used for hypothesis testing, though I have not discussed this.

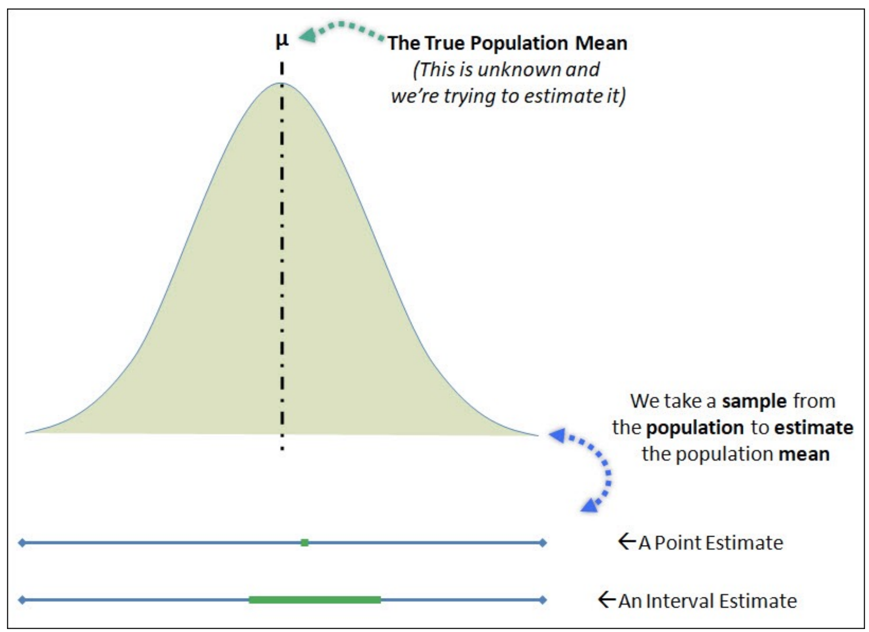
Bootstrapping for Estimation of standard error example

Bootstrap for Estimation

- The bootstrap is amazing and **useful for estimation**.
- It works in almost any situation (if *n* not too small).
- It is approximate, though performs almost as well as parametric methods when assumptions of the parametric methods are met.

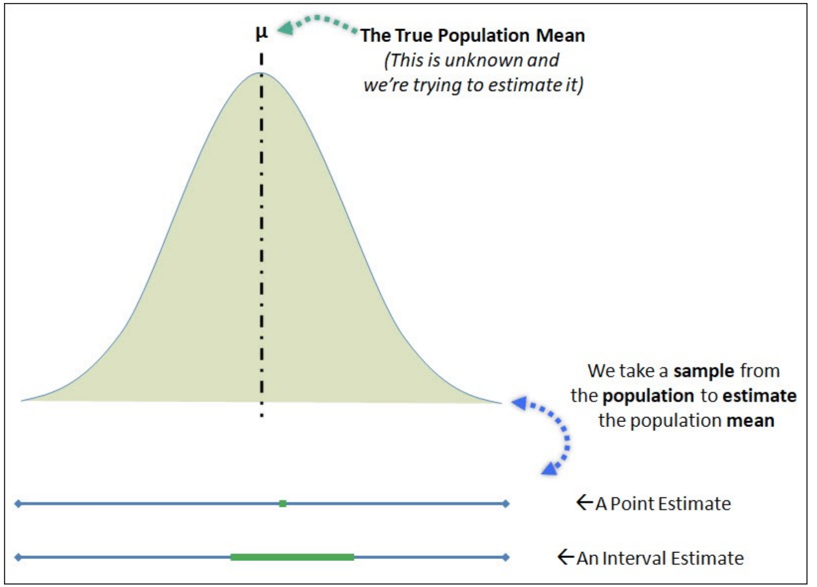
To understand the bootstrap, review how estimation works

- Estimation is the process of inferring a population parameter from sample data.
- The value of a sample *estimate* is almost never the same as the population *parameter* because of random sampling error (chance).



To understand the bootstrap, review how estimation works

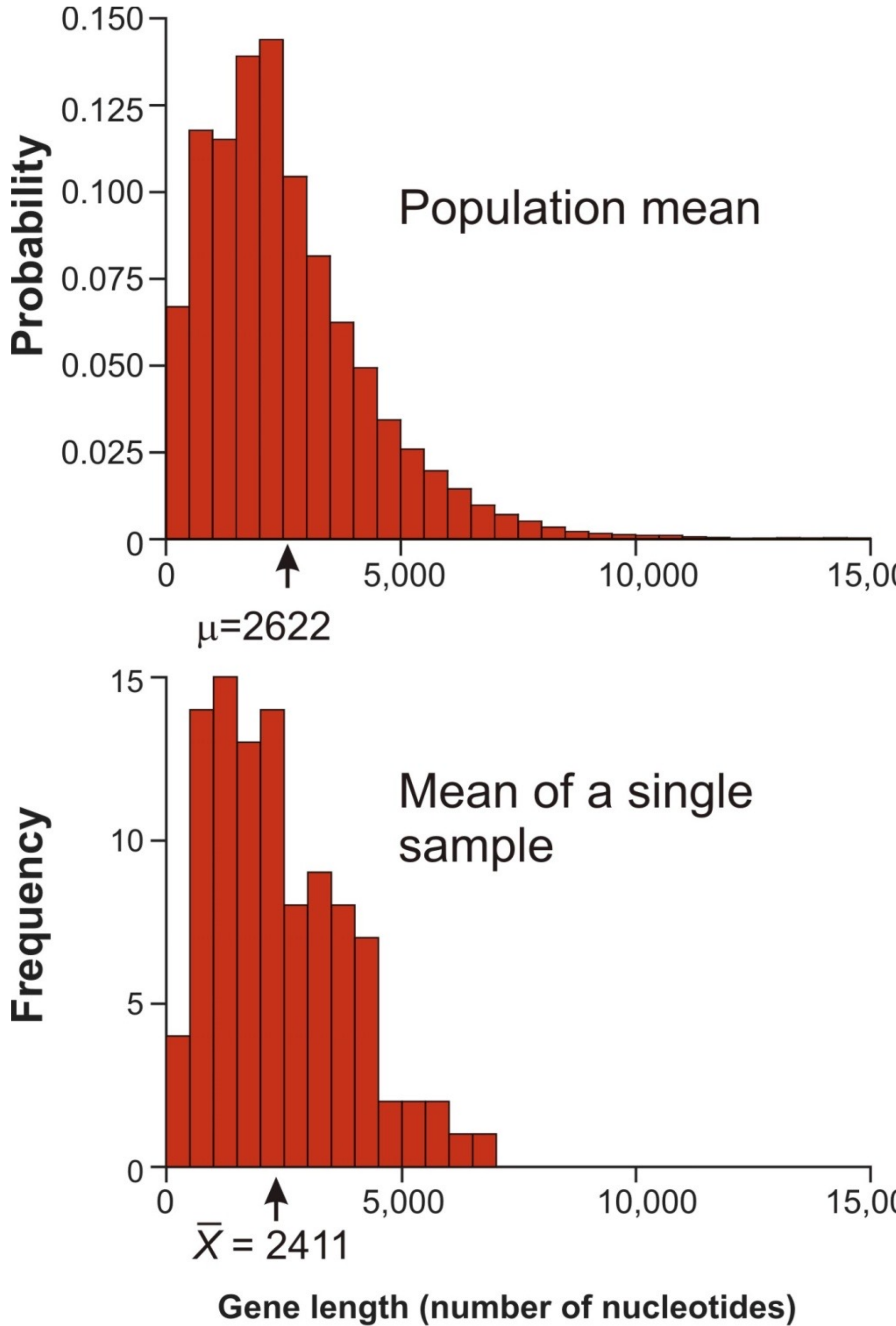
- The sampling distribution of an estimate gives all the values we might have obtained from our sample, and their probabilities of occurrence.
- The standard error of an estimate is the standard deviation of its sampling distribution. Standard error therefore measures the uncertainty of an estimate.



Example: Estimate a mean

- **What we want**:The mean of a variable in the <u>**population**</u> (e.g., the lengths of all the genes in the human genome).

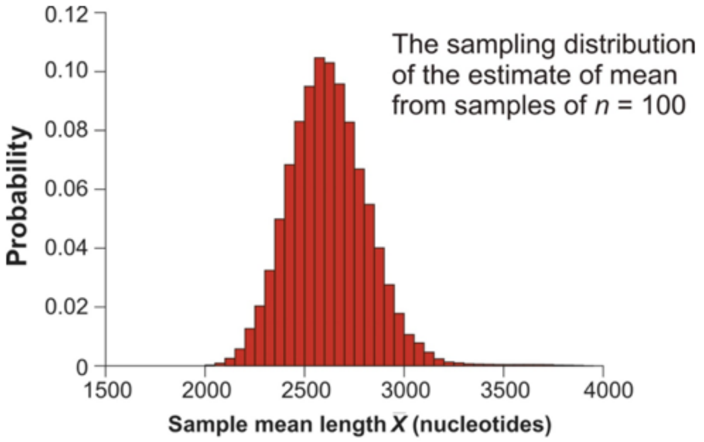
- **What we have instead:**The <u>**sample**</u> mean (e.g., based on a random sample of *n* = 100 genes)



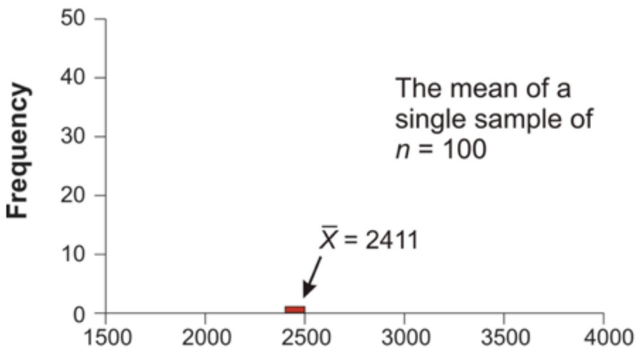
The sampling distribution

We don't know the true mean.

So, we want an approximation of the sampling distribution, the distribution of estimates we *might* obtain from random sampling and their probabilities.

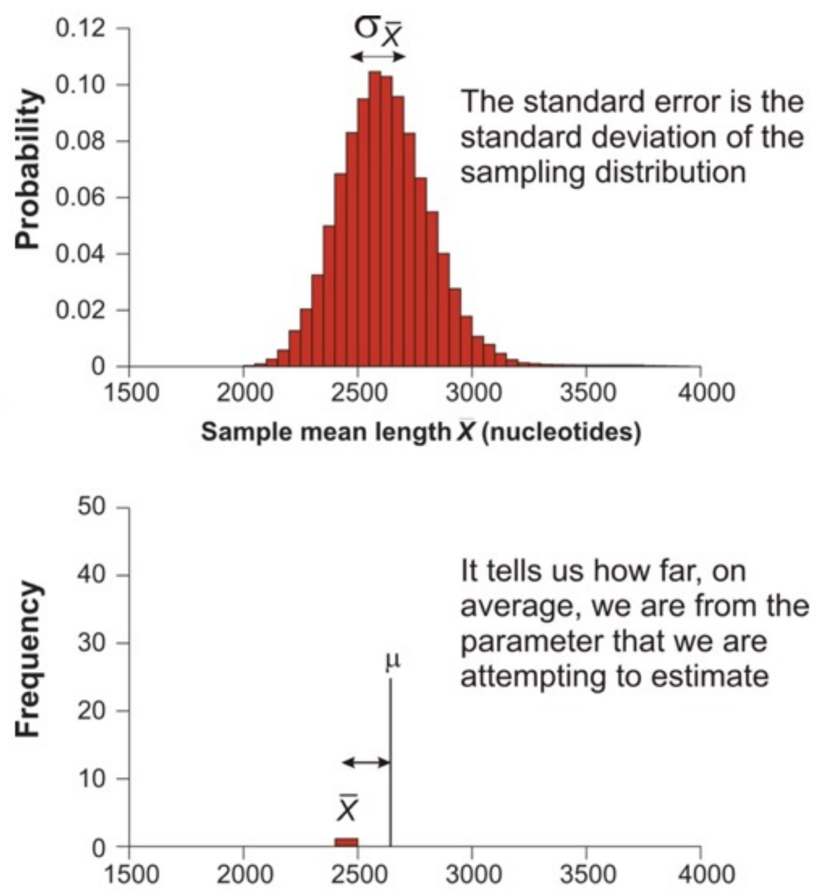


What we have instead:
Just one sample mean



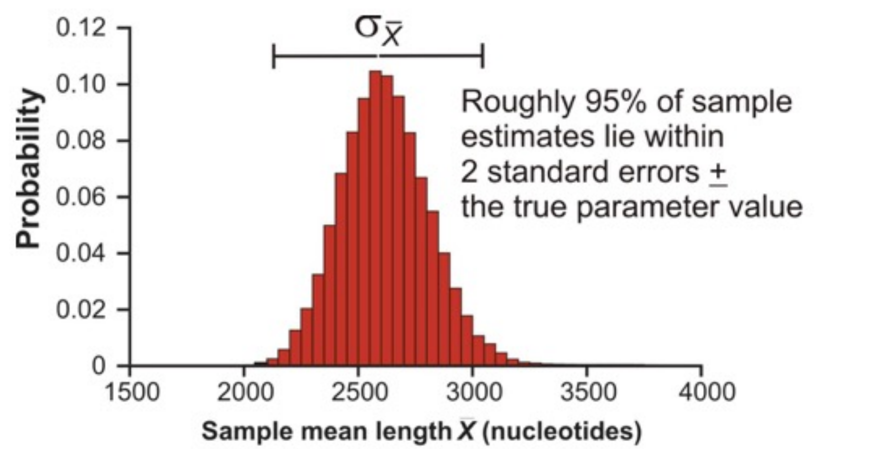
Standard Error

- The standard deviation of the sampling distribution (the standard error) measures the variation of sample estimates around the population parameter.
- Roughly, the standard error tells us **how far we are from the truth, on average**.

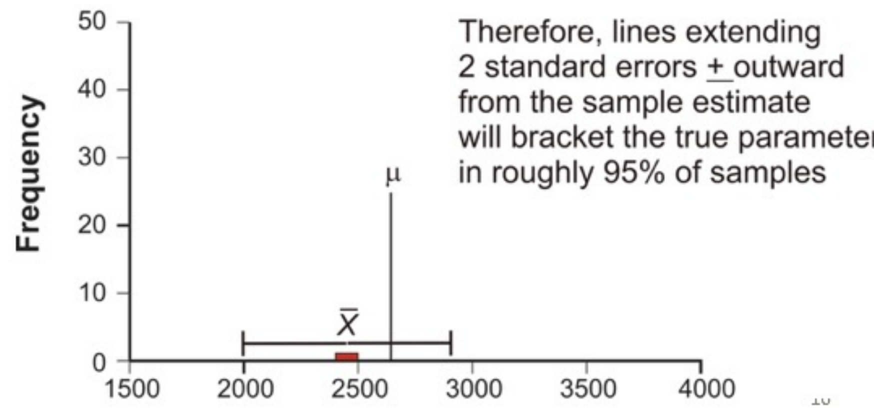


Standard error

If the sampling distribution is approximately bell-shaped, then about 95% of estimates fall within 2 SE's of the population parameter.



Twice the SE therefore provides an approximate 95% confidence interval for the parameter.



Standard error of the sample mean has a remarkable property

It can be estimated from a single sample!

$$\sigma_{\bar{X}} \approx s_{\bar{X}} = \frac{s}{\sqrt{n}}$$

$s_{\bar{X}}$ is the estimated standard error. It is usually called simply the “standard error of the mean” (SE).

This is an unusual feature of $\bar{X}$. No assumptions about normality are required.

The assumption of normality *is* required for the usual formula for the 95% confidence interval.

Standard error of the sample mean has a remarkable property

Sadly, many other kinds of estimates do not have this amazing property. What to do?

One answer: make your own sampling distribution for the estimate using the “bootstrap”.

Method invented by Efron (1979).

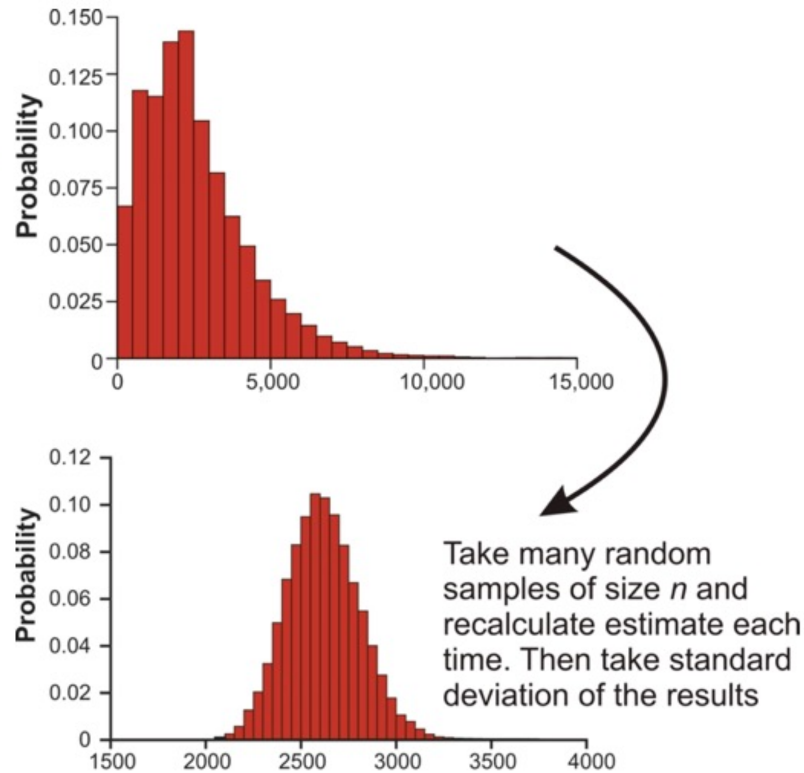
The real sampling distribution

To get real sampling distribution, take many random samples from the same population and estimate the parameter each time.

Then calculate SE as the standard deviation of the resulting sampling distribution

But in reality we only have one sample, and so only one estimate.

Rats!

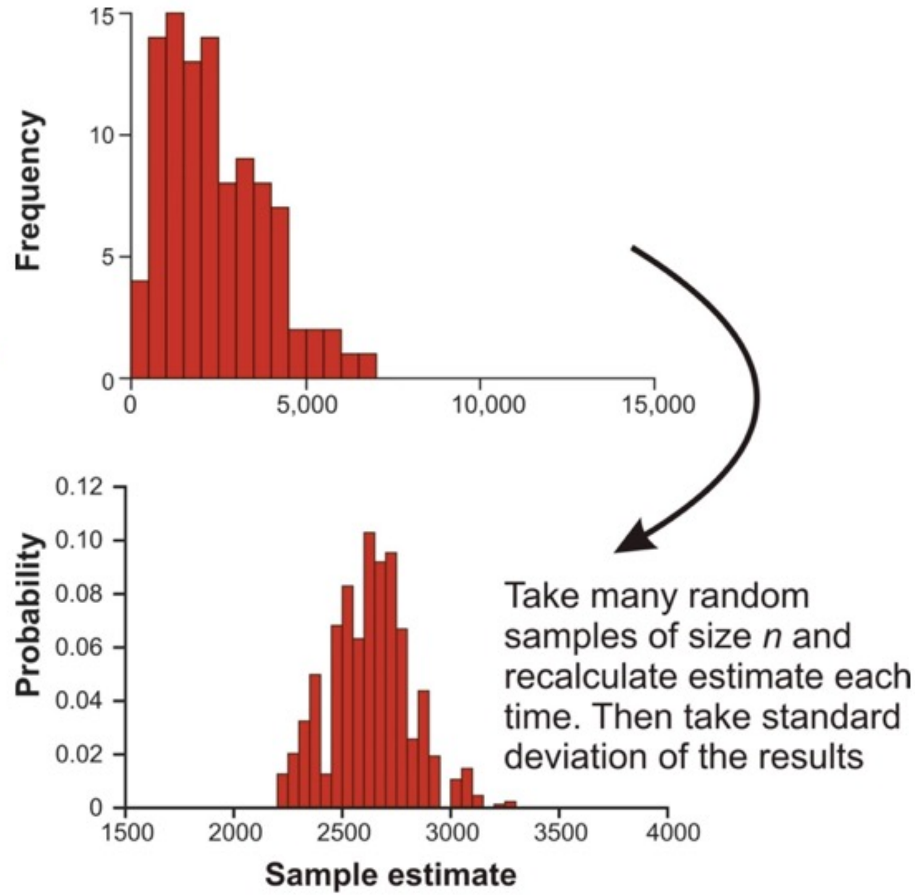


The bootstrap sampling distribution is the next best thing

Pretend the data represent the population. Sample many times from this pretend "population" instead.

Sampling is with replacement so each new bootstrap sample is missing some values from the data and has duplicates of others.

The standard deviation of resulting distribution yields the bootstrap standard error



The bootstrap procedure summary

1. **Random sample with replacement:** Use the computer to take a random sample of *n* individuals from the original data. The bootstrap sample should contain the same number of individuals as the original data: *n*. Each time an observation is chosen, it is left available in the data set to be sampled again ("sampling with replacement").
2. **Calculate the statistic this sample** (estimate) of interest using the measurements in the bootstrap sample from step 1. This is the first **bootstrap replicate estimate**.
3. **Repeat:** Repeat steps 1 and 2 many times (e.g. 10,000). The frequency distribution of all bootstrap replicate estimates yields an approximation of the sampling distribution of the estimate.
4. **Calculate on total replicates:** Calculate the sample standard deviation of **all the bootstrap replicate estimates** obtained in step 3.
5. The resulting quantity is called the **bootstrap standard error**.

Bootstrap Confidence intervals

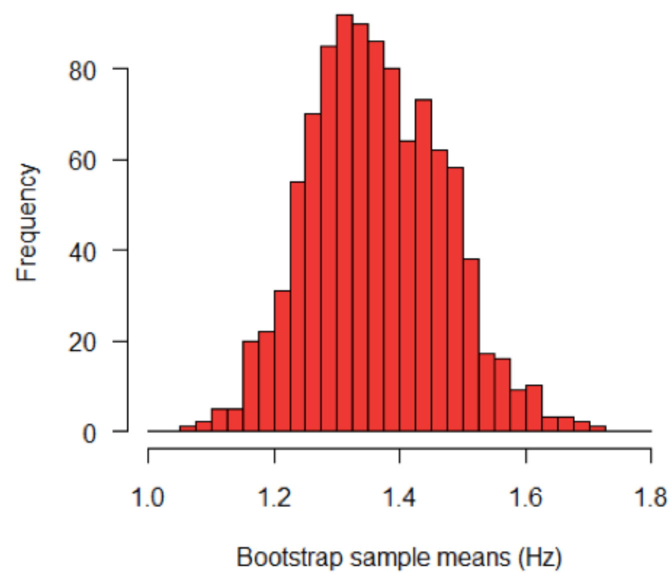
Bootstrap **percentile** confidence intervals

- There is more than 1 way to calculate CI
- Bootstrap percentile CI are more transparent & simpler mathematically compared to other bootstrap approaches
- **The 2.5th and 97.5th percentiles of the bootstrap sampling distribution are an approximate 95% confidence interval**. No transformations or normality assumptions needed.

Bootstrap confidence intervals

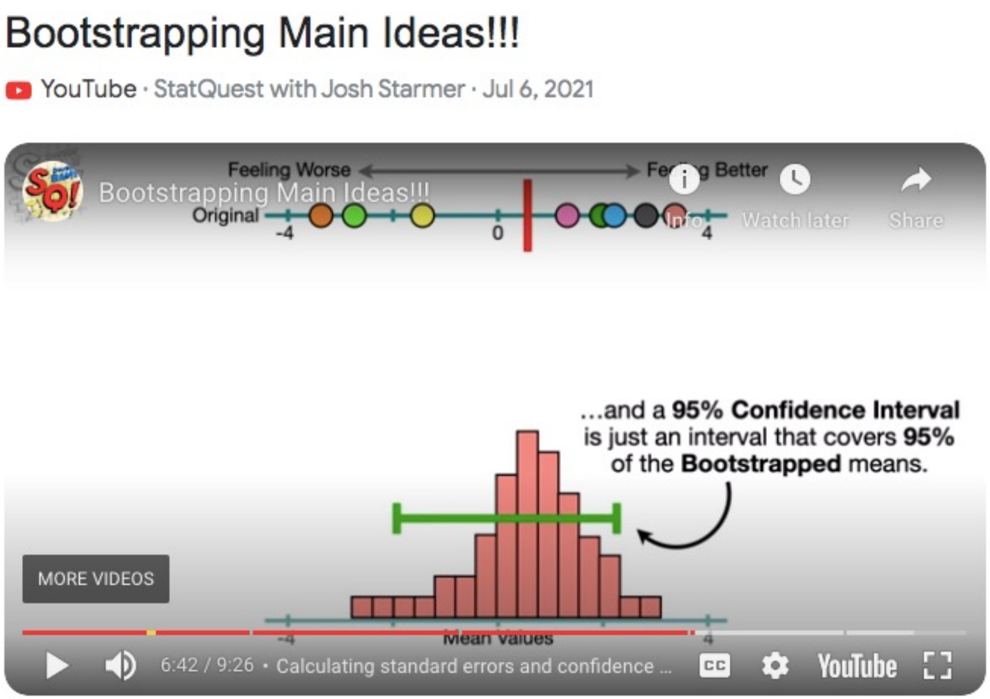
This "percentile" method of obtaining bootstrap confidence intervals works well if the sampling distribution is symmetric and unbiased.

Improved, bias-corrected and accelerated (BCa) confidence intervals improve accuracy by correcting for bias and skew in the bootstrap sampling distribution.



Additional video resource:StatQuest with Josh Starmer

Bootstrap CI is simply an interval that covers 95% of the bootstrapped means



https://www.google.com/search?q=statistics+boostrapping&source=lmns&bih=653&biw=1413&hl=en&sa=X&ved=2ahUKEwjd97H2pOz9AhWhJn0KHdR9CiYQ_AUoAHoECAEQAA#fpstate=ive&vld=cid:b27f5bd4,vid:Xz0x-8-cgaQ

Additional article about Bootstrapped CI

Bootstrapped Confidence Intervals as an Approach to Statistical Inference

MICHAEL WOOD
University of Portsmouth

Confidence intervals are in many ways a more satisfactory basis for statistical inference than hypothesis tests. This article explains a simple method for using bootstrap resampling to derive confidence intervals. This method can be used for a wide variety of statistics—including the mean and median, the difference of two means or proportions, and correlation and regression coefficients. It can be implemented by an Excel spreadsheet, which is available to readers on the Web. The rationale behind the method is transparent, and it relies on almost no sophisticated statistical concepts.

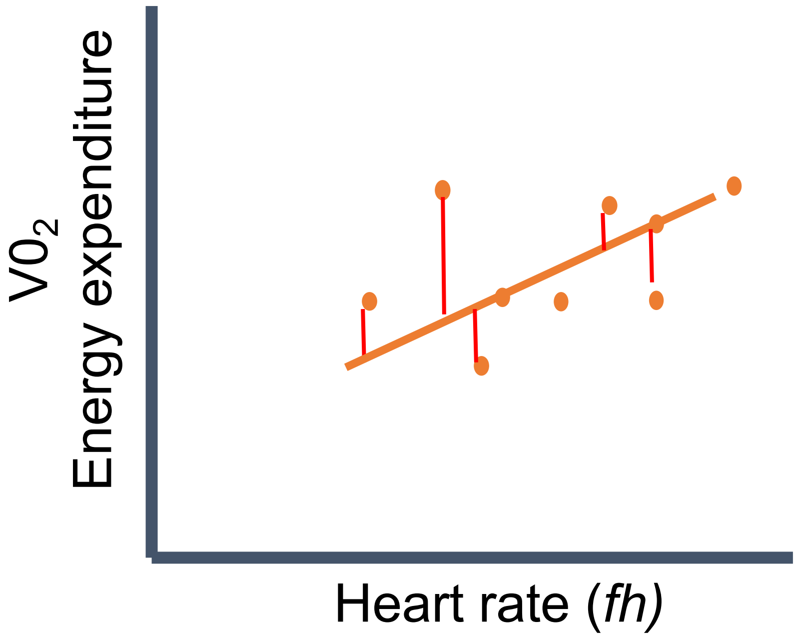
Keywords: *bootstrap; confidence interval; hypothesis test; resampling; statistics*

Wood, Michael. "Bootstrapped confidence intervals as an approach to statistical inference." *Organizational Research Methods* 8.4 (2005): 454-470.

Bootstrapping 95% CI on LME model

1. Fit overall model LME model with stepwise model building
2. Residuals represent **random error** in data that could not be modeled (difference between predicted value and true value)

model1<-lme(vo2~fh,random=~1|animal, method="ML”, data=mydata)



Young, Beth L., et al. "Dive behaviour impacts the ability of heart rate to predict oxygen consumption in Steller sea lions (*Eumetopias jubatus*) foraging at depth." Journal of Experimental Biology 214.13 (2011): 2267-2275.

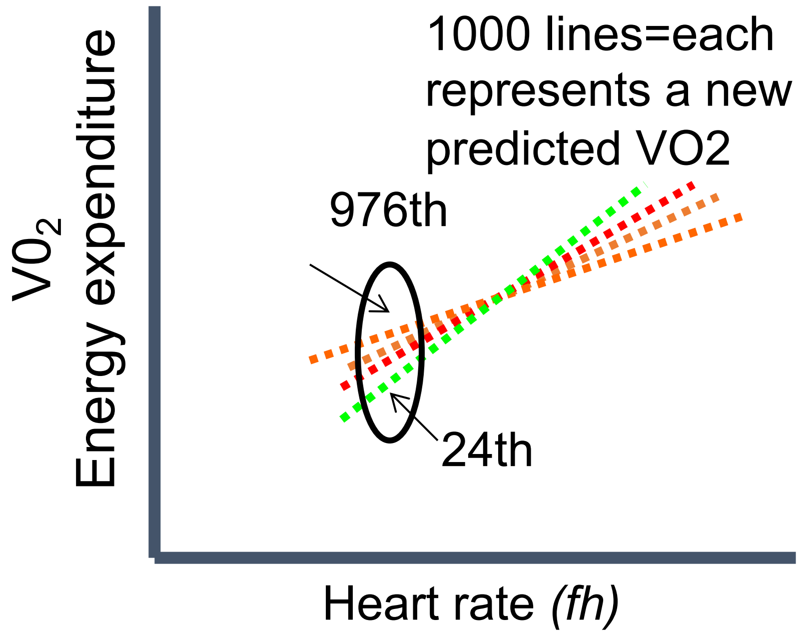
Bootstrapping 95% CI on LME model

3. Took predicted values from model1 and bootstrapped a random sample of the residuals. Reassigned these residuals randomly to the predicted values and refit model1 (preserved structure and repeated measures)

4. Repeat (3) for 1000 replicates on model1

5. Out of the 1000 replicates, order replicates, select specific replicates to represent 95% CI

- Lower CI=24th ordered replicate
  - 2.5%*1000 total replicates=25th replicate, then -1 to be conservative yields 24th replicate
- Upper CI=976th ordered replicate
  - 100-2.5%=97.5%. Then 97.5% of 1000 replicates=975th replicate. Then+1 to be conservative yields 976th replicate



Comparing 2 groups with bootstrapping

Difference between 2 (or more) groups

- Procedure is similar, but now we resample both groups

1. Use the computer to take a random sample of the data (with replacement, same sample sizes) from each group.
2. Calculate the difference between the two bootstrap samples from step 1.
3. Repeat steps 1 and 2 a very large number of times (≥1000)
4. Calculate the sample standard deviation of all the bootstrap replicate estimates obtained in step 3.

The result is the **bootstrap standard error** of the <u>difference</u>

Permutation test
Use for hypothesis testing

What does permutation mean?

- Permutation=rearrangement
- Describes the number of ways a particular dataset can be ordered or arranged.

Permutation Test

- Computer-intensive method for testing hypothesis when data are non-normal as alternative to 2-sample t-test and Mann-Whitney U-test
- Permutation describes the number of ways a particular dataset can be **ordered or re-arranged**.
- A **permutation test** generates a null distribution for a statistic measuring association between two variables (or difference among groups) by repeatedly and randomly rearranging the values of one of the variables.
- Permutation tests assume
  - Random samples
  - Distribution of the variable has same shape in every population

Permutation Test

- Randomly mix up associations among variable
- **Goal:** Gives an idea of what values the test statistic would give if 2 variables were **only** associated by chance
- Used to determine if an observed effect could reasonably been ascribed to the randomness introduced in selecting the sample(s)

Permutation test

- A **permutation test** generates a null distribution for a statistic measuring association between two variables (or difference among groups) by **repeatedly and randomly rearranging** the values of one of the variables.

Permutation Test Assumptions

- Random samples
- To compare between groups, assume that the distribution of the variable has the same shape in every population
- Robust to departures from equal-shape assumptions if n is large
- Lower power than parametric tests when n is small, but more powerful than Mann-Whitney U-test. They have similar power to parametric tests when sample size is large.

Disadvantages and Cautions for Permutation Tests

- Parametric methods provide estimates (with standard errors and confidence interval) of a useful parameter.
- Nonparametric tests, including permutations tests and rank tests, **provide only a P-value**.
- **Do not provide estimates of magnitudes (effect sizes)** with standard errors or confidence intervals.
- They **perpetuate the mistake that the P-value is all you need**, and that the smallness of the P-value indicates the importance of an effect.
  - P-value tells nothing about magnitudes or biological importance.
  - No decision should ever be made on the basis of a *P*-value alone.

Summary

Summary

- The bootstrap is amazing and useful for estimation.
- It works in almost any situation (if n not too small).
- It is approximate, though performs almost as well as parametric methods when assumptions of the parametric methods are met.
- It can also be used for hypothesis testing, though I have not discussed this.
- Permutation tests are useful for obtaining P-value, but …
- Use the bootstrap to estimate magnitudes.

Workshop

Workshop Thurs: Model Selection

Prior to workshop

- R Tips Page on “Resample,Bootstrap”
- Refresh on R code on ‘for loops()’
    - R Workshop “Plan Experiments” has a link to “extra practice with for loops”
    - R Tips Page “Loop,Repeat”

Sections of workshop

- Caribbean Birds Immigration
- Trillium fragmentation
    - Permutation test and nonparametric
- Vampire bat attack
    - No dataset to download