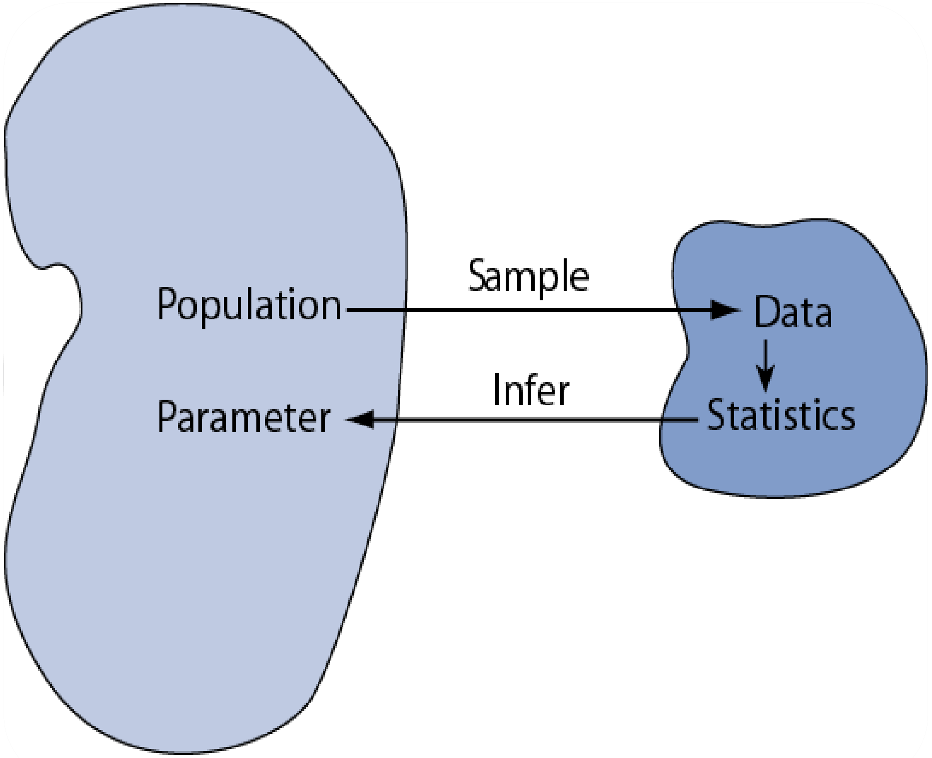
Statistical inference: confidence intervals, p-values

Introduction

Statistical inference is the act of generalizing from a sample to a population with calculated degree of certainty.

We want to learn about population *parameters* …



…but we can only calculate *sample statistics*

Population and sample

Sample statistics

$$\hat{\mu} = \overline{X}_n = \frac{\sum_{i=1}^{n} x}{n}$$

$$\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{X}_n)^2}{n-1}$$

Truth (not observable)

Population parameters

$$\mu = \frac{\sum_{i=1}^{N} x}{N} \qquad \sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

Sample (observation)

*hat notation ^ is often used to indicate "estitmate"

Make guesses about the whole population

Statistics vs. Parameters

Sample Statistic – any summary measure calculated from data; e.g., could be a mean, a difference in means or proportions, an odds ratio, or a correlation coefficient

Population parameter – the true value/true effect in the entire population of interest

Distribution of a statistic…

But the distribution of a statistic is a theoretical construct.

Statisticians ask a thought experiment: how much would the value of the statistic fluctuate if one could repeat a particular study over and over again with different samples of the same size?

By answering this question, statisticians are able to pinpoint exactly how much uncertainty is associated with a given statistic.

Distribution of a statistic

Two approaches to determine the distribution of a statistic:

- 1. Computer simulation
  - Repeat the experiment over and over again virtually!
  - More intuitive; can directly observe the behavior of statistics.
- 2. Mathematical theory
  - Proofs and formulas!
  - More practical; use formulas to solve problems.

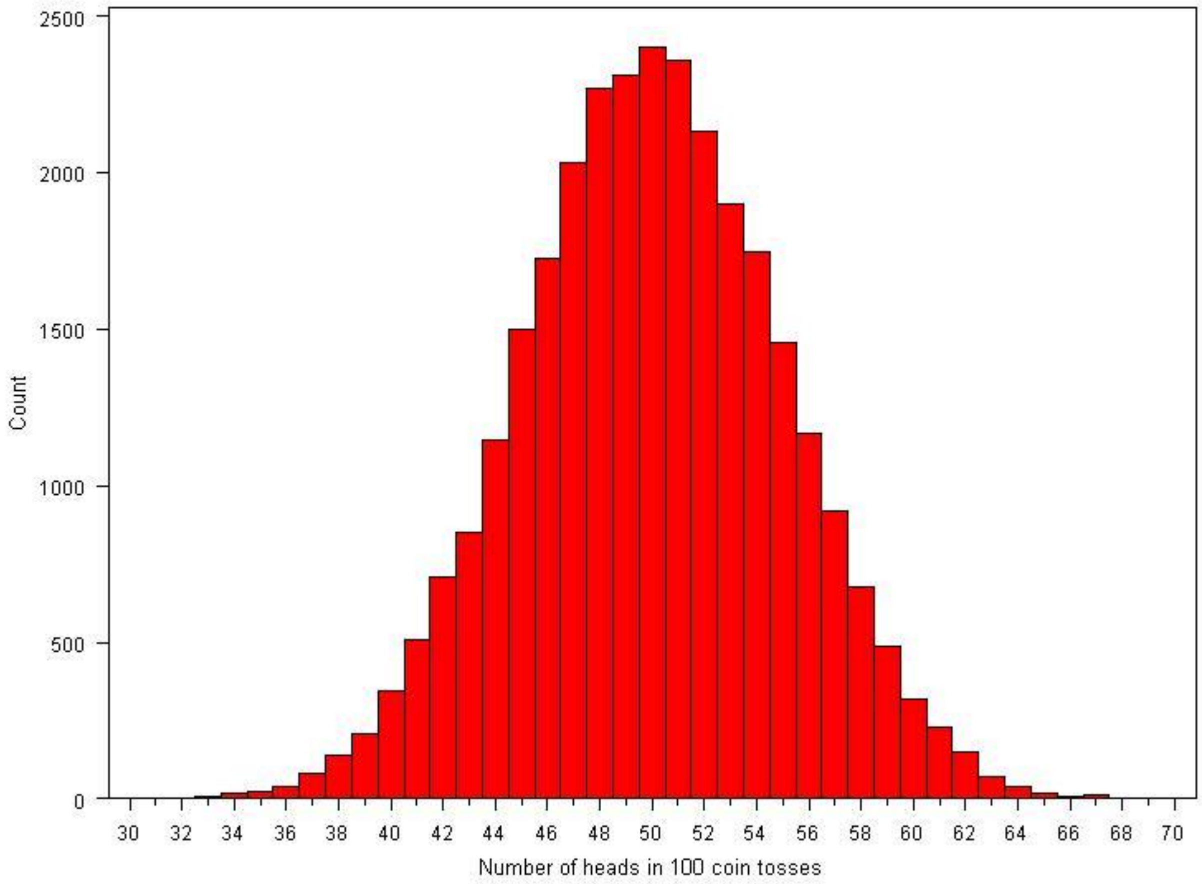
Example of computer simulation…

How many heads come up in 100 coin tosses?

Flip coins virtually

- Flip a coin 100 times; count the number of heads.
- Repeat this over and over again a large number of times (we'll try 30,000 repeats!)
- Plot the 30,000 results.

Coin tosses…



Conclusions:

We usually get between 40 and 60 heads when we flip a coin 100 times.

It's extremely unlikely that we will get 30 heads or 70 heads (didn't happen in 30,000 experiments!).

Symbol Check

$\mu_{\bar{x}}$ **The mean of the sample means.**

$\sigma_{\bar{x}}$ **The standard deviation of the sample means. *Also called "the standard error of the mean."***

Mathematical Proof

If *X* is a random variable from any distribution with known mean, E(*x*), and variance, Var(*x*), then the expected value and variance of the average of *n* observations of X is:

$$E(\bar{X}_n) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{\sum_{i=1}^{n} E(x)}{n} = \frac{nE(x)}{n} = E(x)$$

$$Var(\bar{X}_n) = Var\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{\sum_{i=1}^{n} Var(x)}{n^2} = \frac{nVar(x)}{n^2} = \frac{Var(x)}{n}$$

The Central Limit Theorem:

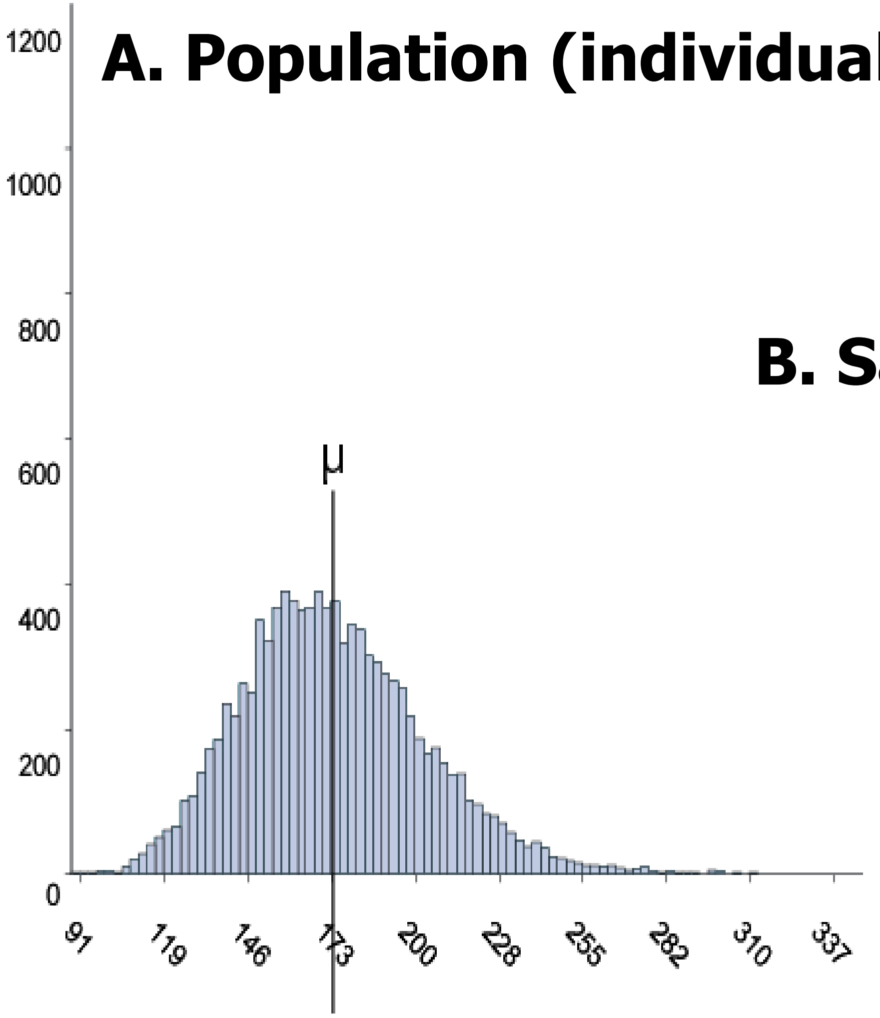
If all possible random samples, each of size *n*, are taken from any population with a mean μ and a standard deviation σ, the sampling distribution of the sample means (averages) will:

1. have mean: $$\mu_{\bar{x}} = \mu$$

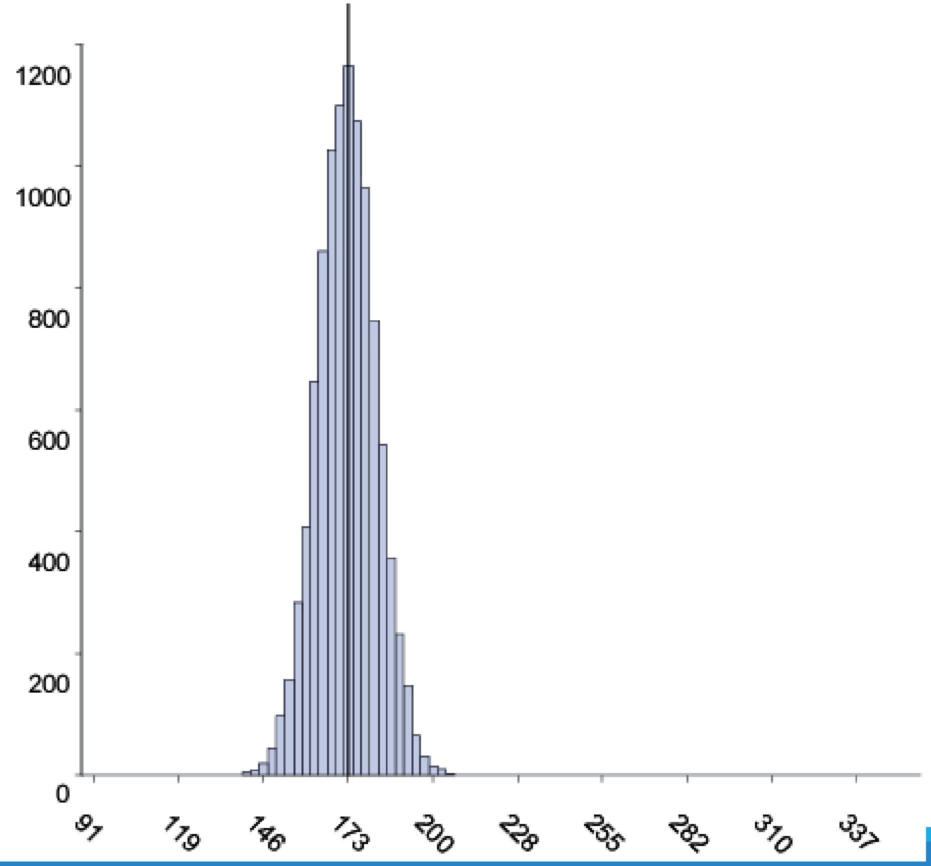
2. have standard deviation: $$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

3. be approximately normally distributed regardless of the shape of the parent population (normality improves with larger *n*)

A. Population (individual values)

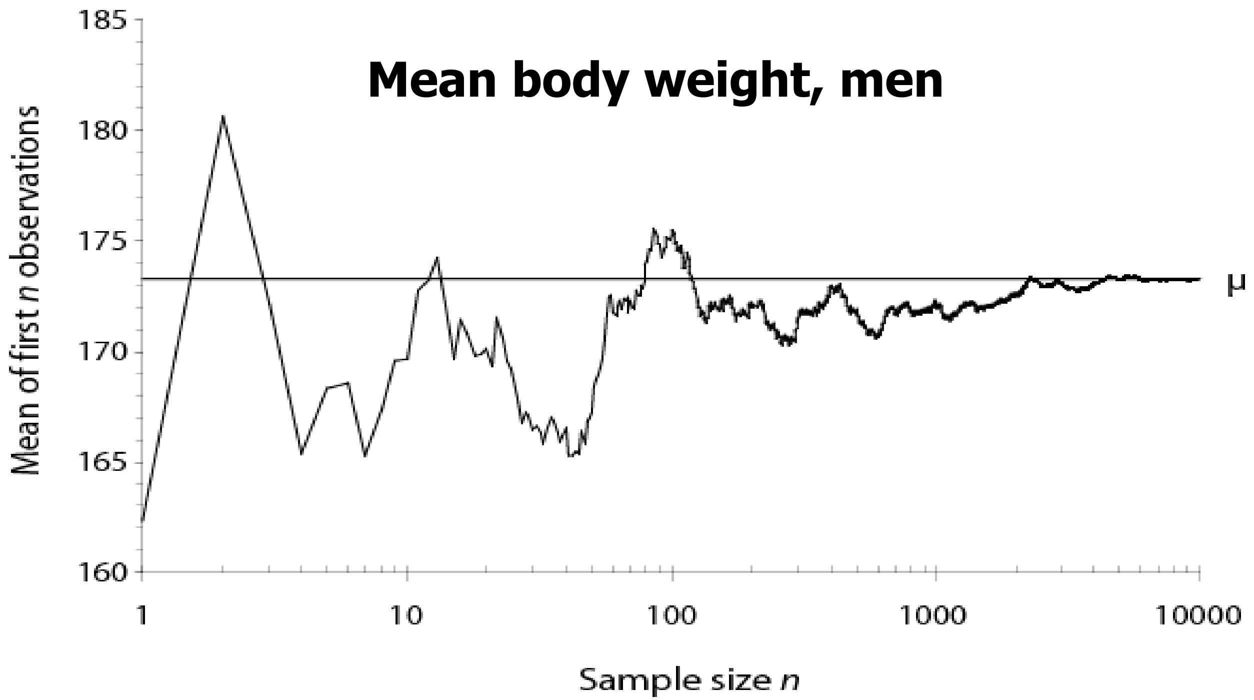


B. Sampling distribution of x-bars



Law of Large Numbers

As a sample gets larger and larger, the x-bar approaches μ. Figure demonstrates results from an experiment done in a population with μ = 173.3



Central Limit Theorem caveats for small samples:

For small samples:

- The sample standard deviation is an imprecise estimate of the true standard deviation (σ); this imprecision changes the distribution to a T-distribution.
  - A t-distribution approaches a normal distribution for large n (≥ 100), but has fatter tails for small n (< 100)
- If the underlying distribution is non-normal, the distribution of the means may be non-normal.

Summary: Single population mean (large n)

Hypothesis test:

$$Z = \frac{\text{observed mean} - \text{null mean}}{\frac{s}{\sqrt{n}}}$$

Confidence Interval

$$\text{confidence interval} = \text{observed mean} \pm Z_{\alpha/2} * (\frac{s}{\sqrt{n}})$$

Single population mean (small n, normally distributed trait)

Hypothesis test:

$$T_{n-1} = \frac{\text{observed mean} - \text{null mean}}{\frac{s}{\sqrt{n}}}$$

Confidence Interval

$$\text{confidence interval} = \text{observed mean} \pm \mathrm{T}_{n-1,\alpha/2} * (\frac{s}{\sqrt{n}})$$