Statistical Inference

Why Use Statistical Inference

- Whenever we collect data, we want our results to be true for the entire population and not just the sample that we used
- But our sample may not be representative of the population
- Inferential statistics allow us to decide if our sample results are probably true for the population
- Inferential statistics also allow us to decide if a treatment probably had an effect

Point Estimates

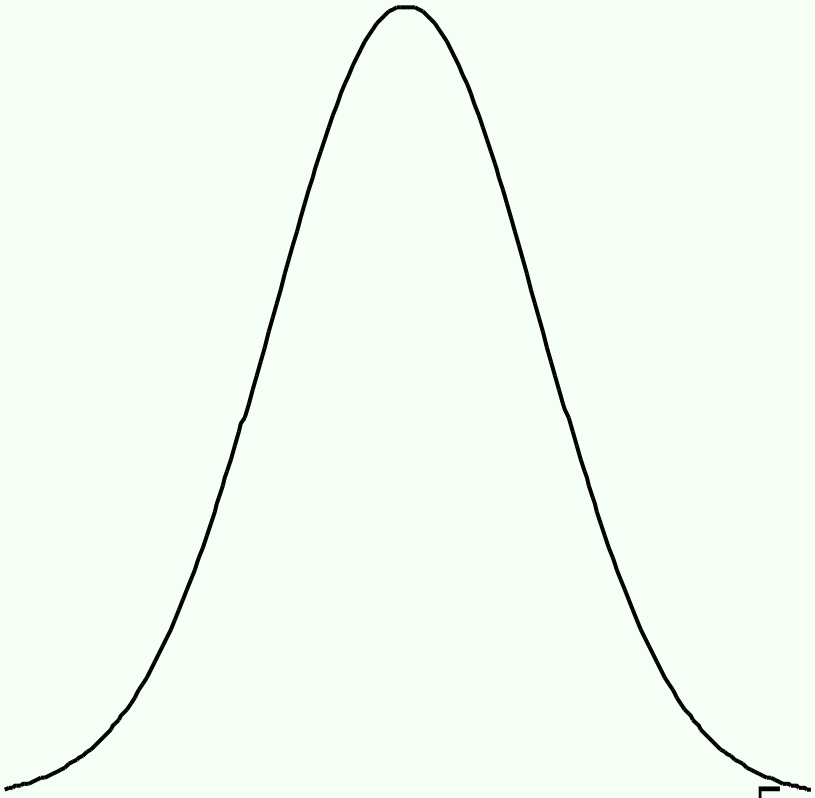
- One of our fundamental questions is: “How well does our sample statistic estimate the value of the population parameter?”
- Equivalently, we may ask “Is our *point estimate* good?”
  - A *point estimate* is a statistic (e.g. $\overline{X}$) that is calculated from sample data in order to estimate the value of the population parameter (e.g. μ)

Point Estimates

- What makes a point estimate "good"?
- First, we must define "good"
  - A good estimate is one that is close to the actual value
  - What statistic is used to calculate how close a value is to another?
    - A difference score, or deviate score $(X - \mu)$
  - What statistic should we use to measure the average "goodness?"
    - Standard deviation

Sampling Distribution

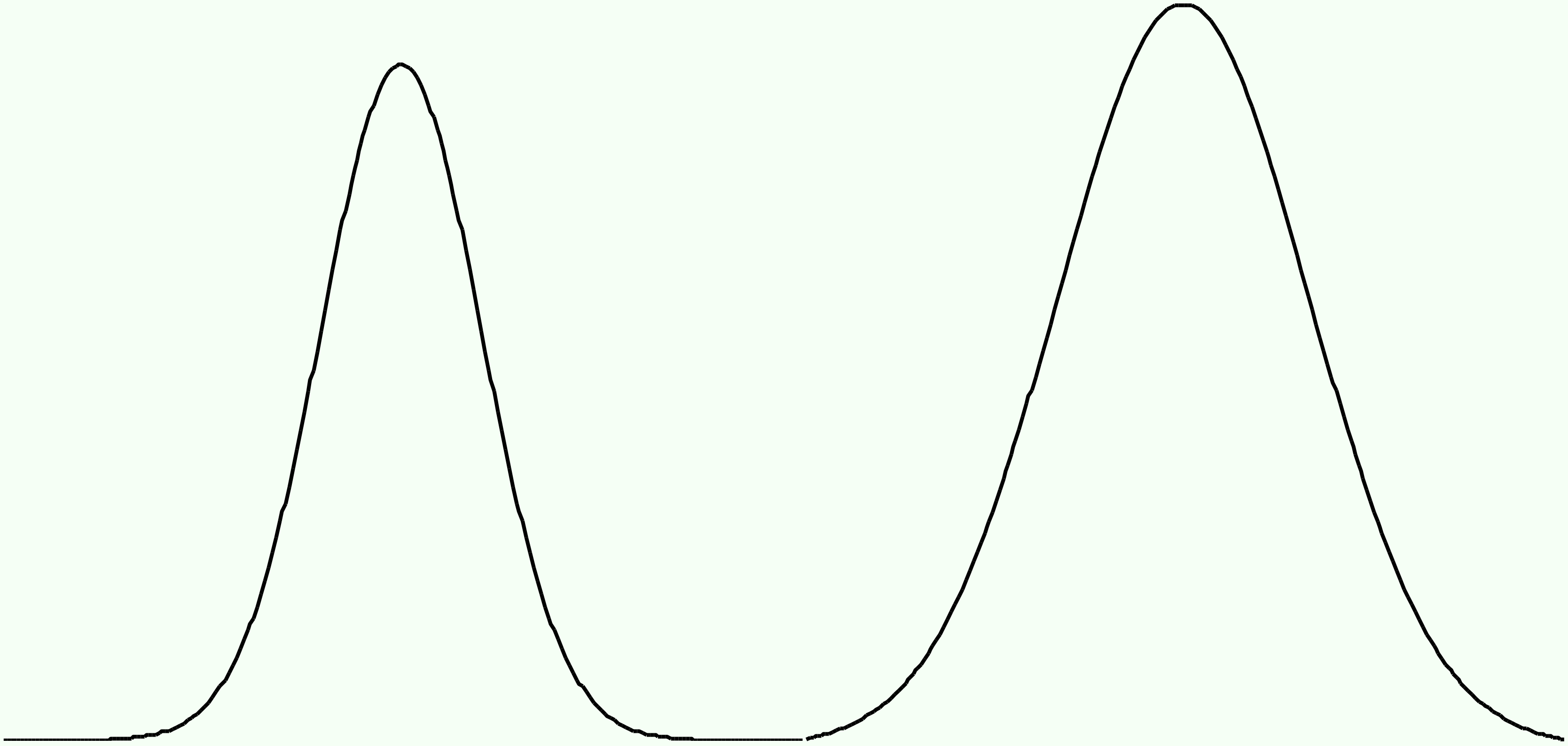
- Draw a sample from the population
- Calculate the point estimate
- Repeat the previous two steps many times
- Draw a frequency distribution of the point estimates
- That distribution is called a *sampling distribution*



Standard Error of the Mean

- The *standard error of the mean* is the standard deviation of the sampling distribution
- Thus, it is measure of how good our point estimate is likely to be
- The symbol $s_{\overline{X}}$ represents the standard error of the mean

Which Sampling Distribution Is Better?



Which sampling distribution is better? Why?

Factors Influencing $s_{\overline{X}}$

- What influences the size of the standard error of the mean?
  - That is, what can you do to make the sample mean closer to the population mean (on average)?
- Increase sample size!
  - A sample mean based on a single observation will not be as accurate as a sample mean based on 10 or 100 observations

Standard Error of the Mean

- The standard error of the mean can be estimated from the standard deviation of the sample:

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}}$$

Central Limit Theorem

- The *central limit theorem* states that the shape of a sampling distribution will be normal (or Gaussian) as long as the sample size is sufficiently large
- The mean of the sampling distribution will equal the mean of the population
- The standard deviation of the sampling distribution (I.e. the standard error of the mean) will equal the standard deviation of the samples divided by the $\sqrt{n}$

Confidence Intervals

- How confident are we in our point estimate of the population mean?
  - The population mean almost always is larger or smaller than the sample mean
- Given the sample mean and standard deviation, we can infer an interval, or range of scores, that probably contain the population mean
- This interval is called the *confidence interval*

Confidence Intervals

- Because of the central limit theorem, the sampling distribution of means is normally distributed, as long as the sample size is sufficiently large
- We can use the table of areas under the normal curve to find a range of numbers that probably contain the population mean

Confidence Intervals

- The area under the normal curve between z-scores of -1 and +1 is .68
- Thus, the 68% confidence interval is given by X ± 1 standard deviation of the sampling distribution
- E.g., X = 4.32 s_X = .57, n = 32
- X ± z x s_X / √n
- 4.32 - .57 / √32 to 4.32 + .57 / √32
- 4.22 to 4.42

Confidence Intervals

- The area under the normal curve between z-scores of -1.96 and +1.96 is .95
- Thus, the 95% confidence interval is given by $\overline{X}$ ± 1.96 standard deviation of the sampling distribution
- E.g., $\overline{X} = 4.32$ $s_X = .57$, $n = 32$
- $\overline{X} \pm z \text{ X } s_X / \sqrt{n}$
- $4.32 - 1.96 \text{ X } .57 / \sqrt{32}$ to $4.32 + 1.96 \text{ X } .57 / \sqrt{32}$
- 4.12 to 4.52

Hypothesis Testing

- *Hypothesis testing* is the procedure by which we infer if two (or more) groups are different from each other
- The first step is to write the *statistical hypotheses* which are expressed in precise mathematical terms
- The statistical hypotheses always come in pairs -- the *null hypothesis* and the *alternative hypothesis*

H_0: The Null Hypothesis

- The *null hypothesis* usually takes the following form:
- $H_0: \mu_1 = \mu_2$
- This is read as: “The null hypothesis is that the mean of condition one equals the mean of condition two”
- Notice that the null hypothesis always deals with population parameters and not the sample statistics

H_0

- The null hypothesis must contain an equal sign of some sort (=, ≥, ≤)
- Statistical tests are designed to reject H_0, never to accept it

H_1: The Alternative Hypothesis

- The alternative hypothesis usually takes the following form:
- $H_1: \mu_1 \neq \mu_2$
- This is read as: “The alternative hypothesis states that the mean of condition one does not equal the mean of condition two”
- As is true for the null, the alternative hypothesis deals with the population parameter and not the sample statistic

H_0 and H_1

- Together, the null and alternative hypotheses must be *mutually exclusive* and *exhaustive*
- Mutual exclusion implies that H_0 and H_1 cannot both be true at the same time
- Exhaustive implies that each of the possible outcomes of the experiment must make either H_0 or H_1 true

Directional vs Non-Directional Hypotheses

- The hypotheses we have been talking about are called *non-directional* hypotheses because they do not specify how the means should differ
  - That is, they do not say that the mean of condition 1 should be larger than the mean of condition 2
  - They only state that the means should differ
- Non-directional hypotheses are sometimes called *two-tailed tests*

Directional vs Non-Diretional Hypotheses

- *Directional hypotheses* include an ordinal relation between the means
  - That is, they state that one mean should be larger than the other mean
- For directional hypotheses, the H_0 and H_1 are written as:
- H_0: $\mu_1 \leq \mu_2$
- H_1: $\mu_1 > \mu_2$
- Directional hypotheses are sometimes called *one-tailed tests*

Converting Word Hypotheses into Statistical Hypotheses

- Convert the following hypothesis into statistical hypotheses:
- Frequently occurring words are easier to recall than words that occur infrequently
- Is this hypothesis directional or non-directional?
  - Directional

Converting Word Hypotheses into Statistical Hypotheses

- Write the relation that we hope to demonstrate. This will be the alternative hypothesis:
- H_1: $\mu_{frequent} > \mu_{infrequent}$
- Write a hypothesis that covers all possibilities that are not covered by the alternative hypothesis. This will be H_0:
- H_0: $\mu_{frequent} \leq \mu_{infrequent}$

Converting Word Hypotheses into Statistical Hypotheses

- Convert the following hypotheses into statistical hypotheses:
- People who eat breakfast will run a race faster or slower than those who do not eat breakfast
- People who own cats will live longer than those who do not own cats
- People who earn an A in statistics are more likely to be admitted to graduate school than those who do not earn an A

Inferential Reasoning

- Statistical inference can never tell us if two means are equal; it can only tell us if the two means are not equal
- Why?
- Statistical inference never proves that two means are not equal; it only tells us if they probably are not equal

Inferential Reasoning

- If two sample means are different from each other, does that imply that the null hypothesis is false?
- NO! Why?
- Sample means are point estimates of the population mean; thus, they are not precise predictors of the population and they change from sample to sample

Inferential Reasoning

- How different do two sample means need to be before we are willing to state that the population means are probably different?
- The answer depends on the distribution of sampling means
  - The more variable the sampling distribution is, the more different the sample means need to be

Inferential Reasoning

- The answer also depends on how willing you are to make an error and incorrectly reject H_0 when, in fact, H_0 is true
  - The less willing you are to make such an error, then the larger the difference needs to be
- This type of error is called a *Type-I* or an α error

α

- The Type-I or α error occurs when you reject H_0 when in fact H_0 is true
- We are free to decide how likely we want to be in making an α error
- The probability of making an α error is given by α
- Psychologists usually set α to either .05 or .01

Inferential Reasoning

- At some point, the sample means are sufficiently different from each other that we are comfortable in concluding that the population means are probably different
- That is, an inferential statistic has told us that the probability of making an α error is less than the α value that we arbitrarily selected

Inferential Reasoning

- When we decide that H_0 is probably not true, we *reject H_0*
- If H_0 is not tenable, then H_1 is the only remaining alternative
- Technically, we never accept H_1 as true; we only reject H_0 as being likely

Inferential Reasoning

- We never accept H_0 as true either
- We only *fail to reject* H_0
- It is always possible that the population means are different, but that the sample means are not sufficiently different

β Error (Type-II Error)

- A second type of error can occur in statistical inference
- A β *error* or *Type-II error* occurs when we fail to reject H_0 when H_0 really is false

Type-I and Type-II Errors

- Ideally, we would like to minimize both Type-I and Type-II errors
- This is not possible for a given sample size
- When we lower the α level to minimize the probability of making a Type-I error, the β level will rise
- When we lower the β level to minimize the probability of making a Type-II error, the α level will rise

Type-I and Type-II Errors