Statistical Estimation and Statistical Inference

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Statistical Estimation and Statistical Inference

Introduction

- In Chapter 5, MWL present a number of key statistical ideas. In this module, I'm going to present many of the same ideas, with a somewhat different topic ordering.

Linear Combination Theory

Basic Definitions

- Suppose you take a course with two exams. Call them MT_1 and MT_2. The final grade in the course weights the second midterm twice as much as the first, i.e.,

$$G = \frac{1}{3}MT_1 + \frac{2}{3}MT_2 \tag{1}$$

Linear Combination Theory

Basic Definitions

- Suppose you take a course with two exams. Call them MT_1 and MT_2. The final grade in the course weights the second midterm twice as much as the first, i.e.,

$$G = \frac{1}{3}MT_1 + \frac{2}{3}MT_2 \tag{1}$$

- Equation 1 is a simple example of a *linear combination*, or weighted sum.

Linear Combination Theory

Basic Definitions

- Suppose you take a course with two exams. Call them MT_1 and MT_2. The final grade in the course weights the second midterm twice as much as the first, i.e.,

$$G = \frac{1}{3}MT_1 + \frac{2}{3}MT_2 \qquad (1)$$

- Equation 1 is a simple example of a *linear combination*, or weighted sum.
- MT_1 and MT_2 are the variables being linearly combined, and $\frac{1}{3}$ and $\frac{2}{3}$ are the *linear weights*.

Linear Combination Theory

Basic Definitions

- Suppose you take a course with two exams. Call them MT_1 and MT_2. The final grade in the course weights the second midterm twice as much as the first, i.e.,

$$G = \frac{1}{3}MT_1 + \frac{2}{3}MT_2 \tag{1}$$

- Equation 1 is a simple example of a *linear combination*, or weighted sum.
- MT_1 and MT_2 are the variables being linearly combined, and $\frac{1}{3}$ and $\frac{2}{3}$ are the *linear weights*.
- In general a linear combination of p variables $X_i, i = 1, p$ is any expression that can be written in the form

$$K = \sum_{i=1}^{p} c_i X_i \tag{2}$$

Linear Combination Theory

Substantive Meaning

- Linear combinations are ubiquitous in statistics.

Linear Combination Theory

Substantive Meaning

- Linear combinations are ubiquitous in statistics.
- Their substantive meaning depends, of course, on context.

Linear Combination Theory

Substantive Meaning

- Linear combinations are ubiquitous in statistics.
- Their substantive meaning depends, of course, on context.
- However, in general, once a set of variables is chosen, a linear combination is essentially *defined by its linear weights*.

The Mean of a Linear Combination

- Consider the linear combination $G = \frac{1}{3}X + \frac{2}{3}Y$

The Mean of a Linear Combination

- Consider the linear combination $G = \frac{1}{3}X + \frac{2}{3}Y$
- The mean of the linear combination can, of course, be obtained by computing all the linear combination scores on G, using the above formula.

The Mean of a Linear Combination

- Consider the linear combination $G = \frac{1}{3}X + \frac{2}{3}Y$
- The mean of the linear combination can, of course, be obtained by computing all the linear combination scores on G, using the above formula.
- However, it is easy to prove that *the mean of a linear combination is the same linear combination of the means of the variables that are linearly combined*.

The Mean of a Linear Combination

- Consider the linear combination $G = \frac{1}{3}X + \frac{2}{3}Y$
- The mean of the linear combination can, of course, be obtained by computing all the linear combination scores on G, using the above formula.
- However, it is easy to prove that *the mean of a linear combination is the same linear combination of the means of the variables that are linearly combined*.
- So $\overline{G}_\bullet = \frac{1}{3}\overline{X}_\bullet + \frac{2}{3}\overline{Y}_\bullet$

The Variance of a Linear Combination

- Consider again the linear combination $G = \frac{1}{3}X + \frac{2}{3}Y$

The Variance of a Linear Combination

- Consider again the linear combination $G = \frac{1}{3}X + \frac{2}{3}Y$
- We saw in the previous slide that we can easily express the mean of G as a function of summary statistics on X and Y, and the linear weights that are used in the combination.

The Variance of a Linear Combination

- Consider again the linear combination $G = \frac{1}{3}X + \frac{2}{3}Y$
- We saw in the previous slide that we can easily express the mean of G as a function of summary statistics on X and Y, and the linear weights that are used in the combination.
- A similar but more complicated situation holds for the variance of G. Here is a *heuristic rule* for generating the variance formula.

The Variance of a Linear Combination

- Consider again the linear combination $G = \frac{1}{3}X + \frac{2}{3}Y$
- We saw in the previous slide that we can easily express the mean of G as a function of summary statistics on X and Y, and the linear weights that are used in the combination.
- A similar but more complicated situation holds for the variance of G. Here is a *heuristic rule* for generating the variance formula.
  1. Write the formula for the linear combination and algebraically square it, i.e.,

$$(\frac{1}{3}X + \frac{2}{3}Y)^2 = \frac{1}{9}(X^2 + 4XY + 4Y^2) \qquad (3)$$

The Variance of a Linear Combination

- Consider again the linear combination $G = \frac{1}{3}X + \frac{2}{3}Y$
- We saw in the previous slide that we can easily express the mean of G as a function of summary statistics on X and Y, and the linear weights that are used in the combination.
- A similar but more complicated situation holds for the variance of G. Here is a *heuristic rule* for generating the variance formula.

1. Write the formula for the linear combination and algebraically square it, i.e.,

$$(\frac{1}{3}X + \frac{2}{3}Y)^2 = \frac{1}{9}(X^2 + 4XY + 4Y^2) \tag{3}$$

2. Apply a "conversion rule" to the result: Leave constants alone, and replace each squared variable by the variance of that variable, and each product of two variables by the covariance of those two variables.

$$\frac{1}{9}(s_X^2 + 4s_{XY} + 4s_Y^2) \tag{4}$$

The Covariance of Two Linear Combinations

- Of course, one may define more than one linear combination on the same set of variables.

The Covariance of Two Linear Combinations

- Of course, one may define more than one linear combination on the same set of variables.
- To compute the covariance between the two linear combinations, you

The Covariance of Two Linear Combinations

- Of course, one may define more than one linear combination on the same set of variables.
- To compute the covariance between the two linear combinations, you
  1. Compute the algebraic cross-product of the two linear combination formuas, and

The Covariance of Two Linear Combinations

- Of course, one may define more than one linear combination on the same set of variables.
- To compute the covariance between the two linear combinations, you
  1. Compute the algebraic cross-product of the two linear combination formuas, and
  2. Apply the same conversion rule used to compute the variance.

The Covariance of Two Linear Combinations

- Of course, one may define more than one linear combination on the same set of variables.
- To compute the covariance between the two linear combinations, you
  1. Compute the algebraic cross-product of the two linear combination formuas, and
  2. Apply the same conversion rule used to compute the variance.
- Since the correlation r_{XY} relates to the variances and covariances via the formula

$$r_{XY} = \frac{s_{XY}}{s_X s_Y} \tag{5}$$

we also have everything we need to compute the correlation between two linear combinations.

Introduction to Sampling Distributions and Point Estimation

- In statistical estimation we use a *statistic* (a function of a sample) to estimate a *parameter*, a numerical characteristic of a statistical population.

Introduction to Sampling Distributions and Point Estimation

- In statistical estimation we use a *statistic* (a function of a sample) to estimate a *parameter*, a numerical characteristic of a statistical population.
- In the preceding discussion of the binomial distribution, we briefly discussed a well-known statistic, the sample proportion p, and how its long-run distribution over repeated samples can be described, using the binomial process and the binomial distribution as models.

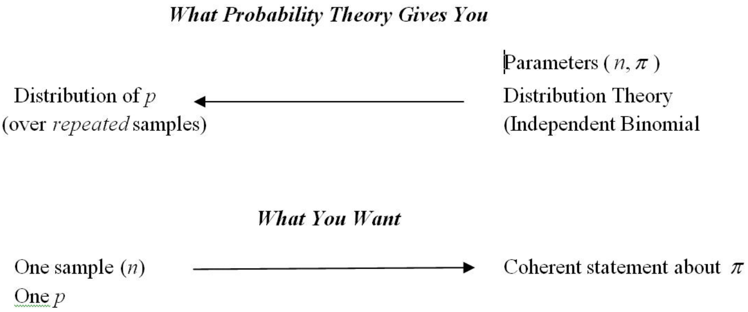
Introduction to Sampling Distributions and Point Estimation

- In statistical estimation we use a *statistic* (a function of a sample) to estimate a *parameter*, a numerical characteristic of a statistical population.
- In the preceding discussion of the binomial distribution, we briefly discussed a well-known statistic, the sample proportion p, and how its long-run distribution over repeated samples can be described, using the binomial process and the binomial distribution as models.
- We found that, if the binomial model is correct, we can describe the exact distribution of p (over repeated samples) if we know N and π, the parameters of the binomial distribution.

Introduction to Sampling Distributions and Point Estimation

- In statistical estimation we use a *statistic* (a function of a sample) to estimate a *parameter*, a numerical characteristic of a statistical population.
- In the preceding discussion of the binomial distribution, we briefly discussed a well-known statistic, the sample proportion p, and how its long-run distribution over repeated samples can be described, using the binomial process and the binomial distribution as models.
- We found that, if the binomial model is correct, we can describe the exact distribution of p (over repeated samples) if we know N and π, the parameters of the binomial distribution.
- Unfortunately, as the diagram demonstrates, what probability theory has given us is not quite what we need.

Introduction to Sampling Distributions and Point Estimation



Introduction to Sampling Distributions and Point Estimation

- Probability theory "gets us halfway" to statistical inference.

Introduction to Sampling Distributions and Point Estimation

- Probability theory "gets us halfway" to statistical inference.
- In the following sections, we'll investigate some approaches to overcoming this problem.

Introduction to Sampling Distributions and Point Estimation

- Probability theory "gets us halfway" to statistical inference.
- In the following sections, we'll investigate some approaches to overcoming this problem.
- To begin with, we should note some characteristics of sampling distributions:

Introduction to Sampling Distributions and Point Estimation

- Probability theory "gets us halfway" to statistical inference.
- In the following sections, we'll investigate some approaches to overcoming this problem.
- To begin with, we should note some characteristics of sampling distributions:
  1. Exact sampling distributions are difficult to derive

Introduction to Sampling Distributions and Point Estimation

- Probability theory "gets us halfway" to statistical inference.
- In the following sections, we'll investigate some approaches to overcoming this problem.
- To begin with, we should note some characteristics of sampling distributions:
  1. Exact sampling distributions are difficult to derive
  2. They are often different in shape from the distribution of the population from which they are sampled

Introduction to Sampling Distributions and Point Estimation

- Probability theory "gets us halfway" to statistical inference.
- In the following sections, we'll investigate some approaches to overcoming this problem.
- To begin with, we should note some characteristics of sampling distributions:
  1. Exact sampling distributions are difficult to derive
  2. They are often different in shape from the distribution of the population from which they are sampled
  3. They often vary in shape (and in other characteristics) as a function of n.

Sampling Error

- Consider any statistic $\hat{\theta}$ used to estimate a parameter θ.

Sampling Error

- Consider any statistic $\hat{\theta}$ used to estimate a parameter θ.
- For any given sample of size n, it is virtually certain that $\hat{\theta}$ will not be equal to θ.

Sampling Error

- Consider any statistic $\hat{\theta}$ used to estimate a parameter θ.
- For any given sample of size *n*, it is virtually certain that $\hat{\theta}$ will not be equal to θ.
- We can describe the situation with the following equation in random variables

$$\hat{\theta} = \theta + \varepsilon \tag{6}$$

where ε is called *sampling error*, and is defined tautologically as

$$\varepsilon = \hat{\theta} - \theta \tag{7}$$

i.e., the amount by which $\hat{\theta}$ is wrong. In most situations, ε can be either positive or negative.

Introduction

- The overriding principle in statistical estimation is that, all other things being equal, we would like ε to be as small as possible.

Introduction

- The overriding principle in statistical estimation is that, all other things being equal, we would like ε to be as small as possible.
- However, other factors intervene—Factors like cost, time, and ethics.

Introduction

- The overriding principle in statistical estimation is that, all other things being equal, we would like ε to be as small as possible.
- However, other factors intervene—Factors like cost, time, and ethics.
- In this section, we discuss some qualities that are considered *in general* to characterize a good estimator.

Introduction

- The overriding principle in statistical estimation is that, all other things being equal, we would like ε to be as small as possible.
- However, other factors intervene—Factors like cost, time, and ethics.
- In this section, we discuss some qualities that are considered *in general* to characterize a good estimator.
- We'll take a quick look at *unbiasedness*, *consistency*, and *efficiency*.

Unbiasedness

- An estimator $\hat{\theta}$ of a parameter θ is *unbiased* if $\mathrm{E}(\hat{\theta}) = \theta$, or, equivalently, if $\mathrm{E}\left(\varepsilon\right) = 0$, where ε is sampling error as defined in Equation 7.

Unbiasedness

- An estimator $\hat{\theta}$ of a parameter θ is *unbiased* if $\mathrm{E}(\hat{\theta}) = \theta$, or, equivalently, if $\mathrm{E}(\varepsilon) = 0$, where ε is sampling error as defined in Equation 7.
- Ideally, we would like the positive and negative errors of an estimator to balance out in the long run, so that, on average, the estimator is neither high (an overestimate) nor low (an underestimate).

Consistency

- We would like an estimator to get better and better as n gets larger and larger, otherwise we are wasting our effort gathering a larger sample.

Consistency

- We would like an estimator to get better and better as n gets larger and larger, otherwise we are wasting our effort gathering a larger sample.
- If we define some error tolerance ϵ, we would like to be sure that sampling error ε is almost certainly less than ϵ if we let n get large enough.

Consistency

- We would like an estimator to get better and better as n gets larger and larger, otherwise we are wasting our effort gathering a larger sample.
- If we define some error tolerance ϵ, we would like to be sure that sampling error ε is almost certainly less than ϵ if we let n get large enough.
- Formally, we say that an estimator $\hat{\theta}$ of a parameter θ is *consistent* if for any error tolerance $\epsilon > 0$, no matter how small, a sequence of statistics $\hat{\theta}_n$ based on a sample of size n will satisfy the following

$$\lim_{n \to \infty} \Pr\left(\left|\hat{\theta}_n - \theta\right| < \epsilon\right) = 1 \tag{8}$$

Consistency

Example (An Unbiased, Inconsistent Estimator)

Consider the statistic $D = (X_1 + X_2)/2$ as an estimator for the population mean. No matter how large n is, D always takes the average of just the first two observations. This statistic has an expected value of μ, the population mean, since

$$
\begin{aligned}
\mathrm{E}\left(\left[\frac{1}{2}X_1 + \frac{1}{2}X_2\right]\right) &= \frac{1}{2}E\left(X_1\right) + \frac{1}{2}E\left(X_2\right) \\
&= \frac{1}{2}\mu + \frac{1}{2}\mu \\
&= \mu
\end{aligned}
$$

but it does not keep improving in accuracy as n gets larger and larger. So it is not consistent.

Efficiency

- All other things being equal, we prefer estimators with a smaller sampling errors. Several reasonable measures of "smallness" suggest themselves:

Efficiency

- All other things being equal, we prefer estimators with a smaller sampling errors. Several reasonable measures of “smallness” suggest themselves:

  1. the average absolute error

Efficiency

- All other things being equal, we prefer estimators with a smaller sampling errors. Several reasonable measures of "smallness" suggest themselves:
  1. the average absolute error
  2. the average squared error

Efficiency

- All other things being equal, we prefer estimators with a smaller sampling errors. Several reasonable measures of "smallness" suggest themselves:
  1. the average absolute error
  2. the average squared error
- Consider the latter. The variance of an estimator can be written

$$\sigma^2_{\hat{\theta}} = \mathrm{E}\left(\hat{\theta} - \mathrm{E}\left(\hat{\theta}\right)\right)^2 \tag{9}$$

and when the estimator is unbiased, $\mathrm{E}\left(\hat{\theta}\right) = \theta$

Efficiency

- All other things being equal, we prefer estimators with a smaller sampling errors. Several reasonable measures of “smallness” suggest themselves:
  1. the average absolute error
  2. the average squared error
- Consider the latter. The variance of an estimator can be written

$$\sigma^2_{\hat{\theta}} = \mathrm{E}\left(\hat{\theta} - \mathrm{E}\left(\hat{\theta}\right)\right)^2 \tag{9}$$

  and when the estimator is unbiased, $\mathrm{E}\left(\hat{\theta}\right) = \theta$
- So the variance becomes

$$\sigma^2_{\hat{\theta}} = \mathrm{E}\left(\hat{\theta} - \theta\right)^2 = E\left(\varepsilon^2\right) \tag{10}$$

  since $\hat{\theta} - \theta = \varepsilon$.

Efficiency

- All other things being equal, we prefer estimators with a smaller sampling errors. Several reasonable measures of "smallness" suggest themselves:
  1. the average absolute error
  2. the average squared error
- Consider the latter. The variance of an estimator can be written

$$\sigma^2_{\hat{\theta}} = \mathrm{E}\left(\hat{\theta} - \mathrm{E}\left(\hat{\theta}\right)\right)^2 \tag{9}$$

  and when the estimator is unbiased, $\mathrm{E}\left(\hat{\theta}\right) = \theta$
- So the variance becomes

$$\sigma^2_{\hat{\theta}} = \mathrm{E}\left(\hat{\theta} - \theta\right)^2 = E\left(\varepsilon^2\right) \tag{10}$$

  since $\hat{\theta} - \theta = \varepsilon$.
- For an unbiased estimator, the sampling variance is also the average squared error, and is a direct measure of how inaccurate the estimator is, on average.

Efficiency

- All other things being equal, we prefer estimators with a smaller sampling errors. Several reasonable measures of "smallness" suggest themselves:
  1. the average absolute error
  2. the average squared error
- Consider the latter. The variance of an estimator can be written

$$\sigma^2_{\hat{\theta}} = \mathrm{E}\left(\hat{\theta} - \mathrm{E}\left(\hat{\theta}\right)\right)^2 \qquad (9)$$

  and when the estimator is unbiased, $\mathrm{E}\left(\hat{\theta}\right) = \theta$
- So the variance becomes

$$\sigma^2_{\hat{\theta}} = \mathrm{E}\left(\hat{\theta} - \theta\right)^2 = E\left(\varepsilon^2\right) \qquad (10)$$

  since $\hat{\theta} - \theta = \varepsilon$.
- For an unbiased estimator, the sampling variance is also the average squared error, and is a direct measure of how inaccurate the estimator is, on average.
- More generally, though, one can think of sampling variance as the randomness, or noise, inherent in a statistic. (The parameter is the "signal.") Such noise is generally to be avoided.

Efficiency

- All other things being equal, we prefer estimators with a smaller sampling errors. Several reasonable measures of "smallness" suggest themselves:
  1. the average absolute error
  2. the average squared error
- Consider the latter. The variance of an estimator can be written

$$\sigma^2_{\hat{\theta}} = \mathrm{E}\left(\hat{\theta} - \mathrm{E}\left(\hat{\theta}\right)\right)^2 \tag{9}$$

  and when the estimator is unbiased, $\mathrm{E}\left(\hat{\theta}\right) = \theta$
- So the variance becomes

$$\sigma^2_{\hat{\theta}} = \mathrm{E}\left(\hat{\theta} - \theta\right)^2 = E\left(\varepsilon^2\right) \tag{10}$$

  since $\hat{\theta} - \theta = \varepsilon$.
- For an unbiased estimator, the sampling variance is also the average squared error, and is a direct measure of how inaccurate the estimator is, on average.
- More generally, though, one can think of sampling variance as the randomness, or noise, inherent in a statistic. (The parameter is the "signal.") Such noise is generally to be avoided.
- Consequently, the *efficiency* of a statistic is inversely related to its sampling variance, i.e.

$$\text{Efficiency}(\hat{\theta}) = \frac{1}{\sigma^2_{\hat{\theta}}} \tag{11}$$

Efficiency

- The *relative efficiency* of two statistics is the ratio of their efficiencies, *which is the inverse of the ratio of their sampling variances.*

Efficiency

- The *relative efficiency* of two statistics is the ratio of their efficiencies, *which is the inverse of the ratio of their sampling variances.*

Example (Relative Efficiency)

Suppose statistic A has a sampling variance of 5, and statistic B has a sampling variance of 10. The relative efficiency of A relative to B is 2.

Sufficiency

- An estimator $\hat{\theta}$ is *sufficient* for estimating θ if it uses all the information about θ available in a sample. The formal definition is as follows:

Sufficiency

- An estimator $\hat{\theta}$ is *sufficient* for estimating θ if it uses all the information about θ available in a sample. The formal definition is as follows:

  1. Recalling that any statistic is a function of the sample, define $\hat{\theta}(S)$ to be a particular value of an estimator $\hat{\theta}$ based on a specific sample S.

Sufficiency

- An estimator $\hat{\theta}$ is *sufficient* for estimating θ if it uses all the information about θ available in a sample. The formal definition is as follows:

  1. Recalling that any statistic is a function of the sample, define $\hat{\theta}(S)$ to be a particular value of an estimator $\hat{\theta}$ based on a specific sample S.
  2. An estimator $\hat{\theta}$ is a sufficient statistic for estimating θ if the conditional distribution of the sample S given $\hat{\theta}(S)$ does not depend on θ.

Sufficiency

- An estimator $\hat{\theta}$ is *sufficient* for estimating θ if it uses all the information about θ available in a sample. The formal definition is as follows:

  1. Recalling that any statistic is a function of the sample, define $\hat{\theta}(S)$ to be a particular value of an estimator $\hat{\theta}$ based on a specific sample S.
  2. An estimator $\hat{\theta}$ is a sufficient statistic for estimating θ if the conditional distribution of the sample S given $\hat{\theta}(S)$ does not depend on θ.

- The fact that once the distribution is conditionalized on $\hat{\theta}$ it no longer depends on θ, shows that all the information that θ might reveal in the sample is captured by $\hat{\theta}$.

Maximum Likelihood

- The likelihood of a sample of n independent observations is simply the product of the probability densities of the individual observations.

Maximum Likelihood

- The likelihood of a sample of n independent observations is simply the product of the probability densities of the individual observations.
- Of course, if you don't know the parameters of the population distribution, you cannot compute the probability density of an observation.

Maximum Likelihood

- The likelihood of a sample of n independent observations is simply the product of the probability densities of the individual observations.
- Of course, if you don't know the parameters of the population distribution, you cannot compute the probability density of an observation.
- The *principle of maximum likelihood* says that the best *estimator* of a population parameter is the one that makes the sample most likely. Deriving estimators by the principle of maximum likelihood often requires calculus to solve the maximization problem, and so we will not pursue the topic here.

Practical vs. Theoretical Considerations

- In any particular situation, depending on circumstances, you may have an overriding consideration that causes you to ignore one or more of the above considerations — for example the need to make as small an error as possible when using your own data.

Practical vs. Theoretical Considerations

- In any particular situation, depending on circumstances, you may have an overriding consideration that causes you to ignore one or more of the above considerations — for example the need to make as small an error as possible when using your own data.
- In some situations, any additional error of estimation can be extremely costly, and practical considerations may dictate a biased estimator if it can be guaranteed that a bias can reduce ε for that sample.

Distribution of the Sample Mean

Sampling Mean and Variance

- From the principles of linear combinations, we saw earlier that, regardless of the shape of the population distribution, the mean and variance of the sampling distribution of the sample mean $\overline{X}_\bullet$ based on n i.i.d observations from a population with mean μ and variance σ^2 are

Distribution of the Sample Mean

Sampling Mean and Variance

- From the principles of linear combinations, we saw earlier that, regardless of the shape of the population distribution, the mean and variance of the sampling distribution of the sample mean $\overline{X}_\bullet$ based on n i.i.d observations from a population with mean μ and variance σ^2 are

    1. $\mathrm{E}(\overline{X}_\bullet) = \mu$

Distribution of the Sample Mean

Sampling Mean and Variance

- From the principles of linear combinations, we saw earlier that, regardless of the shape of the population distribution, the mean and variance of the sampling distribution of the sample mean $\overline{X}_\bullet$ based on n i.i.d observations from a population with mean μ and variance σ^2 are

  1. $\mathrm{E}(\overline{X}_\bullet) = \mu$
  2. $\mathrm{Var}(\overline{X}_\bullet) = \sigma^2/n$

Distribution of the Sample Mean

Shape

- The shape of the distribution of the sample mean depends on several factors.

Distribution of the Sample Mean

Shape

- The shape of the distribution of the sample mean depends on several factors.
- If the population distribution from which the sample was taken is normal, then $\overline{X}_\bullet$ will have a distribution that is exactly normal.

Distribution of the Sample Mean

Shape

- The shape of the distribution of the sample mean depends on several factors.
- If the population distribution from which the sample was taken is normal, then $\overline{X}_\bullet$ will have a distribution that is exactly normal.
- Even if the distribution of the population is not normal, the *Central Limit Theorem* implies that, as n becomes large, $\overline{X}_\bullet$ will still have a distribution that is approximately normal.

Distribution of the Sample Mean

Shape

- The shape of the distribution of the sample mean depends on several factors.
- If the population distribution from which the sample was taken is normal, then $\overline{X}_\bullet$ will have a distribution that is exactly normal.
- Even if the distribution of the population is not normal, the *Central Limit Theorem* implies that, as n becomes large, $\overline{X}_\bullet$ will still have a distribution that is approximately normal.
- This still leaves open the question of "how large is large enough"?

Distribution of the Sample Mean

Shape

- The shape of the distribution of the sample mean depends on several factors.
- If the population distribution from which the sample was taken is normal, then $\overline{X}_{\bullet}$ will have a distribution that is exactly normal.
- Even if the distribution of the population is not normal, the *Central Limit Theorem* implies that, as n becomes large, $\overline{X}_{\bullet}$ will still have a distribution that is approximately normal.
- This still leaves open the question of "how large is large enough"?
- For symmetric distributions, the distribution of the sample mean is often very close to normal with sample sizes as low as $n = 25$.

Distribution of the Sample Mean

Shape

- The shape of the distribution of the sample mean depends on several factors.
- If the population distribution from which the sample was taken is normal, then $\overline{X}_\bullet$ will have a distribution that is exactly normal.
- Even if the distribution of the population is not normal, the *Central Limit Theorem* implies that, as n becomes large, $\overline{X}_\bullet$ will still have a distribution that is approximately normal.
- This still leaves open the question of "how large is large enough"?
- For symmetric distributions, the distribution of the sample mean is often very close to normal with sample sizes as low as $n = 25$.
- For heavily skewed distributions, convergence to a normal shape can take much longer.

The z-Statistic

- As a pedagogical device, we first review the z-statistic for testing a null hypothesis about a single mean when the population variance σ^2 is somehow known.

The z-Statistic

- As a pedagogical device, we first review the z-statistic for testing a null hypothesis about a single mean when the population variance σ^2 is somehow known.
- Suppose that the null hypothesis is

$$H_0 : \mu = \mu_0$$

versus the two-sided alternative

$$H_1 : \mu \neq \mu_0$$

The z-Statistic

- As a pedagogical device, we first review the z-statistic for testing a null hypothesis about a single mean when the population variance σ^2 is somehow known.
- Suppose that the null hypothesis is

$$H_0 : \mu = \mu_0$$

  versus the two-sided alternative

$$H_1 : \mu \neq \mu_0$$

- This will be a two-sided test. The easiest way to devise critical regions for the test is to use a z-statistic, discussed on the next slide.

The z-Statistic

- We realize that, if the population mean is estimated with $\overline{X}_\bullet$ based on a sample of n independent observations, the statistic

$$z = \frac{\overline{X}_\bullet - \mu}{\sigma/\sqrt{n}}$$

will have a mean of zero and a standard deviation of 1.

The z-Statistic

- We realize that, if the population mean is estimated with $\overline{X}_\bullet$ based on a sample of n independent observations, the statistic

$$z = \frac{\overline{X}_\bullet - \mu}{\sigma/\sqrt{n}}$$

  will have a mean of zero and a standard deviation of 1.

- If the population is normal, this statistic will also have a normal distribution, but if the conditions are sufficiently good, convergence to a normal distribution via the Central Limit Theorem effect will occur at a reasonable sample size.

The z-Statistic

- We realize that, if the population mean is estimated with $\overline{X}_\bullet$ based on a sample of n independent observations, the statistic

$$z = \frac{\overline{X}_\bullet - \mu}{\sigma/\sqrt{n}}$$

  will have a mean of zero and a standard deviation of 1.

- If the population is normal, this statistic will also have a normal distribution, but if the conditions are sufficiently good, convergence to a normal distribution via the Central Limit Theorem effect will occur at a reasonable sample size.

- Note that, if the null hypothesis is true, then the *test statistic*

$$z = \frac{\overline{X}_\bullet - \mu_0}{\sigma/\sqrt{n}} \tag{12}$$

  will also have a $N(0, 1)$ distribution.

The z-Statistic

- We realize that, if the population mean is estimated with $\overline{X}_\bullet$ based on a sample of n independent observations, the statistic

$$z = \frac{\overline{X}_\bullet - \mu}{\sigma/\sqrt{n}}$$

will have a mean of zero and a standard deviation of 1.

- If the population is normal, this statistic will also have a normal distribution, but if the conditions are sufficiently good, convergence to a normal distribution via the Central Limit Theorem effect will occur at a reasonable sample size.
- Note that, if the null hypothesis is true, then the *test statistic*

$$z = \frac{\overline{X}_\bullet - \mu_0}{\sigma/\sqrt{n}} \tag{12}$$

will also have a $N(0,1)$ distribution.

- To have a rejection region that controls α at 0.05, we can select the upper and lower 2.5% of the standard normal distribution, i.e. ± 1.96.

The z-Statistic

- We realize that, if the population mean is estimated with $\overline{X}_\bullet$ based on a sample of n independent observations, the statistic

$$z = \frac{\overline{X}_\bullet - \mu}{\sigma/\sqrt{n}}$$

will have a mean of zero and a standard deviation of 1.

- If the population is normal, this statistic will also have a normal distribution, but if the conditions are sufficiently good, convergence to a normal distribution via the Central Limit Theorem effect will occur at a reasonable sample size.
- Note that, if the null hypothesis is true, then the *test statistic*

$$z = \frac{\overline{X}_\bullet - \mu_0}{\sigma/\sqrt{n}} \tag{12}$$

will also have a $N(0,1)$ distribution.

- To have a rejection region that controls α at 0.05, we can select the upper and lower 2.5% of the standard normal distribution, i.e. ± 1.96.
- More generally, the absolute value of the rejection point for this statistic will be, for a test with T tails (either 1 or 2)

$$\Phi^{-1}(1 - \alpha/T) \tag{13}$$

with $\Phi^{-1}()$ the standard normal quantile function.

The z-Statistic

Calculating the Rejection Point

Example (Calculating the Rejection Point)

We can easily calculate the rejection point with a simple R function.

```
> Z1CriticalValue <- function(alpha, tails = 2) {
+     crit = qnorm(1 - alpha/abs(tails))
+     if (tails == 2 || tails == 1)
+         return(crit)
+     if (tails == -1)
+         return(-crit) else return(NA)
+ }
```

To use the function, input the significance level and the number of tails. If the test is one-tailed, enter either 1 or -1 depending on whether the critical region is on the low or high end of the number line relative to μ_0. The default is a two-sided test.

```
> Z1CriticalValue(0.05, 2)
[1] 1.96
> Z1CriticalValue(0.05, -1)
[1] -1.645
```

The Non-Null Distribution of z

- Of course, H_0 need not be true, and in many contexts is almost certainly false.

The Non-Null Distribution of z

- Of course, H_0 need not be true, and in many contexts is almost certainly false.
- The question then becomes one of statistical power.

The Non-Null Distribution of z

- Of course, H_0 need not be true, and in many contexts is almost certainly false.
- The question then becomes one of statistical power.
- Recall that the general approach to power calculation involves first defining the critical region, then determining the distribution of the test statistic under the true state of the world.

The Non-Null Distribution of z

- Of course, H_0 need not be true, and in many contexts is almost certainly false.
- The question then becomes one of statistical power.
- Recall that the general approach to power calculation involves first defining the critical region, then determining the distribution of the test statistic under the true state of the world.
- Suppose that the null hypothesis is that $\mu = \mu_0$, but μ is actually equal to some other value.

The Non-Null Distribution of z

- Of course, H_0 need not be true, and in many contexts is almost certainly false.
- The question then becomes one of statistical power.
- Recall that the general approach to power calculation involves first defining the critical region, then determining the distribution of the test statistic under the true state of the world.
- Suppose that the null hypothesis is that $\mu = \mu_0$, but μ is actually equal to some other value.
- What will the distribution of the z-statistic be?

The Non-Null Distribution of z

- We can derive this easily using our algebra of variances and covariances, because $\overline{X}_\bullet$ is the only random variable in the formula for the z-statistic.

The Non-Null Distribution of z

- We can derive this easily using our algebra of variances and covariances, because $\overline{X}_\bullet$ is the only random variable in the formula for the z-statistic.
- We can easily prove that the z-statistic has a distribution with a mean of $\sqrt(n)E_s$ and a standard deviation of 1.

The Non-Null Distribution of z

- We can derive this easily using our algebra of variances and covariances, because $\overline{X}_\bullet$ is the only random variable in the formula for the z-statistic.
- We can easily prove that the z-statistic has a distribution with a mean of $\sqrt(n)E_s$ and a standard deviation of 1.
- E_s, the "standardized effect size," is defined as

$$E_s = \frac{\mu - \mu_0}{\sigma} \tag{14}$$

and is the amount by which the null hypothesis is wrong, re-expressed in "standard deviation units."

Power Calculation with the One-Sample z

- Now we demonstrate power calculation using the previous results.

Power Calculation with the One-Sample z

- Now we demonstrate power calculation using the previous results.
- Suppose our null hypothesis is one-sided, i.e.

$$H_0 : \mu \leq 70 \quad H_1 : \mu > 70 \tag{15}$$

Power Calculation with the One-Sample z

- Now we demonstrate power calculation using the previous results.
- Suppose our null hypothesis is one-sided, i.e.

$$H_0 : \mu \leq 70 \quad H_1 : \mu > 70 \tag{15}$$

- In this case, then, $\mu_0 = 70$. Assume now that $\sigma = 10$, and that the true state of the world is that $\mu = 75$. What will the power of the z-statistic be if $n = 25$, and we perform the test with the significance level $\alpha = 0.05$?

Power Calculation with the One-Sample z

- In this case, the standardized effect size is

$$E_s = \frac{75 - 70}{10} = 0.50$$

- The mean of the z-statistic is $\sqrt{n}E_s = \sqrt{25} \times 0.50 = 2.50$, and the statistic has a standard deviation of 1.
- The rejection point is one-tailed, and may be calculated using our function as

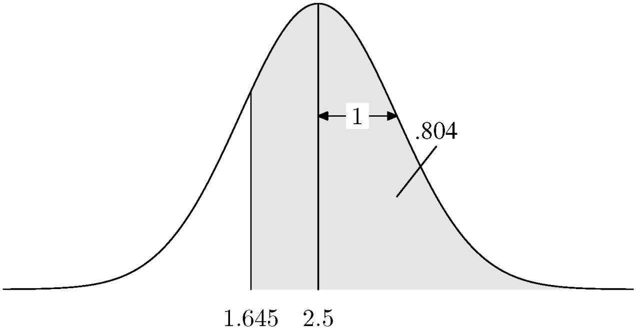
```
> Z1CriticalValue(0.05, 1)

[1] 1.645
```

- The power of the test may be calculated as the probability of exceeding the rejection point of 1.645.

```
> 1 - pnorm(Z1CriticalValue(0.05, 1), 2.5, 1)

[1] 0.8038
```

Power Calculation with the One-Sample z

- Here is a picture of the situation.

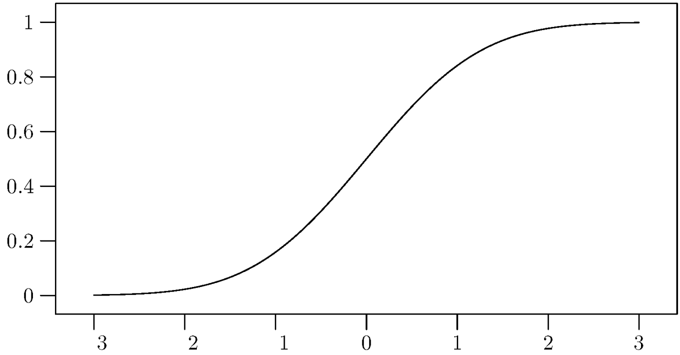


Automating the Power Calculation

- Notice that the power can be calculated by computing the area to the right of 1.645 in a normal distribution with a mean of 2.50 and a standard deviation of 1, but this is also the area to the right of $1.645 - 2.50$ in a normal distribution with a mean of 0 and a standard deviation of 1.
- Since the area to the right of a negative value in a symmetric distribution is equal to the area to the left of its positive equivalent, we arrive at the fact that power in the 1-sample z-test is equal to the area to the left of $\sqrt{n}|E_s| - |R|$ in a standard normal distribution. R is the rejection point.
- Note, we are assuming that the hypothesis test is two-sided, or that the effect is in the hypothesized direction if the hypothesis is one-sided. (Typically, one would not be interested in computing power to detect an effect in the wrong direction!)
- We are also ignoring the miniscule probability of rejection "on the wrong side" with a two-sided test.

Automating the Power Calculation

- Since power is the area to the left of a point on the normal curve, a power chart for the one-sample z-test has the same shape as the cumulative probability curve for the normal distribution.



Automating the Power Calculation

- We can write an R function to compute power of the z-test.

```
> power.onesample.z <- function(mu, mu0, sigma, n, alpha, tails = 2) {
+     Es <- (mu - mu0)/sigma
+     R <- Z1CriticalValue(alpha, tails)
+     m <- sqrt(n) * abs(Es) - abs(R)
+     return(pnorm(m))
+ }
> power.onesample.z(75, 70, 10, 25, 0.05, 1)

[1] 0.8038
```

Sample Size Calculation with the One-Sample z

- We just calculated a power of 0.804 to detect a standardized effect of 0.50 standard deviations.

Sample Size Calculation with the One-Sample z

- We just calculated a power of 0.804 to detect a standardized effect of 0.50 standard deviations.
- Suppose that this power is deemed insufficient for our purposes, that we need a power of 0.95, and that we wish to manipulate power by increasing sample size.

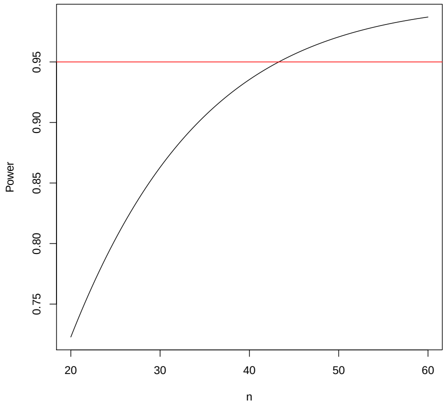
Sample Size Calculation with the One-Sample z

- We just calculated a power of 0.804 to detect a standardized effect of 0.50 standard deviations.
- Suppose that this power is deemed insufficient for our purposes, that we need a power of 0.95, and that we wish to manipulate power by increasing sample size.
- What is the minimum sample size necessary to achieve our desired power?

Sample Size Calculation with the One-Sample z

- We could estimate the power by plotting the power over a range of potential values of n, using our power function.
- The plot shows we need an n of around 42–44.

```
> curve(power.onesample.z(75, 70, 10, x, 0.05, 1), 20, 60, xlab = "n", ylab = "Power")
> abline(h = 0.95, col = "red")
```



Sample Size Calculation with the One-Sample z

- Having narrowed things down, we could then input a vector of possible values of n, and construct a table, thereby discovering the value we need.

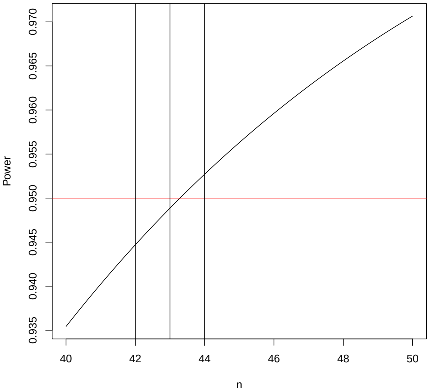
```
> n <- 40:45
> power <- power.onesample.z(75, 70, 10, n, 0.05, 1)
> cbind(n, power)

      n  power
[1,] 40 0.9354
[2,] 41 0.9402
[3,] 42 0.9447
[4,] 43 0.9489
[5,] 44 0.9527
[6,] 45 0.9563
```

- Now it is clear that the minimum n is 44.

Sample Size Calculation with the One-Sample z

- An alternative, graphical approach would be to redraw the graph in a narrower range, and draw vertical lines at the key values.

```
> curve(power.onesample.z(75, 70, 10, x, 0.05, 1), 40, 50, xlab = "n", ylab = "Power")
> abline(h = 0.95, col = "red")
> abline(v = 42)
> abline(v = 43)
> abline(v = 44)
```



Sample Size Calculation with the One-Sample z

- It turns out, we can, in this case, easily derive an analytic formula for the required sample size.

Sample Size Calculation with the One-Sample z

- It turns out, we can, in this case, easily derive an analytic formula for the required sample size.
- Denoting power by P, the normal curve cumulative probability function as $\Phi(\ )$ and the number of tails by T, we saw that power can be written

$$P = \Phi\left(\sqrt{n}|E_s| - |\Phi^{-1}(1 - \alpha/T)|\right) \tag{16}$$

Sample Size Calculation with the One-Sample z

- It turns out, we can, in this case, easily derive an analytic formula for the required sample size.
- Denoting power by P, the normal curve cumulative probability function as $\Phi(\ )$ and the number of tails by T, we saw that power can be written

$$P = \Phi\left(\sqrt{n}|E_s| - |\Phi^{-1}(1-\alpha/T)|\right) \tag{16}$$

- With some trivial algebraic manipulation, we can isolate n on one side of the equation. The key is to remember that $\Phi(\ )$ is an invertible function.

$$n = \text{ceiling}\left(\frac{\Phi^{-1}(P) + \Phi^{-1}(1-\alpha/T)}{|E_s|}\right)^2 \tag{17}$$

Sample Size Calculation with the One-Sample z

- It turns out, we can, in this case, easily derive an analytic formula for the required sample size.
- Denoting power by P, the normal curve cumulative probability function as $\Phi(\ )$ and the number of tails by T, we saw that power can be written

$$P = \Phi\left(\sqrt{n}|E_s| - |\Phi^{-1}(1 - \alpha/T)|\right) \tag{16}$$

- With some trivial algebraic manipulation, we can isolate n on one side of the equation. The key is to remember that $\Phi(\ )$ is an invertible function.

$$n = \text{ceiling}\left(\frac{\Phi^{-1}(P) + \Phi^{-1}(1 - \alpha/T)}{|E_s|}\right)^2 \tag{17}$$

- You may be wondering about the meaning of the term "ceiling" in the above formula.

Sample Size Calculation with the One-Sample z

- It turns out, we can, in this case, easily derive an analytic formula for the required sample size.
- Denoting power by P, the normal curve cumulative probability function as $\Phi(\ )$ and the number of tails by T, we saw that power can be written

$$P = \Phi\left(\sqrt{n}|E_s| - |\Phi^{-1}(1-\alpha/T)|\right) \tag{16}$$

- With some trivial algebraic manipulation, we can isolate n on one side of the equation. The key is to remember that $\Phi(\ )$ is an invertible function.

$$n = \text{ceiling}\left(\frac{\Phi^{-1}(P) + \Phi^{-1}(1-\alpha/T)}{|E_s|}\right)^2 \tag{17}$$

- You may be wondering about the meaning of the term "ceiling" in the above formula.
- Usually n as calculated (without the final application of the ceiling function) in the above will not be an integer, and to exceed the required power, you will need to use the smallest integer that is not less than n as calculated.

Sample Size Calculation with the One-Sample z

- It turns out, we can, in this case, easily derive an analytic formula for the required sample size.
- Denoting power by P, the normal curve cumulative probability function as $\Phi(\ )$ and the number of tails by T, we saw that power can be written

$$P = \Phi\left(\sqrt{n}|E_s| - |\Phi^{-1}(1-\alpha/T)|\right) \tag{16}$$

- With some trivial algebraic manipulation, we can isolate n on one side of the equation. The key is to remember that $\Phi(\ )$ is an invertible function.

$$n = \text{ceiling}\left(\frac{\Phi^{-1}(P) + \Phi^{-1}(1-\alpha/T)}{|E_s|}\right)^2 \tag{17}$$

- You may be wondering about the meaning of the term "ceiling" in the above formula.
- Usually n as calculated (without the final application of the ceiling function) in the above will not be an integer, and to exceed the required power, you will need to use the smallest integer that is not less than n as calculated.
- This value is called ceiling(n).

Sample Size Calculation with the One-Sample z

- The function returns the required n, and the actual power at that sample size.
- The actual power will generally exceed the requested power by a small amount.

```
> Required.n.Z1 <- function(mu, mu0, sigma, alpha, power, tails) {
+     Es <- (mu - mu0)/sigma
+     R <- Z1CriticalValue(alpha, tails)
+     n <- ((qnorm(1 - alpha/tails) + qnorm(power))/(abs(Es)))^2
+     n <- ceiling(n)
+     return(c(n, power.onesample.z(mu, mu0, sigma, n, alpha, tails)))
+ }
> Required.n.Z1(75, 70, 10, 0.05, 0.95, 1)

[1] 44.0000  0.9527
```

Direct Calculations from Standardized Effect Size

- A common misconception in the early days of power analysis was that one actually needed to know μ, μ_0, and σ (or at least have educated guesses) before proceeding.

Direct Calculations from Standardized Effect Size

- A common misconception in the early days of power analysis was that one actually needed to know μ, μ_0, and σ (or at least have educated guesses) before proceeding.
- Among others, Cohen, in his classic book on statistical power analysis, suggested that one can proceed by positing typical values of E_s instead.

Direct Calculations from Standardized Effect Size

- A common misconception in the early days of power analysis was that one actually needed to know μ, μ_0, and σ (or at least have educated guesses) before proceeding.
- Among others, Cohen, in his classic book on statistical power analysis, suggested that one can proceed by positing typical values of E_s instead.
- E_s, after all, is a "metric-free" measure of effect size.

Direct Calculations from Standardized Effect Size

- A common misconception in the early days of power analysis was that one actually needed to know μ, μ_0, and σ (or at least have educated guesses) before proceeding.
- Among others, Cohen, in his classic book on statistical power analysis, suggested that one can proceed by positing typical values of E_s instead.
- E_s, after all, is a "metric-free" measure of effect size.
- Cohen suggested simply using E_s values of 0.20, 0.50, 0.80 as proxies for "small," "medium," and "large" effects.

Direct Calculations from Standardized Effect Size

- A common misconception in the early days of power analysis was that one actually needed to know μ, μ_0, and σ (or at least have educated guesses) before proceeding.
- Among others, Cohen, in his classic book on statistical power analysis, suggested that one can proceed by positing typical values of E_s instead.
- E_s, after all, is a "metric-free" measure of effect size.
- Cohen suggested simply using E_s values of 0.20, 0.50, 0.80 as proxies for "small," "medium," and "large" effects.
- In the case of our power functions, inputting "dummy" values of $\mu_0 = 0$ and $\sigma = 1$ allows you to enter E_s directly as the μ parameter.

Direct Calculations from Standardized Effect Size

Example (Power Calculation using E_s)

Suppose you wish power of 0.90 to detect a "small" effect with a two-sided test with $\alpha = .01$. What is the required n? As shown below, the required n is 372.

```
> Required.n.Z1(0.2, 0, 1, 0.01, 0.9, 2)
[1] 372.0   0.9
```

Direct Calculations from Standardized Effect Size

A Caution

- A number of authors have cautioned against relying too heavily on Cohen's guidelines.

Direct Calculations from Standardized Effect Size

A Caution

- A number of authors have cautioned against relying too heavily on Cohen's guidelines.
- For example, in some situations, a standardized effect size of 0.20 may represent an effect of enormous significance.

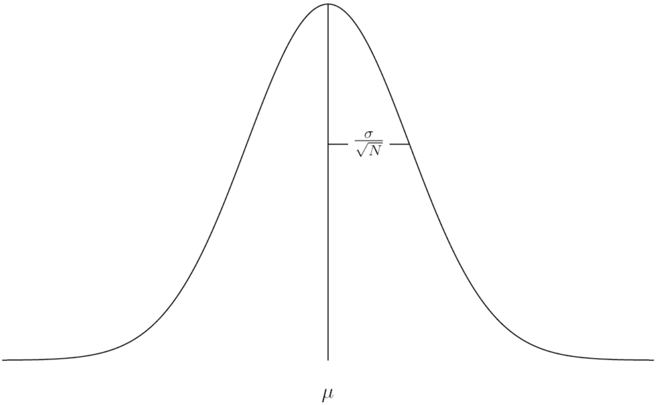
Direct Calculations from Standardized Effect Size

A Caution

- A number of authors have cautioned against relying too heavily on Cohen's guidelines.
- For example, in some situations, a standardized effect size of 0.20 may represent an effect of enormous significance.
- Different measuring instruments may contribute different levels of error variance to different experiments. So σ reflects partly the effect of measurement error, and partly the true individual variation in a measured construct. Consequently, the "same" E_s might reflect different true effects, and vice-versa.

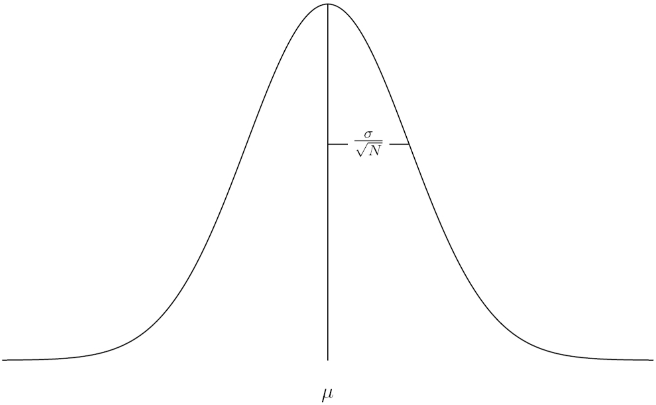
Confidence Interval Estimation

- Assume, for the time being, that we know that the distribution of $\overline{X}_\bullet$ over repeated samples is as pictured below.



Confidence Interval Estimation

- Assume, for the time being, that we know that the distribution of $\overline{X}_\bullet$ over repeated samples is as pictured below.
- This graph demonstrates the distribution of $\overline{X}_\bullet$ over repeated samples. In the above graph, as in any normal curve, 95% of the time a value will be between z-score equivalents of -1.96 and $+1.96$.



Confidence Interval Estimation

- These points are at a raw score that is 1.96 standard deviations below the mean and 1.96 standard deviations above the mean.

Confidence Interval Estimation

- These points are at a raw score that is 1.96 standard deviations below the mean and 1.96 standard deviations above the mean.
- Consequently, if we mark points on the graph at $\mu - 1.96\sigma/\sqrt{n}$ and $\mu + 1.96\sigma/\sqrt{n}$, we will have two points between which $\overline{X}_\bullet$ will occur 95% of the time.

Confidence Interval Estimation

- These points are at a raw score that is 1.96 standard deviations below the mean and 1.96 standard deviations above the mean.
- Consequently, if we mark points on the graph at $\mu - 1.96\sigma/\sqrt{n}$ and $\mu + 1.96\sigma/\sqrt{n}$, we will have two points between which $\overline{X}_\bullet$ will occur 95% of the time.
- We can then say that

$$\Pr\left(\mu - 1.96\frac{\sigma}{\sqrt{n}} \leq \overline{X}_\bullet \leq \mu + 1.96\frac{\sigma}{\sqrt{n}}\right) = .95 \tag{18}$$

Confidence Interval Estimation

- These points are at a raw score that is 1.96 standard deviations below the mean and 1.96 standard deviations above the mean.
- Consequently, if we mark points on the graph at $\mu - 1.96\sigma/\sqrt{n}$ and $\mu + 1.96\sigma/\sqrt{n}$, we will have two points between which $\overline{X}_\bullet$ will occur 95% of the time.
- We can then say that

$$\Pr\left(\mu - 1.96\frac{\sigma}{\sqrt{n}} \leq \overline{X}_\bullet \leq \mu + 1.96\frac{\sigma}{\sqrt{n}}\right) = .95 \qquad (18)$$

- After applying some standard manipulations of inequalities, we can manipulate the μ to the inside of the equality and the $\overline{X}_\bullet$ to the outside, obtaining

$$\Pr\left(\overline{X}_\bullet - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_\bullet + 1.96\frac{\sigma}{\sqrt{n}}\right) = .95 \qquad (19)$$

Confidence Interval Estimation

- Equation 19 implies that, if we take $\overline{X}_\bullet$ and add and subtract the "critical distance" $1.96\sigma/\sqrt{n}$, we obtain an interval that contains the true μ, in the long run, 95% of the time.

Taking a Stroll with Mr. Mu

- Even if you are not familiar with the manipulation of inequalities, there is a way of seeing how Equation 19 follows from Equation 18.

Taking a Stroll with Mr. Mu

- Even if you are not familiar with the manipulation of inequalities, there is a way of seeing how Equation 19 follows from Equation 18.
- The first inequality states that there is a critical distance, $1.96\sigma/\sqrt{n}$. and $\overline{X}_\bullet$ is within that distance of μ 95% of the time, over repeated samples.

Taking a Stroll with Mr. Mu

- Even if you are not familiar with the manipulation of inequalities, there is a way of seeing how Equation 19 follows from Equation 18.
- The first inequality states that there is a critical distance, $1.96\sigma/\sqrt{n}$. and $\overline{X}_\bullet$ is within that distance of μ 95% of the time, over repeated samples.
- Now imagine that you had a friend named Mr. Mu, and you went for a stroll with him.

Taking a Stroll with Mr. Mu

- Even if you are not familiar with the manipulation of inequalities, there is a way of seeing how Equation 19 follows from Equation 18.
- The first inequality states that there is a critical distance, $1.96\sigma/\sqrt{n}$. and $\overline{X}_\bullet$ is within that distance of μ 95% of the time, over repeated samples.
- Now imagine that you had a friend named Mr. Mu, and you went for a stroll with him.
- After a certain length of time, he turned to you and said, "You know, about 95% of the time, you've been walking within 2 feet of me."

Taking a Stroll with Mr. Mu

- Even if you are not familiar with the manipulation of inequalities, there is a way of seeing how Equation 19 follows from Equation 18.
- The first inequality states that there is a critical distance, $1.96\sigma/\sqrt{n}$. and $\overline{X}_\bullet$ is within that distance of μ 95% of the time, over repeated samples.
- Now imagine that you had a friend named Mr. Mu, and you went for a stroll with him.
- After a certain length of time, he turned to you and said, "You know, about 95% of the time, you've been walking within 2 feet of me."
- You could, of course, reply that he has also been within 2 feet of you 95% of the time.

Taking a Stroll with Mr. Mu

- Even if you are not familiar with the manipulation of inequalities, there is a way of seeing how Equation 19 follows from Equation 18.
- The first inequality states that there is a critical distance, $1.96\sigma/\sqrt{n}$. and $\overline{X}_\bullet$ is within that distance of μ 95% of the time, over repeated samples.
- Now imagine that you had a friend named Mr. Mu, and you went for a stroll with him.
- After a certain length of time, he turned to you and said, "You know, about 95% of the time, you've been walking within 2 feet of me."
- You could, of course, reply that he has also been within 2 feet of you 95% of the time.
- The point is, if $\overline{X}_\bullet$ is within a certain distance of μ 95% of the time, it must also be the case (because distances are symmetric) that μ is within the same distance of $\overline{X}_\bullet$ 95% of the time.

Constructing a Confidence Interval

- The confidence interval for μ, when σ is known, is, for a $100(1-\alpha)\%$ confidence level, of the form

$$\overline{X}_{\bullet} \pm \Phi^{-1}(1-\alpha/2)\frac{\sigma}{\sqrt{n}} \tag{20}$$

where $\Phi^{-1}(1-\alpha/2)$ is a critical value from the standard normal curve.

Constructing a Confidence Interval

- The confidence interval for μ, when σ is known, is, for a $100(1-\alpha)\%$ confidence level, of the form

$$\overline{X}_{\bullet} \pm \Phi^{-1}(1-\alpha/2)\frac{\sigma}{\sqrt{n}} \tag{20}$$

  where $\Phi^{-1}(1-\alpha/2)$ is a critical value from the standard normal curve.

- For example, $\Phi^{-1}(.975)$ is equal to 1.96.

Constructing a Confidence Interval

Example (A Simple Confidence Interval)

Suppose you are interested in the average height of Vanderbilt male undergraduates, but you only have the resources to sample about 64 men at random from the general population. You obtain a random sample of size 64, and find that the sample mean is 70.6 inches. Suppose that the population standard deviation is somehow known to be 2.5 inches. What is the 95% confidence interval for μ?

Solution. Simply process the result of Equation 20. We have

$$70.6 \pm 1.96 \frac{2.5}{\sqrt{64}}$$

or

$$70.6 \pm .6125$$

We are 95% confident that the average height for the population of interest is between 69.99 and 71.21 inches.

Interval Estimation, Precision, and Sample Size Planning

- A confidence interval provides an indication of precision of estimation (narrower intervals indicate greater precision), while also indicating the location of the parameter.
- Note that the width of the confidence interval is related inversely to the square root of n, i.e., one must quadruple n to double the precision.
- A number of authors have discussed sample size planning in the context of providing an adequate level of precision, rather than guaranteeing a specific level of power against a specified alternative.
- Power and precision are interrelated, but they are not the same thing. (C.P.)