Recite 2: Review of Basic Statistics

Introduction to Econometrics,Fall 2018

Jing Bu

Business School,Nanjing University

9/18/2018

Population,Parameters and Random Sampling

Population,Sample and i.i.d

A population is a collection of people, items, or events about which you want to make inferences.

- Population always have a probability distribution.

A sample is a subset of population, which draw from population in a certain way.

To represent the population well,a sample should be randomly collected and adequately large.

Population,Sample and i.i.d

A population is a collection of people, items, or events about which you want to make inferences.

- Population always have a probability distribution.

A sample is a subset of population, which draw from population in a certain way.

To represent the population well,a sample should be randomly collected and adequately large.

- Infinite population

Population,Sample and i.i.d

A population is a collection of people, items, or events about which you want to make inferences.

- Population always have a probability distribution.

A sample is a subset of population, which draw from population in a certain way.

To represent the population well,a sample should be randomly collected and adequately large.

- Infinite population
- Finite population

Population,Sample and i.i.d

A population is a collection of people, items, or events about which you want to make inferences.

- Population always have a probability distribution.

A sample is a subset of population, which draw from population in a certain way.

To represent the population well,a sample should be randomly collected and adequately large.

- Infinite population
- Finite population
- With replacement

Population,Sample and i.i.d

A population is a collection of people, items, or events about which you want to make inferences.

- Population always have a probability distribution.

A sample is a subset of population, which draw from population in a certain way.

To represent the population well,a sample should be randomly collected and adequately large.

- Infinite population
- Finite population
- With replacement
- Without replacement: when the population size N is very large,compared with the sample size n, then we could say that they are nearly independent.

Definition

- The r.v.s are called a random sample of size n from the population f(x) if $X_1, ..., X_n$ are mutually independent and have the same p.d.f/p.m.f f(x). Alternatively, $X_1, ..., X_n$ are called independent,and identically distributed random variable with p.d.f/p.m.f ,commonly abbreviated to i.i.d. r.v.s.

Definition

Definition

- The r.v.s are called a random sample of size n from the population f(x) if $X_1, ..., X_n$ are mutually independent and have the same p.d.f/p.m.f f(x). Alternatively, $X_1, ..., X_n$ are called independent,and identically distributed random variable with p.d.f/p.m.f ,commonly abbreviated to i.i.d. r.v.s.
- eg.Random sample of n respondents on a survey question.

Definition

Definition

- The r.v.s are called a random sample of size n from the population f(x) if $X_1, ..., X_n$ are mutually independent and have the same p.d.f/p.m.f f(x). Alternatively, $X_1, ..., X_n$ are called independent,and identically distributed random variable with p.d.f/p.m.f ,commonly abbreviated to i.i.d. r.v.s.
- eg.Random sample of n respondents on a survey question.
- And the joint p.d.f/p.m.f of $X_1, ..., X_n$ is given by $f(x_1, ..., x_n) = f(x_1)...f(x_n) = \prod_{i=1}^{n} f(x_i)$

Definition

Definition

- The r.v.s are called a random sample of size n from the population f(x) if $X_1, ..., X_n$ are mutually independent and have the same p.d.f/p.m.f f(x). Alternatively, $X_1, ..., X_n$ are called independent,and identically distributed random variable with p.d.f/p.m.f ,commonly abbreviated to i.i.d. r.v.s.
- eg.Random sample of n respondents on a survey question.
- And the joint p.d.f/p.m.f of $X_1, ..., X_n$ is given by $f(x_1, ..., x_n) = f(x_1)...f(x_n) = \prod_{i=1}^{n} f(x_i)$

Definition

- $X_1, ..., X_n$ is a random sample of size n from the population f(x). A statistic is a real-valued or vector-valued function fully depended on $X_1, ..., X_n$, thus

$$T = T(X_1, ..., X_n)$$

Definition

- The r.v.s are called a random sample of size n from the population f(x) if $X_1, ..., X_n$ are mutually independent and have the same p.d.f/p.m.f f(x). Alternatively, $X_1, ..., X_n$ are called independent,and identically distributed random variable with p.d.f/p.m.f ,commonly abbreviated to i.i.d. r.v.s.
- eg.Random sample of n respondents on a survey question.
- And the joint p.d.f/p.m.f of $X_1, ..., X_n$ is given by $f(x_1, ..., x_n) = f(x_1)...f(x_n) = \prod_{i=1}^{n} f(x_i)$

Definition

- $X_1, ..., X_n$ is a random sample of size n from the population f(x). A statistic is a real-valued or vector-valued function fully depended on $X_1, ..., X_n$, thus
$$T = T(X_1, ..., X_n)$$
- and the probability distribution of a statistic T is called the sampling distribution of T.

Definition

- The r.v.s are called a random sample of size n from the population f(x) if $X_1, ..., X_n$ are mutually independent and have the same p.d.f/p.m.f f(x). Alternatively, $X_1, ..., X_n$ are called independent,and identically distributed random variable with p.d.f/p.m.f ,commonly abbreviated to i.i.d. r.v.s.
- eg.Random sample of n respondents on a survey question.
- And the joint p.d.f/p.m.f of $X_1, ..., X_n$ is given by $f(x_1, ..., x_n) = f(x_1)...f(x_n) = \prod_{i=1}^{n} f(x_i)$

Definition

- $X_1, ..., X_n$ is a random sample of size n from the population f(x). A statistic is a real-valued or vector-valued function fully depended on $X_1, ..., X_n$, thus

$$T = T(X_1, ..., X_n)$$

- and the probability distribution of a statistic T is called the sampling distribution of T.
- A statistic is only a function of the sample.

Sample Mean and Sample Variance

Definition

- The sample average or sample mean X of the n observations $X_1, ..., X_n$ is

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The sample variance is the statistic defined by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})$$

Sample Mean and Sample Variance

Definition

- The sample average or sample mean X of the n observations $X_1, ..., X_n$ is

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The sample variance is the statistic defined by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})$$

- we could assume that the sample mean has some certain probability functions to describe its distributions

Sample Mean and Sample Variance

Definition

- The sample average or sample mean X of the n observations $X_1, ..., X_n$ is

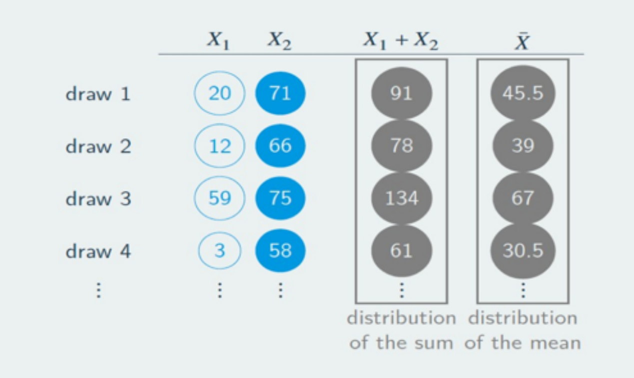
$$\bar{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  The sample variance is the statistic defined by

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})$$

- we could assume that the sample mean has some certain probability functions to describe its distributions
- what is the expectation, variance or p.d.f./c.d.f of this distribution?

A simple case of sample mean



- Let $X_n \in [1, 100]$, assume n = 2, thus only X_1 and X_2

Large-Sample Approximations to Sampling Distributions

Sampling Distributions

1. There are two approaches to characterizing sampling distributions:

Sampling Distributions

1. There are two approaches to characterizing sampling distributions:
   - exact/finite sample distribution: The sampling distribution that exactly describes the distribution of X for any n is called the exact/finite sample distribution of X.

Sampling Distributions

1. There are two approaches to characterizing sampling distributions:
   - exact/finite sample distribution: The sampling distribution that exactly describes the distribution of X for any n is called the exact/finite sample distribution of X.
   - approximate/asymptotic distribution: when the sample size n is large, the sample distribution approximates to a certain distribution function.

Sampling Distributions

1. There are two approaches to characterizing sampling distributions:
   - exact/finite sample distribution: The sampling distribution that exactly describes the distribution of X for any n is called the exact/finite sample distribution of X.
   - approximate/asymptotic distribution: when the sample size n is large, the sample distribution approximates to a certain distribution function.
2. Two key tools used to approximate sampling distributions when the sample size is large,assume that $n \to \infty$

Sampling Distributions

1. There are two approaches to characterizing sampling distributions:
   - exact/finite sample distribution: The sampling distribution that exactly describes the distribution of X for any n is called the exact/finite sample distribution of X.
   - approximate/asymptotic distribution: when the sample size n is large, the sample distribution approximates to a certain distribution function.
2. Two key tools used to approximate sampling distributions when the sample size is large,assume that $n \to \infty$
   - The Law of Large Numbers(L.L.N.):when the sample size is large,X will be close to μ_Y the population mean with very high probability

Sampling Distributions

1. There are two approaches to characterizing sampling distributions:
   - exact/finite sample distribution: The sampling distribution that exactly describes the distribution of X for any n is called the exact/finite sample distribution of X.
   - approximate/asymptotic distribution: when the sample size n is large, the sample distribution approximates to a certain distribution function.
2. Two key tools used to approximate sampling distributions when the sample size is large,assume that $n \to \infty$
   - The Law of Large Numbers(L.L.N.):when the sample size is large,X will be close to μ_Y the population mean with very high probability
   - The Central Limit Theorem(C.L.T.): when the sample size is large,the sampling distribution of the standardized sample average, $\frac{(\bar{Y}-\mu_Y)}{\sigma_Y}$ is approximately normal.

Convergence in probability

Definition

- Let $X_1, ..., X_n$ be an random variables or sequence, is said to converge in probability to a value b if for every $\varepsilon > 0$,

$$P(\|X_n - b\| > \varepsilon) \to 0$$

as $n \to \infty$ We write this $X_n \longrightarrow^p b$ or $plim(X_n) = b$

Convergence in probability

Definition

- Let $X_1, ..., X_n$ be an random variables or sequence, is said to converge in probability to a value b if for every $\varepsilon > 0$,

$$P(\|X_n - b\| > \varepsilon) \to 0$$

as $n \to \infty$ We write this $X_n \longrightarrow^p b$ or $plim(X_n) = b$

- it is similar to the concept of a limitation in a probability way.

the Law of Large Numbers

Theorem

- Let $X_1, ..., X_n$ be an i.i.d draws from a distribution with mean μ and finite variance σ (a population) $\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean then

$$\bar{X} \to \mu$$

the Law of Large Numbers

Theorem

- Let $X_1, ..., X_n$ be an i.i.d draws from a distribution with mean μ and finite variance σ (a population) $\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean then

$$\bar{X} \to \mu$$

- Intuition:the distribution of ??collapses?? $\bar{X}$ on μ

A simple case

Example

- Suppose X has a Bernoulli distribution if it have a binary values $X \varepsilon 0, 1$ and its probability mass function is

$$P(X = x) = \begin{cases} 0.78, & ifx = 1 \\ 0.22, & ifx = 0 \end{cases}$$

A simple case

Example

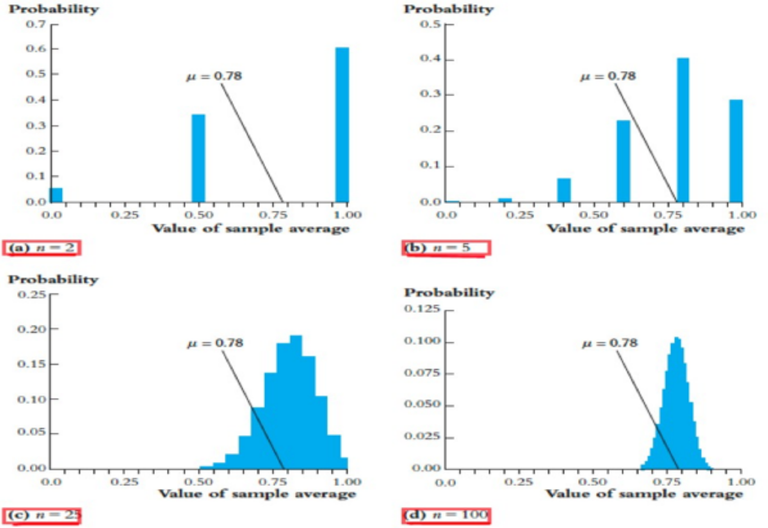
- Suppose X has a Bernoulli distribution if it have a binary values $X \varepsilon 0, 1$ and its probability mass function is

$$P(X = x) = \begin{cases} 0.78, & ifx = 1 \\ 0.22, & ifx = 0 \end{cases}$$

- Then E(X)=p=0.78 and Var(X)=p(1-p)=0.1716

Convergence in Distribution

FIGURE 2.8 Sampling Distribution of the Sample Average of n Bernoulli Random Variables



Definition

- Let $X_1, ..., X_n$ be a sequence of r.v.s, and for n = 1,2,... let $F_n(x)$ be the c.d.f of X_n.Then it is said that $X_1, X_2, ...$ converges in distribution to r.v.W with c.d.f, F_W if

$$\lim_{n \longrightarrow \infty} F_n(x) = F_W(x)$$

which we write as $X_n \to^p \mu$

Definition

- Let $X_1, ..., X_n$ be a sequence of r.v.s, and for n = 1,2,... let $F_n(x)$ be the c.d.f of X_n.Then it is said that $X_1, X_2, ...$ converges in distribution to r.v.W with c.d.f, F_W if

$$\lim_{n \longrightarrow \infty} F_n(x) = F_W(x)$$

which we write as $X_n \rightarrow^p \mu$

- Basically: when n is big, the distribution of X_n is very similar to the distribution of w.

Definition

- Let $X_1, ..., X_n$ be a sequence of r.v.s, and for n = 1,2,... let $F_n(x)$ be the c.d.f of X_n.Then it is said that $X_1, X_2, ...$ converges in distribution to r.v.W with c.d.f, F_W if

$$\lim_{n \longrightarrow \infty} F_n(x) = F_W(x)$$

which we write as $X_n \rightarrow^p \mu$

- Basically: when n is big, the distribution of X_n is very similar to the distribution of w.
- Common to standardize a r.v. by subtracting its expectation and dividing by its standard deviation

$$Z = \frac{X - E[X]}{Var[X]}$$

The Central Limit Theorem

Theorem

- Let $X_1, ..., Xn$ be an i.i.d draws from a distribution with sample size n with mean μ and $0 < \sigma^2 < \infty$, then

$$\frac{\bar{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0, 1)$$

The Central Limit Theorem

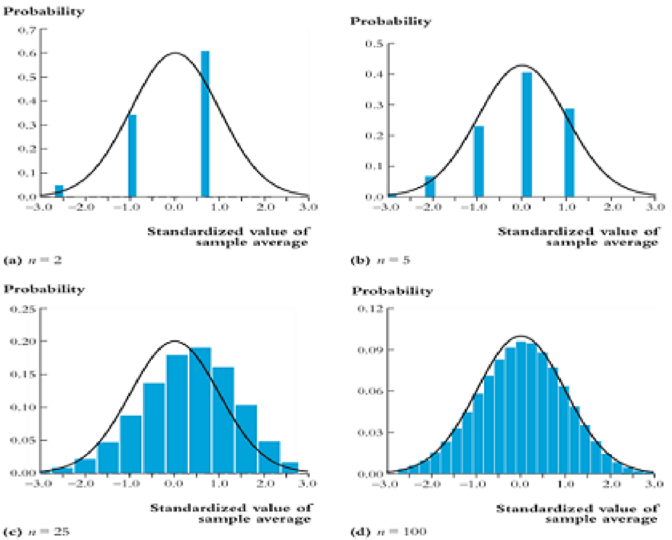
Theorem

- Let $X_1, ..., Xn$ be an i.i.d draws from a distribution with sample size n with mean μ and $0 < \sigma^2 < \infty$, then

$$\frac{\bar{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0,1)$$

- Because we don??t have to make specific assumption about the distribution of X_i, so whatever the distribution of X_i, when n is big the standardized $\bar{X_n} \sim N(\mu, \frac{\sigma^2}{n})$

FIGURE 2.9 Distribution of the Standardized Sample Average of n Bernoulli Random Variables with $p = 0.78$



How large is ??large enough???

1. How large is ??large enough???

How large is ??large enough???

1. How large is ??large enough???
    - how large must n be for the distribution of Y to be approximately normal?

How large is ??large enough???

1. How large is ??large enough???
   - how large must n be for the distribution of Y to be approximately normal?
2. The answer:it depends.

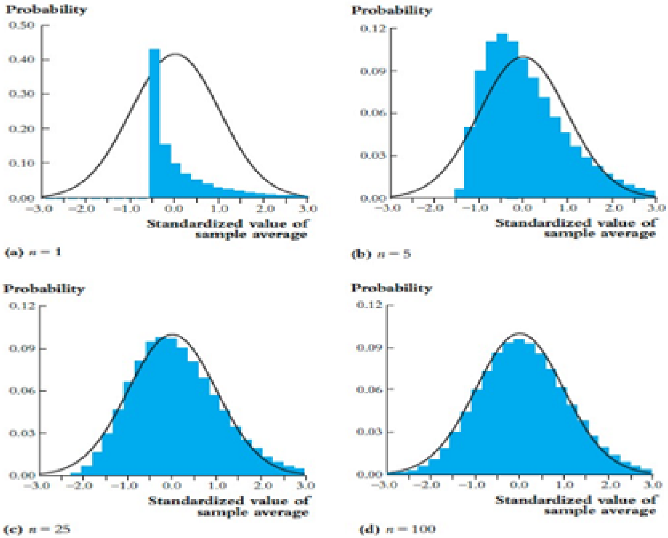
How large is ??large enough???

1. How large is ??large enough???
   - how large must n be for the distribution of Y to be approximately normal?
2. The answer:it depends.
   - if Y_i are themselves normally distributed,then Y is exactly normally distributed for all n.

How large is ??large enough???

1. How large is ??large enough???
   - how large must n be for the distribution of Y to be approximately normal?
2. The answer:it depends.
   - if Y_i are themselves normally distributed,then Y is exactly normally distributed for all n.
   - if Y_i themselves have a distribution that is far from normal,then this approximation can require $n = 30$ or even more.

FIGURE 2.10 Distribution of the Standardized Sample Average of n Draws from a Skewed Distribution



Statistical Inference: Estimation, Confident Intervals and Testing

Statistical Inference

Inference

- What is our best guess about some quantity of interest?

Statistical Inference

Inference

- What is our best guess about some quantity of interest?
- What are a set of plausible values of the quantity of interest?

Statistical Inference

Inference

- What is our best guess about some quantity of interest?
- What are a set of plausible values of the quantity of interest?
- Compare estimators,such as in an experiment

Statistical Inference

Inference

- What is our best guess about some quantity of interest?
- What are a set of plausible values of the quantity of interest?
- Compare estimators,such as in an experiment
- we use simple difference in sample meansor the post-stratification estimator, where we estimate the estimate the difference among two subsets of the data (male and female,for instance) and then take the weighted average of the two variable

Inference:from Samples to Population

- Our focus: $\{Y_1, Y_2, ..., Y_n\}$ are i.i.d.draws from $f(y)$ or $F(Y)$, thus population distribution.

Inference:from Samples to Population

- Our focus: $\{Y_1, Y_2, ..., Y_n\}$ are i.i.d.draws from $f(y)$ or $F(Y)$, thus population distribution.
- Statistical inference or learning is using samples to infer $f(y)$.

Inference:from Samples to Population

- Our focus: $\{Y_1, Y_2, ..., Y_n\}$ are i.i.d.draws from $f(y)$ or $F(Y)$, thus population distribution.
- Statistical inference or learning is using samples to infer $f(y)$.
- two ways: parametric and Non-parametric

Point estimation

- Point estimation: providing a single ??best guess?? as to the value of some fixed, unknown quantity of interest,θ which is is a feature of the population distribution $f(y)$.

Example

Point estimation

- Point estimation: providing a single ??best guess?? as to the value of some fixed, unknown quantity of interest,θ which is is a feature of the population distribution $f(y)$.

Example

- $\mu = E[Y]$

Point estimation

- Point estimation: providing a single ??best guess?? as to the value of some fixed, unknown quantity of interest,θ which is is a feature of the population distribution $f(y)$.

Example

- $\mu = E[Y]$
- $\sigma^2 = Var[Y]$

Point estimation

- Point estimation: providing a single ??best guess?? as to the value of some fixed, unknown quantity of interest,θ which is is a feature of the population distribution $f(y)$.

Example

- $\mu = E[Y]$
- $\sigma^2 = Var[Y]$
- $\mu_y - \mu_x = E[Y] - E[X]$

Estimator and Estimate

Definition

Given a random sample$\{Y_1, Y_2, ..., Y_n\}$ drawn from a population distribution that depends on an unknown parameter θ, and an estimator $\hat{\theta}$ is a function of the sample: thus $\hat{\theta}_n = h(Y_1, Y_2, ..., Y_n)$

- An estimator is a r.v.because it is a function of r.v.s.

Estimator and Estimate

Definition

Given a random sample$\{Y_1, Y_2, ..., Y_n\}$ drawn from a population distribution that depends on an unknown parameter θ, and an estimator $\widehat{\theta}$ is a function of the sample: thus $\widehat{\theta}_n = h(Y_1, Y_2, ..., Y_n)$

- An estimator is a r.v.because it is a function of r.v.s.
- $\{\widehat{\theta}_1, \widehat{\theta}_2, ..., \widehat{\theta}_n\}$ is a sequence of r.v.s, so it has convergence in probability/distribution.

Estimator and Estimate

Definition

Given a random sample$\{Y_1, Y_2, ..., Y_n\}$ drawn from a population distribution that depends on an unknown parameter θ, and an estimator $\hat{\theta}$ is a function of the sample: thus $\hat{\theta}_n = h(Y_1, Y_2, ..., Y_n)$

- An estimator is a r.v.because it is a function of r.v.s.
- $\{\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_n\}$ is a sequence of r.v.s, so it has convergence in probability/distribution.
- Question: what is the difference between an estimator and an statistic?

Estimator and Estimate

Definition

An estimate is the numerical value of the estimator when it is actually computed using data from a specific sample. Thus if we have the actual data $\{y_1, y_2, ..., y_n\}$,then $\hat{\theta} = h(y_1, y_2, ..., y_n)$

Three Characteristics of an Estimator

Let $\hat{\mu}_Y$ denote some estimator of μ_Y and $E(\hat{\mu}_Y)$ is the mean of the sampling distribution of $\hat{\mu}_Y$,

1. **Unbiasedness:** the estimator of μ_Y is unbiased if $E(\hat{\mu}_Y) = \hat{\mu}_Y$

Three Characteristics of an Estimator

Let $\hat{\mu}_Y$ denote some estimator of μ_Y and $E(\hat{\mu}_Y)$ is the mean of the sampling distribution of $\hat{\mu}_Y$,

1. **Unbiasedness:** the estimator of μ_Y is unbiased if $E(\hat{\mu}_Y) = \hat{\mu}_Y$
2. **Consistency:** the estimator of μ_Y is consistent if $E(\hat{\mu}_Y) \rightarrow \hat{\mu}_Y$

Three Characteristics of an Estimator

Let $\hat{\mu}_Y$ denote some estimator of μ_Y and $E(\hat{\mu}_Y)$ is the mean of the sampling distribution of $\hat{\mu}_Y$,

1. **Unbiasedness:** the estimator of μ_Y is unbiased if $E(\hat{\mu}_Y) = \hat{\mu}_Y$
2. **Consistency:** the estimator of μ_Y is consistent if $E(\hat{\mu}_Y) \rightarrow \hat{\mu}_Y$
3. **Efficiency:** Let $\tilde{\mu}_Y$ be another estimator of μ_Y and suppose that both $\tilde{\mu}_Y$ and $\hat{\mu}_Y$ are unbiaesd.Then $\hat{\mu}_Y$ is said to be more efficient than $\tilde{\mu}_Y$

$$Var(\hat{\mu}_Y) < Var(\tilde{\mu}_Y)$$

Three Characteristics of an Estimator

Let $\hat{\mu}_Y$ denote some estimator of μ_Y and $E(\hat{\mu}_Y)$ is the mean of the sampling distribution of $\hat{\mu}_Y$,

1. **Unbiasedness:** the estimator of μ_Y is unbiased if $E(\hat{\mu}_Y) = \hat{\mu}_Y$
2. **Consistency:** the estimator of μ_Y is consistent if $E(\hat{\mu}_Y) \to \hat{\mu}_Y$
3. **Efficiency:** Let $\tilde{\mu}_Y$ be another estimator of μ_Y and suppose that both $\tilde{\mu}_Y$ and $\hat{\mu}_Y$ are unbiaesd.Then $\hat{\mu}_Y$ is said to be more efficient than $\tilde{\mu}_Y$

$$Var(\hat{\mu}_Y) < Var(\tilde{\mu}_Y)$$

- Comparing variances is difficult if we do not restrict our attention to unbiased estimators because we could always use a trivial estimator with variance zero that is biased.

Properties of the sample mean

Let μ_Y and σ_Y^2 denote the mean and variance of Y, then

$$E(\bar{Y}) = \frac{1}{n}\sum_{i=1}^{N} E(Y_i) = \mu_Y$$

so Y is an **unbiased** estimator of μ_Y

- Based on the L.L.N.,$\bar{Y} \rightarrow \mu_Y$ so $\bar{Y}$ is also consistent.

The standard deviation of the sample mean is $\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}}$

Properties of the sample mean

Let μ_Y and σ_Y^2 denote the mean and variance of Y, then

$$E(\bar{Y}) = \frac{1}{n}\sum_{i=1}^{N} E(Y_i) = \mu_Y$$

so Y is an **unbiased** estimator of μ_Y

- Based on the L.L.N.,$\bar{Y} \to \mu_Y$ so $\bar{Y}$ is also consistent.
- The variance of sample mean

$$Var(\bar{Y}) = var(\frac{1}{n}\sum_{i=1}^{N} Y_i) = \frac{1}{n^2}\sum_{i=1}^{N} Var(Y_i) = \frac{\sigma_Y^2}{n}$$

The standard deviation of the sample mean is $\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}}$

Properties of the sample mean

Because efficiency entails a comparison of estimators,we need to specify the estimator or estimators to which Y is to be compared.

- Let $\widetilde{Y} = \frac{1}{n}(\frac{1}{2}Y_1 + \frac{3}{2}Y_2 + \frac{1}{2}Y_3 + \frac{3}{2}Y_4 + ... + \frac{1}{2}Y_{n-1} + \frac{3}{2}Y_n)$

Thus $\bar{Y}$ is more efficient than $\widetilde{Y}$.

Properties of the sample mean

Because efficiency entails a comparison of estimators,we need to specify the estimator or estimators to which Y is to be compared.

- Let $\widetilde{Y} = \frac{1}{n}(\frac{1}{2}Y_1 + \frac{3}{2}Y_2 + \frac{1}{2}Y_3 + \frac{3}{2}Y_4 + ... + \frac{1}{2}Y_{n-1} + \frac{3}{2}Y_n)$
- $Var(\widetilde{Y}) = 1.25\frac{\sigma_Y^2}{n} > \frac{\sigma_Y^2}{n} = Var(\bar{Y})$

Thus $\bar{Y}$ is more efficient than $\widetilde{Y}$.

Confidence Interval and Interval Estimation

Interval Estimation

- A point estimate provides no information about how close the estimate is ??likely?? to be to the population parameter.

Interval Estimation

- A point estimate provides no information about how close the estimate is ??likely?? to be to the population parameter.
- We cannot know how close an estimate for a particular sample is to the population parameter because the population is unknown.

Interval Estimation

- A point estimate provides no information about how close the estimate is ??likely?? to be to the population parameter.
- We cannot know how close an estimate for a particular sample is to the population parameter because the population is unknown.
- A different (complementary) approach to estimation is to produce a range of values that will contain the truth with some fixed probability.

What is a Confidence Interval?

Definition

A $100(1-\alpha)\%$ confidence interval for a population parameter θ is an interval $C_n = (a, b)$,where $a = a(Y_1, Y_2, ..., Y_n)$ and $b = b(Y_1, Y_2, ..., Y_n)$ are functions of the data such that

$$P(a < \theta < b) = 1 - \alpha$$

- In general,this confidence level is $1-\alpha$; where α is called **significance level**.

Interval Estimation and Condence Intervals

Suppose the population has a normal distribution $N(\mu, \sigma^2)$ and let $Y_1, Y_2, ..., Y_n$ be a random sample from the population.

- Then the sample mean has a normal distribution: $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$

Interval Estimation and Condence Intervals

Suppose the population has a normal distribution $N(\mu, \sigma^2)$ and let $Y_1, Y_2, ..., Y_n$ be a random sample from the population.

- Then the sample mean has a normal distribution: $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$
- The standardized sample mean $\bar{Z}$ is given by: $\bar{Z} = \frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$

Interval Estimation and Condence Intervals

Suppose the population has a normal distribution $N(\mu, \sigma^2)$ and let $Y_1, Y_2, ..., Y_n$ be a random sample from the population.

- Then the sample mean has a normal distribution: $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$
- The standardized sample mean $\bar{Z}$ is given by: $\bar{Z} = \frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
- then $\theta = \bar{Z}$,then $P(a < \theta < b) = 1 - \alpha$ turns into

$$a < \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} < b$$

Interval Estimation and Condence Intervals

Suppose the population has a normal distribution $N(\mu, \sigma^2)$ and let $Y_1, Y_2, ..., Y_n$ be a random sample from the population.

- Then the sample mean has a normal distribution: $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$
- The standardized sample mean $\bar{Z}$ is given by: $\bar{Z} = \frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
- then $\theta = \bar{Z}$,then $P(a < \theta < b) = 1 - \alpha$ turns into

$$a < \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} < b$$

- then it follows that

$$P(\bar{Y} - a\frac{\sigma}{\sqrt{n}} < \mu < \bar{Y} + b\frac{\sigma}{\sqrt{n}}) - 1 - \alpha$$

Interval Estimation and Condence Intervals

Suppose the population has a normal distribution $N(\mu, \sigma^2)$ and let $Y_1, Y_2, ..., Y_n$ be a random sample from the population.

- Then the sample mean has a normal distribution: $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n})$
- The standardized sample mean $\bar{Z}$ is given by: $\bar{Z} = \frac{\bar{Y}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
- then $\theta = \bar{Z}$,then $P(a < \theta < b) = 1 - \alpha$ turns into

$$a < \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}} < b$$

- then it follows that

$$P(\bar{Y} - a\frac{\sigma}{\sqrt{n}} < \mu < \bar{Y} + b\frac{\sigma}{\sqrt{n}}) - 1 - \alpha$$

- The random interval contains the population mean with a probability $1 - \alpha$

Interval Estimation and Condence Intervals

Two cases:σ is known and unknown - When σ is known, for example, $\sigma = 1$,thus $Y \sim N(\mu, 1)$ then $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n} = \frac{1}{n})$

- From this we can standardize $\bar{Y}$,and because the standardized version of $\bar{Y}$ has a standard normal distribution, and we let $\alpha = 0.05$,then we have

$$P(-1.96 < \frac{\bar{Y} - \mu}{\frac{1}{\sqrt{n}}} < 1.96) = 1 - 0.05$$

Interval Estimation and Condence Intervals

Two cases:σ is known and unknown - When σ is known, for example, $\sigma = 1$,thus $Y \sim N(\mu, 1)$ then $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n} = \frac{1}{n})$

- From this we can standardize $\bar{Y}$,and because the standardized version of $\bar{Y}$ has a standard normal distribution, and we let $\alpha = 0.05$,then we have

$$P(-1.96 < \frac{\bar{Y} - \mu}{\frac{1}{\sqrt{n}}} < 1.96) = 1 - 0.05$$

- The event in parentheses is identical to the event $\bar{Y} - \frac{1.96}{\sqrt{n}} < \mu < \bar{Y} + \frac{1.96}{\sqrt{n}}$

Interval Estimation and Condence Intervals

Two cases:σ is known and unknown - When σ is known, for example, $\sigma = 1$,thus $Y \sim N(\mu, 1)$ then $\bar{Y} \sim N(\mu, \frac{\sigma^2}{n} = \frac{1}{n})$

- From this we can standardize $\bar{Y}$,and because the standardized version of $\bar{Y}$ has a standard normal distribution, and we let $\alpha = 0.05$,then we have

$$P(-1.96 < \frac{\bar{Y} - \mu}{\frac{1}{\sqrt{n}}} < 1.96) = 1 - 0.05$$

- The event in parentheses is identical to the event $\bar{Y} - \frac{1.96}{\sqrt{n}} < \mu < \bar{Y} + \frac{1.96}{\sqrt{n}}$
- So $P(\bar{Y} - \frac{1.96}{\sqrt{n}} < \mu < \bar{Y} + \frac{1.96}{\sqrt{n}}) = 0.95$

Interval Estimation and Condence Intervals

- The interval estimate of μ may be written as $(\bar{Y} - \frac{1.96}{\sqrt{n}}, \bar{Y} + \frac{1.96}{\sqrt{n}})$

Interval Estimation and Condence Intervals

- The interval estimate of μ may be written as $(\bar{Y} - \frac{1.96}{\sqrt{n}}, \bar{Y} + \frac{1.96}{\sqrt{n}})$
- when σ is unknown, we must use estimate S, denote the sample standard deviation,replacing unknown σ

$$P(\bar{Y} - S\frac{1.96}{\sqrt{n}}, \bar{Y} + S\frac{1.96}{\sqrt{n}})$$

Interval Estimation and Condence Intervals

- The interval estimate of μ may be written as $(\bar{Y} - \frac{1.96}{\sqrt{n}}, \bar{Y} + \frac{1.96}{\sqrt{n}})$
- when σ is unknown, we must use estimate S, denote the sample standard deviation,replacing unknown σ

$$P(\bar{Y} - S\frac{1.96}{\sqrt{n}}, \bar{Y} + S\frac{1.96}{\sqrt{n}})$$

- This could not work because S is not a constant but a r.v.

Interval Estimation and Condence Intervals

Definition

- The t-statistics or t-ration

$$\frac{\bar{Y} - \mu}{SE(\bar{Y})} \sim t_{n-1}$$

Interval Estimation and Condence Intervals

Definition

- The t-statistics or t-ration

$$\frac{\bar{Y} - \mu}{SE(\bar{Y})} \sim t_{n-1}$$

- To construct a 95% confidence interval, let c denote the $97.5^t h$ percentile in the t_{n-1} distribution.

$$P(-c < t \leq c) = 0.95$$

where $c_{\frac{\alpha}{2}}$ is the critical value of the t distribution.

Interval Estimation and Condence Intervals

Definition

- The t-statistics or t-ration

$$\frac{\bar{Y}-\mu}{SE(\bar{Y})} \sim t_{n-1}$$

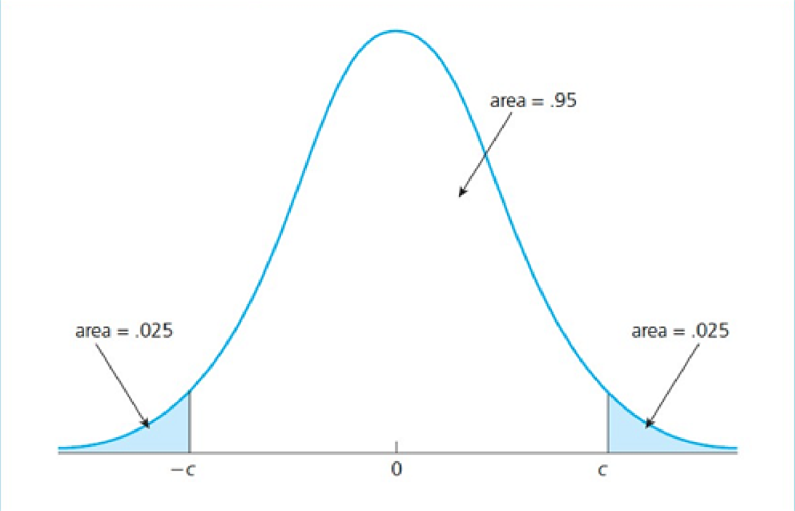
- To construct a 95% confidence interval, let c denote the 97.5^th percentile in the t_{n-1} distribution.

$$P(-c < t \leq c) = 0.95$$

where $c_{\frac{\alpha}{2}}$ is the critical value of the t distribution.

- The condence interval may be written as $[Y \pm c_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}]$

FIGURE C.4 The 97.5th percentile, c, in a t distribution.



A simple rule of thumb for a 95% confidence interval

Caution! An often recited, but incorrect interpretation of a confidence interval is the following:

- "I calculated a 95% confidence interval of $[0.05, 0.13]$, which means that there is a 95% chance that the true means is in that interval.??

The probabilistic interpretation comes from the fact that for 95% of all random samples,the constructed confidence interval will contain μ.

A simple rule of thumb for a 95% confidence interval

Caution! An often recited, but incorrect interpretation of a confidence interval is the following:

- "I calculated a 95% confidence interval of $[0.05, 0.13]$, which means that there is a 95% chance that the true means is in that interval.??
- This is **WRONG**. actually μ either is or is not in the interval.

The probabilistic interpretation comes from the fact that for 95% of all random samples,the constructed confidence interval will contain μ.

TABLE 3.1 Trends in Hourly Earnings in the United States of Working College Graduates, Ages 25–34, 1992 to 2008, in 2008 Dollars

	Men			Women			Difference, Men vs. Women		
Year	$\overline{Y}_m$	s_m	n_m	$\overline{Y}_w$	s_w	n_w	$\overline{Y}_m - \overline{Y}_w$	$SE(\overline{Y}_m - \overline{Y}_w)$	95% Confidence Interval for d
1992	23.27	10.17	1594	20.05	7.87	1368	3.22**	0.33	2.58–3.88
1996	22.48	10.10	1379	18.98	7.95	1230	3.50**	0.35	2.80–4.19
2000	24.88	11.60	1303	20.74	9.36	1181	4.14**	0.42	3.32–4.97
2004	25.12	12.01	1894	21.02	9.36	1735	4.10**	0.36	3.40–4.80
2008	24.98	11.78	1838	20.87	9.66	1871	4.11**	0.35	3.41–4.80

These estimates are computed using data on all full-time workers aged 25–34 surveyed in the Current Population Survey conducted in March of the next year (for example, the data for 2008 were collected in March 2009). The difference is significantly different from zero at the **1% significance level.

Hypothesis Testing

Hypothesis Testing

Definition

- A hypothesis is a statement about a population parameter,thus θ.Formally,we want to test whether is significantly different from a certain value μ_0

$$H_0 : \theta = \mu_0$$

Hypothesis Testing

Definition

- A hypothesis is a statement about a population parameter,thus θ.Formally,we want to test whether is significantly different from a certain value μ_0

$$H_0 : \theta = \mu_0$$

- which is called null hypothesis.The alternative hypothesis is

$$H_0 : \theta \neq \mu_0$$

Hypothesis Testing

Definition

- A hypothesis is a statement about a population parameter,thus θ.Formally,we want to test whether is significantly different from a certain value μ_0

$$H_0 : \theta = \mu_0$$

- which is called null hypothesis.The alternative hypothesis is

$$H_0 : \theta \neq \mu_0$$

- If the value μ_0 does not lie within the calculated confidence interval,then we **reject** the null hypothesis.

Hypothesis Testing

Definition

- A hypothesis is a statement about a population parameter,thus θ.Formally,we want to test whether is significantly different from a certain value μ_0

$$H_0 : \theta = \mu_0$$

- which is called null hypothesis.The alternative hypothesis is

$$H_0 : \theta \neq \mu_0$$

- If the value μ_0 does not lie within the calculated confidence interval,then we **reject** the null hypothesis.
- If the value μ_0 lie within the calculated confidence interval,then we **fail to reject** the null hypothesis.

General framework

- A hypothesis test chooses whether or not to reject the null hypothesis based on the data we observe.

General framework

- A hypothesis test chooses whether or not to reject the null hypothesis based on the data we observe.
- Rejection based on a test statistic

$$T_n = T(Y_1, Y_2, ..., Y_n)$$

General framework

- A hypothesis test chooses whether or not to reject the null hypothesis based on the data we observe.
- Rejection based on a test statistic

$$T_n = T(Y_1, Y_2, ..., Y_n)$$

- The null/reference distribution is the distribution of T under the null.We??ll write its probabilities as $P_0(T_n \leq T)$

Two Type Errors

In both cases, there is a certain risk that our conclusion is wrong.

Type I Error

- A Type I error is when we reject the null hypothesis when it is in fact true.(??left-wing??)

Type II Error

Two Type Errors

In both cases, there is a certain risk that our conclusion is wrong.

Type I Error

- A Type I error is when we reject the null hypothesis when it is in fact true.(??left-wing??)
- We say that the Lady is discerning when she is just guessing(null hypothesis: she is just guessing)

Type II Error

Two Type Errors

In both cases, there is a certain risk that our conclusion is wrong.

Type I Error

- A Type I error is when we reject the null hypothesis when it is in fact true.(??left-wing??)
- We say that the Lady is discerning when she is just guessing(null hypothesis: she is just guessing)

Type II Error

- A Type II error is when we fail to reject the null hypothesis when it is false.(??right-wing??)

General framework

- A hypothesis test chooses whether or not to reject the null hypothesis based on the data we observe.

General framework

- A hypothesis test chooses whether or not to reject the null hypothesis based on the data we observe.
- Rejection based on a test statistic.

$$T_n = T(Y_1, Y_2, ..., Y_n)$$

General framework

- A hypothesis test chooses whether or not to reject the null hypothesis based on the data we observe.
- Rejection based on a test statistic.

$$T_n = T(Y_1, Y_2, ..., Y_n)$$

- The null/reference distribution is the distribution of T under the null.

General framework

- A hypothesis test chooses whether or not to reject the null hypothesis based on the data we observe.
- Rejection based on a test statistic.

$$T_n = T(Y_1, Y_2, ..., Y_n)$$

- The null/reference distribution is the distribution of T under the null.
- We??ll write its probabilities as $P_0(T_n \leq T)$

P-value

- To provide additional information, we could ask the question: What is the largest significance level at which we could carry out the test and still fail to reject the null hypothesis?

We can consider the **p-value** of a test

where denotes the standard normal c.d.f.(we assume that n is large enough)

P-value

- To provide additional information, we could ask the question: What is the largest significance level at which we could carry out the test and still fail to reject the null hypothesis?

We can consider the **p-value** of a test

- Calculate the t-statistic

where denotes the standard normal c.d.f.(we assume that n is large enough)

P-value

- To provide additional information, we could ask the question: What is the largest significance level at which we could carry out the test and still fail to reject the null hypothesis?

We can consider the **p-value** of a test

- Calculate the t-statistic
- The largest significance level at which we would fail to reject H_0 is the significance level associated with using t as our critical value.

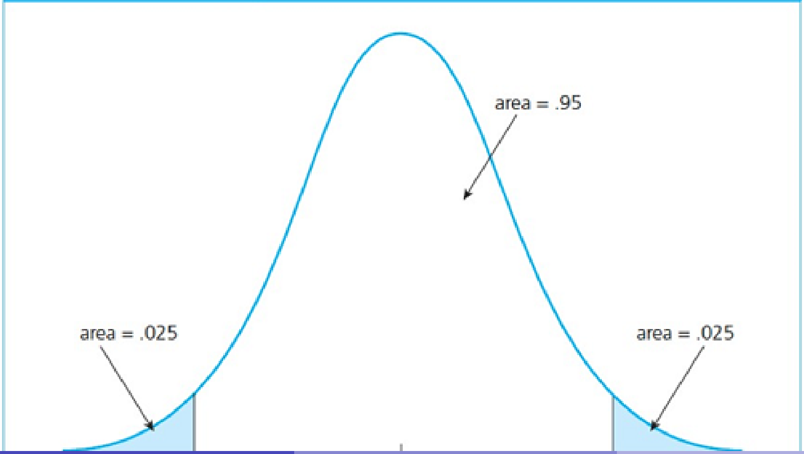
$$p - value = 1 - \Phi(t)$$

where denotes the standard normal c.d.f.(we assume that n is large enough)

- Suppose that $t = 1.52$, then we can find the largest significance level at which we would fail to reject H_0

$$p - value = P(T > 1.52|H_0) = 1 - \Phi(1.52) = 0.065$$

FIGURE C.4 The 97.5[th] percentile, c, in a t distribution.



For example,I said that I shot very well, and I can hit an average of 8 rings. How can I verify it? Let me do a few shots to see how my level is. First you choose to believe me, assuming I didn't brag, my score is near the 8 rings (this is the null hypothesis). As an 8-ring level player, the number of rings in the shot should be subject to a distribution with an average of 8

Comparing Means from Different Populations

An Example:Comparing Means from Different Populations

- Do recent male and female college graduates earn the same amount on average? This question involves comparing the means of two different population distributions.

Difference in mean=$\bar{Y}_{treat} - \bar{Y}_{control}$

An Example:Comparing Means from Different Populations

- Do recent male and female college graduates earn the same amount on average? This question involves comparing the means of two different population distributions.
- In an RCT, we would like to estimate the average causal effects over the population.

$$ATE = ATT = E\{Y_i(1) - Y_i(0)\}$$

Difference in mean=$\bar{Y}_{treat} - \bar{Y}_{control}$

An Example:Comparing Means from Different Populations

- Do recent male and female college graduates earn the same amount on average? This question involves comparing the means of two different population distributions.
- In an RCT, we would like to estimate the average causal effects over the population.

$$ATE = ATT = E\{Y_i(1) - Y_i(0)\}$$

- We only have random samples and random assignment to treatment, then what we can estimate instead

Difference in mean=$\bar{Y}_{treat} - \bar{Y}_{control}$

An Example:Comparing Means from Different Populations

- Do recent male and female college graduates earn the same amount on average? This question involves comparing the means of two different population distributions.
- In an RCT, we would like to estimate the average causal effects over the population.

$$ATE = ATT = E\{Y_i(1) - Y_i(0)\}$$

- We only have random samples and random assignment to treatment, then what we can estimate instead

Difference in mean=$\bar{Y}_{treat} - \bar{Y}_{control}$

- Under randomization, difference-in-means is a good estimate for the ATE.

Hypothesis Tests for the Difference Between Two Means

- To illustrate a test for the difference between two means, let mw be the mean hourly earning in the population of women recently graduated from college and let mm be the population mean for recently graduated men.

Hypothesis Tests for the Difference Between Two Means

- To illustrate a test for the difference between two means, let mw be the mean hourly earning in the population of women recently graduated from college and let mm be the population mean for recently graduated men.
- Then the null hypothesis and the two-sided alternative hypothesis are

$$H_0 : \mu_m = \mu_w$$

$$H_1 : \mu_m \neq \mu_w$$

Hypothesis Tests for the Difference Between Two Means

- To illustrate a test for the difference between two means, let mw be the mean hourly earning in the population of women recently graduated from college and let mm be the population mean for recently graduated men.
- Then the null hypothesis and the two-sided alternative hypothesis are

$$H_0 : \mu_m = \mu_w$$

$$H_1 : \mu_m \neq \mu_w$$

- Consider the null hypothesis that mean earnings for these two populations differ by a certain amount, say d_0. The null hypothesis that men and women in these populations have the same mean earnings corresponds to $H_0 : d_0 = \mu_m - \mu_w = 0$

The Difference Between Two Means

- Suppose we have samples of n_m men and n_w women drawn at random from their populations.Let the sample average annual earnings be $\bar{Y}_m$ for men and for $\bar{Y}_w$ for women.Then an estimator of $\mu_m - \mu_w$ is $\bar{Y}_m - \bar{Y}_w$

The Difference Between Two Means

- Suppose we have samples of n_m men and n_w women drawn at random from their populations.Let the sample average annual earnings be $\bar{Y_m}$ for men and for $\bar{Y_w}$ for women.Then an estimator of $\mu_m - \mu_w$ is $\bar{Y_m} - \bar{Y_w}$
- Let us discuss the distribution of $\bar{Y_m} - \bar{Y_w} \sim N(\mu_m - \mu_w, \frac{\sigma_m^2}{n_m} + \frac{\sigma_w^2}{n_w})$

The Difference Between Two Means

- Suppose we have samples of n_m men and n_w women drawn at random from their populations.Let the sample average annual earnings be $\bar{Y_m}$ for men and for $\bar{Y_w}$ for women.Then an estimator of $\mu_m - \mu_w$ is $\bar{Y_m} - \bar{Y_w}$
- Let us discuss the distribution of $\bar{Y_m} - \bar{Y_w} \sim N(\mu_m - \mu_w, \frac{\sigma_m^2}{n_m} + \frac{\sigma_w^2}{n_w})$
- if σ_m^2 and σ_w^2 are known, then the this approximate normal distribution can be used to compute p-values for the test of the null hypothesis.In practice,however, these population variances are typically unknown so they must be estimated.

The Difference Between Two Means

- Suppose we have samples of n_m men and n_w women drawn at random from their populations.Let the sample average annual earnings be $\bar{Y_m}$ for men and for $\bar{Y_w}$ for women.Then an estimator of $\mu_m - \mu_w$ is $\bar{Y_m} - \bar{Y_w}$
- Let us discuss the distribution of $\bar{Y_m} - \bar{Y_w} \sim N(\mu_m - \mu_w, \frac{\sigma_m^2}{n_m} + \frac{\sigma_w^2}{n_w})$
- if σ_m^2 and σ_w^2 are known, then the this approximate normal distribution can be used to compute p-values for the test of the null hypothesis.In practice,however, these population variances are typically unknown so they must be estimated.
- Thus the standard error of $\bar{Y_m} - \bar{Y_w}$ is

$$SE(\bar{Y_m} - \bar{Y_w}) = \sqrt{\frac{s_m^2}{n_m} + \frac{s_w^2}{n_w}}$$

The Difference Between Two Means

- The t-statistic for testing the null hypothesis is constructed analogously to the t-statistic for testing a hypothesis about a single population mean,thus t-statistic for comparing two means is

$$t = \frac{\bar{Y_m} - \bar{Y_w} - d_0}{SE(\bar{Y_m} - \bar{Y_w})}$$

The Difference Between Two Means

- The t-statistic for testing the null hypothesis is constructed analogously to the t-statistic for testing a hypothesis about a single population mean,thus t-statistic for comparing two means is

$$t = \frac{\bar{Y_m} - \bar{Y_w} - d_0}{SE(\bar{Y_m} - \bar{Y_w})}$$

- If both n_m and n_w are large, then this t-statistic has a standard normal distribution when the null hypothesis is true.

Confidence Intervals for the Difference Between Two Population Means

the 95% two-sided confidence interval for d consists of those values of d within ± 1.96 standard errors of $\bar{Y_m} - \bar{Y_w}$ thus $d = \mu_m - \mu_w$ is:

$$(\bar{Y_m} - \bar{Y_w}) \pm 1.96 SE(\bar{Y_m} - \bar{Y_w})$$

Thank you