# Introduction to Econometrics

*Lecture 2 : Causal Inference and Random Control Trails(RCT)*

**Zhaopeng Qu**

**Business School,Nanjing University**

Sep. 18th, 2017

# Outlines

1. Random Experiment as the Research Design : Program Evaluation Econometrics
2. Review of Statistics
   - Population, Parameters and Random Sampling
   - Large-Sample Approximations to Sampling Distributions
   - Statistical Inference: Estimation, Confident Intervals and Testing
   - Interval Estimation and Confidence Intervals
   - Hypothesis Testing
   - Comparing Means from Different Populations

# Outlines

# Random Experiment as the Research Design : Program Evaluation Econometrics

# Random Assignment Solves the Selection Problem

- Think of causal effects in terms of comparing counterfactuals or potential outcomes. However, we can never observe both counterfactuals —fundamental problem of causal inference.
- To construct the counterfactuals, we could use two broad categories of empirical strategies.
    - Random Controlled Trials/Experiments:
        - it can eliminates selection bias which is the most important bias arises in empirical research. If we could observe the counterfactual directly, then there is no evaluation problem, just simply difference.
    - The various approaches using naturally-occurring data provide alternative methods of constructing the proper counterfactual.

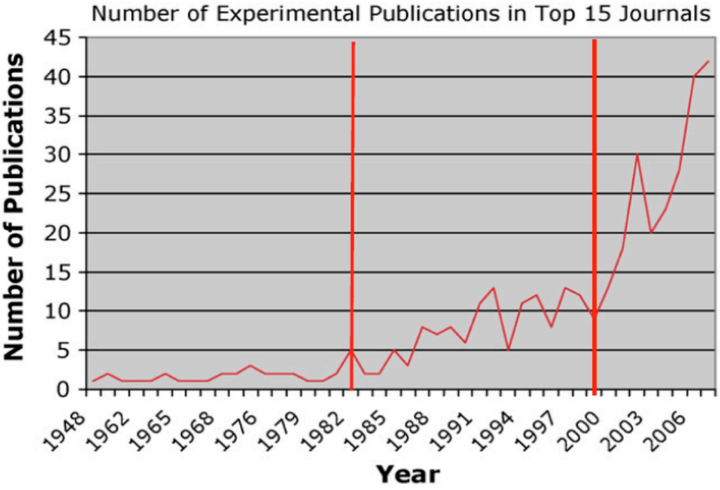
# Randomized Controlled Trials(RCT)

- First recorded RCT was done in 1747 by James Lind,who was a Scottish physician in the Royal Navy.
- Scurvy is a terrible disease caused by Vitamin C deficiency. Serious issue during long sea voyages.
- Lind took 12 sailors with scurvy and split them into six groups of two.
- Groups were assigned:
  - (1) 1 qt cider(苹果酒) (2) 25 drops of vitriol(硫酸 ) (3) 6 spoonfuls of vinegar, (4) 1/2 pt of sea water, (5) garlic, mustard(芥末)and barley water ( 大麦汤 ) , (6) 2 oranges and 1 lemon
- Only Group 6 (citrus fruit) showed substantial improvement.

# Types of RCT

- Lab Experiments
  - eg: computer game for gamble in Lab
- Field Experiments
  - eg: the role of women in household's decision or fake resumes in job application
- Quasi-Experiment or Natural Experiments: some unexpected institutional change or natural shock
  - eg: Germany reunion, Great Famine in China and U.S Bombing in Vietnam.

Figure:


Number of Experimental Publications in Top 15 Journals
Number of Publications
45
40
35
30
25
20
15
10
5
0
1948
1962
1965
1968
1976
1979
1982
1985
1988
1991
1994
1997
2000
2003
2006
Year

# RCT are far from perfect!

- High Costs, Long Duration
- Potential Ethical Problems: "Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomized controlled trials."
  - Milgram Experiment
  - Stanford Prison Experiment
  - Monkey Experiment
- Limited generalizability
- RCTs allow us to gain knowledge about causal effects without knowing the mechanism.

# Potential Problems in Practice

- Small sample: Student Effect
- Hawthorne effect : The subjects are in an experiment can change their behavior.
- Attrition ( 样本流失 ): It refers to subjects dropping out of the study after being randomly assigned to the treatment or control group.
- Failure to randomize or failure to follow treatment protocol: People don't always do what they are told.
  - Wearing glasses program in Western Rural China.

# Program Evaluation Econometrics(项目评估计量经济学)

- Question: How to do empirical research scientifically when we can not do experiments? It means that we always have selection bias in our data, or in term of "endogeneity".
- Answer: Build a reasonable counterfactual world by naturally occurring data to find a proper control group is the core of econometrical methods.
- Here you Furious Seven Weapons in Applied Econometrics(七种盖世武器)
  1. Random Trials and OLS (随机实验)
  2. Decomposition (分解)
  3. Instrumental Variable (工具变量)
  4. Differences in Differences (双差分)
  5. Matching and Propensity Score (匹配)
  6. Regression Discontinuity (断点回归)
  7. Synthetic Control (合成控制法)

- These Furious Seven are the most basic and popular methods in applied econometrics and so powerful that
  - even if you just master one, you may finish your empirical paper and get a good score.
  - if you master several ones, you could have opportunity to publish your paper.
  - If you master all of them, you might to teach applied econometrics class just as what I am doing now.
- We will introduce the essentials of these methods in the class as many as possible. Let's start our journey together.

# 微信签到 (实验)，请扫码！

**计量经济学 曲兆鹏**

签到公示地址 http://dsj.nju.edu.cn/wx/?cid=11

# Review of Statistics

# 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.*

# Random Sample and i.i.d

## 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.
- $X_i \perp X_j$ for all $i \neq j$
- $f_{X_i}(x)$ is the same for all $i$.
- And the joint p.d.f/p.m.f of $X_1, ..., X_n$ is given by

$$f(x_1, ..., x_n) = f(x_1)...f(x_n) = \prod_{i=1}^{n} f(x_i)$$

# Statistic and Sampling Distribution

## 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**, $\overline{X}$, of the $n$ observation $X_1, ..., X_n$ is

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

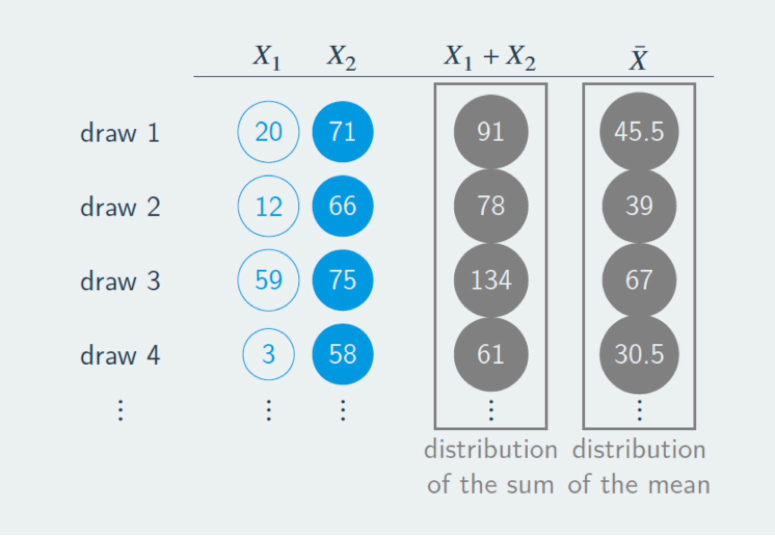
The **sample variance** is the statistic defined by

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

- if $X_i$ is a r.v., then $\sum X_i$ is also a r.v.
- the sample mean and the sample variance are also a function of sums, so they are a r.v. too.
    - 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 $\{X_1, X_n\} \in [1, 100]$ , assume $n = 2$, thus only $X_1$and $X_2$



# Sampling Distributions

- 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 $\overline{X}$.
  - *approximate/asymptotic* distribution: when the sample size $n$ is large, the sample distribution approximates to a certain distribution function.
- 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, $\overline{X}$ will be close to $\mu_Y$ , 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, $(\overline{Y}-\mu_Y)/\sigma_{\overline{Y}}$ , 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 $\varepsilon > 0$,

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

as $n \to \infty$. We write this $X_n \xrightarrow{p} b$ or $plim(X_n) = b$.

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

**Theorem**

*Let $X_1, ..., X_n$ be an i.i.d draws from a distribution with mean $\mu$ and finite variance $\sigma^2$(a population) and $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean, then*

$$\overline{X} \xrightarrow{p} \mu$$

- Intuition: the distribution of $\overline{X}_n$ "collapses" on $\mu$.

# A simple case

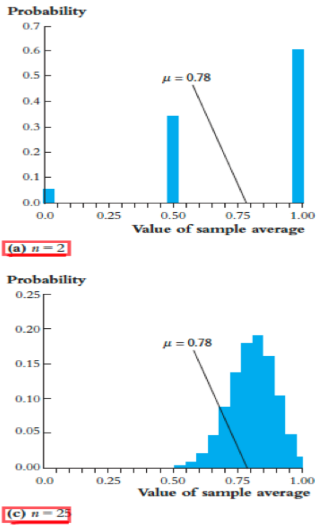
**Example**

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

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

- then $E(X) = p = 0.78$ and $Var(X) = p(1 - p) = 0.1716$.

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



# Convergence in Distribution

## Definition

Let $X_1, X_{2,}...$ 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

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

which we write as $X_n \xrightarrow{d} W$.

- 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

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

# The Central Limit Theorem

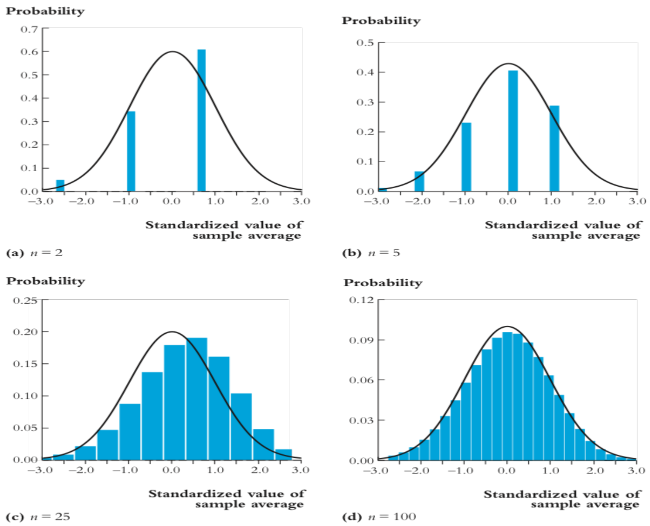
**Theorem**

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

$$\frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} N(0, 1)$$

- 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 $\overline{X}_n \sim N(0, 1)$
  - $\overline{X}_n \sim N(\mu, \frac{\sigma^2}{n})$

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



- How large is large enough ?
  - how large must $n$ be for the distribution of $\overline{Y}$ to be approximately normal?
- The answer: it depends.
  - if $Y_i$ are themselves normally distributed, then $\overline{Y}$ 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.

# How large is "large enough" ?

- How large is large enough ?
  - how large must $n$ be for the distribution of $\overline{Y}$ to be approximately normal?
- The answer: it depends.
  - if $Y_i$ are themselves normally distributed, then $\overline{Y}$ 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.

# How large is "large enough" ?

- How large is large enough ?
  - how large must $n$ be for the distribution of $\overline{Y}$ to be approximately normal?
- The answer: it depends.
  - if $Y_i$ are themselves normally distributed, then $\overline{Y}$ 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.

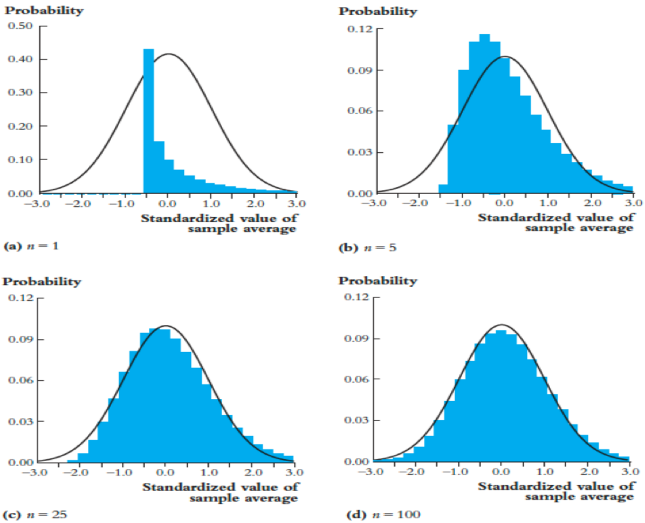
# How large is "large enough" ?

- How large is large enough ?
  - how large must $n$ be for the distribution of $\overline{Y}$ to be approximately normal?
- The answer: it depends.
  - if $Y_i$ are themselves normally distributed, then $\overline{Y}$ 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.

# How large is "large enough" ?

- How large is large enough ?
  - how large must $n$ be for the distribution of $\overline{Y}$ to be approximately normal?
- The answer: it depends.
  - if $Y_i$ are themselves normally distributed, then $\overline{Y}$ 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

- 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 means?
  - or 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
  - which is better? how could we know?

# 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
  - Non-parametric

# Point estimation

- Point estimation: providing a single "best guess"as to the value of some fixed, unknown quantity of interest, $\theta$, which is is a feature of the population distribution, $f(y)$.
- Examples
  - $\mu = E[Y]$
  - $\sigma^2 = Var[Y]$
  - $\mu_y - \mu_x = E[Y] - E[X]$

# 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 $\theta$, 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?

## 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 $\hat{\theta} = h(y_1, y_2, ..., y_n)$

## Example

# Three Characteristics of an Estimator

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

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

   - 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

1. Let $\mu_Y$ and $\sigma_Y^2$ denote the mean and variance of $Y_i$, then

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

   so $\overline{Y}$ is an *unbiased* estimator of $\mu_Y$.

2. Based on the L.L.N., $\overline{Y} \xrightarrow{p} \mu_Y$, so $\overline{Y}$ is also *consistent*.
3. the variance of sample mean

$$Var(\overline{Y}) = var\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(Y_i) = \frac{\sigma_Y^2}{n}$$

4. the standard deviation of the sample mean is $\sigma_{\overline{Y}} = \frac{\sigma_Y}{\sqrt{n}}$

# 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 $\widetilde{Y} = \frac{1}{n}\left(\frac{1}{2}Y_1 + \frac{3}{2}Y_2 + \frac{1}{2}Y_3 + \frac{3}{2}Y_4 + ... + \frac{1}{2}Y_{n-1} + \frac{3}{2}Y_n\right)$
    - $Var(\widetilde{Y}) = 1.25\frac{\sigma_Y^2}{n} > \frac{\sigma_Y^2}{n} = Var(\overline{Y})$
    - Thus $\overline{Y}$ is more efficient than $\widetilde{Y}$

# Properties of the Sample Variance

- Let $\mu_Y$ and $\sigma_Y^2$ denote the mean and variance of $Y_i$ , then the sample variance : $S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$

1. $E(S_Y^2) = \sigma_Y^2$, thus $S^2$ is an *unbiased* estimator of $\sigma_Y^2$. It is also the reason why the average uses the divisor $n-1$ instead of $n$.
2. $S_Y^2 \xrightarrow{P} \sigma_Y^2$, thus the sample variance is a consistent estimator of the population variance.
    - Because $\sigma_{\overline{Y}} = \frac{\sigma_Y}{\sqrt{n}}$, so the statement above justifies using $\frac{S_Y}{\sqrt{n}}$ as an estimator of the standard deviation of the sample mean, $\sigma_{\overline{Y}}$.
    - It is called **the standard error** of the sample mean and it dented $SE[\overline{Y}]$ or $\hat{\sigma}_{\overline{Y}}$.

# 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 $\theta$ is an interval $C_n = (a, b)$ , where $a = a(Y_1, ..., Y_n)$ and $b = b(Y_1, ..., Y_n)$ are functions of the data such that

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

- In general, this confidence level is $1-\alpha$ ; where $\alpha$ 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: $\overline{Y} \sim N(\mu, \frac{\sigma^2}{n})$
  - The standardized sample mean $\overline{Z}$ is given by: $\overline{Z} = \frac{\overline{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- Then $\theta = \overline{Z}$, then $P(a < \theta < b) = 1 - \alpha$ turns into

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

then it follows that

$$P(\overline{Y} - a\sigma/\sqrt{n} < \mu < \overline{Y} + b\sigma/\sqrt{n}) = 1 - \alpha$$

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

# Interval Estimation and Condence Intervals

- Two cases: $\sigma$ is known and unknown
- When $\sigma$ is known, for example,$\sigma = 1$, thus $Y \sim N(\mu, 1)$,
- then $\overline{Y} \sim N(\mu, \frac{\sigma^2}{n} = \frac{1}{n})$
- From this, we can standardize $\overline{Y}$, and, because the standardized version of $\overline{Y}$ has a standard normal distribution, and we let $\alpha = 0.05$, then we have

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

- The event in parentheses is identical to the event $\overline{Y} - 1.96/\sqrt{n} \leq \mu \leq \overline{Y} + 1.96/\sqrt{n}$, so

$$P(\overline{Y} - 1.96/\sqrt{n} \leq \mu \leq \overline{Y} + 1.96/\sqrt{n}) = 0.95$$

- The interval estimate of $\mu$ may be written as $[\overline{Y} - 1.96/\sqrt{n}, \overline{Y} + 1.96/\sqrt{n}]$

# Interval Estimation and Condence Intervals

- When $\sigma$ is unknown, we must use an estimate $S$ , denote the sample standard deviation, replacing unknown $\sigma$

$$P(\overline{Y} - 1.96 S/\sqrt{n} \leq \mu \leq \overline{Y} + 1.96 S/\sqrt{n}) = 0.95$$

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

**Definition**

The **t-statistic** or **t-ratio**:

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

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

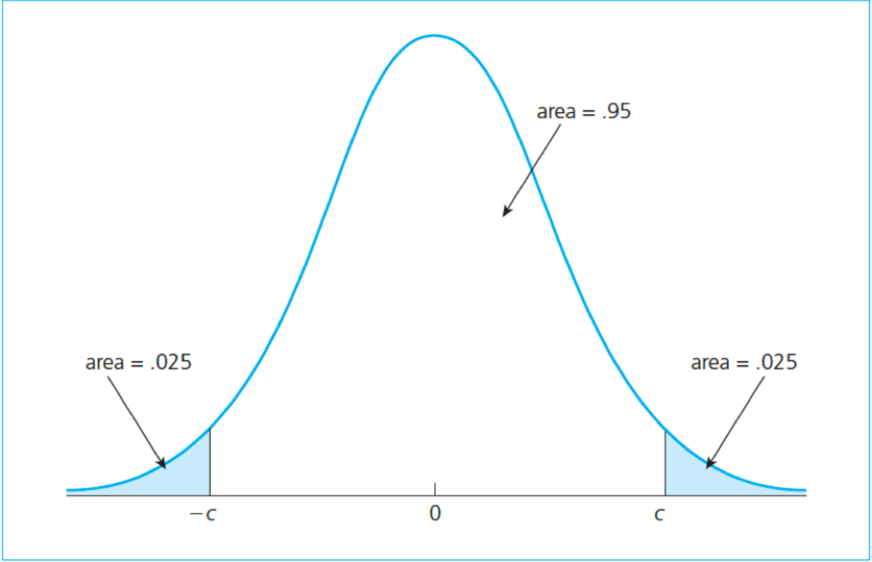
$$P(-c < t \leq c) = 0.95$$

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

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

# Interval Estimation and Condence Intervals

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."
  - This is WRONG. actually $\mu$ 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 $\mu$.

# Interpreting the 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 $\mu$ 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 $\mu$.

# Hypothesis Testing

## Definition

A hypothesis is a statement about a population parameter, thus $\theta$. Formally, we want to test whether is significantly different from a certain value $\mu_0$

$$H_0 : \theta = \mu_0$$

which is called **null hypothesis**. **The alternative hypothesis** is

$$H_1 : \theta \neq \mu_0$$

- If the value $\mu_0$ does not lie within the calculated condence interval, then we **reject** the null hypothesis.
- If the value $\mu_0$ lie within the calculated condence 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.
- Rejection based on a test statistic

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

- 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")

- We say that the Lady is discerning when she is just guessing(null hypo: 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.
- Rejection based on a test statistic

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

- 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
  1. Calculate the t-statistic t
  2. 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

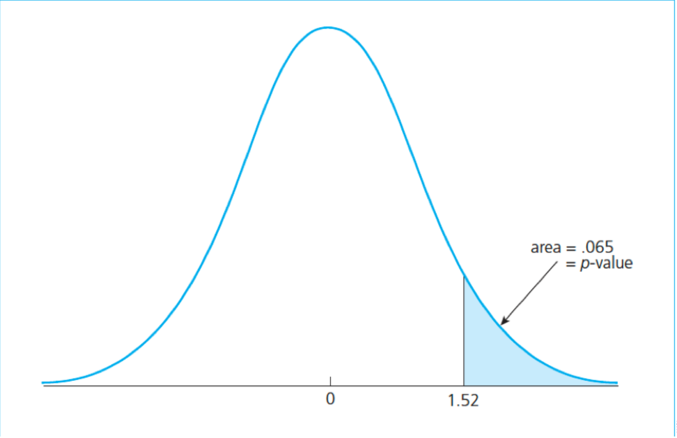
$$p - value = 1 - \Phi(t)$$

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 $H_0$

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

**FIGURE C.7** The $p$-value when $t = 1.52$ for the one-sided alternative $\mu$ . $\mu_0$.



# 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

$$difference\ in\ mean = \overline{Y}_{treated} - \overline{Y}_{control}$$

- 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.
- Then the **null hypothesis** and **the two-sided alternative hypothesis** are

$$\begin{aligned} H_0 : \mu_m &= \mu_w \\ H_1 : \mu_m &\neq \mu_w \end{aligned}$$

- 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 $H_0$ : $H_0 : d_0 = \mu_m - \mu_w = 0$

# 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 $\overline{Y}_m$ for men and $\overline{Y}_w$ for women. Then an estimator of $\mu_m - \mu_w$ is $\overline{Y}_m - \overline{Y}_w$ .
- Let us discuss the distribution of $\overline{Y}_m - \overline{Y}_w$ .

$$\sim N(\mu_m - \mu_w, \frac{\sigma_m^2}{n_m} + \frac{\sigma_w^2}{n_w})$$

- if $\sigma_m^2$and $\sigma_w^2$ 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 $\overline{Y}_m - \overline{Y}_w$ is

$$SE(\overline{Y}_m - \overline{Y}_w) = \sqrt{\frac{s_m^2}{n_m} + \frac{s_w^2}{n_w}}$$

# 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

$$t = \frac{\overline{Y}_m - \overline{Y}_w - d_0}{SE(\overline{Y}_m - \overline{Y}_w)}$$

- If both $n_m$and $n_m$ 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 $\pm 1.96$ standard errors of $\overline{Y}_m - \overline{Y}_w$ , thus $d = \mu_m - \mu_w$ is

$$(\overline{Y}_m - \overline{Y}_w) \pm 1.96 SE(\overline{Y}_m - \overline{Y}_w)$$