Applied Econometrics 122B
Introduction & Review

Giuseppe Ragusa

Luiss University

gragusa@luiss.it
http://gragusa.org/

June 26, 2017

University of California, Irvine

Econometrics

- The use of mathematical and statistical methods:
  - to verify economic theories
  - to fit economic models to real data
  - to forecast future values of economic quantities
- Econometric techniques are also used by: sociologists, political scientists and other social scientists.

Why should you study econometrics?

Policy evaluation

Assist in evaluating effects of policies both before and after implementation

Financial markets

Forecasting, CAPM, APT, etc.

Strategic management

Inventory management and firm performances, analysis of divenstiture, etc.

Marketing

Demand functions for industries, study of consumer behavior, etc.

Macroeconomics

Models and business cycles, models of the monetary policy, growth forecast, etc.

Industrial organization

Price discrimination theories, estimation of market power, etc.

Why studying econometrics?

> *One hurdle is a talent and skills gap. The United States alone, McKinsey projects, will need* **140,000** *to* **190,000** *more people with "deep analytical" skills, typically experts in statistical methods and data-analysis technologies.*
>
> *McKinsey says the nation will also need 1.5 million more data-literate managers [. . .]. [. . .] the need for a sweeping change in business to adapt a new way of managing and making decisions that relies more on data analysis.*

Source: McKinsey (2011), *Big data: The next frontier for innovation, competition, and productivity*

Policy evaluation

Question: how much does cigarettes consumption respond to a change in price?

Policy evaluation

Question: how much does cigarettes consumption respond to a change in price?

- We know from consumer theory that price elasticities are negative

$$\epsilon = \frac{dq}{dp}\frac{p}{q} < 0$$

Policy evaluation

Question: how much does cigarettes consumption respond to a change in price?

- We know from consumer theory that price elasticities are negative

$$\epsilon = \frac{dq}{dp}\frac{p}{q} < 0$$

- but theory does not tell us its exact value

Policy evaluation

Question: how much does cigarettes consumption respond to a change in price?

- We know from consumer theory that price elasticities are negative

$$\epsilon = \frac{dq}{dp}\frac{p}{q} < 0$$

- but theory does not tell us its exact value
- yet, from a policy point of view it is fundamental to know the exact magnitude of the elasticity.

Financial market

Testing the validity of various versions of CAPMs attracts a lot attention in empirical finance. Various testing procedures and statistical methods have been proposed and studied.

CAPM

Fama and French (1993) form a three factor model to explain the expected excessive returns of assets. Broadly speaking, the three factors are:

- market index
- value equity of firms
- book-to-market value

Other examples

- What is the quantitative effect of reducing class size on student achievement?
- How does another year of education change earnings?
- Are “better” performing CEO payed more?
- What is the effect on output growth of a 1 percentage point increase in interest rates by the European Central Bank?
- What is the effect on housing prices on the environment?

In this course you will:

- Learn methods for estimating causal effects using observational data
- Learn some tools that can be used for other purposes, for example forecasting using time series data;
- Focus on applications - theory is used only as needed to understand the "why"s of the methods;
- Learn to evaluate the regression analysis of others - this means you will be able to read/understand empirical economics papers in other econ courses;
- Get some hands-on experience with regression analysis in your problem sets.

This course is "mostly" about using data to measure causal effects.

- Ideally, we would like an experiment
  - what would be an experiment to estimate the effect of class size on standardized test scores?
- But almost always we only have observational (nonexperimental) data.
  - returns to education
  - cigarette prices
  - monetary policy
- Most of the course deals with difficulties arising from using observational to estimate causal effects
  - confounding effects (omitted factors)
  - simultaneous causality
  - **correlation does not imply causation**

All you need is "data" ...

Figure 1: The cover of The Economist, November 2010

Types of data: Time Series, Cross Section, Panel Data

Time series

data for a single entity (person, firm, country) collected at multiple time periods

Cross section

data on different entity entity (workers, consumers, firms, etc.) collected at a single time period

Panel data

data for multiple entities in which each entry is observed at two or more time periods.

Time series

```
##              SP500
## 2016-03-31 2059.74
## 2016-04-29 2065.30
## 2016-05-31 2096.96
## 2016-06-30 2098.86
## 2016-07-29 2173.60
## 2016-08-31 2170.95
## 2016-09-30 2168.27
## 2016-10-31 2126.15
## 2016-11-30 2198.81
## 2016-12-30 2238.83
## 2017-01-31 2278.87
## 2017-02-28 2363.64
## 2017-03-31 2362.72
## 2017-04-28 2384.20
## 2017-05-31 2411.80
## 2017-06-23 2438.30
```

Cross section

```
salary    pcsalary  sales     roe       pcroe     ros       indus     finance
consprod  utility   lsalary   lsales

Obs:   208
  1. salary                  1990 salary, thousands $
  2. pcsalary                % change salary, 89-90
  3. sales                   1990 firm sales, millions $
  4. roe                     return on equity, 88-90 avg
  5. pcroe                   % change roe, 88-90
  6. ros                     return on firm's stock, 88-90
  7. indus                   =1 if industrial firm
  8. finance                 =1 if financial firm
  9. consprod                =1 if consumer product firm
 10. utility                 =1 if transport. or utilties
```

Cross section

```
dataset: CASchools
district school county      grades   students teachers calworks
lunch  computer expenditure income   english  read math testscore
Obs: 420
  1 school:      School name
  2 county:      County name
  3 grades:      Grade span of district
  4 students:    Student enrollment
  5 teachers:    Number of teachers
  6 calworks:    % of qualifying for CalWorks (income assistance)
  7 lunch:       % qualifying for reduced-price lunch
  8 computer:    Number of computers
  9 expenditure: Expenditure per student
 10 income:      District average income (in USD 1,000)
 11 english:     % of English learners
 12 read:        read test score
 13 math:        math test score
 14 testscore:   average of math and read
 15 str:         students/teachers
```

Panel Data

```
dataset: Cigarettes
state      year       cpi        population packs      income
tax        price      taxs
```

1. state: State
2. year: Year
3. cpi: Consumer price index
4. population: State population
5. packs: Number of packs per capita
6. income: State personal income (total, nominal)
7. tax: Average state, federal and average local excise taxes for fiscal year
8. price: Average price during fiscal year, including sales tax
9. taxs Average excise taxes for fiscal year, including sales tax

Panel Data

	state	year	cpi	population	packs	income	tax	price	taxs
01	AL	1985	1.076	3973000	116.48	46014968	32.5	102.18	33.34
02	AR	1985	1.076	2327000	128.53	26210736	37.0	101.47	37.00
03	AZ	1985	1.076	3184000	104.52	43956936	31.0	108.57	36.17
[...]									
49	AL	1995	1.524	4262731	101.085	83903280	40.5	158.37	41.90
50	AR	1995	1.524	2480121	111.042	45995496	55.5	175.54	63.85
51	AZ	1995	1.524	4306908	71.954	88870496	65.3	198.60	74.79
[...]									
94	WI	1995	1.524	5137004	92.466	115959680	62.0	201.38	71.58
95	WV	1995	1.524	1820560	115.568	32611268	41.0	166.51	50.42
96	WY	1995	1.524	478447	112.238	10293195	36.0	158.54	36.00

Formal definition of data

The data are

- a sample of size n, denoted

$$\{Y_1, Y_2, \ldots, Y_n\}$$

  - Y_1 is the first observation, Y_2 is the second observation, etc.
  - for cross section the typical observation is the i^{th} observation, denoted Y_i
  - for time series the typical observation is customarly denoted by Y_t

- if we have data on more than one variable, we have a multivariate sample,

$$\{(Y_1, X_1), (Y_2, X_2), \ldots, (Y_n, X_n)\}$$

Data summary

Central tendency

(sample) mean, (sample) median

Dispersion

(sample) variance, (sample) standard deviation

Position

(sample) Percentiles, (sample) deciles, and (sample) quartiles

Central tendency

- The leading measure of central tendency is the sample mean, which is the arithmetic average of the data
- For a sample of size n, the sample mean

$$\bar{Y} = (Y_1 + Y_2 + \ldots + Y_n)/n$$

- Often, this formula is abbreviated using the summation convention

$$\bar{Y} = \sum_{i=1}^{n} Y_i/n$$

Central tendency

- The other leading indicator of central tendency is the (sample) median, which is the value of the sample that divides the data after ordering into two halves, the median bein the midpoint
- The median is relatively easy to calculate:

- Odd Number of Data Values (n is odd)
  1. arrange data in order from smallest to largest
  2. Find the data value in the **exact** middle

- Even Number of Data Values (n is even)
  1. arrange data in order from smallest to largest
  2. Find the mean of the **two** middle numbers

Dispersion

- the sample variance
  - is the average of the deviation of the data from the mean

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

- the sample standard deviation
  - is the square root of the sample variance

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

Measure of symmetry (skewness)

The skewness measures the symmetry of the distribution:

$$skew = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - \bar{Y}}{s}\right)^3$$

- A **positive skewed** or **right-skewed** data have a much longer tail on the right (skew¿0)
- A **negative skewed** or **left-skewed** data have a much longer tail on the right (skew¡0)

Measure of peakedness (kurtosis)

The peakedness of the distribution and and fatness of the tails is measured by the kurtosis

$$kurt = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i - \bar{Y}}{s}\right)^4 - 3$$

- **positive kurtosis** indicates data that is relatively flat
- **negative kurtosis** indicates data that is relatively peaked

Position

The sample quartiles divides the data in 4 parts:

- the lower quartile (Q_1) is that point where one-quarter of the observed sample lies below and three-quarters of the order sample lies above
- the middle quartile (Q_2) is the sample median
- the upper quartile (Q_3) is that point where three-quarters of the ordered sample lies below and one-quarter of the order sample lies above.

Position

Even more detailed divisions of the sample are possible.

- Deciles split the ordered sample into tenth and are used, for example, to summarize the distribution of individual income
- Percentiles split the order sample into hundredths. The p^{th} percentile is the value for which p percent of the observed values are equal to or less than the value

Summarizing data

Table 1: Descriptive Statistics, California Schools Dataset.

Statistic	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
students	2,629.0	3,913.0	379	950.5	3,008
teachers	129.1	187.9	19.7	48.6	146.4
calworks	13.2	11.5	4.4	10.5	19.0
lunch	44.7	27.1	23.3	41.8	66.9
computer	303.4	441.3	46	117.5	375.2
expenditure	5,312.0	633.9	4,906.0	5,214.0	5,601.0
income	15.3	7.2	10.6	13.7	17.6
english	15.8	18.3	1.9	8.8	23.0
read	655.0	20.1	640.4	655.8	668.7
math	653.3	18.8	639.4	652.4	665.8

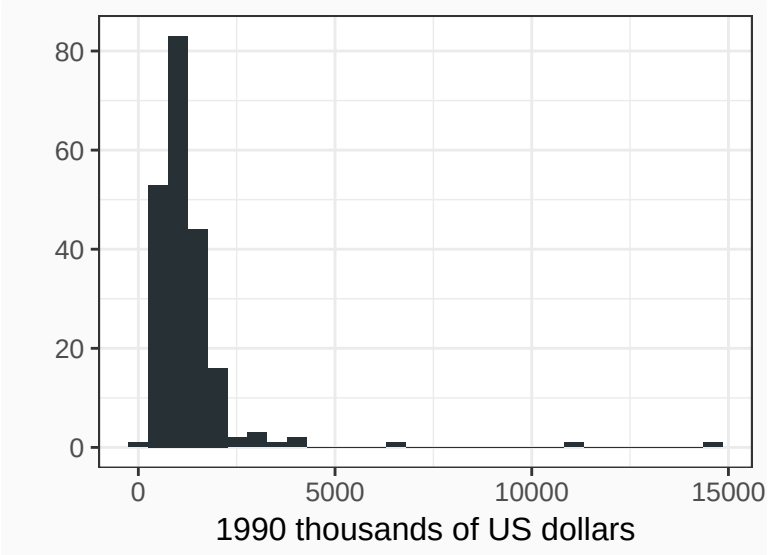
Graphical representation of data

Graphical methods used vary with the type of univariate data

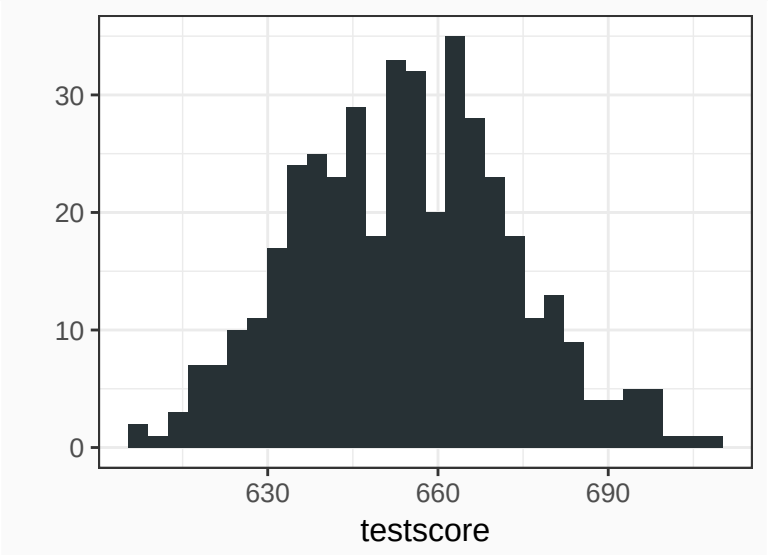
Cross-section
histogram, boxplot, pie-chart

Time series
line chart

Cross section



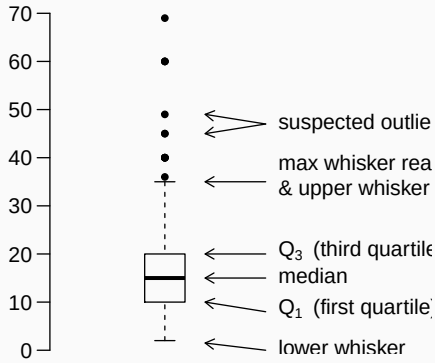
Cross section



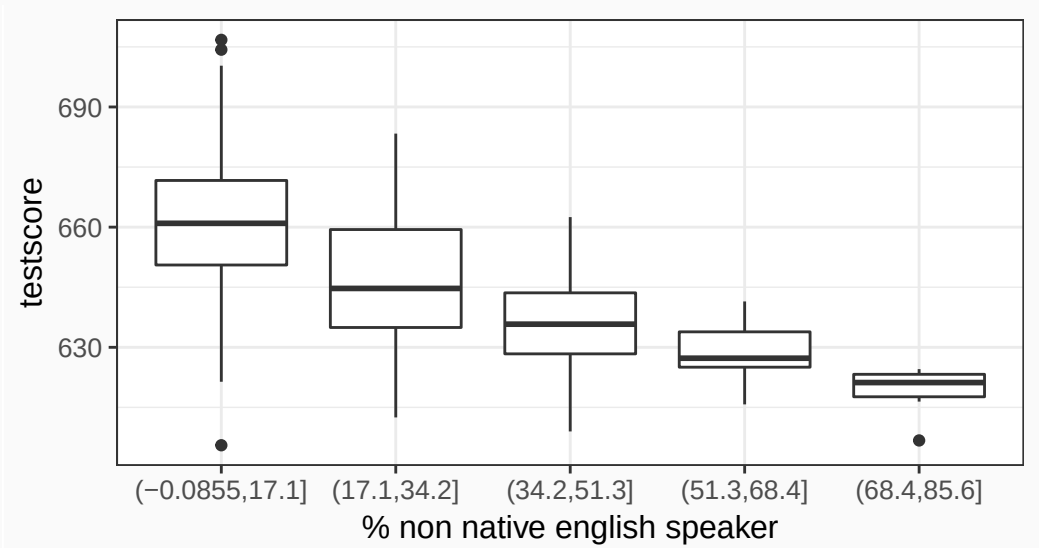
Box plots

A box plot or boxplot is a convenient way of graphically depicting groups of numerical data through their quartiles.

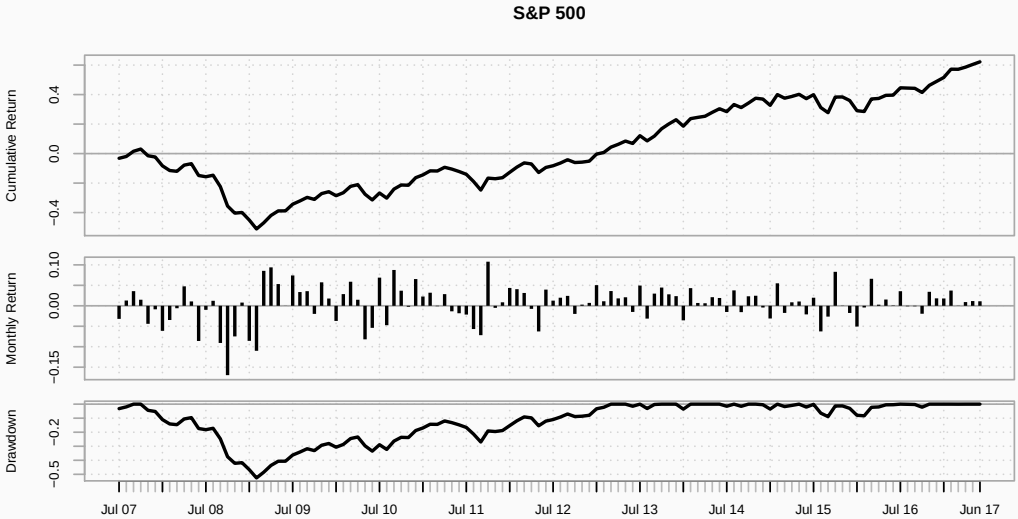
Box plots have lines extending vertically from the boxes (**whiskers**) indicating variability outside the upper and lower quartiles, hence the terms box-and-whisker plot.



Cross section



Time series


S&P 500
Cumulative Return
0.4
0.0
−0.4
Monthly Return
0.10
0.00
−0.15
Drawdown
−0.2
−0.5
Jul 07
Jul 08
Jul 09
Jul 10
Jul 11
Jul 12
Jul 13
Jul 14
Jul 15
Jul 16
Jun 17

Bivariate data analysis

Bivariate data analysis considers the relationship between two variables, such as education and income, or price and house size, or testscore and str.

Data summery tools:

- scatterplot (graphical)
- covariance and correlation (numerical)

Scatter plot

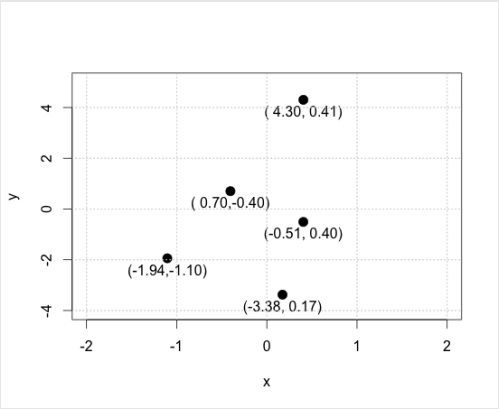
The data is displayed as a collection of points.

Example (Data)

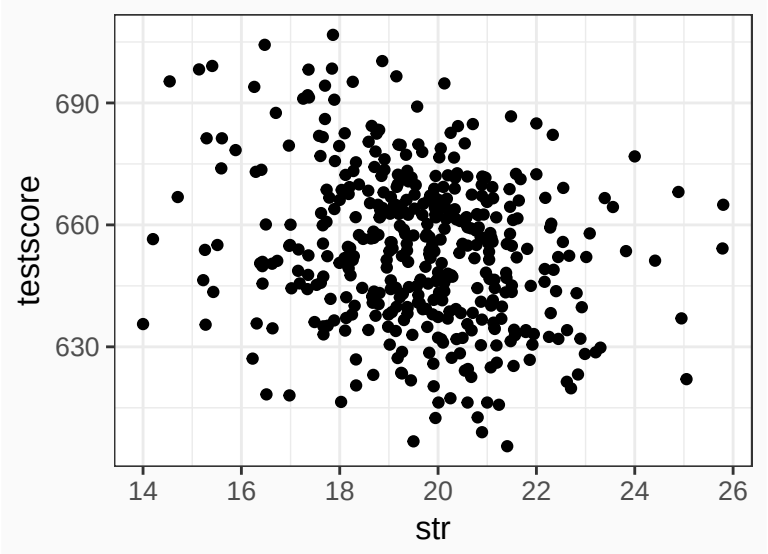
Suppose to have the following bivariate sample:

	Y	X
1	-0.51	0.40
2	0.70	-0.40
3	-3.38	0.17
4	-1.94	-1.10
5	4.30	0.41

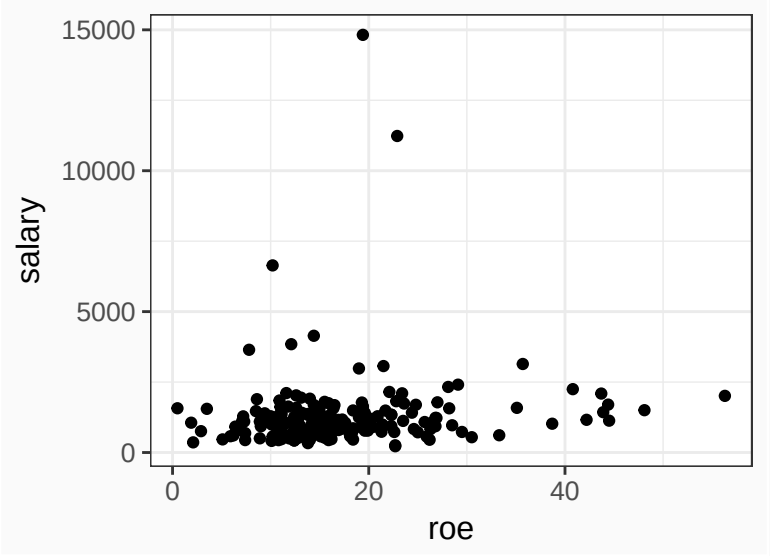
Scatterplot



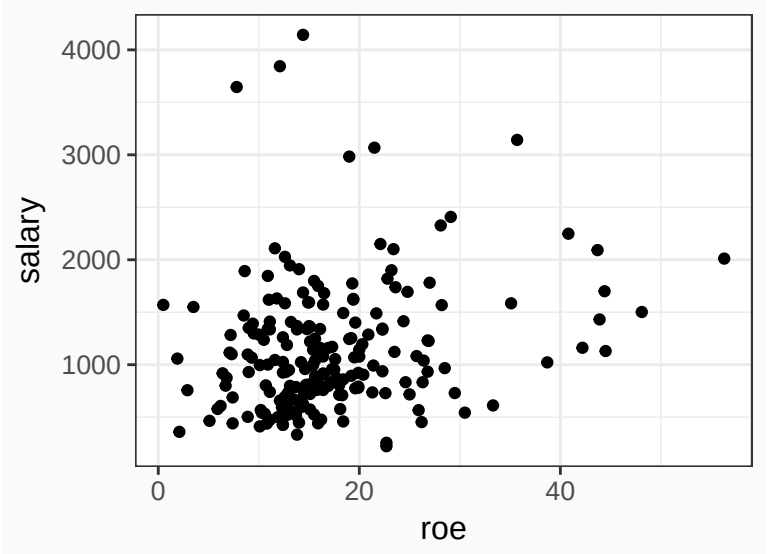
Scatter plot: `testscore` and `str`



Scatter plot: `salary` and `roe`



Scatter plot: `salary` and `roe`



Covariance

A measure of association between two variables, say x and y is the covariance:

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

Covariance

A measure of association between two variables, say x and y is the covariance:

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

- $s_{XY} > 0$: if x and y tend to move together in the **same** direction

Covariance

A measure of association between two variables, say x and y is the covariance:

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

- $s_{XY} > 0$: if x and y tend to move together in the **same** direction
- $s_{XY} < 0$: if X and Y tend to move together in the **opposite** direction

Covariance

A measure of association between two variables, say x and y is the covariance:

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

- $s_{XY} > 0$: if x and y tend to move together in the **same** direction
- $s_{XY} < 0$: if X and Y tend to move together in the **opposite** direction
- $s_{XY} = 0$: no association

Covariance

A measure of association between two variables, say x and y is the covariance:

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

- $s_{XY} > 0$: if x and y tend to move together in the **same** direction
- $s_{XY} < 0$: if X and Y tend to move together in the **opposite** direction
- $s_{XY} = 0$: no association

The covariance measures only linear association between y and x

Covariance between `testscore` and `str`

The covariance between Test Score and str is negative:

```
## [1] -8.159323
```

- the unit of measure of the covariance are **difficult** to interpret

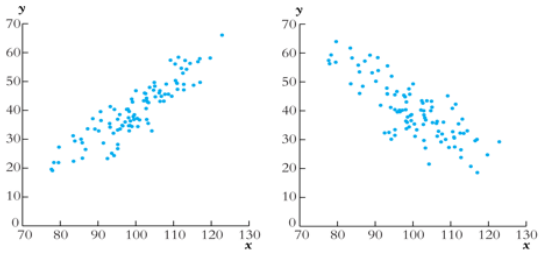
Correlation coefficient

The correlation coefficient is defined in terms of the covariance:

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

- $-1 \leq r_{XY} \leq 1$
- $r_{XY} = 1$ mean perfect **positive** linear association
- $r_{XY} = -1$ means perfect **negative** linear association
- $r_{XY} = 0$ means **no** linear association

The correlation coefficient measures linear association



(a) Correlation = +0.9

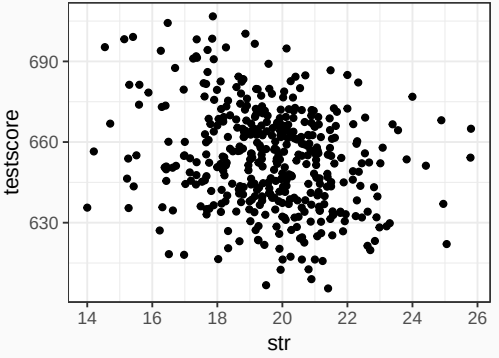
(b) Correlation = −0.8

(c) Correlation = 0.0

(d) Correlation = 0.0 (quadratic)

Correlation `testscore` vs. `str`

How big do you think it is?




```
## [1] -0.2263627
```

Correlation coefficient

The correlation coefficient is defined as

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

where s_X and s_Y are the sample standard deviations of X and Y, respectively.

Remarks

- treat X and Y symmetrically: $r_{XY} = r_{YX}$
- while r_{XY} detects (linear) association, it is neutral on whether it is X that is causing Y or Y that is causing X
- it can be shown that r_{XY} measure the number of standard deviations that Y changes by when X changes by one standard deviation

Empirical problem: Class size and educational output

- Policy question: What is the effect on test scores (or some other outcome measure) of reducing class size by one student per class? by 8 students/class?
- We must use data to find out (is there any way to answer this without data?)

The California Test Score Data Set

All K-6 and K-8 California school districts (n = 420)

Variables:

- 5th grade test scores (Stanford-9 achievement test, combined math and reading), district average
- Student-teacher ratio (STR) = no. of students in the district divided by no. full-time equivalent teachers

Initial look at the data:

(You should already know how to interpret this table)

TABLE 4.1 Summary of the Distribution of Student–Teacher Ratios and Fifth-Grade Test Scores for 420 K–8 Districts in California in 1998

			Percentile						
	Average	Standard Deviation	10%	25%	40%	50% (median)	60%	75%	90%
Student–teacher ratio	19.6	1.9	17.3	18.6	19.3	19.7	20.1	20.9	21.9
Test score	665.2	19.1	630.4	640.0	649.1	654.5	659.4	666.7	679.1

Figure 6: This table doesn't tell us anything about the relationship between test scores and the STR.

Do districts with smaller classes have higher test scores?

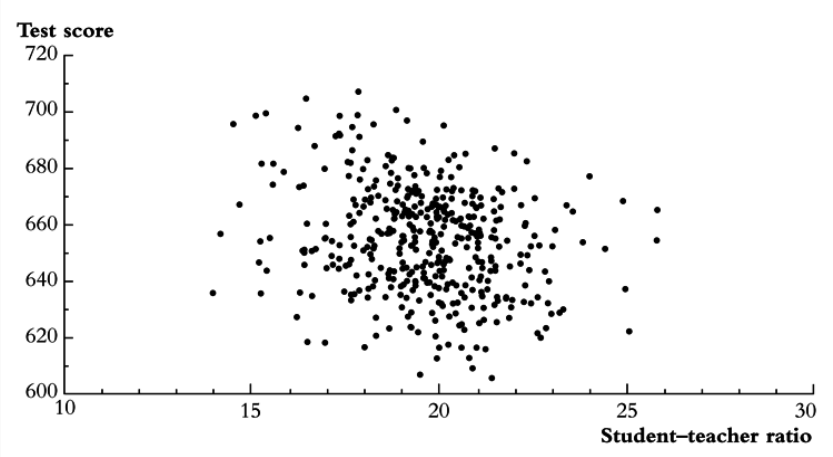


Figure 7: Scatterplot str and testscore

Approach

We need to get some numerical evidence on whether districts with low STRs have higher test scores - but how?

"Estimation"

Compare average test scores in districts with low STRs to those with high STRs

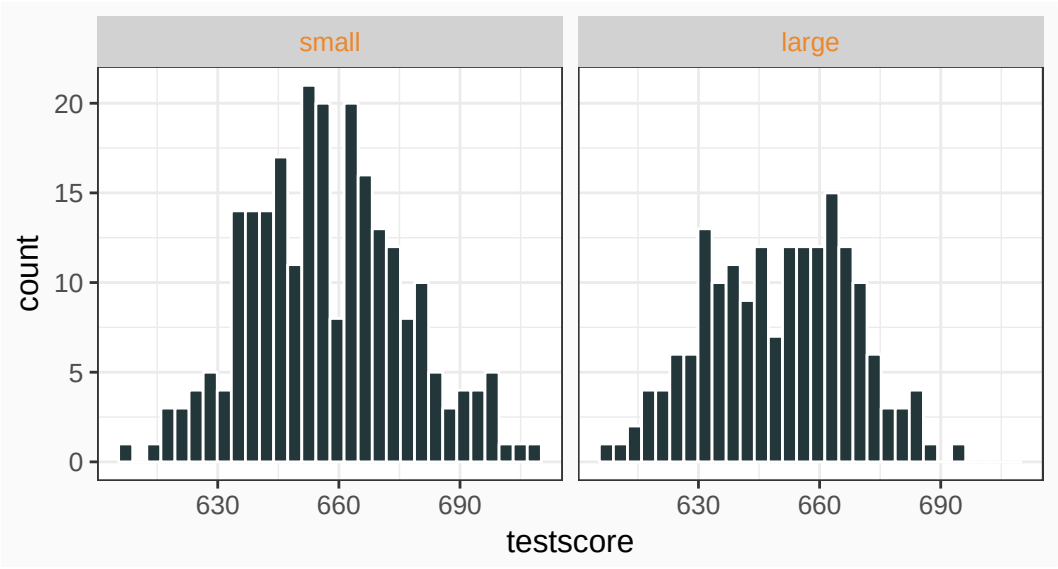
"Hypothesis testing"

Test the "null" hypothesis that the mean test scores in the two types of districts are the same, against the "alternative" hypothesis that they differ

"Confidence interval"

Estimate an interval for the difference in the mean test scores, high v. low STR districts

Initial data analysis



Initial data analysis

		testscore	
str	n	mean	sd
small	239	657.25	19.39
large	181	650.08	17.85
all	420	654.16	19.05

Steps

1. Estimation of $\Delta = \bar{Y}_{small} - \bar{Y}_{large}$ (difference between group means)
2. Test the hypothesis that $\Delta = 0$
3. Construct a confidence interval for Δ

Estimation

$$\begin{aligned}\bar{Y}_{small} - \bar{Y}_{large} =& \frac{1}{n_{small}} \sum_{i \in \text{small}} Y_i - \frac{1}{n_{large}} \sum_{i \in \text{large}} Y_i \\ =& 657.25 - 650.08 \\ =& 7.17\end{aligned}$$

Estimation

$$\begin{aligned}\bar{Y}_{small} - \bar{Y}_{large} =& \frac{1}{n_{small}} \sum_{i \in \text{small}} Y_i - \frac{1}{n_{large}} \sum_{i \in \text{large}} Y_i \\ =& 657.25 - 650.08 \\ =& 7.17\end{aligned}$$

- **Question:** Is this a large difference in a real-world sense?

Estimation

$$\begin{aligned}\bar{Y}_{small} - \bar{Y}_{large} =& \frac{1}{n_{small}} \sum_{i \in \text{small}} Y_i - \frac{1}{n_{large}} \sum_{i \in \text{large}} Y_i \\ =& 657.25 - 650.08 \\ =& 7.17\end{aligned}$$

- **Question:** Is this a large difference in a real-world sense?
  - Standard deviation across districts = 19.05

Estimation

$$\begin{aligned}\bar{Y}_{small} - \bar{Y}_{large} =& \frac{1}{n_{small}} \sum_{i \in \text{small}} Y_i - \frac{1}{n_{large}} \sum_{i \in \text{large}} Y_i \\ =& 657.25 - 650.08 \\ =& 7.17\end{aligned}$$

- **Question:** Is this a large difference in a real-world sense?
  - Standard deviation across districts $= 19.05$
  - Difference between 60th and 75th percentiles of test score distribution is $667.6 - 659.4 = 8.2$

Estimation

$$\begin{aligned}\bar{Y}_{small} - \bar{Y}_{large} =& \frac{1}{n_{small}} \sum_{i \in \text{small}} Y_i - \frac{1}{n_{large}} \sum_{i \in \text{large}} Y_i \\ =& 657.25 - 650.08 \\ =& 7.17\end{aligned}$$

- **Question:** Is this a large difference in a real-world sense?
  - Standard deviation across districts $= 19.05$
  - Difference between 60th and 75th percentiles of test score distribution is $667.6 - 659.4 = 8.2$
  - This is a big enough difference to be important for school reform discussions, for parents, or for a school committee?

Hypothesis testing

Difference-in-means test: compute the t-statistic:

$$t = \frac{\bar{Y}_{small} - \bar{Y}_{large}}{\sqrt{\frac{s^2_{small}}{n_{small}} + \frac{s^2_{large}}{n_{large}}}} = \frac{\bar{Y}_{small} - \bar{Y}_{large}}{SE(Y_{small} - Y_{large})}$$

where $SE(Y_{small} - Y_{large})$ is the standard error of $\bar{Y}_{small} - \bar{Y}_{large}$ and

$$s_{small} = \frac{1}{n_{small} - 1} \sum_{i \in small} (Y_i - \bar{Y})^2, \quad s_{large} = \frac{1}{n_{large} - 1} \sum_{i \in large} (Y_i - \bar{Y})^2.$$

Compute the difference-of-means t-statistic:

		testscore	
str	n	mean	sd
large	239	657.25	19.39
small	181	650.08	17.85
All	420	654.16	19.05

$$t = \frac{\bar{Y}_{small} - \bar{Y}_{large}}{SE(Y_{small} - Y_{large})} = \frac{657.25 - 650.08}{\sqrt{\frac{19.39^2}{239} + \frac{17.85^2}{182}}} = \frac{7.17}{1.82} = 3.93$$

t-test

$|t| > 1.96$, so reject (at the 5% significance level) the null hypothesis that the two means are the same.

Confidence interval

A 95% confidence interval for the difference between the means is,

$$(\bar{Y}_{small} - \bar{Y}_{large}) \pm 1.96 \times SE(\bar{Y}_{small} - \bar{Y}_{large}) = 7.17 \pm 1.96 \times 1.82 = (3.6, 10.7)$$

- Two equivalent statements:
  1. The 95% confidence interval for $\bar{Y}_{small} - \bar{Y}_{large}$ doesn't include 0;
  2. The null hypothesis that $\bar{Y}_{small} - \bar{Y}_{large} = 0$ vs. a dual sided alternative is rejected at the 5% significance level.

What comes next...

- The mechanics of estimation, hypothesis testing, and confidence intervals should be familiar
- These concepts extend directly to regression and its variants
- Before turning to regression, however, we will review some of the underlying theory of estimation, hypothesis testing, and confidence intervals:
  - why do these procedures work, and why use these rather than others?
  - So we will review the intellectual foundations of statistics and econometrics

Review of Statistical Theory

1. **The probability framework for statistical inference**
2. Estimation
3. Testing
4. Confidence Intervals

The probability framework for statistical inference

- Population, sample
- Random variable, and distribution
- Moments of a distribution (mean, variance, standard deviation, covariance, correlation)
- Conditional distributions and conditional means
- Distribution of a sample of data drawn randomly from a population: $Y_1, ..., Y_n$

Population and sample

Population

- The group or collection of all possible entities of interest (school districts)
- We will think of populations as infinitely large (∞ is an approximation to "very big")

Sample

A sample is a **subset** selected from the population

Population and sample

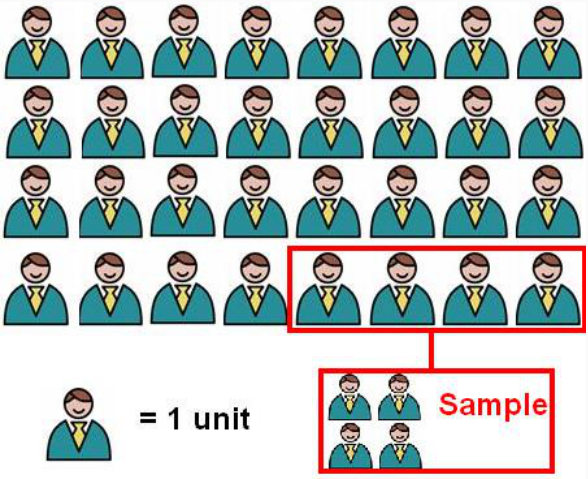


Figure 8: Population and sample

Random variables and probability distributions

Random variable X

- Numerical summary of a random outcome (district average test score, district str)

Random variables and probability distributions

Random variable X

- Numerical summary of a random outcome (district average test score, district str)

Probability distribution of X

- The probabilities of different values of Y that occur in the population, for ex. $\Pr[X = 650]$ (when X is discrete)
- or: The probabilities of sets of these values, for ex. $\Pr[640 \leq Y \leq 660]$ (when X is continuous)
  - in this case the probability is expressed through probability density fincion (p.d.f.)

Probability distribution

If X is continuous, the probability of X is expressed as

$$\Pr[a \leq X \leq b] = \int_a^b f(x)dx,$$

where $f(x)$ is the p.d.f. of X.

Notation

If the random variable X has a normal distribution, we say write

$$X \sim N(\mu, \sigma^2).$$

Probability distribution

A very important distribution is the normal (or Gaussian) distribution. The normal distribution has a bell-shaped p.d.f. which is formally given by:

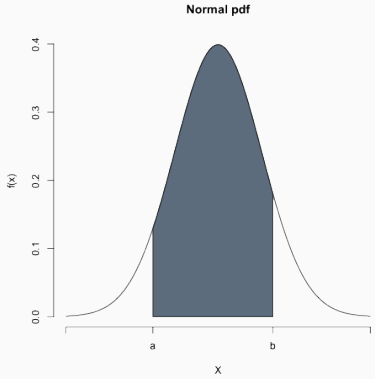
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

where μ and σ are parameters that we will see have an important interpretation.

Probability distribution

The probability is the area under the bell shaped p.d.f.

$$\Pr[a \leq X \leq b] = \int_a^b f(x)dx$$



Probability distribution

An other important distribution is the chi-squared distribution:

$$f(x) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2}, & x \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

where

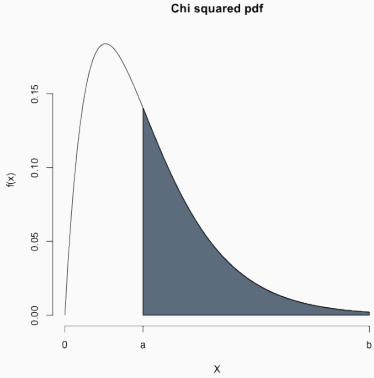
- $\Gamma(\cdot)$ is a complicated function(called Gamma function)
- ν is a parameter (this parameter indicated the "degrees of freedom" of the χ^2 distribution—we often say that $X \sim \chi^2_d$

Notation

If the random variable X has a chi-squared distribution with ν degrees of freedom, we write

$$X \sim \chi^2_\nu.$$

Probability distribution



The probability is the area under the p.d.f.

$$\Pr[a \leq X \leq b] = \int_a^b f(x)dx$$

Probability distribution

An other important distribution is the t-student distribution:

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

where

- $\Gamma(\cdot)$ is a complicated function (called Gamma function)
- ν is a parameter (this parameter denotes the "degrees of freedom"—we often say that $X \sim t(\nu)$

Notation

If the random variable X has a t-student distribution with ν degrees of fredom, we write

$$X \sim t(\nu).$$

Probability distribution

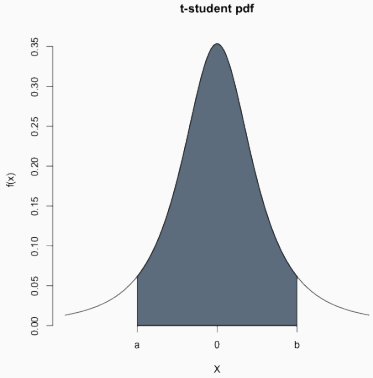


Figure 10:

The probability is the area under the p.d.f.

$$\int^{b}$$

Moments of a population distribution

mean (long-run average value of Y over repeated realizations)

$$E(X) := \int xf(x)dx$$

The shorthand for the expected value of a r.v. X is μ_X.

Moments of a population distribution

mean (long-run average value of Y over repeated realizations)

$$E(X) := \int xf(x)dx$$

The shorthand for the expected value of a r.v. X is μ_X.

variance (measure of the squared spread of the distribution)

$$E(X - \mu_X)^2 := \int (x - \mu_x)^2 f(x)dx$$

The shorthand for the variance of a r.v. X is σ_X^2.

Moments, cont'd

skewness (measure of asymmetry of a distribution)

$$\frac{E[(Y - \mu_Y)^3]}{\sigma_Y^3}$$

- skewness $= 0$: distribution is symmetric
- skewness $>$ ($<$) 0: distribution has long right (left) tail

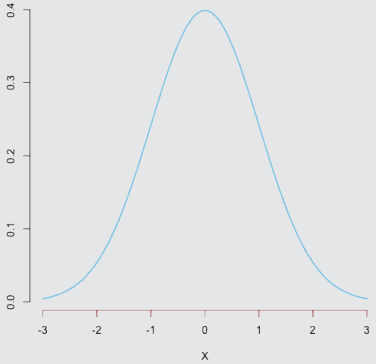
kurtosis (measure of mass in tails)

$$\frac{E[(Y - \mu_Y)^4]}{\sigma_Y^4}$$

- kurtosis $= 3$: normal distribution
- skewness ¿ 3: heavy tails ("leptokurtotic")

Moments, cont'd

$X \sim N(0, 1)$

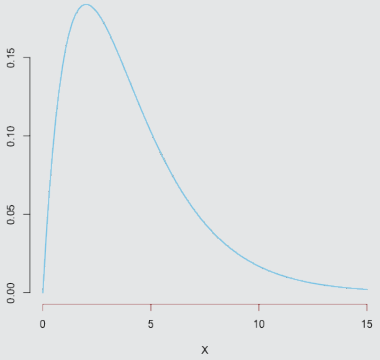


Moments

$$\mu_X = E(X) = 0$$
$$\sigma_X^2 = \mathrm{Var}(X) = 1$$
$$\sigma_X = \sqrt{\mathrm{Var}(X)} = 1$$
$$\mathrm{skew}(X) = \frac{E(X - \mu_X)^3}{\sigma_X^3} = 0$$
$$\mathrm{kurt}(X) = \frac{E(X - \mu_X)^4}{\sigma_X^4} = 3$$

Moments, cont'd

$X \sim \chi^2_\nu$

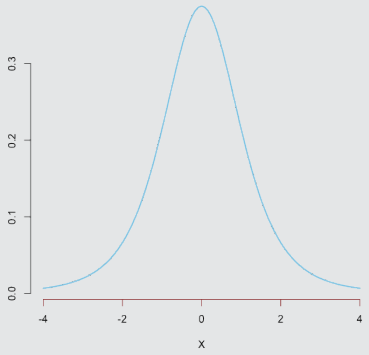


Moments

$$
\begin{aligned}
\mu_X &= E(X) = \nu \\
\sigma_X^2 &= \mathrm{Var}(X) = 2\nu \\
\sigma_X &= \sqrt{\mathrm{Var}(X)} = \sqrt{2\nu} \\
\mathrm{skew}(X) &= \frac{E(X-\mu_X)^3}{\sigma_X^3} = \sqrt{8/\nu} \\
\mathrm{kurt}(X) &= \frac{E(X-\mu_X)^4}{\sigma_X^4} = 12/\nu
\end{aligned}
$$

Moments, cont'd

$X \sim t(\nu)$



Moments

$$
\begin{aligned}
\mu_X &= E(X) = 0, \text{ if } \nu > 1 \\
\sigma_X^2 &= \mathrm{Var}(X) = \nu/(\nu-2) \\
\sigma_X &= \sqrt{\mathrm{Var}(X)} = \sqrt{\nu/(\nu-2)} \\
\mathrm{skew}(X) &= \frac{E(X-\mu_X)^3}{\sigma_X^3} = 0 \\
\mathrm{kurt}(X) &= \frac{E(X-\mu_X)^4}{\sigma_X^4} = 6/(\nu-4)
\end{aligned}
$$

Random variables: joint distributions and covariance

- Random variables X and Z have a joint distribution
- The covariance between X and Z is

$$cov(X, Z) = E[(X - \mu_X)(Z - \mu_Z)] = \sigma_{XZ}$$

- The covariance is a measure of the linear association between X and Z; its units are units of X and units of Z
- $cov(X, Z) > 0$ means a positive relation between X and Z
- If X and Z are independently distributed, then $cov(X, Z) = 0$ (but not vice versa!!)
- The covariance of a r.v. with itself is its variance:

$$cov(X, X) = E[(X - \mu_X)(X - \mu_X)] = E[(X - \mu_X)^2]$$

Conditional distributions and conditional means

Conditional distributions

The distribution of Y, given value(s) of some other random variable, X

Conditional expectations and conditional moments

- conditional mean = mean of conditional distribution

$$E(Y|X = x) = \int yf(y|X = x)dy$$

- conditional variance = variance of conditional distribution

$$Var(Y|X = x) = \int yf(y|X = x)dy$$

Example (Example:)

$E(Testscores|STR < 20)$ = the mean of test scores among districts with small class sizes

Difference in (conditional) mean

The difference in means is the difference between the means of two conditional distributions:

$$\Delta = E[testscore|str < 20] - E[testscore|str \geq 20]$$

Other examples of conditional means:

- Wages of all female workers (Y = wages, X = gender)
- Mortality rate of those given an experimental treatment (Y live/die; X = treated/not treated)

Important fact: mean independence

Take two random variables, say U and X. Then is

$$E[U|X = x] = \text{constant}, \quad \text{for all } x$$

then

$$cov(U, X) = 0, \quad E[U] = \text{constant}.$$

We say in this case that U is conditional mean independent from X.

Notice that, $cov(X, U) = 0$ does not imply $E[U|X] = \text{constant}$.

Distribution of a sample drawn randomly from a population

Let Y denote a variable of interest, for instance

$$Y = \{\text{net wage of italian full time employees}\}$$

.

Think of $(Y_1, Y_2, \ldots, Y_n)$ as the collection of wages of n workers drawn from the population

- **Prior to sample selection**, the wages $(Y_1, \ldots, Y_n)$ are random variables because the workers are randomly selected
- **Once the worker is selected** and the value of Y is observed, then $(Y_1, \ldots, Y_n)$ are just an array of numbers - not random

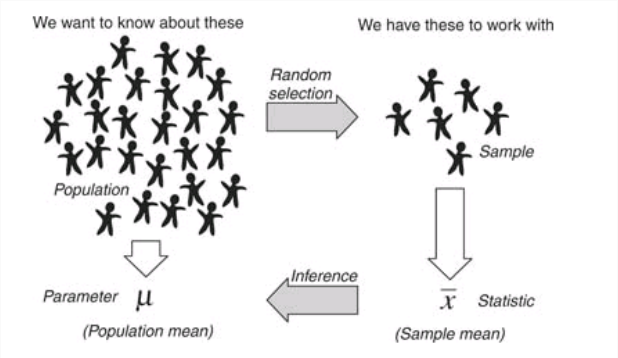
Simple random sampling (or iid)

We will assume simple random sampling, that is, entities (district, entity) are drawn at random from the population.

In this case we will say that $(Y_1, \ldots, Y_n)$ is a family of *independent, identically distributed* (i.i.d.) random variables.

- Y_j and Y_k are independent, that is, the value Y_j has no information content for Y_k (independently)
- The probability distribution of each r.v. is the same (identically)

Sampling distribution



Framework

This framework allows rigorous statistical inferences about moments of population distributions using a sample of data from that population

1. The probability framework for statistical inference
2. **Estimation**
3. Testing
4. Confidence Intervals

Estimation

$\bar{Y}$ is the natural estimator of the expected value of Y, μ_Y. But:

1. What are the properties of $\bar{Y}$?
2. Why should we use $\bar{Y}$ rather than some other estimator?
   - y_1 (the first observation)
   - maybe unequal weights - not simple average
   - $median(Y_1, ..., Y_n)$

The sampling distribution of $\bar{Y}$

The sampling distribution of $\bar{Y}$ is a random variable, and its properties are determined by the sampling distribution of $\bar{Y}$

- The sample is i.i.d.

The sampling distribution of $\bar{Y}$

The sampling distribution of $\bar{Y}$ is a random variable, and its properties are determined by the sampling distribution of $\bar{Y}$

- The sample is i.i.d.
- Thus the values of $(Y_1, \ldots, Y_n)$ are random

The sampling distribution of $\bar{Y}$

The sampling distribution of $\bar{Y}$ is a random variable, and its properties are determined by the sampling distribution of $\bar{Y}$

- The sample is i.i.d.
- Thus the values of $(Y_1, \ldots, Y_n)$ are random
- Thus functions of $(Y_1, \ldots, Y_n)$, such as $\bar{Y}$, are random:
  - had a different sample been drawn, they would have taken on a different value

The sampling distribution of $\bar{Y}$

The sampling distribution of $\bar{Y}$ is a random variable, and its properties are determined by the sampling distribution of $\bar{Y}$

- The sample is i.i.d.
- Thus the values of $(Y_1, \ldots, Y_n)$ are random
- Thus functions of $(Y_1, \ldots, Y_n)$, such as $\bar{Y}$, are random:
  - had a different sample been drawn, they would have taken on a different value
- The distribution of $\bar{Y}$ over different possible samples of size n is called the sampling distribution of $\bar{Y}$

The sampling distribution of $\bar{Y}$

The sampling distribution of $\bar{Y}$ is a random variable, and its properties are determined by the sampling distribution of $\bar{Y}$

- The sample is i.i.d.
- Thus the values of $(Y_1, \ldots, Y_n)$ are random
- Thus functions of $(Y_1, \ldots, Y_n)$, such as $\bar{Y}$, are random:
  - had a different sample been drawn, they would have taken on a different value
- The distribution of $\bar{Y}$ over different possible samples of size n is called the sampling distribution of $\bar{Y}$
- The mean and variance of $\bar{Y}$ are the mean and variance of its sampling distribution, $E(\bar{Y})$ and $var(\bar{Y})$.

The sampling distribution of $\bar{Y}$

The sampling distribution of $\bar{Y}$ is a random variable, and its properties are determined by the sampling distribution of $\bar{Y}$

- The sample is i.i.d.
- Thus the values of $(Y_1, \ldots, Y_n)$ are random
- Thus functions of $(Y_1, \ldots, Y_n)$, such as $\bar{Y}$, are random:
  - had a different sample been drawn, they would have taken on a different value
- The distribution of $\bar{Y}$ over different possible samples of size n is called the sampling distribution of $\bar{Y}$
- The mean and variance of $\bar{Y}$ are the mean and variance of its sampling distribution, $E(\bar{Y})$ and $var(\bar{Y})$.
- The concept of the sampling distribution underpins all of econometrics.

Example: Bernoulli distribution

Suppose Y takes on 0 or 1 (a Bernoulli random variable) with

$$Y = \begin{cases} 0 & p = .22 \\ 1 & p = .78 \end{cases}$$

Then

$$E(Y) = p \times 1 + (1 - p) \times 0 = p = .78$$

and

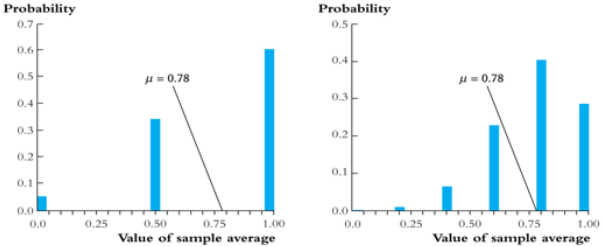
$$\sigma_Y^2 = E[Y - E(Y)]^2 = p(1 - p) = .78 \times (1 - .78) = 0.1716$$

The sampling distribution of $\bar{Y}$ depends on n.

Consider n = 2. The sampling distribution of $\bar{Y}$ is,

- $\Pr(\bar{Y} = 0) = .22^2 = .0484$
- $\Pr(\bar{Y} = 1/2) = 2 \times .22 \times .78 = .3432$
- $\Pr(\bar{Y} = 1) = .78^2 = .6084$

The sampling distribution of $\bar{Y}$ when Y is Bernoulli ($p = .78$):



Things we want to know about the sampling distribution:

- What is the mean of $\bar{Y}$?
  - If $E(\bar{Y}) = \mu = .78$, then $\bar{Y}$ is an unbiased estimator of μ

Things we want to know about the sampling distribution:

- What is the mean of $\bar{Y}$?
  - If $E(\bar{Y}) = \mu = .78$, then $\bar{Y}$ is an unbiased estimator of μ
- What is the variance of $\bar{Y}$?
  - How does $var(\bar{Y})$ depend on n?
  - Does $\bar{Y}$ become close to μ when n is large?
    - Law of large numbers: $\bar{Y}$ is a consistent estimator of μ?

Things we want to know about the sampling distribution:

- What is the mean of $\bar{Y}$?
  - If $E(\bar{Y}) = \mu = .78$, then $\bar{Y}$ is an unbiased estimator of μ
- What is the variance of $\bar{Y}$?
  - How does $var(\bar{Y})$ depend on n?
  - Does $\bar{Y}$ become close to μ when n is large?
    - Law of large numbers: $\bar{Y}$ is a consistent estimator of μ?
- $\bar{Y} - \mu$ appears bell shaped for n large... is this generally true?
  - In fact, $\bar{Y} - \mu$ is approximately normally distributed for n large (Central Limit Theorem)

Mean and variance of sampling distribution of $\bar{Y}$

$$E(\bar{Y}) = \mu_Y$$

and

$$var(\bar{Y}) = \frac{\sigma_Y^2}{n}$$

Implications:

1. $\bar{Y}$ is an unbiased estimator of μ_Y, (that is, $E(\bar{Y}) = \mu_Y$)
2. $var(\bar{Y})$ is inversely proportional to n
   - the spread of the sampling distribution is proportional to $1/n$
   - Thus the sampling uncertainty associated with is proportional to 1/n (larger samples, less uncertainty, but square-root law)

The sampling distribution of $\bar{Y}$ when n is large

For small sample sizes, the distribution of $\bar{Y}$ is complicated, but if n is large, the sampling distribution is simple!

1. As n increases, the distribution of becomes more tightly centered around μ_Y (the Law of Large Numbers)
2. Moreover, the distribution of $\bar{Y} - \mu_Y$ becomes normal (the Central Limit Theorem)

The Law of Large Numbers (LLN)

An estimator is consistent if the probability that its falls within an interval of the true population value tends to one as the sample size increases.

Theorem (LLN)

If $(Y1, \ldots, Yn)$ *are i.i.d. and* $\sigma_Y^2 < \infty$, *then* $\bar{Y}$ *is a consistent estimator of* μ_Y, *that is,*

$$\Pr[|\bar{Y} - \mu_Y| < \epsilon] \to 1 \text{ as } n \to \infty$$

which can be written, $\bar{Y} \xrightarrow{p} \mu_Y$

The Central Limit Theorem (CLT):

If $(Y_1, \ldots, Y_n)$ are i.i.d. and $0 < \sigma_Y^2 < \infty$, then when n is large the distribution of $\bar{Y}$ is well approximated by a normal distribution.

- $\bar{Y}$ is approximately distributed $N(\mu_Y, \frac{\sigma_Y^2}{n})$ ("normal distribution with mean μ_Y and variance σ^2/n)

The Central Limit Theorem (CLT):

If $(Y_1, \ldots, Y_n)$ are i.i.d. and $0 < \sigma_Y^2 < \infty$, then when n is large the distribution of $\bar{Y}$ is well approximated by a normal distribution.

- $\bar{Y}$ is approximately distributed $N(\mu_Y, \frac{\sigma_Y^2}{n})$ ("normal distribution with mean μ_Y and variance σ^2/n)
- $\sqrt{n}(\bar{Y} - \mu_Y)/\sigma_Y$ is approximately distributed $N(0, 1)$ (standard normal)

The Central Limit Theorem (CLT):

If $(Y_1, \ldots, Y_n)$ are i.i.d. and $0 < \sigma_Y^2 < \infty$, then when n is large the distribution of $\bar{Y}$ is well approximated by a normal distribution.

- $\bar{Y}$ is approximately distributed $N(\mu_Y, \frac{\sigma_Y^2}{n})$ ("normal distribution with mean μ_Y and variance σ^2/n)
- $\sqrt{n}(\bar{Y} - \mu_Y)/\sigma_Y$ is approximately distributed $N(0,1)$ (standard normal)
- $\sqrt{n}(\bar{Y} - \mu_Y)/s_Y$ is approximately distributed $N(0,1)$ (standard normal)

The Central Limit Theorem (CLT):

If $(Y_1, \ldots, Y_n)$ are i.i.d. and $0 < \sigma_Y^2 < \infty$, then when n is large the distribution of $\bar{Y}$ is well approximated by a normal distribution.

- $\bar{Y}$ is approximately distributed $N(\mu_Y, \frac{\sigma_Y^2}{n})$ ("normal distribution with mean μ_Y and variance σ^2/n)
- $\sqrt{n}(\bar{Y} - \mu_Y)/\sigma_Y$ is approximately distributed $N(0,1)$ (standard normal)
- $\sqrt{n}(\bar{Y} - \mu_Y)/s_Y$ is approximately distributed $N(0,1)$ (standard normal)
- The larger is n, the better are these approximations.

Summary: The Sampling Distribution of $\bar{Y}$

For $Y_1, \ldots, Y_n$ i.i.d. with $0 < \sigma_Y^2 < \infty$

- The exact (finite sample) sampling distribution of has mean μ_Y and variance σ_Y^2/n
- Other than its mean and variance, the exact distribution of is complicated and depends on the distribution of Y
- When n is large, the sampling distribution simplifies:
  - $$\bar{Y} \xrightarrow{p} \mu_Y, \text{ (Law of large numbers)}$$
  - $$\frac{\sqrt{n}(\bar{Y} - \mu_Y)}{\sigma_Y} \text{ is approximately N(0,1), (CLT)}$$

Why use $\bar{Y}$ to estimate μ_Y?

- is unbiased: $E(\bar{Y}) = \mu_Y$
- is consistent: $\bar{Y} \xrightarrow{p} \mu_Y$
- is the "least squares" estimator of μ_Y; $\bar{Y}$ solves

$$\min_m \sum_{i=1}^{n} (Y_i - m)^2$$

$\bar{Y}$ minimizes the sum of squared "residuals"

Set derivative to zero and denote optimal value of m by

$$\frac{d}{dm} \sum_{i=1}^{n} (Y_i - m)^2 = \sum_{i=1}^{n} \frac{d}{dm} (Y_i - m)^2 = 2 \sum_{i=1}^{n} (y_i - m).$$

Setting the derivative to zero $m = \frac{1}{n} \sum_{i=1} Y_i = \bar{Y}$.

Why Use $\bar{Y}$ To Estimate μ_Y?, ctd.

- $\bar{Y}$ has a smaller variance than all other linear unbiased estimators:

Example

consider the estimator, $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} a_i Y_i$, where $\{a_i\}$ are such that $\bar{\mu}$ is unbiased;

- then $\mathrm{var}(\hat{\mu}) \geq \mathrm{var}(\bar{Y})$

Estimator of the variance of Y

A good estimator of σ_Y^2 is the sample variance of Y

$$s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

Facts:

If $(Y_1, \ldots, Y_n)$ are i.i.d. and $E(Y_4) < \infty$, then $s_Y^2 \xrightarrow{p} \sigma_Y^2$ and, also,

$$s_Y \xrightarrow{p} \sigma_Y$$

Why does the law of large numbers apply?

- Because s_Y^2 is a sample average (of $(Y_i - \bar{Y})^2$)
- Technical note: we assume $E(Y^4) < \infty$ because here the average is not of Y_i, but of its square

Actually:

population quantity	alternative notation	sample quantity
$E(Y)$	μ_Y	$\bar{Y}$
$\mathrm{Var}(Y)$	σ_Y^2	s_Y^2
$\sqrt{\mathrm{Var}(Y)}$	σ_Y	s_Y
$\mathrm{cov}(Y, X)$		s_{YX}
$\mathrm{corr}(Y, X)$		ρ_{XY}

Sample ¡-¿ Quantities

All these sample quantities are all "good" estimators of the population quantities, in the sense that they are all consistent.

Where are we?

1. The probability framework for statistical inference
2. Estimation
3. **Hypothesis Testing**
4. Confidence intervals

Where are we?

1. The probability framework for statistical inference
2. Estimation
3. **Hypothesis Testing**
4. Confidence intervals

Hypothesis Testing

The hypothesis testing problem (for the mean): make a provisional decision, based on the evidence at hand, whether a null hypothesis is true, or instead that some alternative hypothesis is true. That is, test

$$H0 : E(Y) = \mu_{Y,0} \text{ vs. } H1 : E(Y) > \mu_{Y,0} \text{ (1-sided, } >)$$

$$H0 : E(Y) = \mu_{Y,0} \text{ vs. } H1 : E(Y) < \mu_{Y,0} \text{ (1-sided, } <)$$

$$H0 : E(Y) = \mu_{Y,0} \text{ vs. } H1 : E(Y) \neq \mu_{Y,0} \text{ (2-sided, } \neq)$$

Some terminology for testing statistical hypotheses:

The significance level of a test

is a pre-specified probability of incorrectly rejecting the null, when the null is true.

Testing (Two sided)

$$H_0 : \mu_Y = \mu_{Y,0} \quad \text{vs.} \quad H_1 : \mu_Y \neq \mu_{Y,0}$$

We reject the null hypothesis at the **5%** significance level if:

$$\frac{\bar{Y} - \mu_{Y,0}}{s_Y/\sqrt{n}} > 1.96 \qquad \text{or} \qquad \frac{\bar{Y} - \mu_{Y,0}}{s_Y/\sqrt{n}} < -1.96$$

or, more compactly, if

$$\left| \frac{\bar{Y} - \mu_{Y,0}}{s_Y/\sqrt{n}} \right| > 1.96$$

Testing (Two sided)

$$H_0 : \mu_Y = \mu_{Y,0} \quad \text{vs.} \quad H_1 : \mu_Y \neq \mu_{Y,0}$$

We reject the null hypothesis at the **10%** significance level if:

$$\frac{\bar{Y} - \mu_{Y,0}}{s_Y/\sqrt{n}} > 1.64 \qquad \text{or} \qquad \frac{\bar{Y} - \mu_{Y,0}}{s_Y/\sqrt{n}} < -1.64$$

or, more compactly, if

$$\left| \frac{\bar{Y} - \mu_{Y,0}}{s_Y/\sqrt{n}} \right| > 1.64$$

Testing (one sided)

$$H_0 : \mu_Y = \mu_{Y,0} \quad \text{vs.} \quad H_1 : \mu_Y > \mu_{Y,0}$$

We reject the null hypothesis at the **5%** significance level if:

$$\frac{\bar{Y} - \mu_{Y,0}}{s_Y/\sqrt{n}} \geq 1.64$$

$$H0 : \mu_Y = \mu_{Y,0} \text{ vs. } H1 : \mu_Y < \mu_{Y,0}$$

We reject the null hypothesis at the **5%** significance level if:

$$\frac{\bar{Y} - \mu_{Y,0}}{s_Y/\sqrt{n}} \leq 1.64$$

Terminology

The quantity:

$$t = \frac{\bar{Y} - \mu_{Y,0}}{s_Y/\sqrt{n}}$$

is referred to as the (Student's) t-statistics.

The same quantity can be equivalently expressed as:

$$t = \frac{\bar{Y} - \mu_{Y,0}}{SE(\bar{Y})}$$

where $SE(\bar{Y}) = s_Y/\sqrt{n}$ is called the standard error of $\bar{Y}$.

The p-value

p-value

probability of drawing a statistic (e.g. $\bar{Y}$) at least as adverse to the null as the value actually computed with your data, assuming that the null hypothesis is true. Calculating the p-value based on $\bar{Y}$:

$$p - value = \Pr[|\bar{Y} - \mu_{Y,0}| > |\bar{Y}^{act} - \mu_{Y,0}|]$$

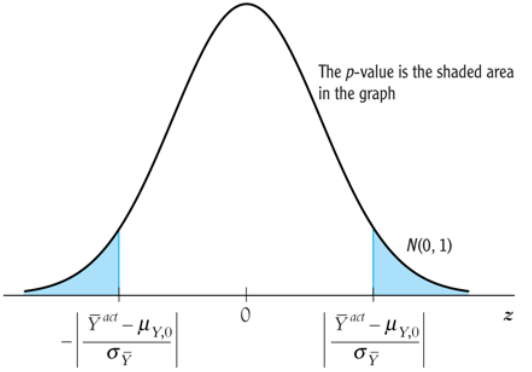
where $\bar{Y}^{act}$ is the value of $\bar{Y}$ actually observed (nonrandom)

Calculating the p-value, ctd.

- To compute the p-value, you need the to know the sampling distribution of $\bar{Y}$, which is complicated if n is small.
- If n is large, you can use the normal approximation (CLT):

$$
\begin{aligned}
p-value &= \Pr_{H_0}[|\bar{Y} - \mu_{Y,0}| > |\bar{Y}^{act} - \mu_{Y,0}|] \\
&= \Pr_{H_0}\left[\frac{|\bar{Y} - \mu_{Y,0}|}{\sigma_Y/\sqrt{n}} > \frac{|\bar{Y}^{act} - \mu_{Y,0}|}{\sigma_Y/\sqrt{n}}\right] \\
&\approx \text{probability under left+right of } N(0,1) \text{ density}
\end{aligned}
$$

Calculating the p-value with σ_Y known:



- For large n, p-value = the probability that a N(0,1) random variable falls outside $\sqrt{n}\frac{|(\bar{Y}-\mu_{Y,0})|}{\sigma_Y}$
- In practice, is unknown - it must be **estimated**

Computing the p-value with σ_Y estimated:

$$
\begin{aligned}
p-value &= \Pr_{H_0}[|\bar{Y}-\mu_{Y,0}| > |\bar{Y}^{act}-\mu_{Y,0}|] \\
&= \Pr_{H_0}[\frac{|\bar{Y}-\mu_{Y,0}|}{s_Y/\sqrt{n}} > \frac{|\bar{Y}^{act}-\mu_{Y,0}|}{s_Y/\sqrt{n}}] \\
&\approx \text{probability under left+right N(0,1) density}
\end{aligned}
$$

so,

$$
\begin{aligned}
p-value &= \Pr_{H_0}[|t| > |t^{act}|] \qquad (\sigma_Y \text{ estimated}) \\
&= \text{probability under normal N(0,1) tails outside } |t^{act}|
\end{aligned}
$$

where t is the **t-statistic** seen as a random variable.

What is the link between the p-value and the significance level?

Computer programs often communicate the p-value since the p-value contains more information.

For example, if the prespecified significance level is 5%,

- you reject the null hypothesis if $|t| > 1.96$
- equivalently, you reject H_0 if $p - value < 0.05$.

In general:

If $p - value < (\alpha \times 100)\%$ we reject the null hypothesis at $(\alpha \times 100)\%$.

t-test and t-table

At this point, you might be wondering,...

What happened to the t-table and the degrees of freedom?

Digression: the Student t distribution

If Y_i, $i = 1, \ldots, n$ is i.i.d. $N(\mu_Y, \sigma_Y^2)$, then the t-statistic has the Student t-distribution with n-1 degrees of freedom.

- The critical values of the Student t-distribution is tabulated in the back of all statistics books.

Comments on the Student t-distribution

1. The theory of the t-distribution was one of the early triumphs of mathematical statistics. It is astounding, really: if Y is i.i.d. normal, then you can know the exact, finite-sample distribution of the t-statistic - it is the Student t. So, you can construct confidence intervals (using the Student t critical value) that have exactly the right coverage rate, no matter what the sample size. This result was really useful in times when "computer" was a job title, data collection was expensive, and the number of observations was perhaps a dozen. It is also a conceptually beautiful result, and the math is beautiful too - which is probably why stats profs love to teach the t-distribution. But....

Comments on Student t distribution, ctd.

1. If the sample size is moderate (several dozen) or large (hundreds or more), the difference between the t-distribution and N(0,1) critical values are negligible. Here are some 5% critical values for 2-sided tests:

degrees of freedom (n - 1)	5% t-distribution critical value
10	2.23
20	2.09
30	2.04
60	2.00
∞	1.96

Comments on Student t distribution, ctd.

1. So, the Student-t distribution is only relevant when the sample size is very small; but in that case, for it to be correct, you must be sure that the population distribution of Y is normal. In economic data, the normality assumption is rarely credible. Here are the distributions of some economic data.
2. Do you think earnings are normally distributed?
3. Suppose you have a sample of $n = 10$ observations from one of these distributions — would you feel comfortable using the Student t distribution?

Comments on Student t distribution, ctd.

1. You might not know this. Consider the t-statistic testing the hypothesis that two means (groups s, l) are equal:

$$t = \frac{\bar{Y}_s - \bar{Y}_l}{\sqrt{\frac{s_s^2}{n_s} + \frac{s_l^2}{n_l}}} = \frac{\bar{Y}_s - \bar{Y}_l}{SE(\bar{Y}_s - \bar{Y}_l)}$$

   - Even if the population distribution of Y in the two groups is normal, this statistic doesn't have a Student t distribution!
   - There is a statistic testing this hypothesis that has a normal distribution, the "pooled variance" t-statistic - see SW (Section 3.6) - however the pooled variance t-statistic is only valid if the variances of the normal distributions are the same in the two groups.
   - Would you expect this to be true, say, for men's v. women's wages?

The Student-t distribution - summary

- The assumption that Y is distributed $N(\mu_Y, \sigma_Y^2)$ is rarely plausible in practice (income? number of children?)
- For n ¿ 30, the t-distribution and $N(0,1)$ are very close (as n grows large, the $t(n-1)$ distribution converges to $N(0,1)$)
- The t-distribution is an artifact from days when sample sizes were small and "computers" were people
- For historical reasons, statistical software typically uses the t-distribution to compute p-values - but this is irrelevant when the sample size is moderate or large.
- For these reasons, in this class we will focus on the large-n approximation given by the CLT

Where are we?

1. The probability framework for statistical inference
2. Estimation
3. Testing
4. **Confidence intervals**

Confidence Intervals

Definition

- A 95% confidence interval for μ_Y is an interval that contains the true value of $\bar{Y}$ in 95% of repeated samples.

Confidence Intervals

Definition

- A 95% confidence interval for μ_Y is an interval that contains the true value of $\bar{Y}$ in 95% of repeated samples.
- In general, a $(\alpha \times 100)\%$ confidence interval for μ_Y is an interval that contains the true value of $\bar{Y}$ in $(\alpha \times 100)\%$ of repeated samples.

Confidence intervals, ctd.

Digression:

- What is random here? The values of $Y_1, \ldots, Y_n$ and thus any functions of them - including the confidence interval.

Confidence intervals, ctd.

Digression:

- What is random here? The values of $Y_1, \ldots, Y_n$ and thus any functions of them - including the confidence interval.
- The confidence interval it will differ from one sample to the next.

Confidence intervals, ctd.

Digression:

- What is random here? The values of $Y_1, \ldots, Y_n$ and thus any functions of them - including the confidence interval.
- The confidence interval it will differ from one sample to the next.
- The population parameter, μ_Y, is not random, we just don't know it.

Confidence intervals, ctd.

A **95% confidence interval** has the followig form:

$$\left\{\mu_Y : \left|\frac{Y - \mu_Y}{s_Y/\sqrt{n}}\right| > 1.96\right\} = \left\{\mu_Y \in \left(\bar{Y} - 1.96 \times \frac{s_Y}{\sqrt{n}}, \bar{Y} + 1.96 \times \frac{s_Y}{\sqrt{n}},\right)\right\}$$

This confidence interval relies on the large-n results that $\bar{Y}$ is **approximately normally** distributed and $\sigma_Y \xrightarrow{p} \sigma_Y$.

Confidence intervals, ctd.

A **90% confidence interval** has the followig form:

$$\left\{\mu_Y : \left|\frac{Y - \mu_Y}{s_Y/\sqrt{n}}\right| > 1.64\right\} = \left\{\mu_Y \in \left(\bar{Y} - 1.64 \times \frac{s_Y}{\sqrt{n}}, \bar{Y} + 1.64 \times \frac{s_Y}{\sqrt{n}},\right)\right\}$$

Summary:

From the two assumptions of:

1. simple random sampling of a population, that is, $\{Y_i, i = 1, \ldots, n\}$ are i.i.d.
2. $0 < E(Y^2) < \infty$

we developed, for large samples (large n):

- Theory of estimation (sampling distribution of)
- Theory of hypothesis testing (large-n distribution of t-statistic and computation of the p-value)
- Theory of confidence intervals (constructed by inverting test statistic)

Are assumptions 1. & 2. plausible in practice? Yes