Lecture II: Review of Regression and Inference

Stanislao Maldonado

Universidad del Rosario
stanislao.maldonado@urosario.edu.co

Impact Evaluation
Universidad del Rosario
January 31th, 2017

1. Motivation

- Standard introductions to econometrics are based on the classical linear regression model
- For instance, Greene (2012) introduces the classical regression model as follows:

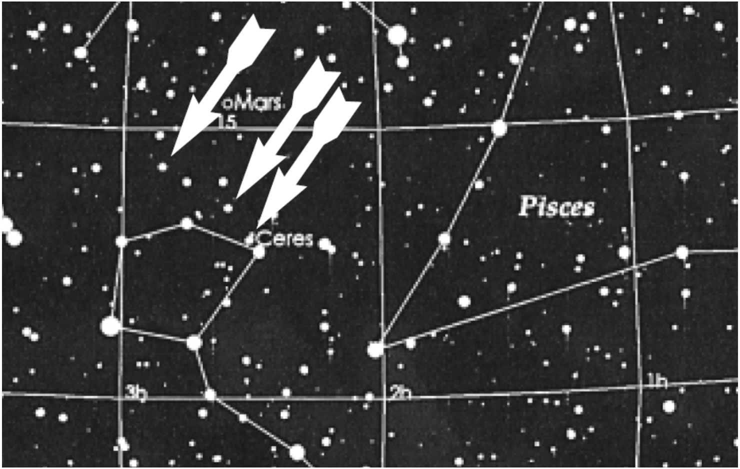
Assumptions of the Linear Regression Model (Greene 2012)

- A.1. Linearity: $y = X\beta + \epsilon$
- A.2. Full rank: X is an n x K matrix with rank K
- A.3. Exogeneity: $\mathbb{E}[\epsilon|X] = 0$
- A.4. Homoscedasticity and nonautocorrelation: $\mathbb{E}[\epsilon\epsilon^{'}|X] = \sigma^2 I$
- A.5. Data generation: X may be fixed or random.
- A.6. Normal distribution: $\epsilon|X \sim N[0, \sigma^2 I]$

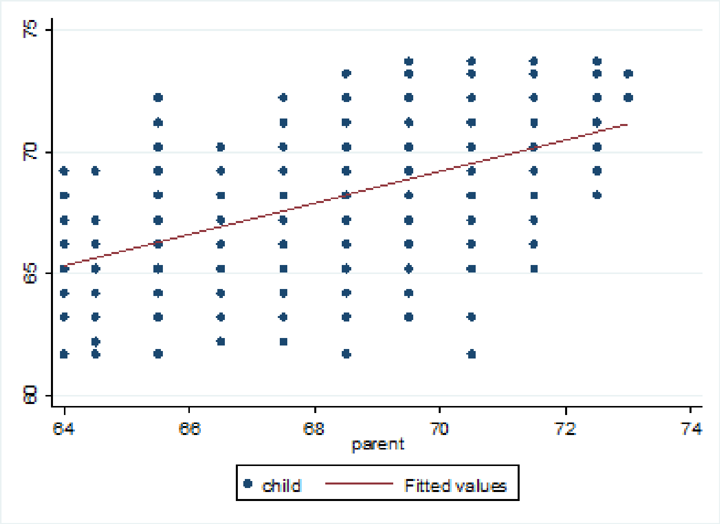
- These introductions to regression implicitly assume a structural approach

- However, regression was originally developed without any causal pretension:
  - Least squares appeared for the first time published in 1805 in a paper by the mathematician Adrien Legendre (1752-1833)
  - Friedrich Gauss (1777-1855) used regression to predict the orbit of the asteroid Ceres
- Francis Galton (1822-1911) was known for his innovative work on heredity and famous for coining the term “regression”:
  - Question: relationship between children body heights and the average body height of both parents
  - Result: approximate linear trend of 2/3 implying that children of tall (short) parents are usually shorter (taller)
  - Tendency towards the population average (*regression* towards the mean)

Regression for predicting the orbit of Ceres



Galton and regression



- Regression can be just seen as a data reduction tool without any causal connotation
- Most of the textbook assumptions are not needed if we are willing to interpret regression coefficients without any causal flavor:

  > *"Such requirements are very stringent, and have a causal flavor that is not part of the explicit specification of classic regression (CR) model. An alternative position is less stringent and free of causal language. Nothing in the CR model itself requires an exhaustive list of explanatory variables nor any assumption about the direction of causality. We have in mind a joint probability distribution, in which any conditional expectation function is conceivably of interest. . . It may be true that causal relations are the most interesting ones, but that is a matter of economics rather than of statistics."* (Goldberger 1991, page 173).

- Lack of clarity about this is a source of confusion among students and scholars
- We will discuss in this class regression as a statistical tool to approximate the **conditional expectation function (CEF)** without causal implications

2. Agnostic Regression

- Another look at regression: regression as an statistical rather than econometric tool
- Conditions needed to run a regression are fairly simple provided you interpret the results appropriately
- We will see that most of the assumptions you were told to believe in order to run a regression are not needed (and there is nothing wrong with that!)
- Example: schooling and earnings
    - On average, people with more schooling tend to earn more than people with less schooling
    - Education predicts earnings in a narrow statistical sense

2.1 Conditional Expectation Function (CEF)

- We are interested in the relationship between the dependent variable y and the explanatory variables X
- Some reasons:
  - **Description**: how is the observed relationship between y and X?
  - **Prediction**: can we use X to create a good predictor of y?
  - **Causality**: what happens to y if we experimentally manipulate X?
- If we are not interested in causality, we may be interested in studying the expected value of y conditional on X; $\mathbb{E}(y/x)$
- This relationship is given by the **CEF**

Example: Labor income distribution in Argentina

- Consider the distribution of labor income in Argentina in 2006
- We can describe labor income using a probability distribution:

$$F(u) = Pr(income \leq u) \tag{1}$$

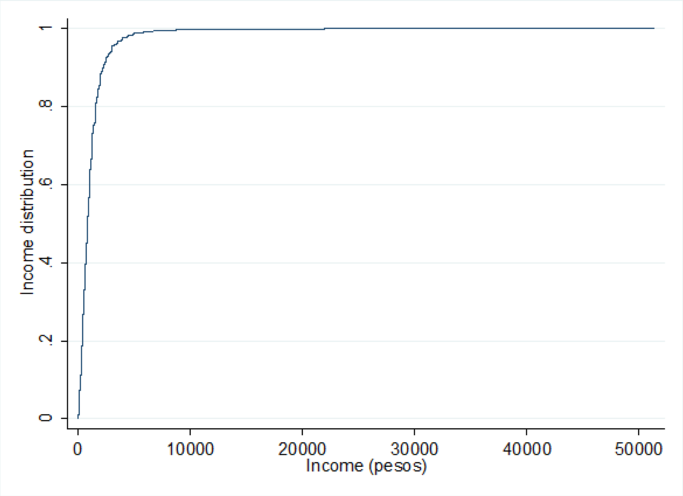
- Assuming that F is differentiable, the probability density function (PDF) is:

$$f(u) = \frac{dF(u)}{du} \tag{2}$$

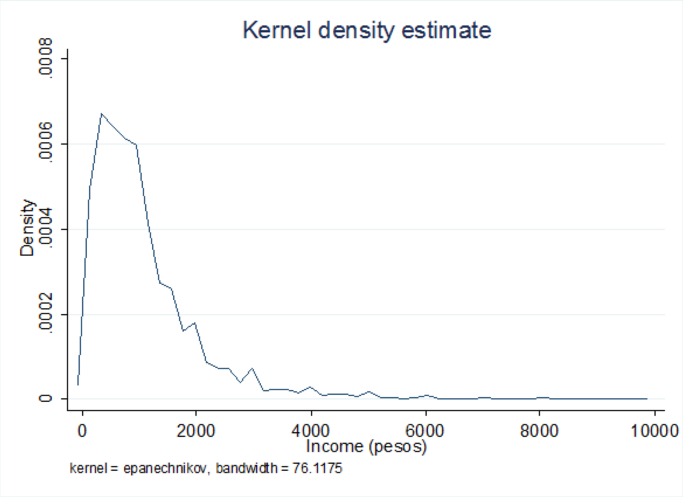
- The expectation of a random variable y with density f is:

$$\mu = \mathbb{E}(y) = \int uf(u)\, du \tag{3}$$

- The empirical labor income distribution in 2006 (Encuesta Permanente de Hogares):



- The (kernel) density function:



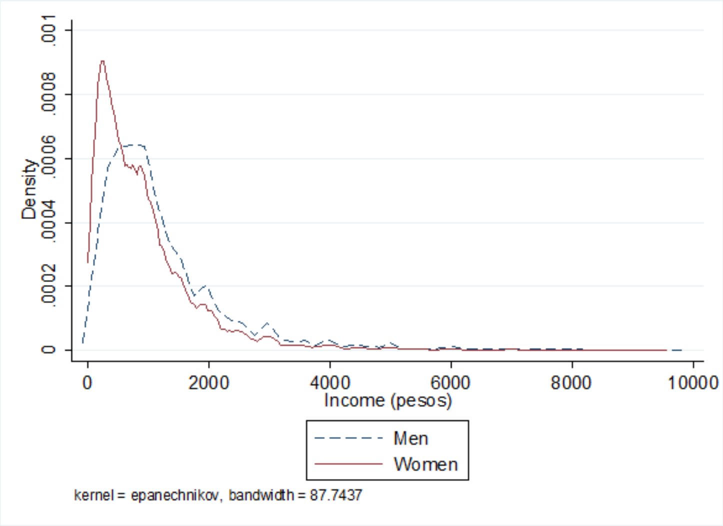
- The average labor income: 1,116 pesos
- Is labor income distribution the same for all type of workers? does labor income differ across subpopulations?
- The average labor income is higher for males (1,257 pesos) than females (923 pesos)
- These means are known as **conditional means** or **conditional expectations**. The specific values are:

$$\mathbb{E}(income|gender = men) = 1,257$$
$$\mathbb{E}(income|gender = women) = 923$$

- Typically, the difference of conditional means is usually of interest in empirical work:

$$\mathbb{E}(income|gender = men) - \mathbb{E}(income|gender = women) = 334$$

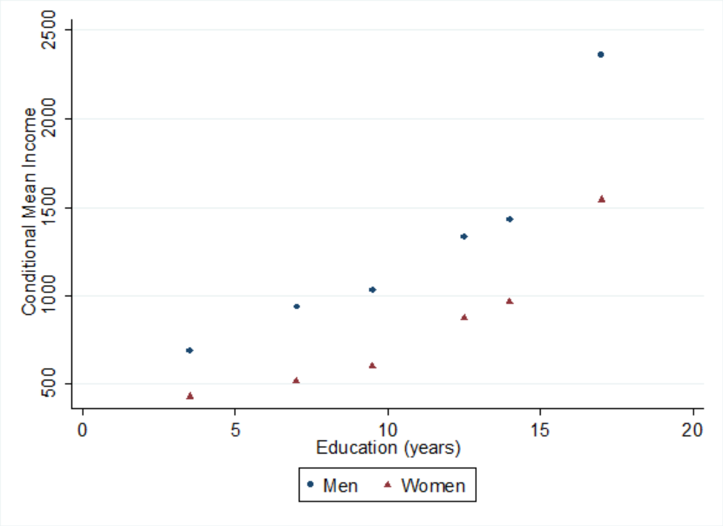
- The (kernel) density function by gender:



- Focusing on conditional means reduces complex distributions into single summary measures that facilitates comparisons across groups
- Extending the analysis to incorporate more dimensions (e.g. education) leads to the CEF
- The conditional expectation would be the following:

  $$\mathbb{E}(income|gender = men, education = 17) = 2,364$$

- More generally, we can display the conditional means by gender as a function of education:



- Formally, the CEF can be defined as follows:

$$\mathbb{E}(y|x_1, x_2, ..., x_k) = m(x_1, x_2, ..., x_k) \tag{4}$$

- Alternatively:

$$\mathbb{E}(y|\mathbf{x}) = m(\mathbf{x}) \tag{5}$$

- Before continuing, recall:

Law of Iterated Expectations

- **Simple Law of Iterated Expectations**:
  If $\mathbb{E}(y) < \infty$, then for any random vector $\mathbf{x}$, $\mathbb{E}(\mathbb{E}(y|\mathbf{x})) = \mathbb{E}(y)$.
- **Law of Iterated Expectations**:
  If $\mathbb{E}(y) < \infty$, then for any random vectors x_1 and x_2,
  $\mathbb{E}(\mathbb{E}(y|x_1, x_2)|x_1) = \mathbb{E}(y|x_1)$.

2.2 CEF Error

- We can define the **CEF Error** as follows:

$$e = y - m(\mathbf{x}) \tag{6}$$

- By construction:

$$y = m(\mathbf{x}) + e \tag{7}$$

- Notice that e is derived from the joint distribution of $(y, \mathbf{x})$
- The CEF error has a zero conditional mean. To see that, we need another key theorem:

Conditioning Theorem

If $\mathbb{E}|g(\mathbf{x})y| < \infty$, then
$\mathbb{E}(g(\mathbf{x})y|\mathbf{x}) = g(\mathbf{x})\mathbb{E}(y|\mathbf{x})$ and $\mathbb{E}(g(\mathbf{x})y) = \mathbb{E}(g(\mathbf{x})\mathbb{E}(y|\mathbf{x}))$.

The conditional mean of the CEF Error is zero: $\mathbb{E}(e|\mathbf{x}) = 0$

$$
\begin{aligned}
\mathbb{E}(e|\mathbf{x}) &= \mathbb{E}((y - m(\mathbf{x}))|\mathbf{x}) \\
&= \mathbb{E}(y|\mathbf{x}) - \mathbb{E}(m(\mathbf{x})|\mathbf{x}) \\
&= m(\mathbf{x}) - m(\mathbf{x}) \\
&= 0
\end{aligned}
$$

- Using the Law of Iterated Expectations (LIE) it can be also shown that the unconditional mean is also zero:

The unconditional mean of the CEF Error is zero: $\mathbb{E}(e) = 0$

$$\mathbb{E}(e) = \mathbb{E}(\mathbb{E}(e|\mathbf{x})) = \mathbb{E}(0) = 0$$

- Key to understand that both results are true by definition
- Alternatively, y can be written in the following way:

$$y = \mathbb{E}(y|\mathbf{x}) + e \tag{8}$$

- This leads to an alternative derivation of the zero conditional mean property:

CEF Decomposition Property (Angrist and Pischke 2009)

$$y = \mathbb{E}(y|\mathbf{x}) + e,$$

Where:
(i) e is mean-independent of $\mathbf{x}$, i.e., $\mathbb{E}(e|\mathbf{x}) = 0$
(ii) e is uncorrelated with any function of $\mathbf{x}$.

- (i) is similar to the proof above
- (ii) is based on the LIE and by the mean independence as follows:

Proof of second part of the CEF decomposition property

Let $h(\mathbf{x})$ be any function of $\mathbf{x}$. Then:

$$\begin{aligned}\mathbb{E}(h(\mathbf{x})e) &= \mathbb{E}(h(\mathbf{x})\mathbb{E}(e|\mathbf{x})) \\ &= \mathbb{E}(h(\mathbf{x})0) \\ &= 0.\end{aligned}$$

- Key: any random variable can be decomposed into a fraction explained by $\mathbf{x}$ (The CEF) and a fraction left over which is uncorrelated with any function of $\mathbf{x}$.

2.3 CEF as the best predictor

- CEF is a good summary of the relationship between y and $\mathbf{x}$ but it can also play a predictive role
- Assume we want to predict y. Any predictor can be written as $g(\mathbf{x})$. Then, the prediction error is: $y - g(\mathbf{x})$
- The prediction error is the square of its expectation: $\mathbb{E}(y - g(\mathbf{x}))^2$
- The best predictor is defined as the function $g(\mathbf{x})$ that minimizes $\mathbb{E}(y - g(\mathbf{x}))^2$. We will show that the best predictor is the CEF: $m(\mathbf{x})$

CEF Prediction Property (Angrist and Pischke 2009)

Let $m(\mathbf{x})$ be any function of $\mathbf{x}$. The CEF solves $\mathbb{E}(y|\mathbf{x}) =$ ArgMin $\mathbb{E}[(y - m(\mathbf{x}))^2]$, so it is the Minimum Mean Squared Error predictor of y given $\mathbf{x}$.

Proof of the CEF Prediction Property

$$
\begin{aligned}
\mathbb{E}(y - g(\mathbf{x}))^2 &= \mathbb{E}(e + m(\mathbf{x}) - g(\mathbf{x}))^2 \\
&= \mathbb{E}e^2 + 2\mathbb{E}(e(m(\mathbf{x}) - g(\mathbf{x})))^2 + \mathbb{E}(m(\mathbf{x}) - g(\mathbf{x}))^2 \\
&= \mathbb{E}e^2 + \mathbb{E}(m(\mathbf{x}) - g(\mathbf{x}))^2 \\
&>= \mathbb{E}e^2 \\
&= \mathbb{E}(y - m(\mathbf{x}))^2
\end{aligned}
$$

2.4 Linear regression and the CEF

- We know that CEF has nice properties, but what is its connection with the linear regression and why we want to run a linear regression?
- Regression is closely linked with the CEF and the CEF provides a natural summary of empirical relationships
- We are concerned with the vector of **population regression coefficients** (solution to the least squares problem)
- There are three reasons why the population regression coefficients are of interest:
  - If the CEF is linear, the population regression function is it
  - The population regression function is the best linear predictor of y
  - Even if the CEF is non-linear, regression provides the best linear approximation to it

A. Linear CEF

- Important case is when CEF is linear:

$$\begin{aligned} m(\mathbf{x}) &= x_1\beta_1 + x_2\beta_2 + ... + x_{k-1}\beta_{k-1} + \beta_k \\ &= \mathbf{x}'\beta \end{aligned}$$

- This is usually known as the **linear regression model**. One appealing feature its the interpretation of the coefficients as marginal changes
- We have the following theorem:

The Linear CEF Theorem (Angrist and Pischke 2009)

Suppose the CEF is linear, then the population regression function is it.

Proof of the Linear CEF Theorem

Suppose $\mathbb{E}(y|\mathbf{x}) = \mathbf{x}'\beta^*$ for a kx1 vector of coefficients, β^*. Recall that $\mathbb{E}(\mathbf{x}(y - \mathbb{E}(y|\mathbf{x}))) = 0$ by the CEF decomposition property. Substitute using $\mathbb{E}(y|\mathbf{x}) = \mathbf{x}'\beta^*$ to find that $\beta^* = (\mathbb{E}(\mathbf{x}\mathbf{x}'))^{-1}\mathbb{E}(\mathbf{x}y) = \beta$

- Under what circumstances a CEF is linear?
  - Joint normality
  - Saturated model

B. Best Linear Predictor

- The linear CEF model is empirically unlikely to be accurate unless $\mathbf{x}$ is discrete and low dimensional. It is more realistic to view the linear specification as an approximation
- $m(\mathbf{x})$ is the best predictor of y in MMSE sense. By extension, an approximation to the CEF can be defined by the linear function with the lowest mean squared error among all linear predictors
- Some regularity conditions are required:
  - $\mathbb{E}y^2 < \infty$
  - $\mathbb{E}||\mathbf{x}||^2 < \infty$
  - $\mathbb{E}(\mathbf{x}\mathbf{x}')$ is positive definite

- A linear predictor of y given $\mathbf{x}$ is a function of the form $\mathbf{x}'\beta$ for some $\beta \in R^k$. The mean squared prediction error is:

$$S(\beta) = \mathbb{E}(y - \mathbf{x}'\beta)^2 \tag{9}$$

- The best linear predictor of y given $\mathbf{x}$, written $P(y|\mathbf{x})$, is found by selecting the vector β to minimize $S(\beta)$

Best Linear Predictor

The best linear predictor of y given $\mathbf{x}$ is $P(y|\mathbf{x}) = \mathbf{x}'\beta$, where β minimizes the mean squared prediction error $S(\beta) = \mathbb{E}(y - \mathbf{x}'\beta)^2$. The minimizer $\beta = \text{ArgMin } S(\beta) = (\mathbb{E}(\mathbf{x}\mathbf{x}'))^{-1}\mathbb{E}(\mathbf{x}y)$ is called the **Linear Projection Coefficient**

- Notice that the minimizer solves the least squares problem from your basic econometric class

- The **best linear predictor** or **linear projection** of y on $\mathbf{x}$:

$$P(y|\mathbf{x}) = \mathbf{x}^{'}(\mathbb{E}(\mathbf{x}\mathbf{x}^{'}))^{-1}\mathbb{E}(\mathbf{x}y) \tag{10}$$

- The **projection error** is:

$$e = y - \mathbf{x}^{'}\beta \tag{11}$$

- This equals the error from the regression equation when the conditional mean is linear in $\mathbf{x}$
- We can re-write the previous equation to obtain a decomposition of y into linear predictor and error:

$$y = \mathbf{x}^{'}\beta + e \tag{12}$$

- $\mathbf{x}'\beta$ is usually called the **best predictor** of y given $\mathbf{x}$
- An important property of the prediction error is the following:

$$\mathbb{E}(\mathbf{x}e) = 0 \tag{13}$$

Proof of $\mathbb{E}(\mathbf{x}e) = 0$

$$\begin{aligned}
\mathbb{E}(\mathbf{x}e) &= \mathbb{E}(\mathbf{x}(y - \mathbf{x}'\beta)) \\
&= \mathbb{E}(\mathbf{x}y) - \mathbb{E}(\mathbf{x}\mathbf{x}')(\mathbb{E}(\mathbf{x}\mathbf{x}'))^{-1}\mathbb{E}(\mathbf{x}y) \\
&= \mathbf{0}.
\end{aligned}$$

- We can state the following theorem summarizing these results:

The Best Linear Predictor Theorem (Angrist and Pischke 2009)

The function $\mathbf{x}'\beta$ is the best linear predictor of y given $\mathbf{x}$ in a MMSE sense.
Proof: $\beta = (\mathbb{E}(\mathbf{x}\mathbf{x}'))^{-1}\mathbb{E}(\mathbf{x}y)$ solves the least squares problem.

- The linear model in (12) exists quite generally based on fair assumptions. However, it cannot be interpreted as a conditional mean or a parameter of a causal economic model

C. Best Linear Approximation

- Although prediction has some value in economic research, economists are usually more interested in the distribution of y instead of predicting a specific value of y
- How to use a regression to approximate $m(\mathbf{x})$?
- We define the mean-square approximation error of $\mathbf{x}'\beta$ to $m(\mathbf{x})$ as follows:

$$d(\beta) = \mathbb{E}(m(\mathbf{x}) - \mathbf{x}'\beta)^2 \quad (14)$$

- We can define the best linear approximation to the conditional $m(\mathbf{x})$ as the function $\mathbf{x}'\beta$ obtained by selecting β to minimize $d(\beta)$.

- Following previous results, we derive the following solution:

$$\beta = \textbf{argmin } d(\beta) = (\mathbb{E}(\mathbf{x}\mathbf{x}'))^{-1}\mathbb{E}(\mathbf{x}y) \tag{15}$$

The Regression-CEF Theorem (Angrist and Pischke 2009)

The function $\mathbf{x}'\beta$ provides the MMSE linear approximation to $m(\mathbf{x})$; that is,

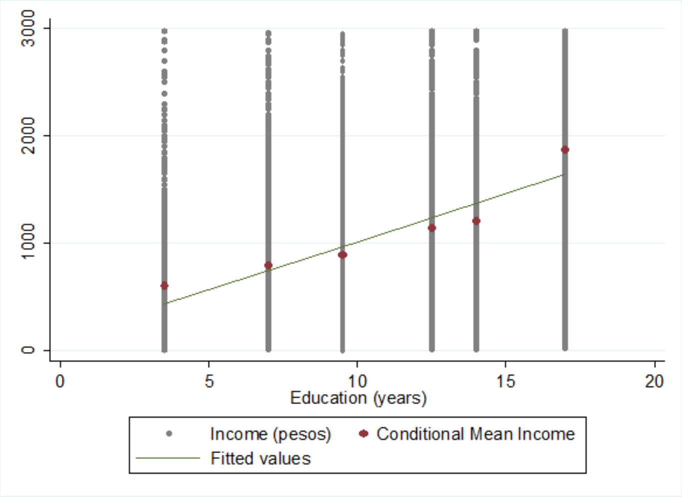
$$\beta = \textbf{argmin } \mathbb{E}(m(\mathbf{x}) - \mathbf{x}'\beta)^2$$

Proof: Solution is similar to the previous theorem and has the same structure as the population least square problem

- Regression provides the best linear approximation to the CEF even when the CEF is non-linear

- Notice that these results depends almost on nothing!
  - Whether your data is i.i.d.
  - Whether your explanatory variables are fixed or random
  - Whether your regressors are correlated with the CEF residuals
  - Whether the CEF is linear or not
  - Whether your dependent variable is continuous, discrete, etc.

- Let's return to the Argentinean example: The regression-CEF approximation for labor income and education

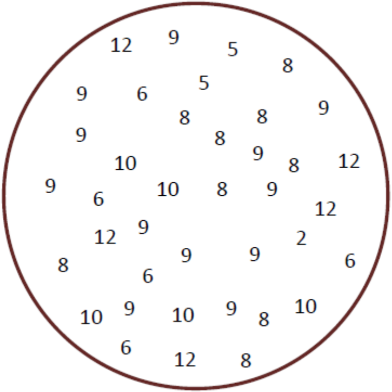


3. Inference

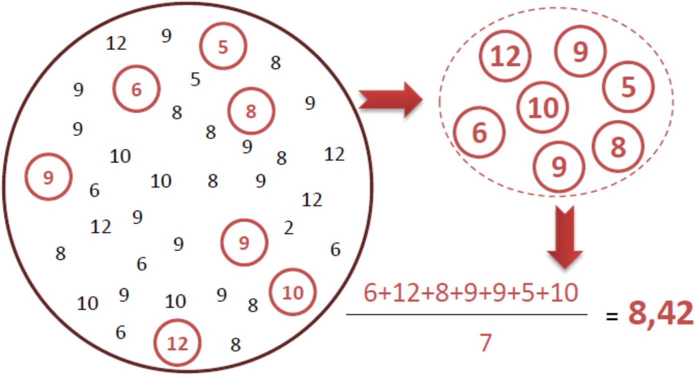
- Evaluating the impact of an intervention is typically based on samples from treatment and control groups
- We will offer a review the problem of inference that emphasize the logic behind the standard large sample approach that you should be already familiar with
- The key messages are the following
  - Distinguish sampling variability from true impacts is the business of inference
  - Statistical theory come to rescue to establish conditions to deal with sampling variability based on the use of statistical tests

3.1 Sampling variability

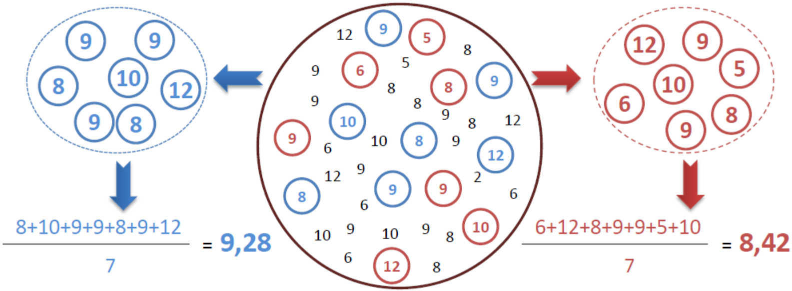
- Assume we are interested in estimating the number of days workers are unemployed in a month. We collect a random sample of individuals:



- The first sample:

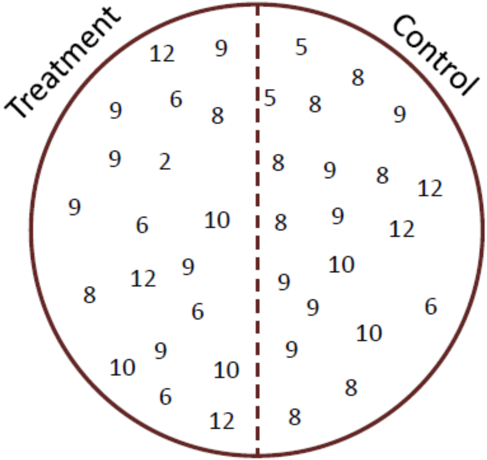


- The second sample:

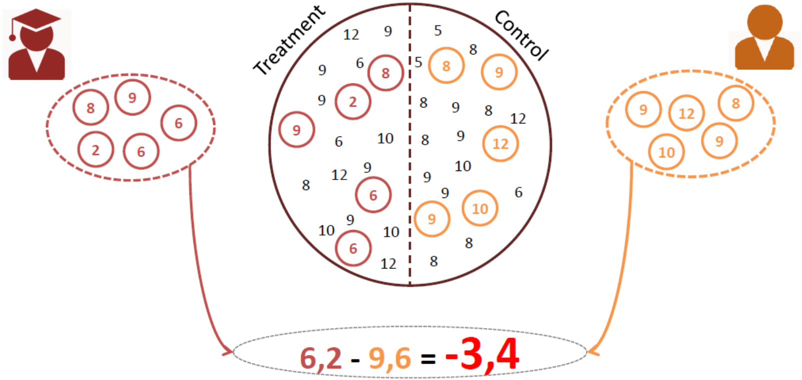


- Different samples from the same population produce different results!
- This phenomenon is known as **sampling variability**
- Consider now the case of an impact evaluation in which a comparison between treatment and control units needs to be made:
  - We want to evaluate the impact of a large training program on the number of days unemployed
  - The program was randomly assigned to individuals
  - We collect two samples from treatment and control groups

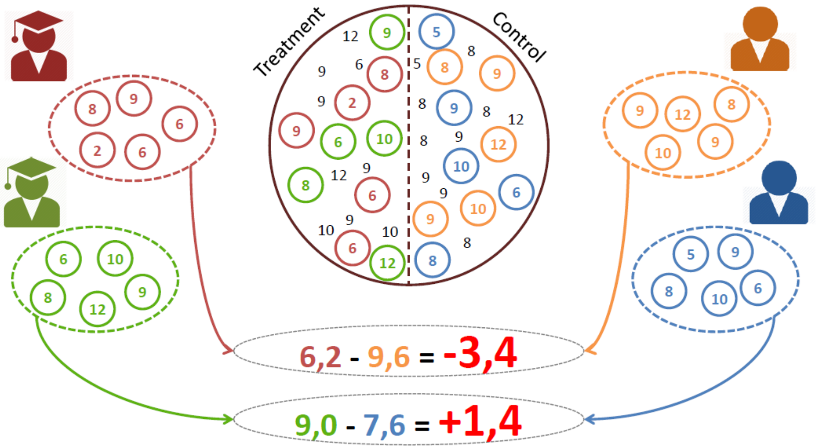
- Population: days of unemployment in a month



- First sample:



- Second sample:



- Evaluating the impact of an intervention is based on a **single sample**, just one from all the samples that can be collected from the same population
- In all cases, analyzing this single sample delivers a particular estimate that may reflect two cases:
  - The result reflects the **true impact** of the intervention
  - The result is due to **sampling variability**
- What can we do to distinguish between these two cases when **only one sample** is available for the analysis?

3.2 Large sample approach to inference

- If we know that sampling variability exists, why do we trust in results coming from a single sample?
- Significant test are used to make a probability argument (probability of observing a given outcome by chance is very small)
- Understanding randomness implies to move away our attention from individual predictions to averages
- Quote from Cramer (1974):
  "*In spite of the irregular behavior of individual results, the average of many results show a striking regularity*"
- Statisticians refer to this property as **statistical regularity**

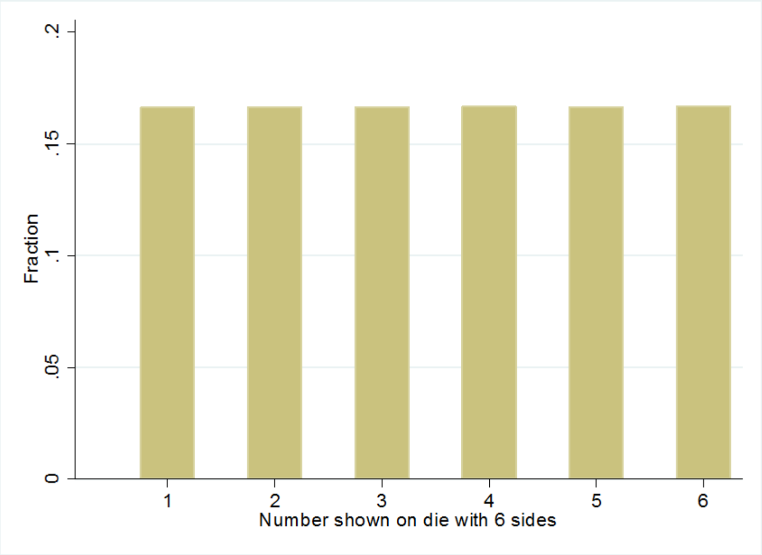
- What's an average? Consider the following experiments:
  - If you were to roll a dice, what would be the expected result (average)?

  - If you were to roll two dices, what would be the expected result (average)?

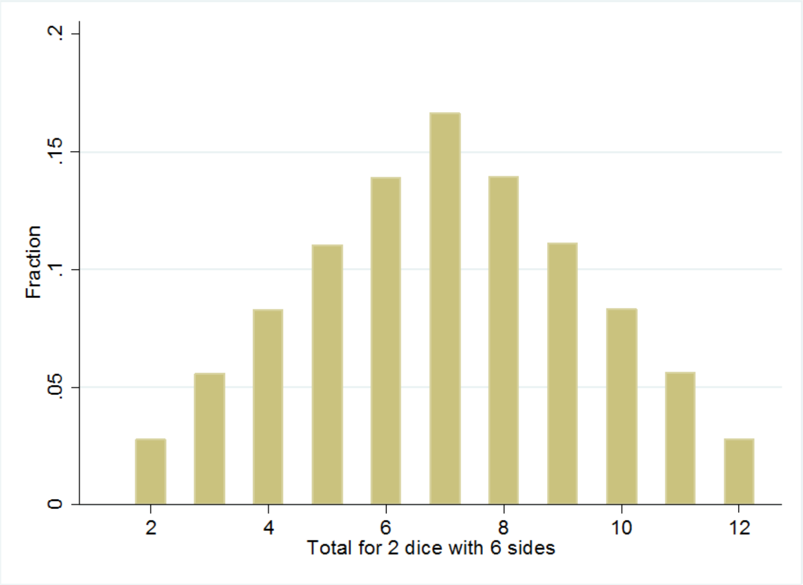
- How this experiment looks like when we increase the number of dice?

One dice

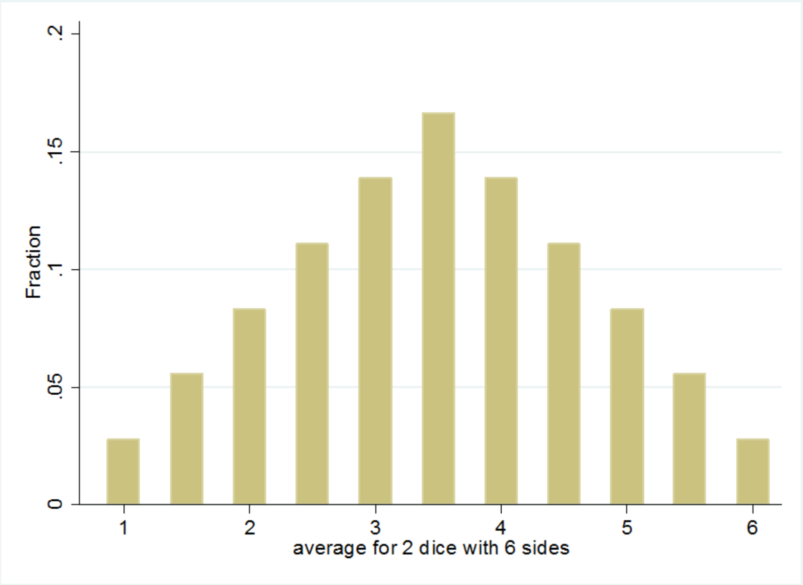


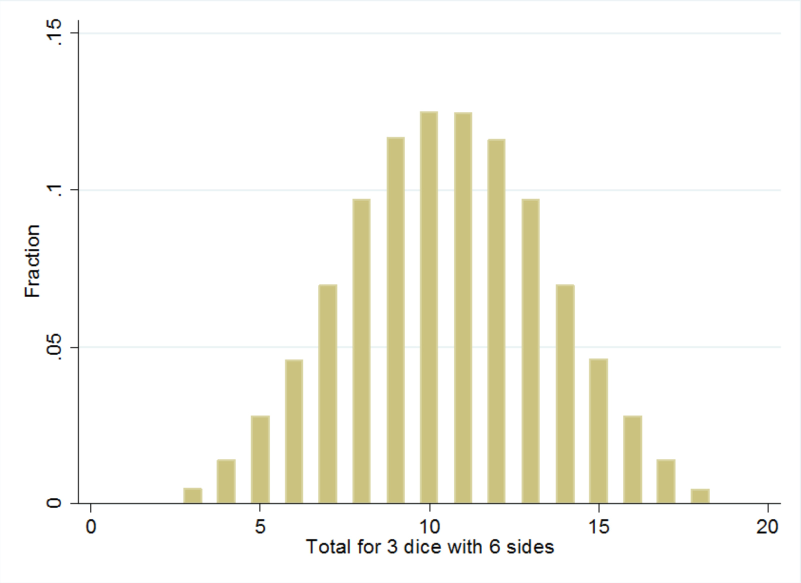
Totals and permutations in 2-dice experiment

		Dice 1					
		⚀	⚁	⚂	⚃	⚄	⚅
Dice 2	⚀	2	3	4	5	6	**7**
	⚁	3	4	5	6	**7**	8
	⚂	4	5	6	**7**	8	9
	⚃	5	6	**7**	8	9	10
	⚄	6	**7**	8	9	10	11
	⚅	**7**	8	9	10	11	12

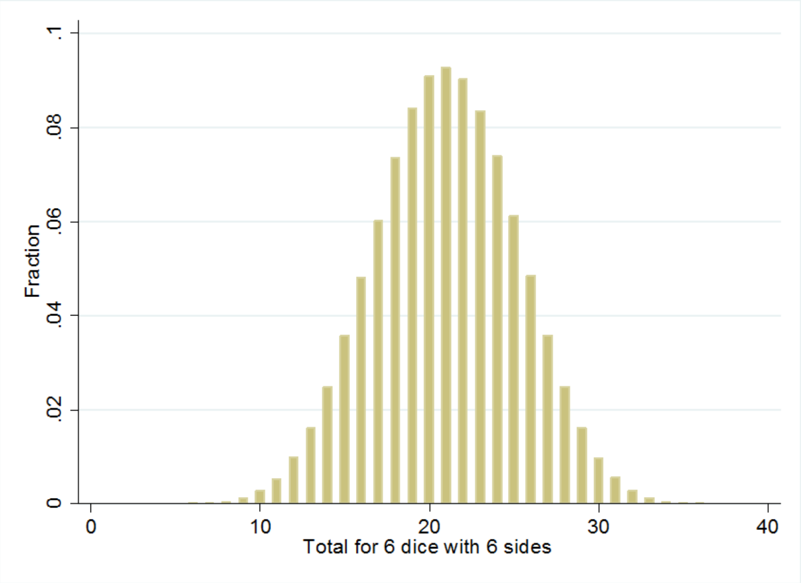


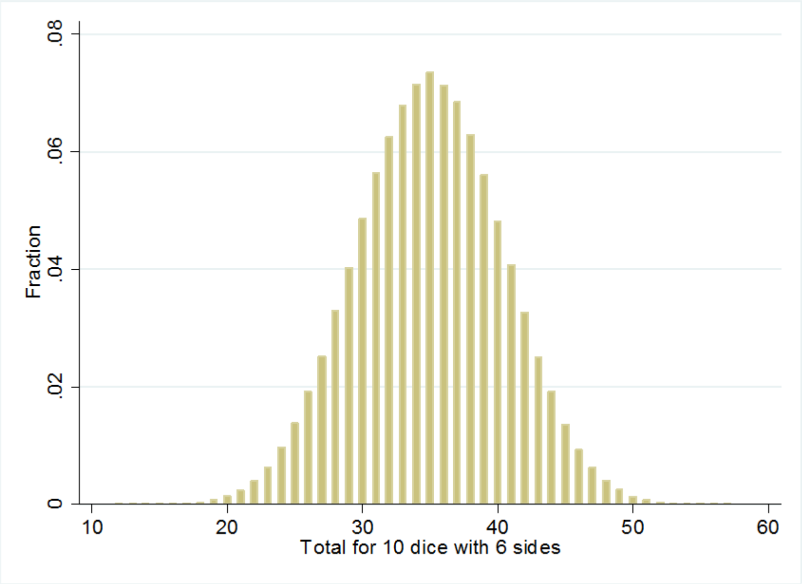
Fraction
0
.05
.1
.15
.2
2
4
6
8
10
12
Total for 2 dice with 6 sides

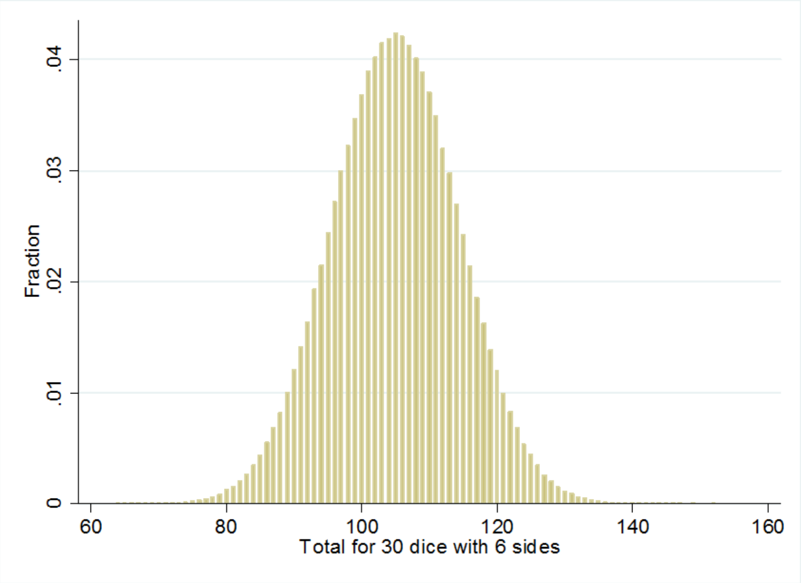

Fraction
0
.05
.1
.15
.2
1
2
3
4
5
6
average for 2 dice with 6 sides

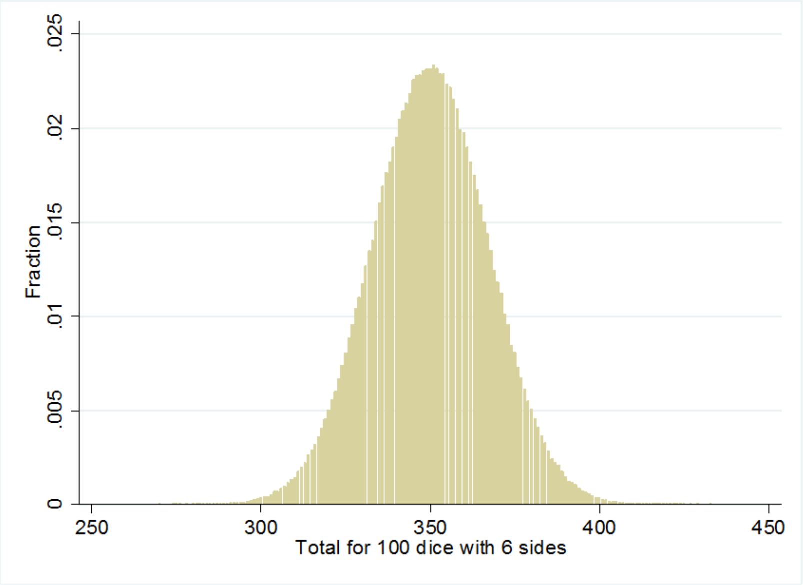


Fraction
0
.05
.1
.15
0
5
10
15
20
Total for 3 dice with 6 sides

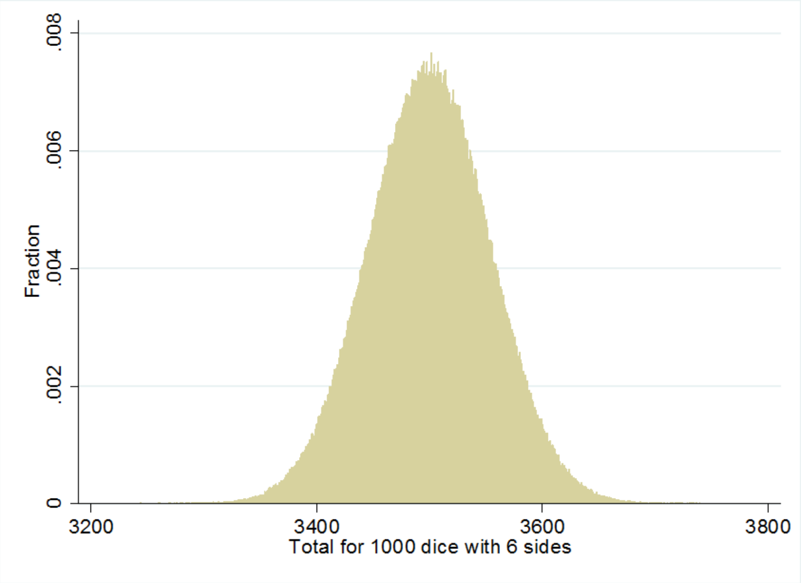

Fraction
0
.02
.04
.06
.08
.1
0
10
20
30
40
Total for 6 dice with 6 sides


Fraction
0
.02
.04
.06
.08
10
20
30
40
50
60
Total for 10 dice with 6 sides


Fraction
0
.01
.02
.03
.04
60
80
100
120
140
160
Total for 30 dice with 6 sides



Fraction
0
.005
.01
.015
.02
.025
250
300
350
400
450
Total for 100 dice with 6 sides


Fraction
0
.002
.004
.006
.008
3200
3400
3600
3800
Total for 1000 dice with 6 sides

Lessons from the rolling experiment

- Sampling (and hypotheses testing) relies in two fundamental properties illustrated by the rolling dice experiment:
  - **The Law of Large Numbers**:
    The more dice you roll, the closer the average to the true average (distribution gets "tighter")
  - **The Central Limit Theorem**:
    The more dice you roll, the more the distribution of possible averages (sampling distribution) looks like a bell curve

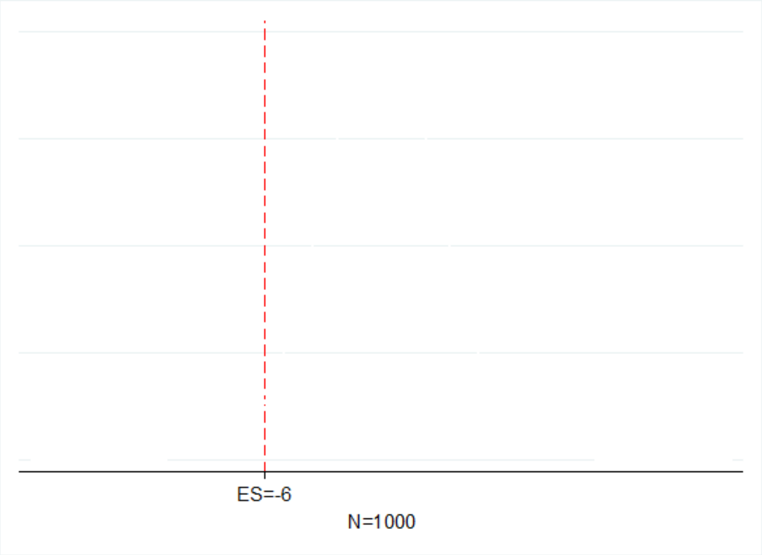
- The property of converging to a specific values comes partly from the fact that the distribution of a larger sample tends to be closer to the distribution of the sampled population
- The property of the "noise" shrinking as the sample size increases is partly arithmetic:

$$\bar{X} = \frac{\sum X}{n} = \frac{\sum_{i=1}^{n-1} X_i}{(n-1)+1} = \frac{5000+10}{1000+1} = 5.0004995$$

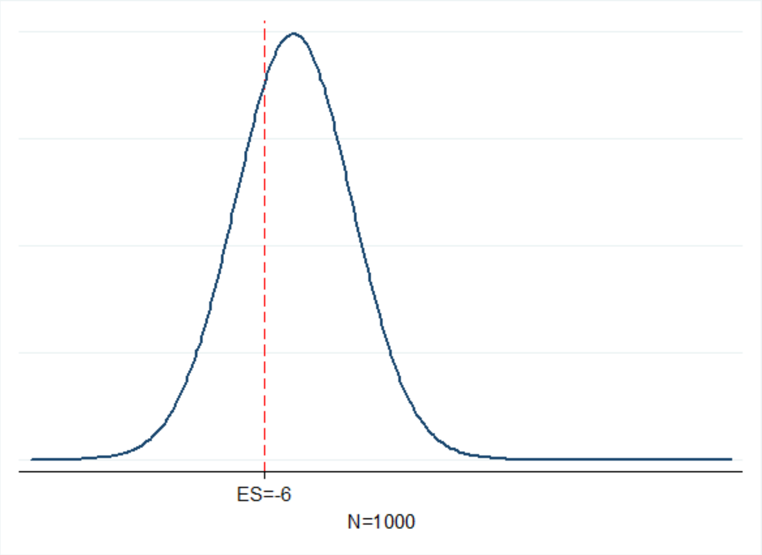
- Key point: the behavior of the mean is predictable!

- Let's return to impact evaluation
- Assume that the sampling estimate of the impact of the training program on unemployment is -6 days
- Due to the law of large numbers and the central limit theorem, what do we know about -6?
    - -6 has a high likelihood of being close to the true impact
    - -6 belongs to a normal distribution

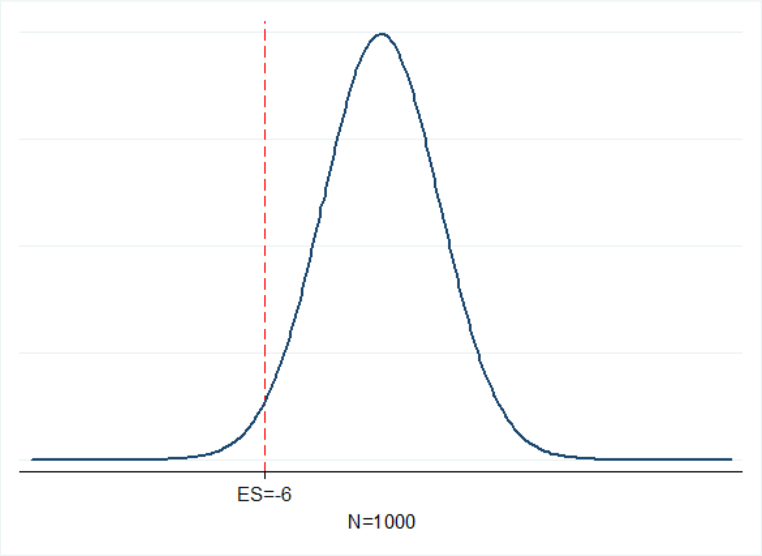
Impact of a training program on unemployment



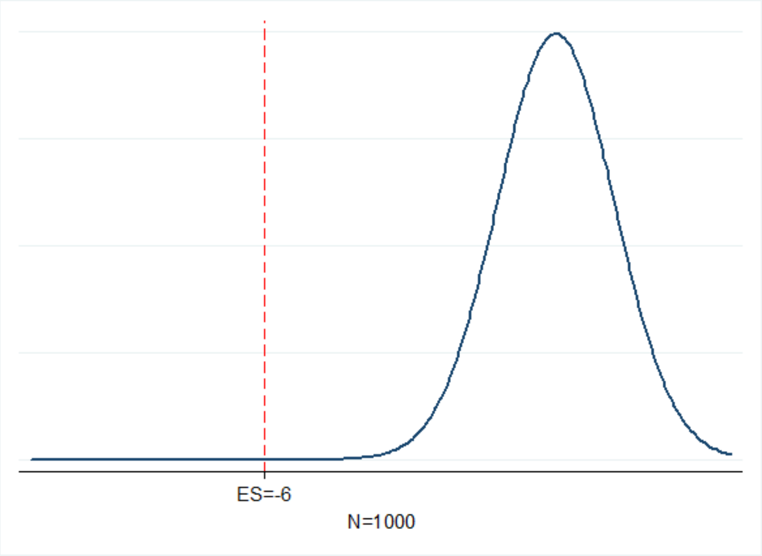
Impact of a training program on unemployment



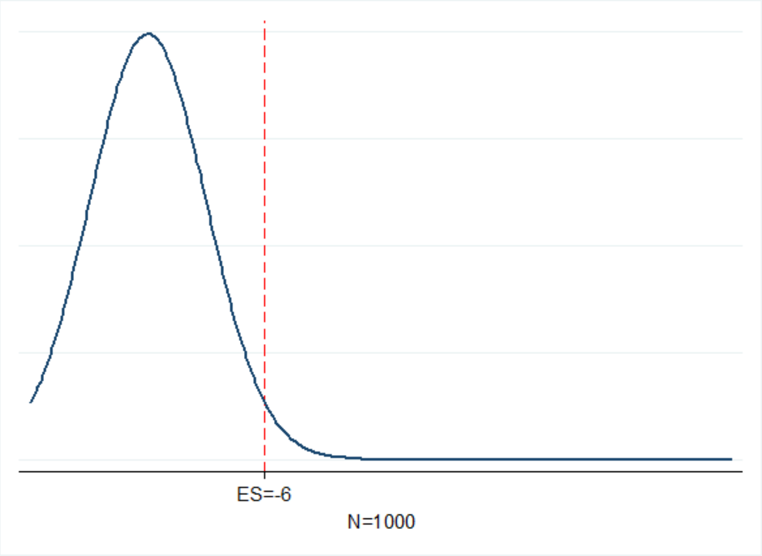
Impact of a training program on unemployment



Impact of a training program on unemployment

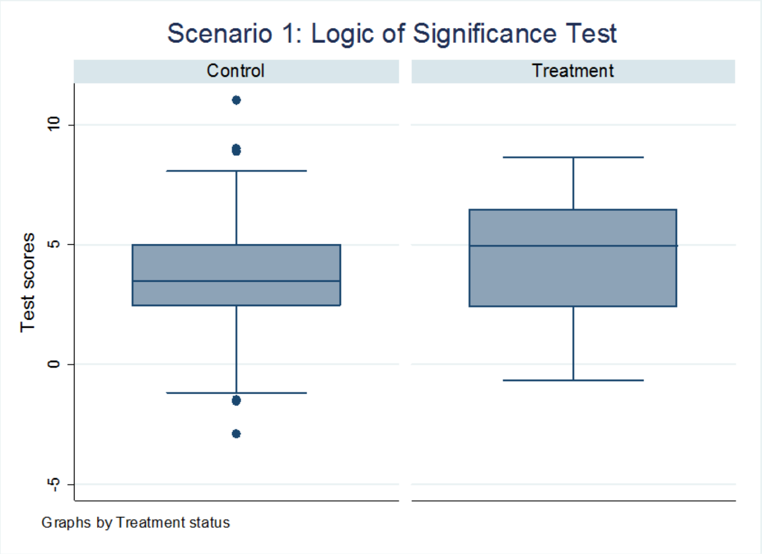


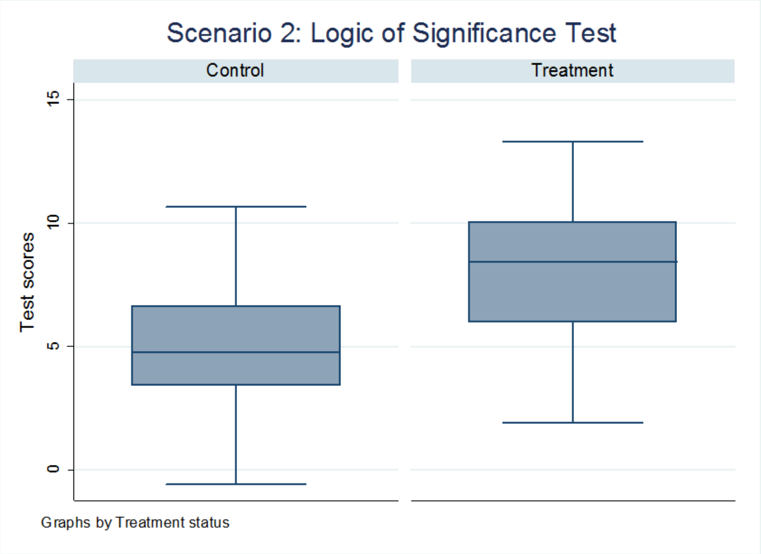
Impact of a training program on unemployment

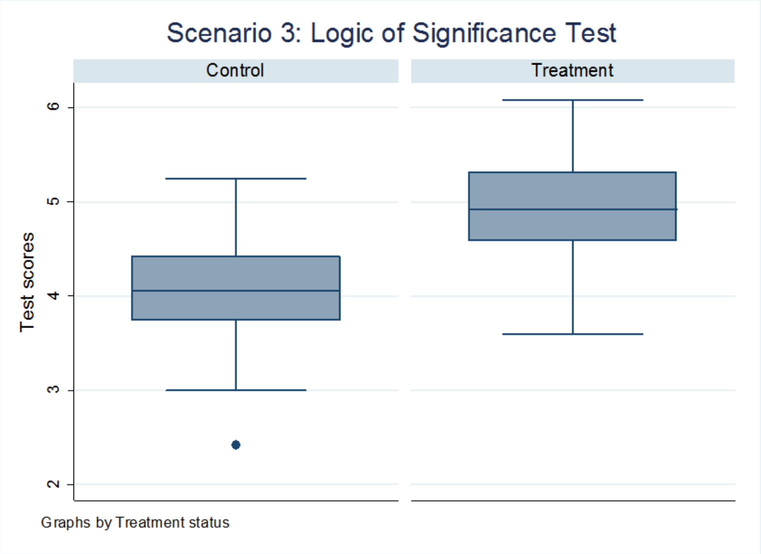


3.3 Logic of significance tests

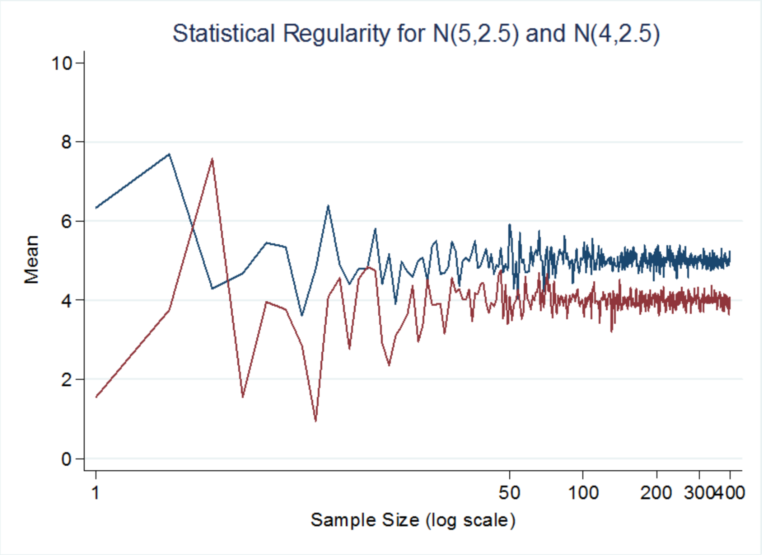
- We are concerned with **sampling variability**
- Significance test calculate the probability that the result was due to sampling variability
- Suppose the following scenarios:
  - **Scenario 1**: control (3.9, 2.9) – treatment (4.0, 2.3)
  - **Scenario 2**: control (4.6, 2.5) – treatment (7.5, 2.9)
  - **Scenario 3**: control (4.0, 0.5) – treatment (5.1, 0.5)
- What we can conclude?


Scenario 1: Logic of Significance Test
Control
Treatment
Test scores
10
5
0
-5
Graphs by Treatment status


Scenario 2: Logic of Significance Test
Control
Treatment
Test scores
0
5
10
15
Graphs by Treatment status


Scenario 3: Logic of Significance Test
Control
Treatment
Test scores
6
5
4
3
2
Graphs by Treatment status

- We have just compared the separation of the means (effect or signal) to the background scatter of individual responses (variability or noise). This forms an **effect-to-variability ratio** or a **signal-to-noise ratio**
- This **signal-to-noise ratio** becomes more clear as sample size increases (see graph below)
- We need just a single number that summarizes these 3 criteria: **sample size**, **effect** and **variability**. That's exactly the role of a **test statistic**


Statistical Regularity for N(5,2.5) and N(4,2.5)
Mean
10
8
6
4
2
0
1
50
100
200
300400
Sample Size (log scale)

- A formula for a two-sample t test for comparing means:

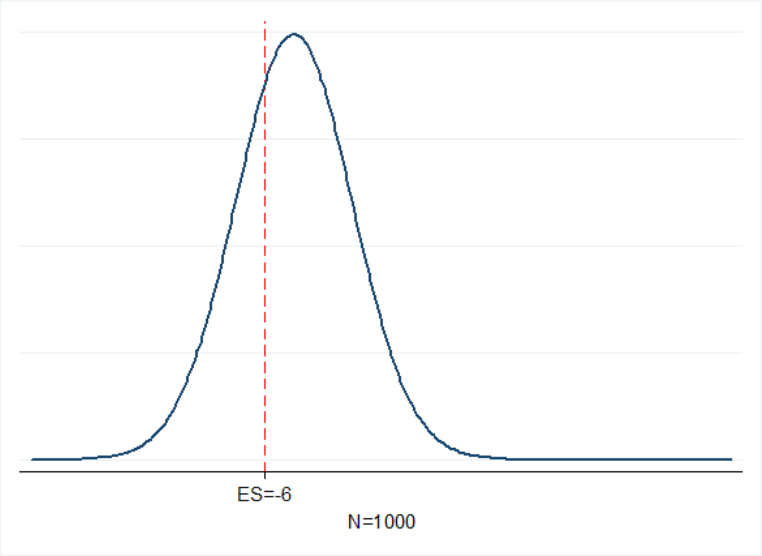
$$t = \frac{\bar{X}_T - \bar{X}_C}{\frac{s_T}{\sqrt{n_T}} + \frac{s_C}{\sqrt{n_C}}} \tag{16}$$

- We use this test statistics to compute the probability of chance, or sampling variability

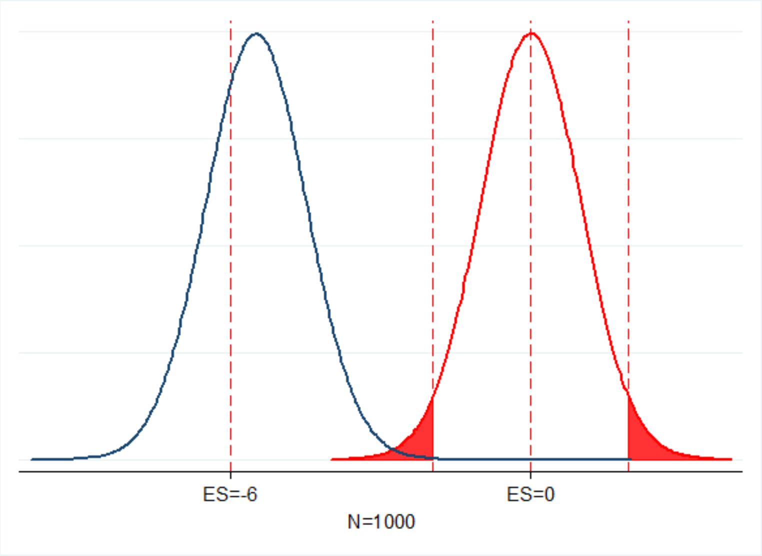
3.4 Sampling distribution of a test statistic

- Motivating example: impact of training program (-6 impact)

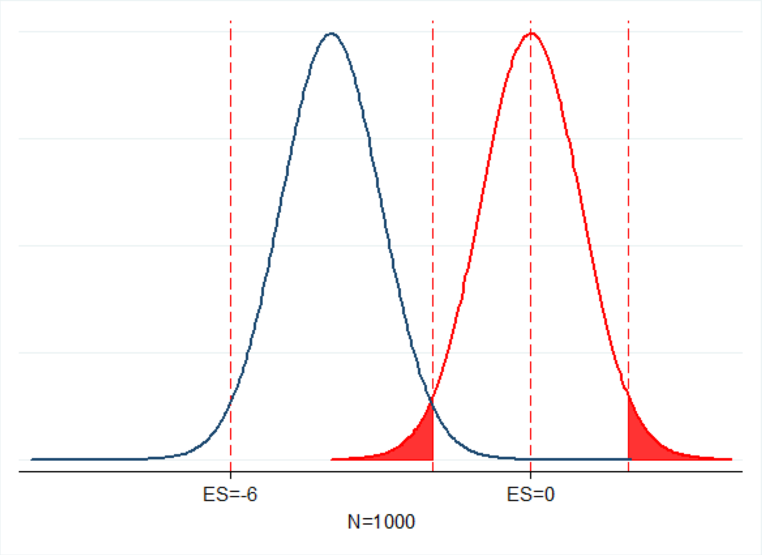
The null hypothesis



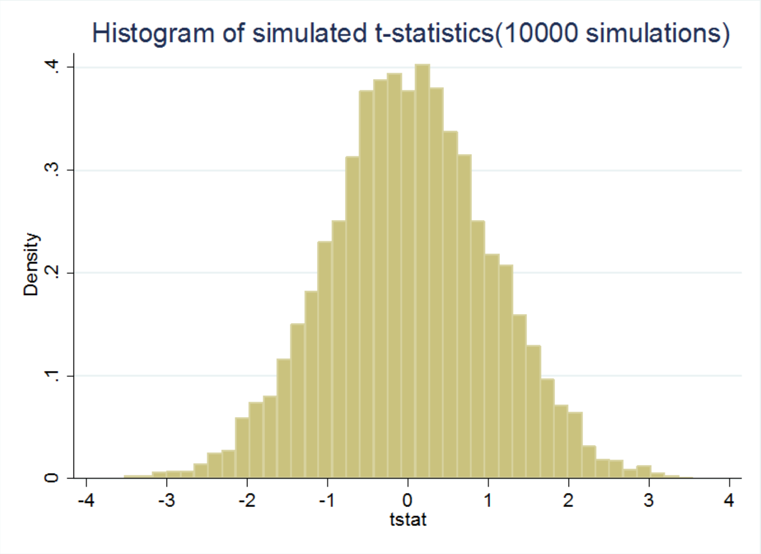
The null hypothesis

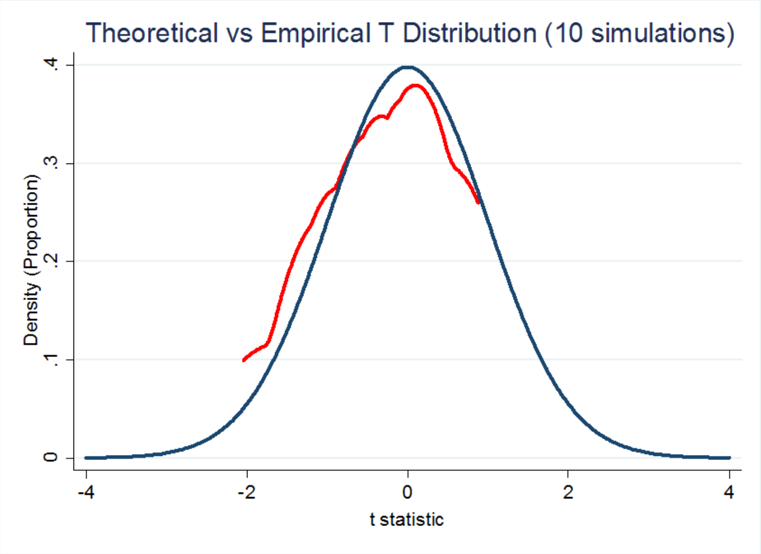


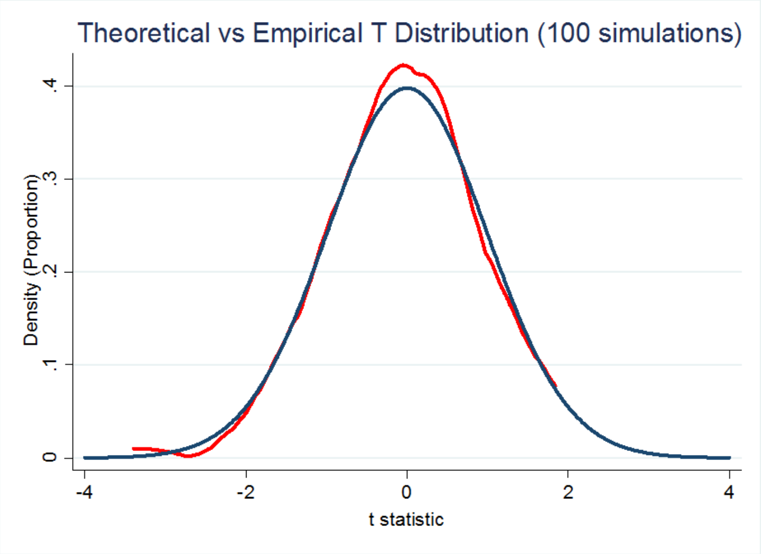
The null hypothesis



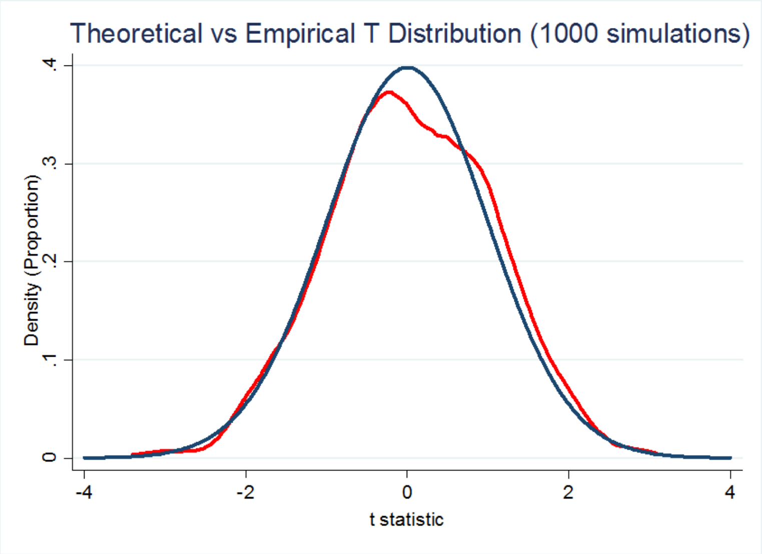
- We create a distribution of the test statistic values that occur when **no study effect** exists (when the two population means are equal)
- A middle 95% of this sampling distribution are considered consistent with no effect (or an observed result that was due to sampling variation)
- A theoretical distribution (t-student) has been derived for this t-test
- We can also simulate empirically this distribution

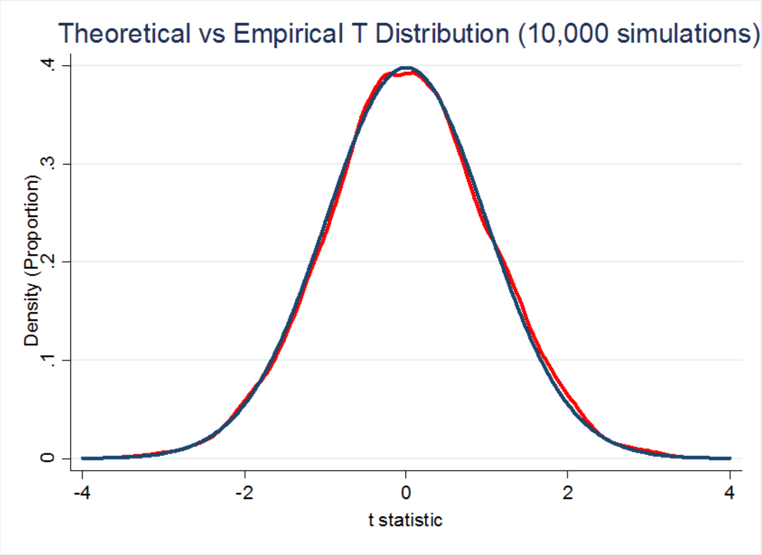

Histogram of simulated t-statistics(10000 simulations)
Density
0
.1
.2
.3
.4
-4
-3
-2
-1
0
1
2
3
4
tstat


Theoretical vs Empirical T Distribution (10 simulations)
Density (Proportion)
0
.1
.2
.3
.4
-4
-2
0
2
4
t statistic



Theoretical vs Empirical T Distribution (100 simulations)
Density (Proportion)
0
.1
.2
.3
.4
-4
-2
0
2
4
t statistic


Theoretical vs Empirical T Distribution (1000 simulations)
Density (Proportion)
0
.1
.2
.3
.4
-4
-2
0
2
4
t statistic


Theoretical vs Empirical T Distribution (10,000 simulations)
Density (Proportion)
0
.1
.2
.3
.4
-4
-2
0
2
4
t statistic

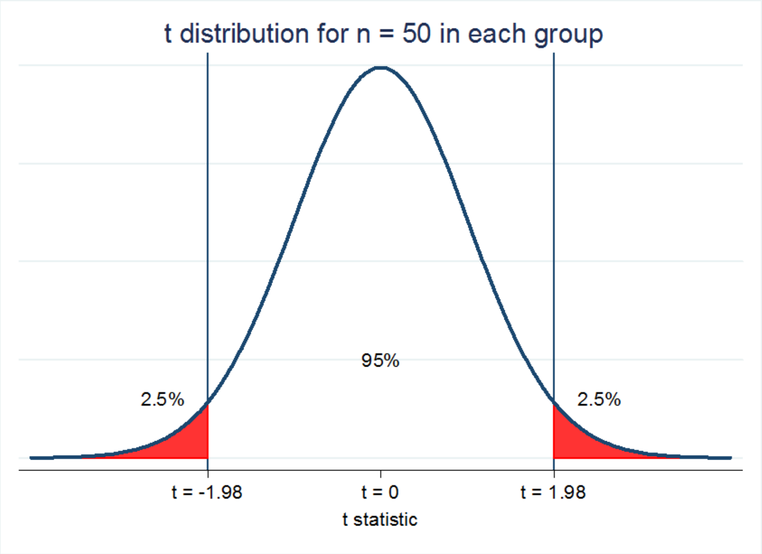
- Applying the usual standards, we use the middle 95% (90% and 99% are also common) of the t-statistics values to represent values consistent with no study effect (null hypothesis):

$$H_0 : \mu_T = \mu_C$$

- Outer 5% of these t-statistics are considered consistent with a study effect:

$$H_A : \mu_T \neq \mu_C$$

- Using theoretical values, this can be represented in this way:


t distribution for n = 50 in each group
95%
2.5%
2.5%
t = -1.98
t = 0
t = 1.98
t statistic

- We quantify the result of our t test with a p-value:

Definition: p-value

p-value=Prob(obs. sample effect/ $H_0 : \mu_T = \mu_C$)
p-value=Prob(obs. sample effect/ due to sampling variability)

- P-value is the area under the sampling distribution curve (the t distribution)
- Statistical significance is understood as obtaining an extreme test statistic inconsistent with the null hypothesis of no difference

Concluding Remarks

- What we have learned?