Georgia Institute of Technology

Statistics for Economics, Lecture 20

jim.crozier@gatech.edu, 2023

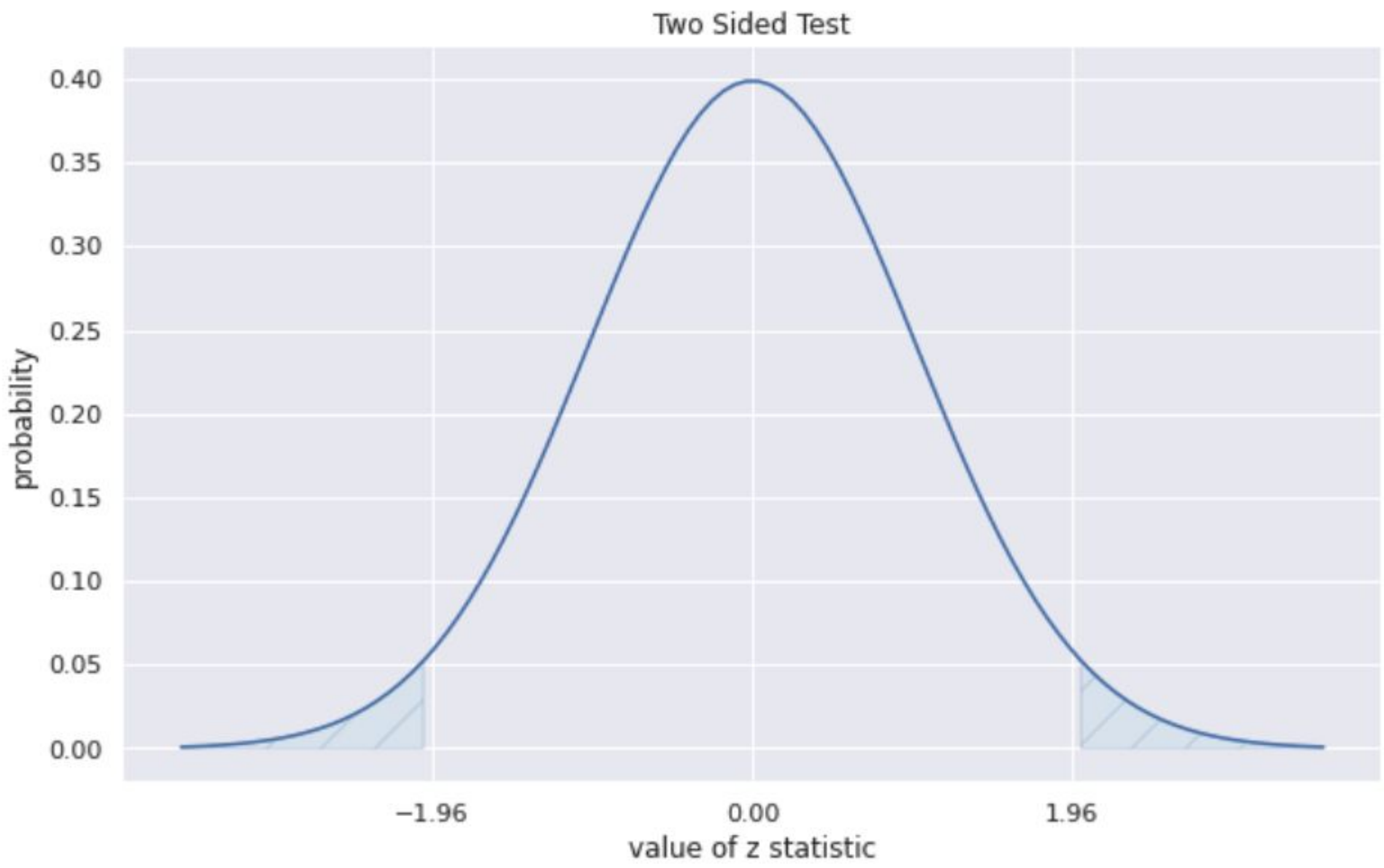
Today

- Review homework 9
- Review t-test with varying sample size and variance
- Review regression line t-test
- Introduce confidence intervals
- Introduce Central Limit Theorem

What are the three things that economists are looking for in applied analysis?

Review homework 9

desired alpha level	two-sided test	one-sided test
0.100	1.644854	1.281552
0.050	1.959964	1.644854
0.010	2.575829	2.326348
0.001	3.290527	3.090232


Two Sided Test
probability
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−1.96
0.00
1.96
value of z statistic

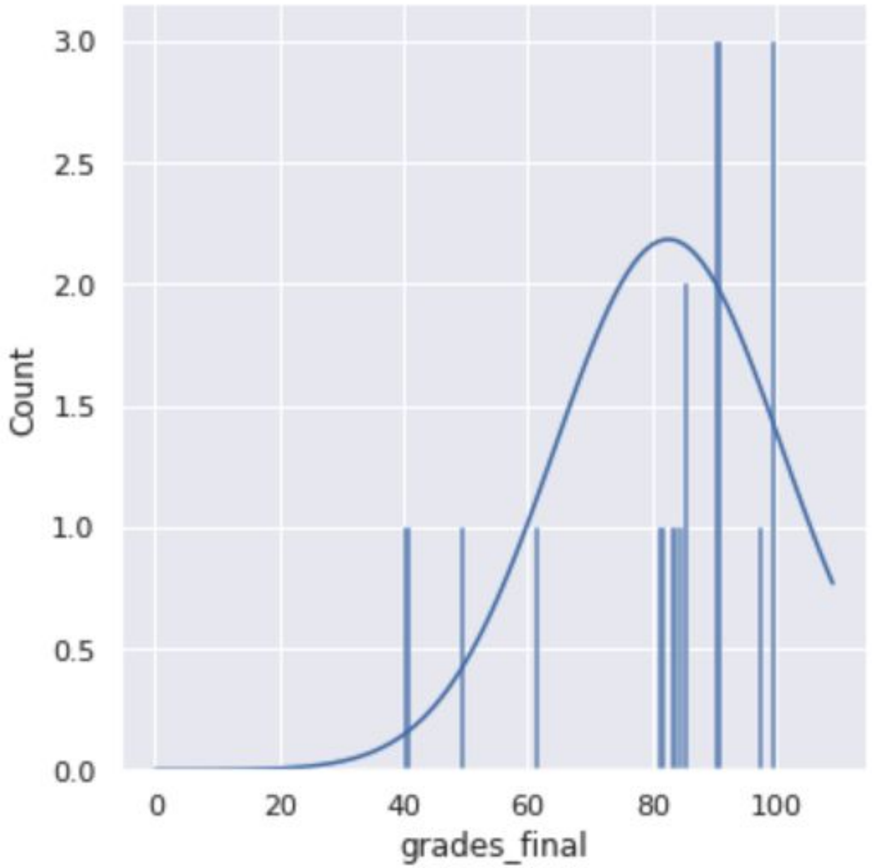
(simple) T Tests = z-test

$$z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{N}}$$

#1. We can make a normal with the mean and the standard deviation

$$\bar{x} = \frac{\sum x}{15} = \frac{1238}{15} = 82$$

$$\sigma = \sqrt{\frac{(\sum x - \bar{x})^2}{N-1}} = \sqrt{\frac{4672}{14}} = 18.2$$

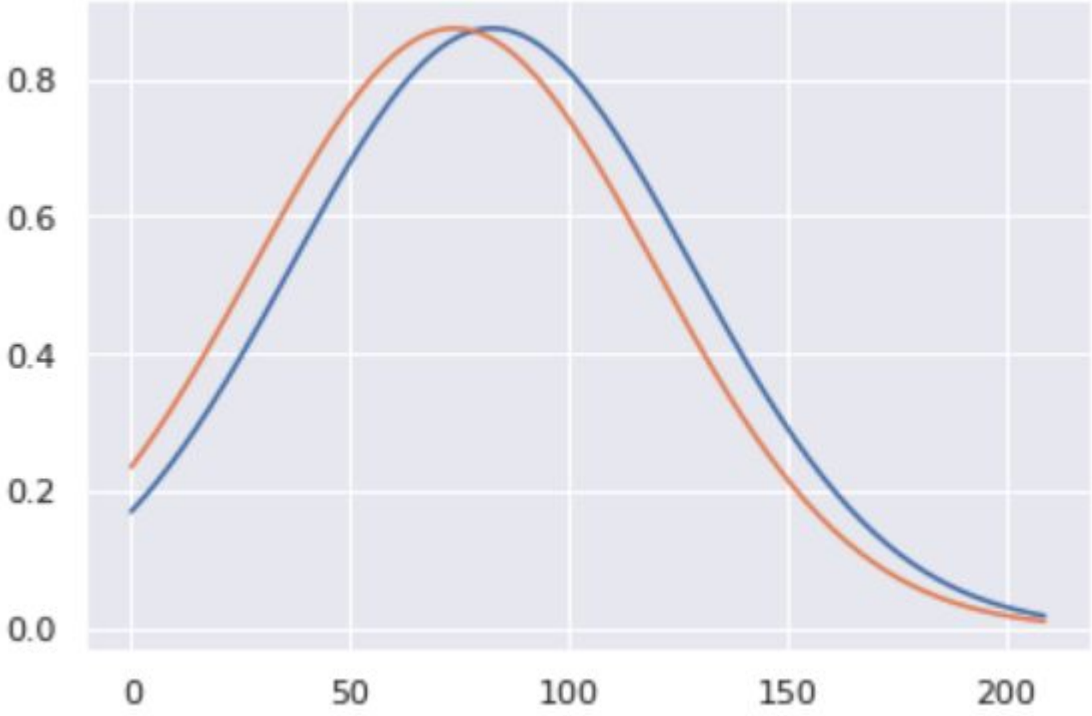


Can use the abstraction to compare to other abstractions:

Compare 2 classes:

Class1: 82 (18)

Class2: 72 (18)



Can extend this to further abstraction:

Compare 2 classes:

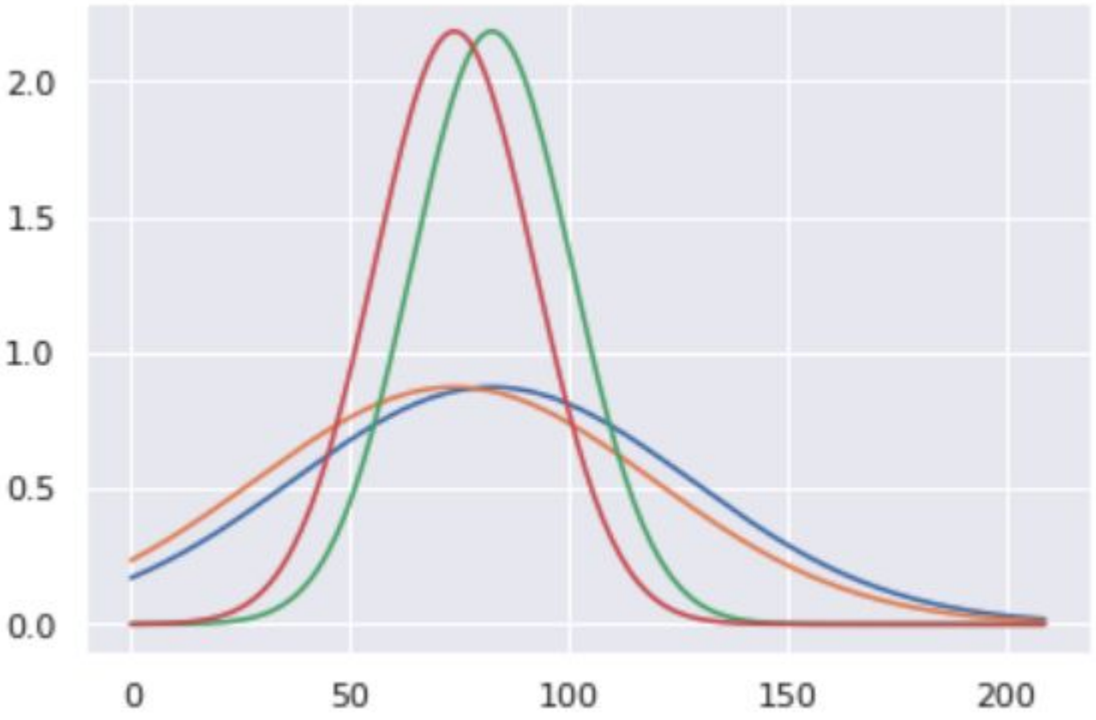
Class1: 82 (18) (green)

Class2: 72 (18)(red)

Or even

Class1: 82 (18) (green)

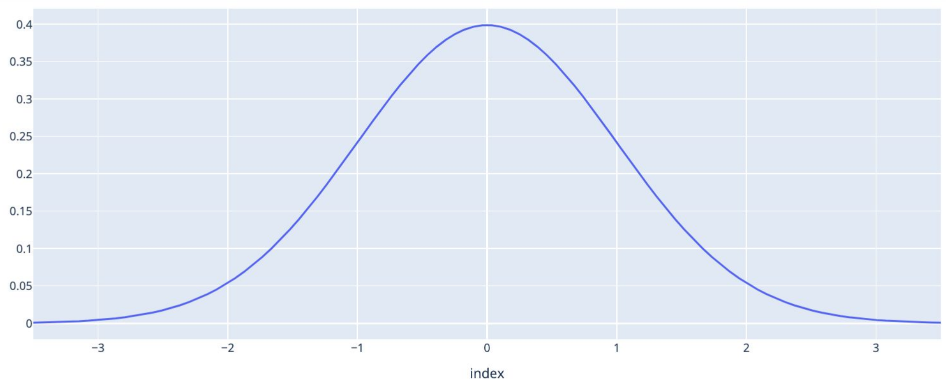
Class2: 72 (36)(orange)

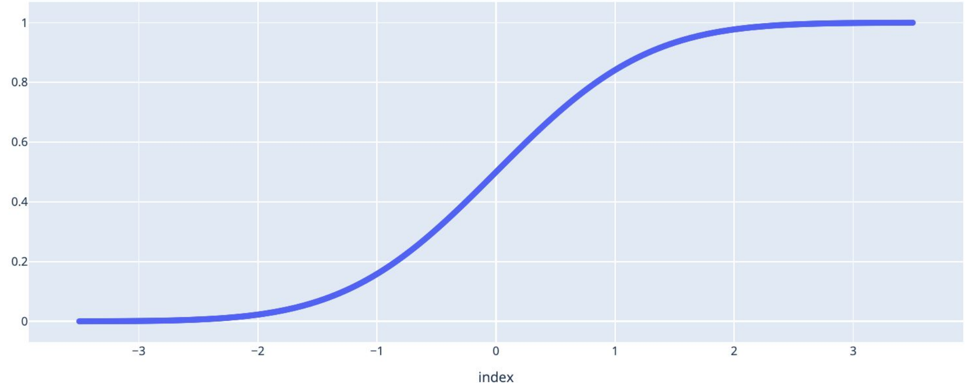


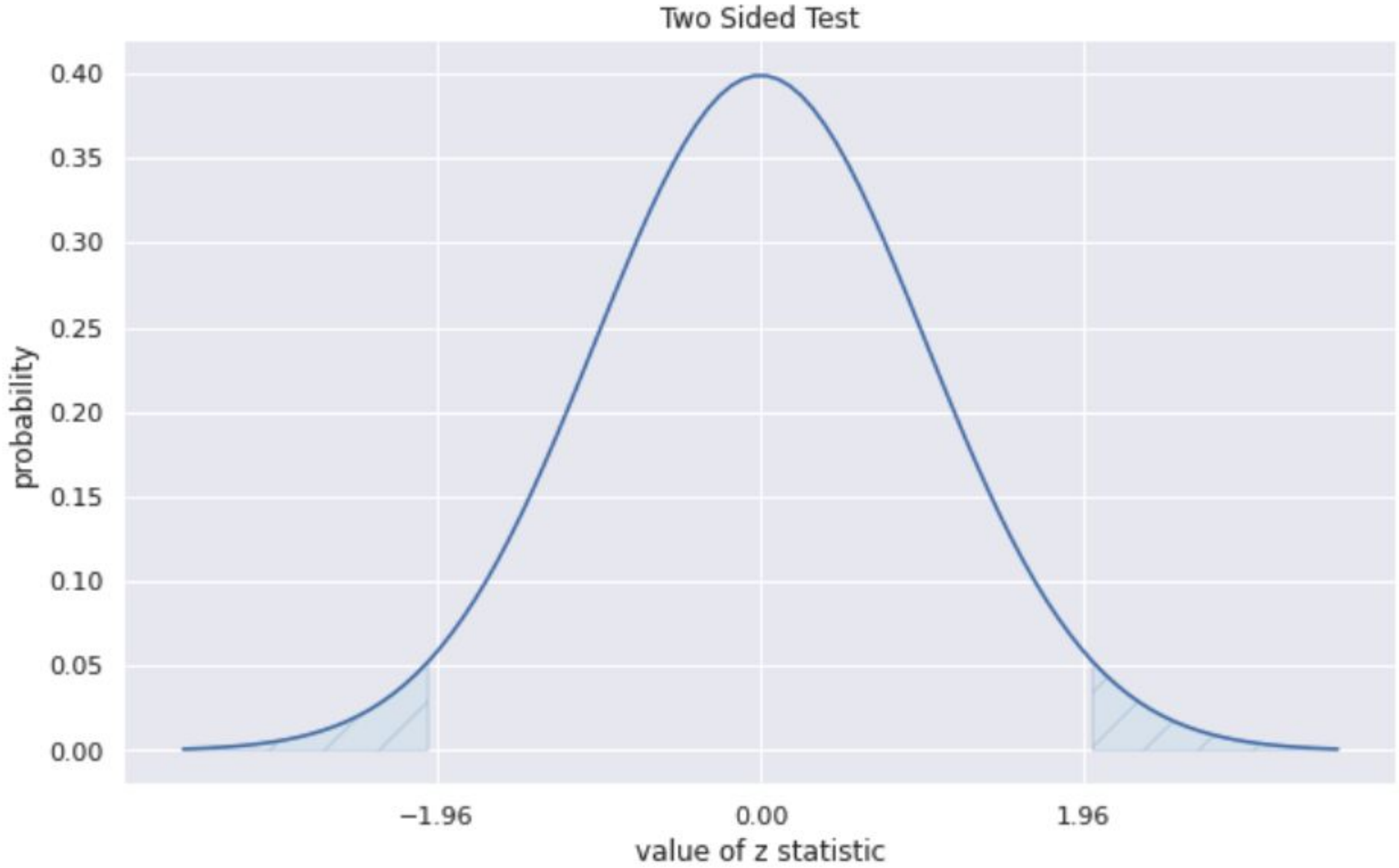
Center any normal to standard normal (0,1)

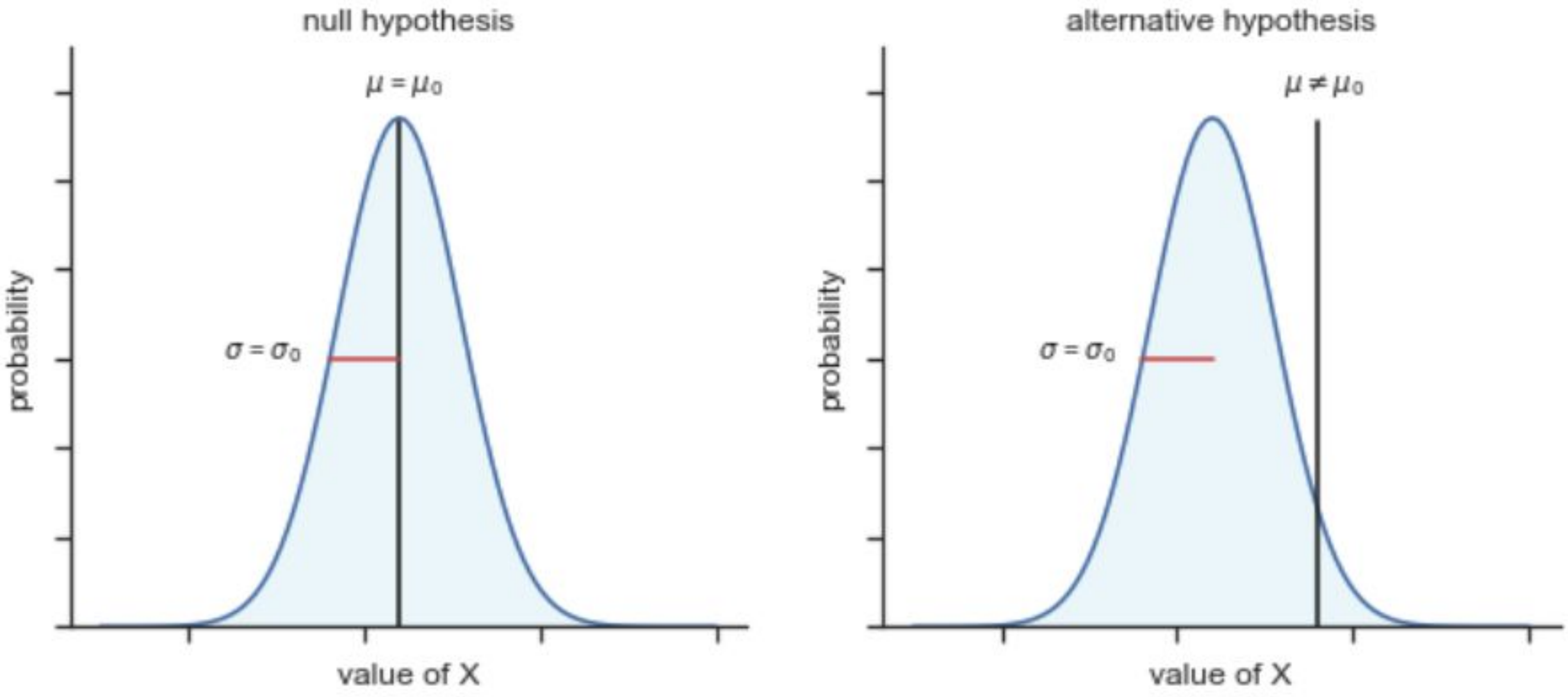
- Subtract mean
- Divide by variation

Remember, we can use the CDF is find how much of the area is bounded under curve






Two Sided Test
probability
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−1.96
0.00
1.96
value of z statistic


null hypothesis
μ = μ0
σ = σ0
probability
value of X
alternative hypothesis
μ ≠ μ0
σ = σ0
probability
value of X

Simple T-Test Case 1

Class1: 82 (18)

Class2: 72 (18)

H0: Difference - 0

H1: Not H0

$$z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{N}}$$

$$Z = \frac{82 - 72}{18/\sqrt{15}} = 2.27$$

Simple T-Test Case 2: smaller delta

Class1: 82 (18)

Class2: 76 (18)

H0: Difference - 0

H1: Not H0

$$z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{N}}$$

$$Z = \frac{82 - 76}{18/\sqrt{15}} = 1.3$$

Simple T-Test Case 3: higher variance

Class1: 82 (28)

Class2: 72 (28)

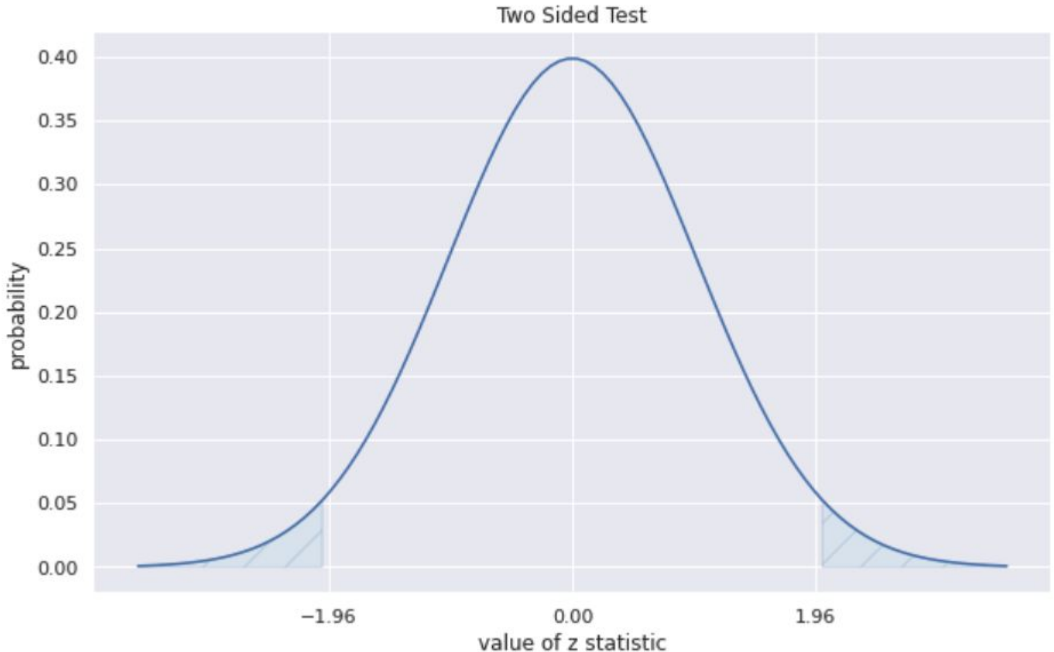
H0: Difference - 0

H1: Not H0

$$z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{N}}$$

$$Z = \frac{82 - 72}{28/\sqrt{15}} = 1.4$$

desired alpha level	two-sided test	one-sided test
0.100	1.644854	1.281552
0.050	1.959964	1.644854
0.010	2.575829	2.326348
0.001	3.290527	3.090232



T-test

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\text{SE}}$$

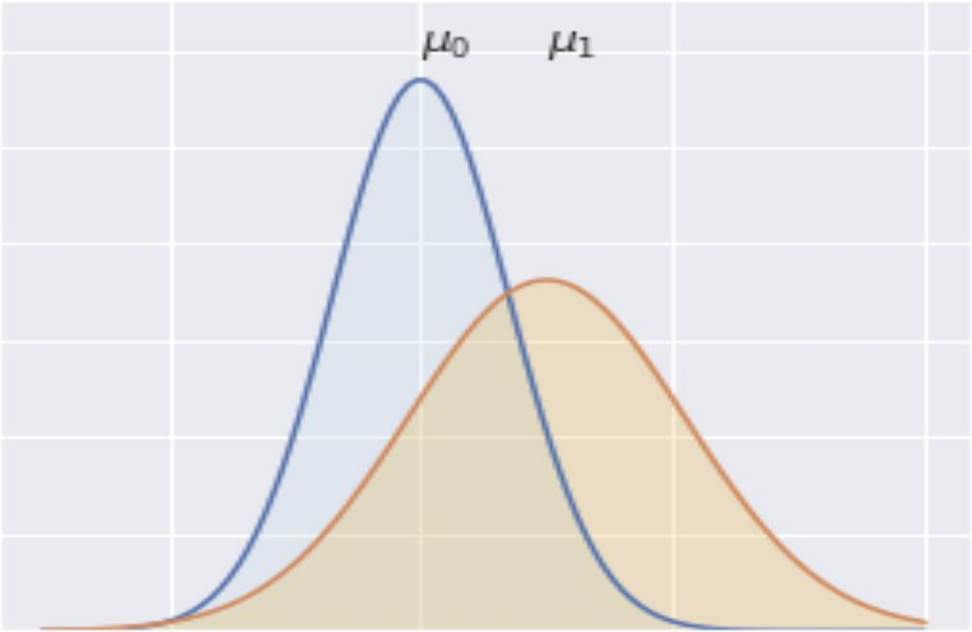
where

$$SE = \sqrt{\frac{\sigma_1^2}{N1} + \frac{\sigma_2^2}{N2}}$$

Difference between two groups with different sample sizes and variance

$$H_0 : \quad \mu_1 = \mu_2$$
$$H_1 : \quad \mu_1 \neq \mu_2$$

- Classroom 1
  - Avg: 82
  - Stdev: 18
  - Size: 15
- Classroom 2
  - Avg: 72
  - Stdev: 22
  - Size: 30



- Classroom 1
  - Avg: 82
  - Stdev: 18
  - Size: 15
- Classroom 2
  - Avg: 72
  - Stdev: 22
  - Size: 24

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\mathrm{SE}}$$

where

$$SE = \sqrt{\frac{\sigma_1^2}{N1} + \frac{\sigma_2^2}{N2}}$$

$$SE = \sqrt{\frac{18^2}{15} + \frac{24^2}{30}} = 6.3$$

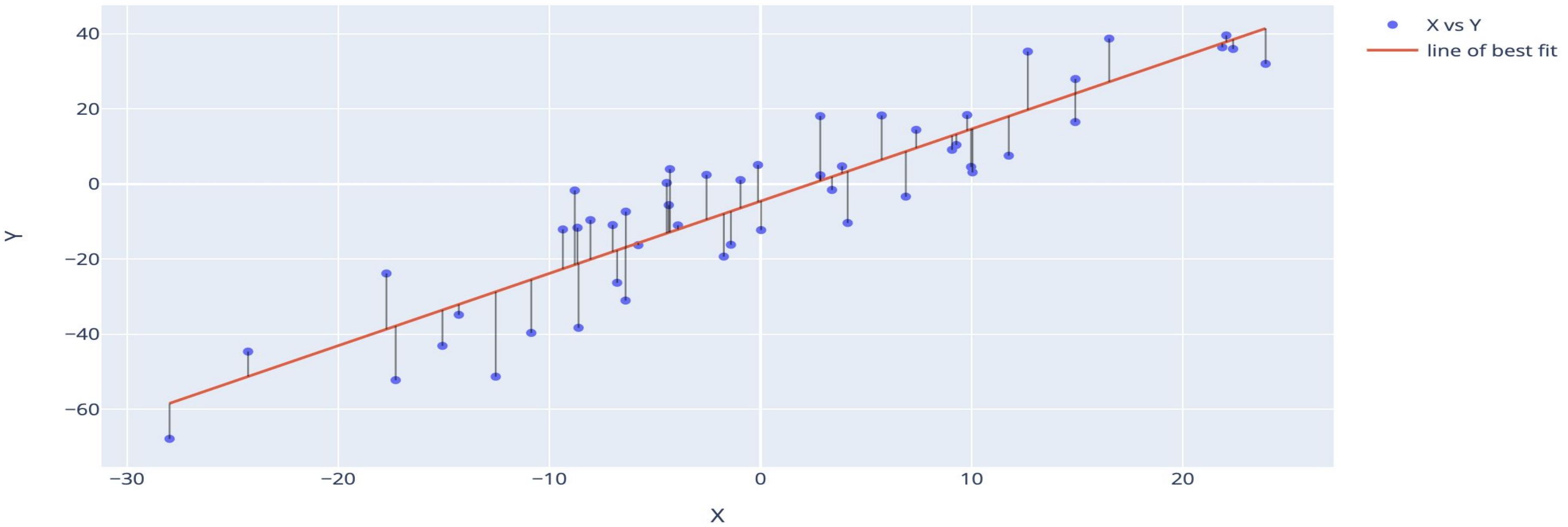
$$T = \frac{82-72}{6.3} = 1.56$$

Linear Regression

$$y_i = a + b * x_i + u_i$$

$$\hat{y}_i = \hat{a} + \hat{b} * x_i$$

$$\hat{u}_i = y_i - \hat{y}_i$$



$$\hat{y}_i = \hat{a} + \hat{b} * x_i$$

$$\hat{b} = \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)} = \frac{\sum(x_i - \bar{x}) \sum(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$\hat{a} = \bar{y} - \hat{b} * \bar{x}$$

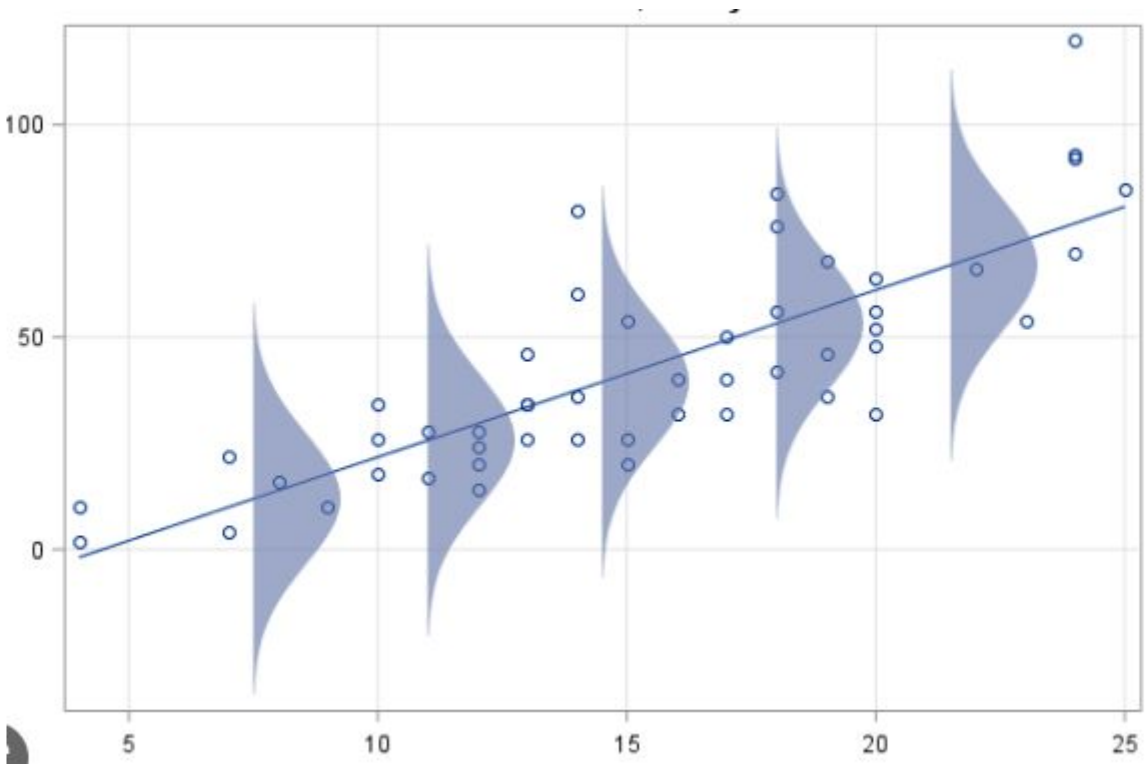
Walk through example

								yhat	y-yhat
x	y	demean_x	demean_x_sq	demean_y	demean_y_sq	demean_x*demean_y		b0 + b1*x	error
3	6	-2.4	5.76	1.6	2.56	-3.84		4.90	1.10
5	9	-0.4	0.16	4.6	21.16	-1.84		4.48	4.52
2	2	-3.4	11.56	-2.4	5.76	8.16		5.11	-3.11
8	1	2.6	6.76	-3.4	11.56	-8.84		3.85	-2.85
9	4	3.6	12.96	-0.4	0.16	-1.44		3.65	0.35
			37.2		**41.2**	-7.8	sum		
						-1.95	sum/(n-1)		

mean_y	4.4
mean_x	5.4

numerator	**-1.95**	-1.95	**-0.199**	**correlation**
denom	sqrt(**37.2/4** * **41.2/4)**	9.79		

slope	-0.210
intercept	5.53


100
50
0
5
10
15
20
25

Confidence Interval

H_0: b_1= 0 (we don't believe there is a slope (relationship))

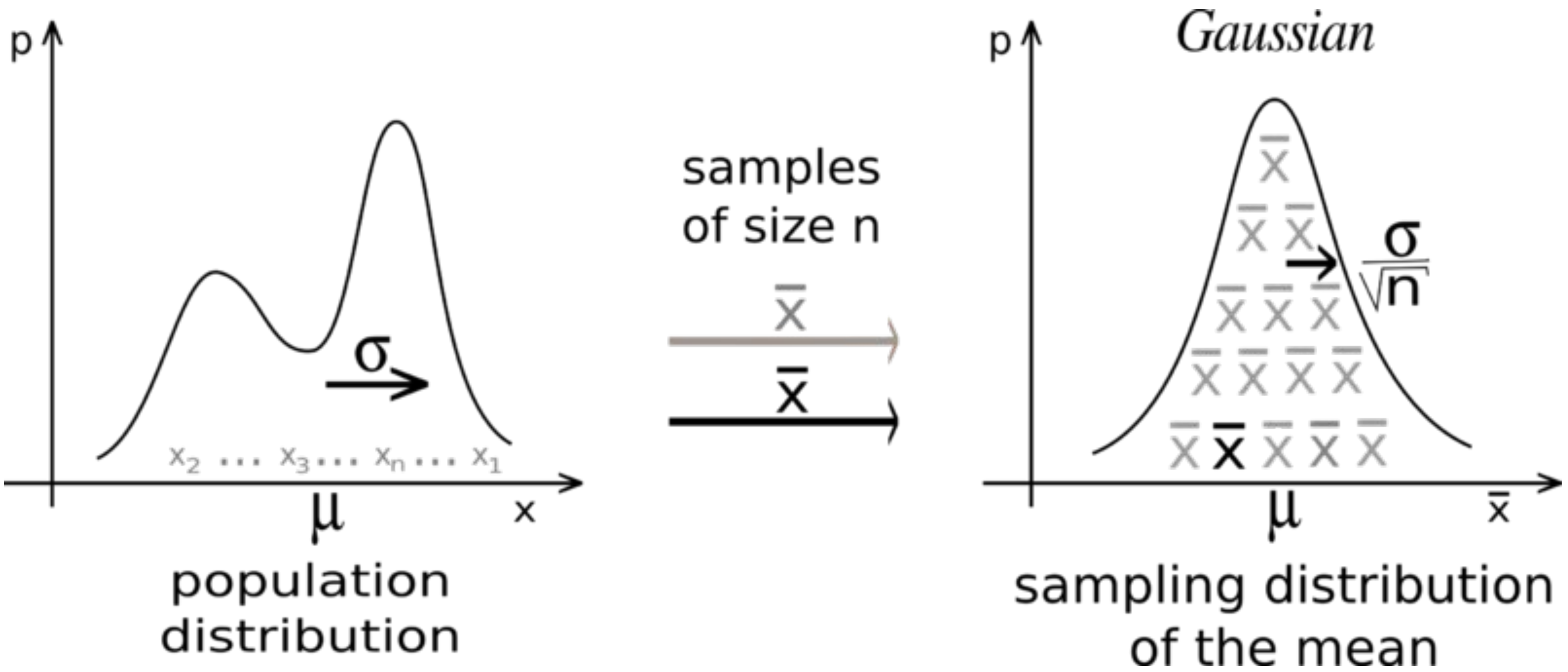
H_A: b_1 not equal 0 (we believe there is a slope (relationship))

$$t = \frac{b_1 - 0}{SE_{b_1}} = \frac{b_1}{SE_{b_1}}$$

Confidence interval is b1 +/- CV* SE, where CV is critical value. This is easier to show than anything else (see notebook).

Central Limit Theorem

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{a} \mathcal{N}\left(0, \sigma^2\right)$$


Gaussian
p
p
x
μ
μ
population distribution
samples of size n
sampling distribution of the mean