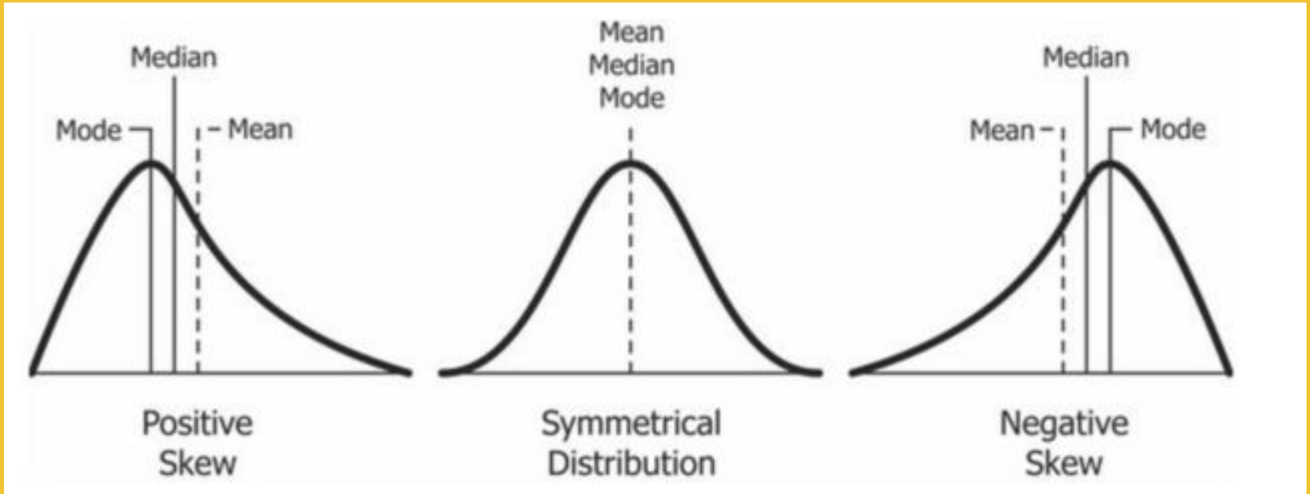


Econ 2250: Stats for Econ

Fall 2022

[Source for pic stats above.]

Variance

$$V(X) = E((X - E(X))^2)$$

Variance

Theorem. **Computational formula for the variance:**

$$\operatorname{Var}(X) = E\left[X^2\right] - \left(EX\right)^2.$$

Proof:

$$\begin{aligned}
\operatorname{Var}(X) &= E\left[(X - \mu_X)^2\right] \\
&= E\left[X^2 - 2\mu_X X + \mu_X^2\right] \\
&= E\left[X^2\right] - 2E\left[\mu_X X\right] + E\left[\mu_X^2\right] \quad \text{by linearity of expectation.} \\
&= E\left[X^2\right] - 2\mu_X^2 + \mu_X^2 \\
&= E\left[X^2\right] - \mu_X^2.
\end{aligned}$$

From the example above

$$\sigma^2 = \frac{1}{n}\sum(x_i - \mu_x)^2$$

x	mu	demean	demean_sq
3	6.3	-3.3	11.1
12	6.3	5.7	32.1
4	6.3	-2.3	5.4

Sum of demean_sq = 11.1 + 32.1 + 5.4 = 48.6

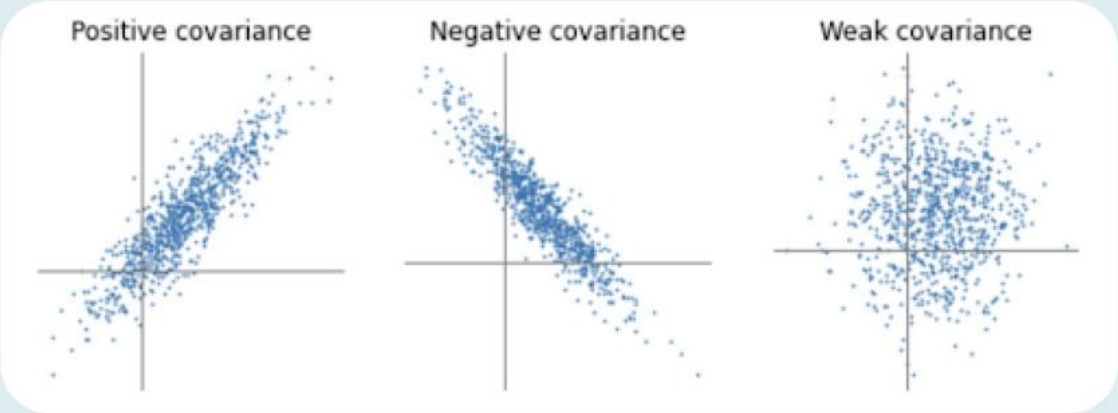
48.6/3 = 16.222

Above we showed Var(X) = E(X^2) - mu_x^2

E(X^2) =⅓ *9 + ⅓ * 144 + ⅓ 16 = 56.33

Mu_x^2 = (6.333)^2 = 40.111

E(X^2) - Mu_x^2 = 56.333 - 40.111 = 16.222



Covariance

$Cov(X, Y)$ = E[(X-E(X)(Y-E(Y)]

image

Covariance

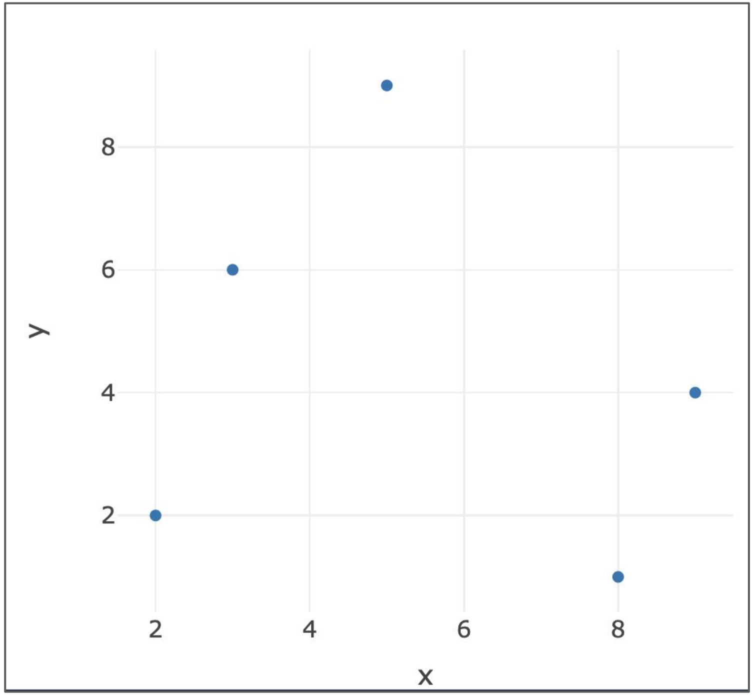
$$
\begin{aligned}
Cov(x, y) &= E[(X - E(X))(Y - E(Y))] \\
&= \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n - 1}
\end{aligned}
$$

Example covariance

$$\frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n - 1}$$

x	y	demean_x	demean_y	demean_x*demean_y	
3	6	-2.4	1.6	-3.84	
5	9	-0.4	4.6	-1.84	
2	2	-3.4	-2.4	8.16	
8	1	2.6	-3.4	-8.84	
9	4	3.6	-0.4	-1.44	
				-7.8	sum
				-1.95	sum/(n-1)

mean_y	4.4
mean_x	5.4



Linear Regression

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

$$\hat{y}_i = \text{best guess intercept} + \text{best guess slope} * x_i$$

$$\hat{a} = \text{best guess intercept}$$

$$\hat{b} = \text{best guess slope}$$

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

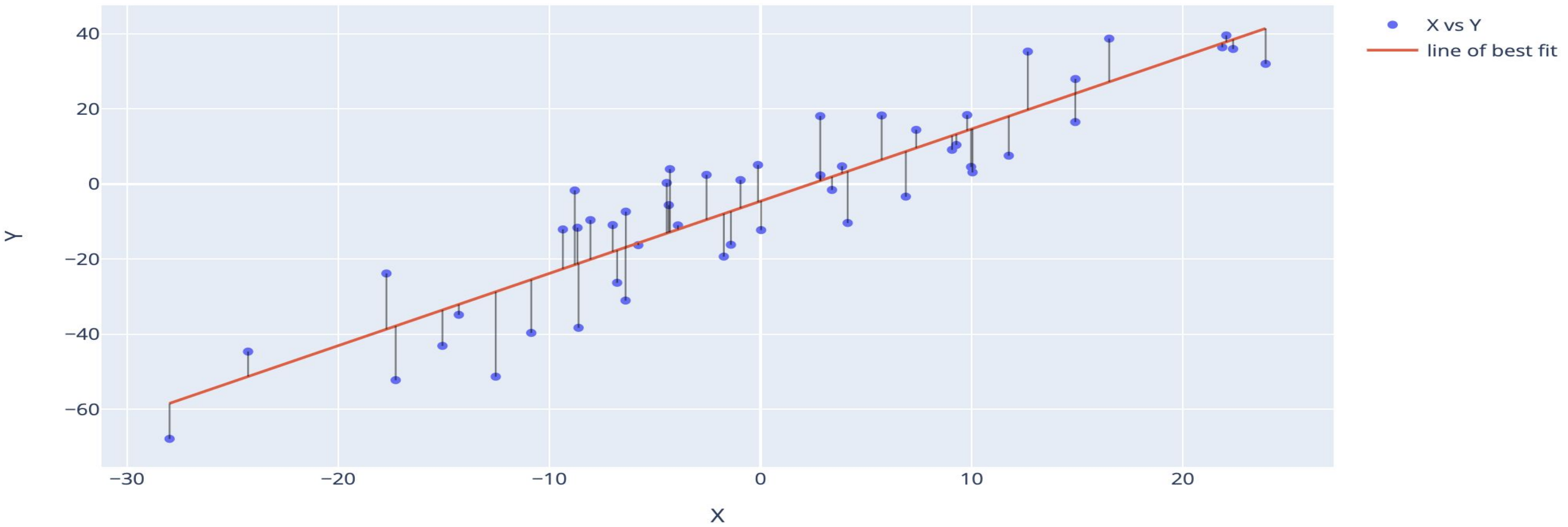
$$\hat{y}_i = \hat{a} + \hat{b} * x_i$$

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

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

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

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



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		

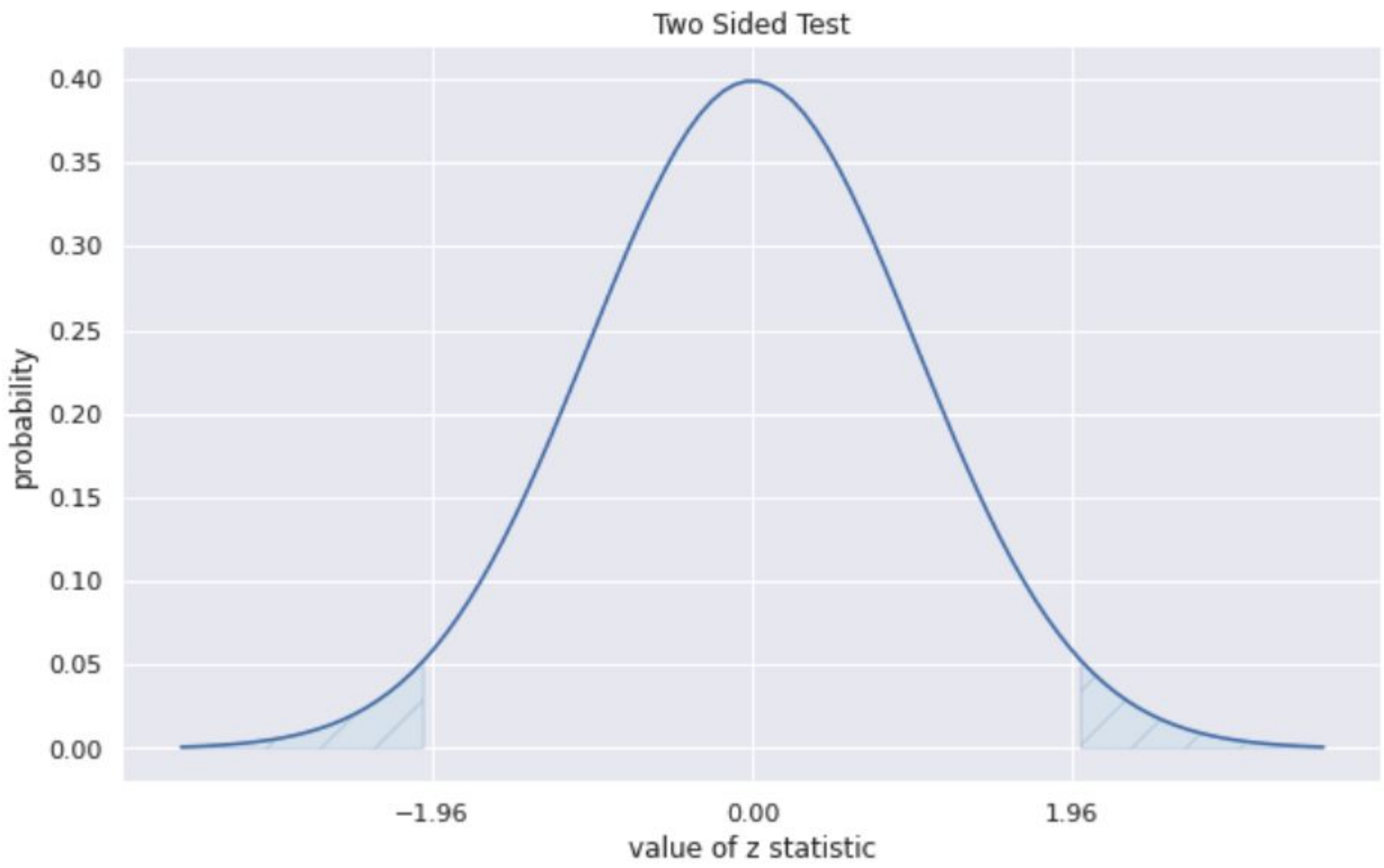
slope	-0.210
intercept	5.53

Look at Colab

```
df = pd.DataFrame({'x': x, 'y':y})
df['x_minus_xbar'] = df['x'] - mean_x
df['y_minus_ybar'] = df['y'] - mean_y
df['demaned_x_and_y'] = df['x_minus_xbar'] * df['y_minus_ybar']
df['demaned_x_sq'] =  df['x_minus_xbar']**2
df.head()
```

	x	y	x_minus_xbar	y_minus_ybar	demaned_x_and_y	demaned_x_sq
0	2.912054	1.320250	1.773713	4.257710	7.551954	3.146056
1	5.665337	7.127269	4.526996	10.064729	45.562987	20.493692
2	5.035918	4.597814	3.897577	7.535275	29.369309	15.191103
3	2.852957	0.649218	1.714616	3.586678	6.149775	2.939907
4	4.842881	6.043560	3.704540	8.981020	33.270548	13.723617

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



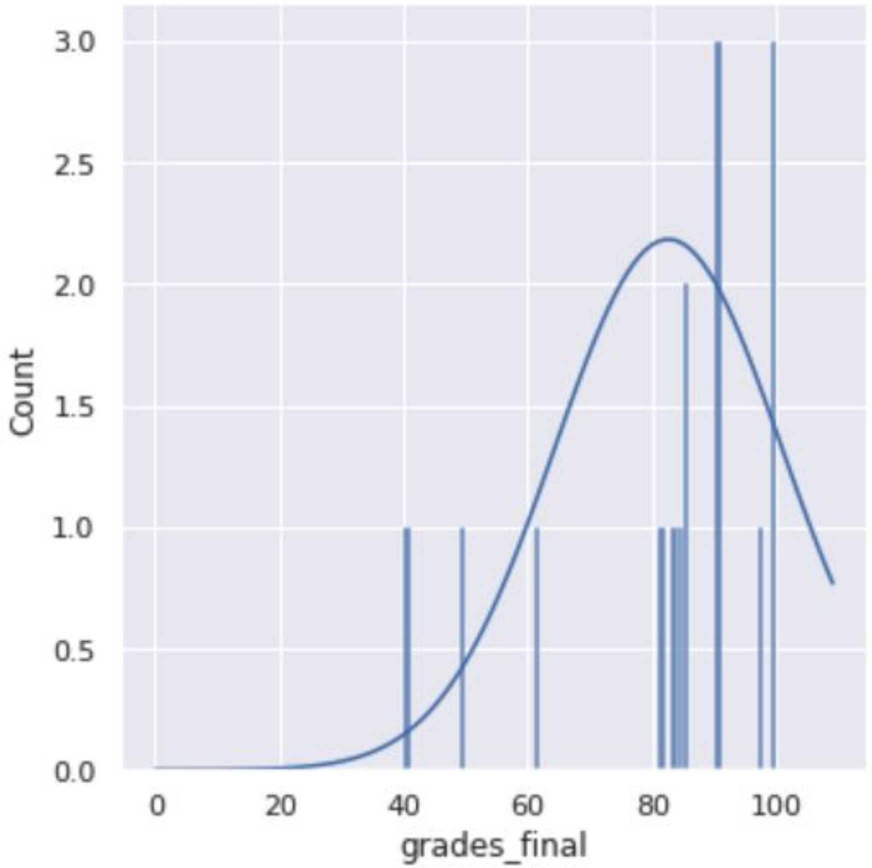
Two Sided Test
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
probability
−1.96
0.00
1.96
value of z statistic

(simple) T Tests = z-test

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

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

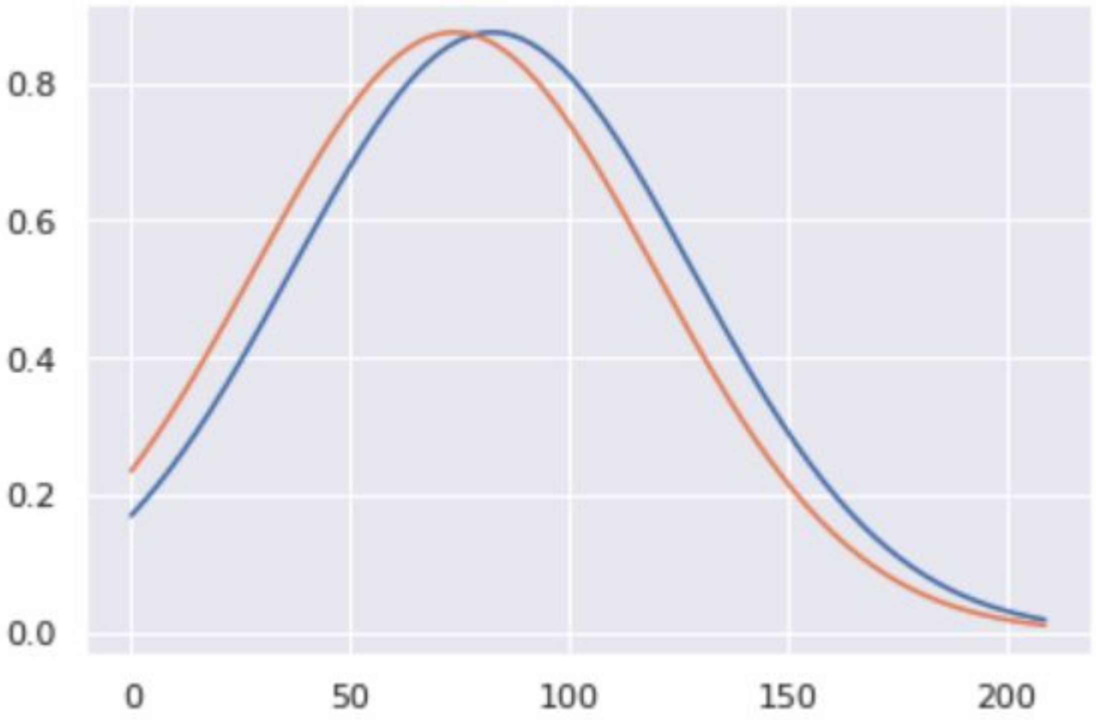
$$\sigma = \sqrt{\frac{(\sum x - \bar{x})^2}{N - 1}} = \sqrt{\frac{4672}{14}} = 18.2$$



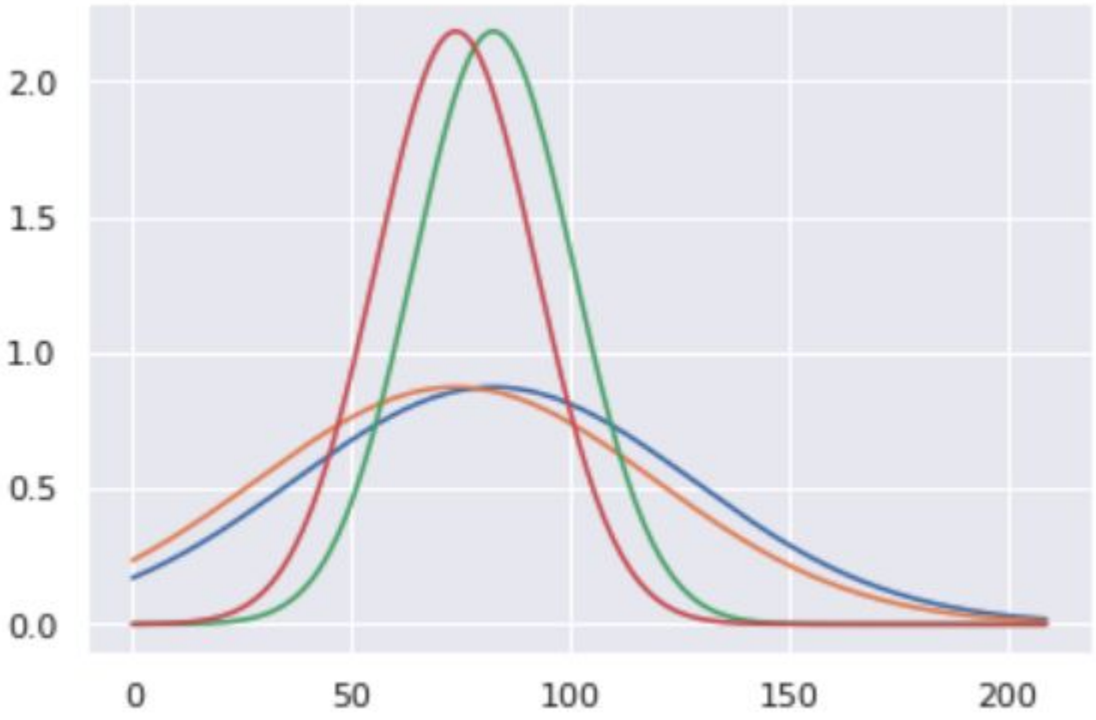
How to tell if significant difference between 2 classes?

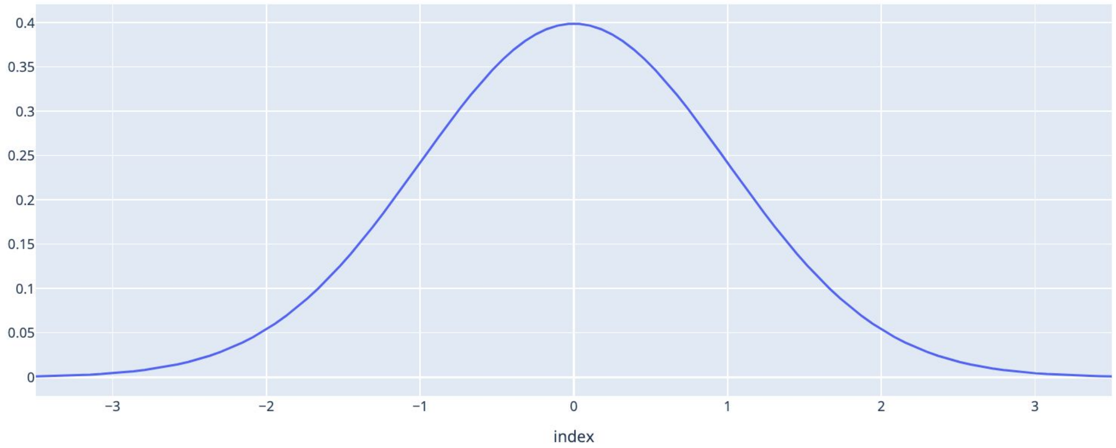
Class1: 82 (18)

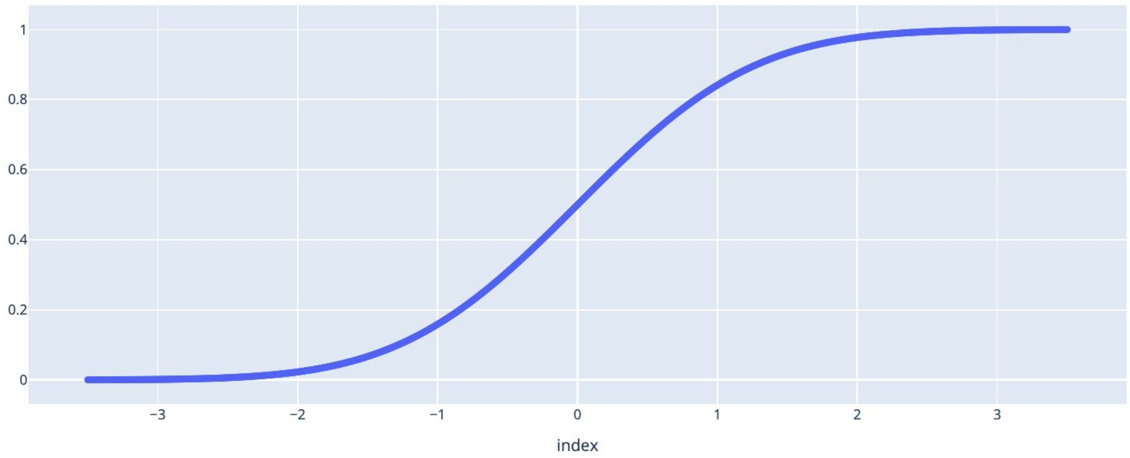
Class2: 72 (18)

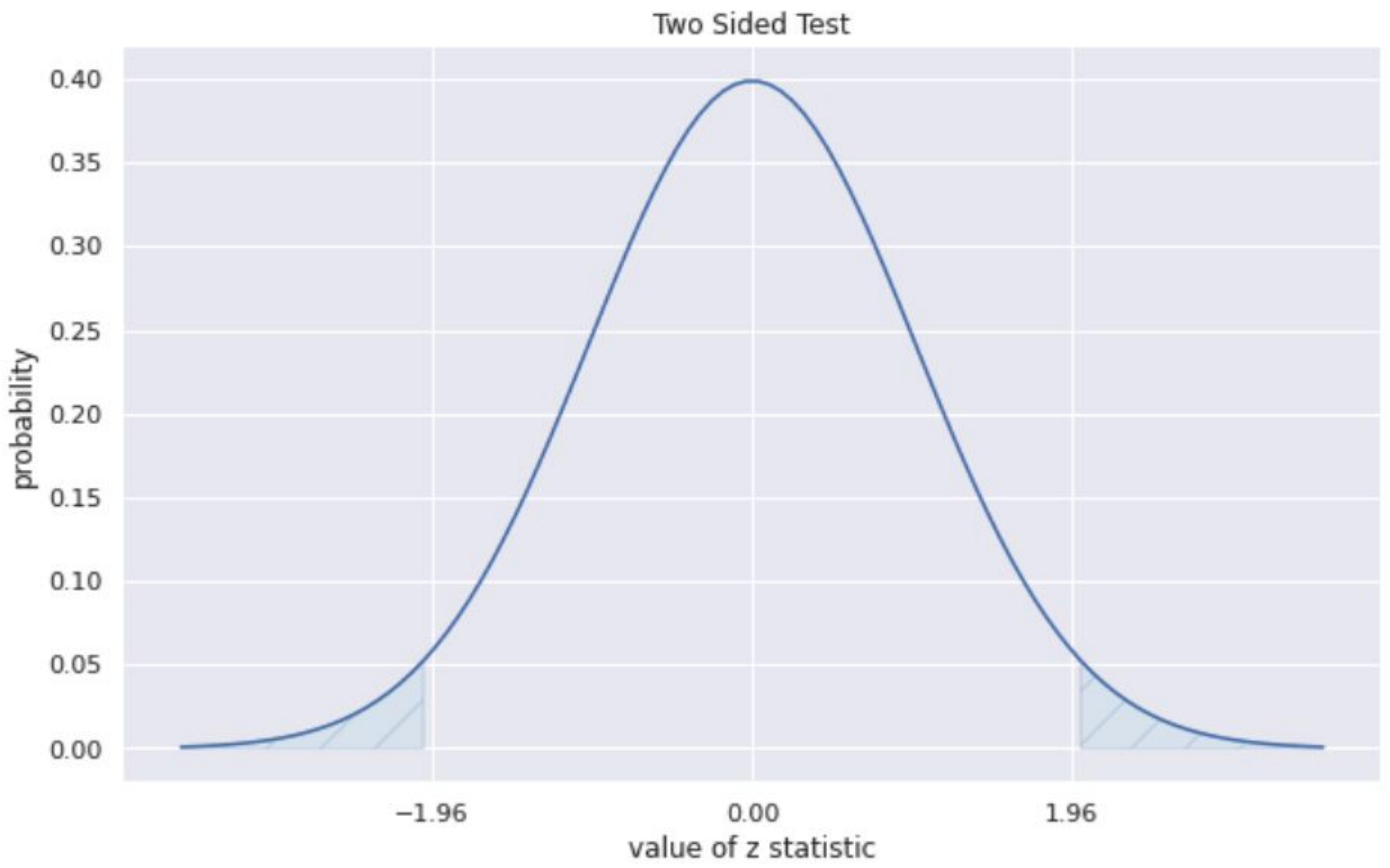


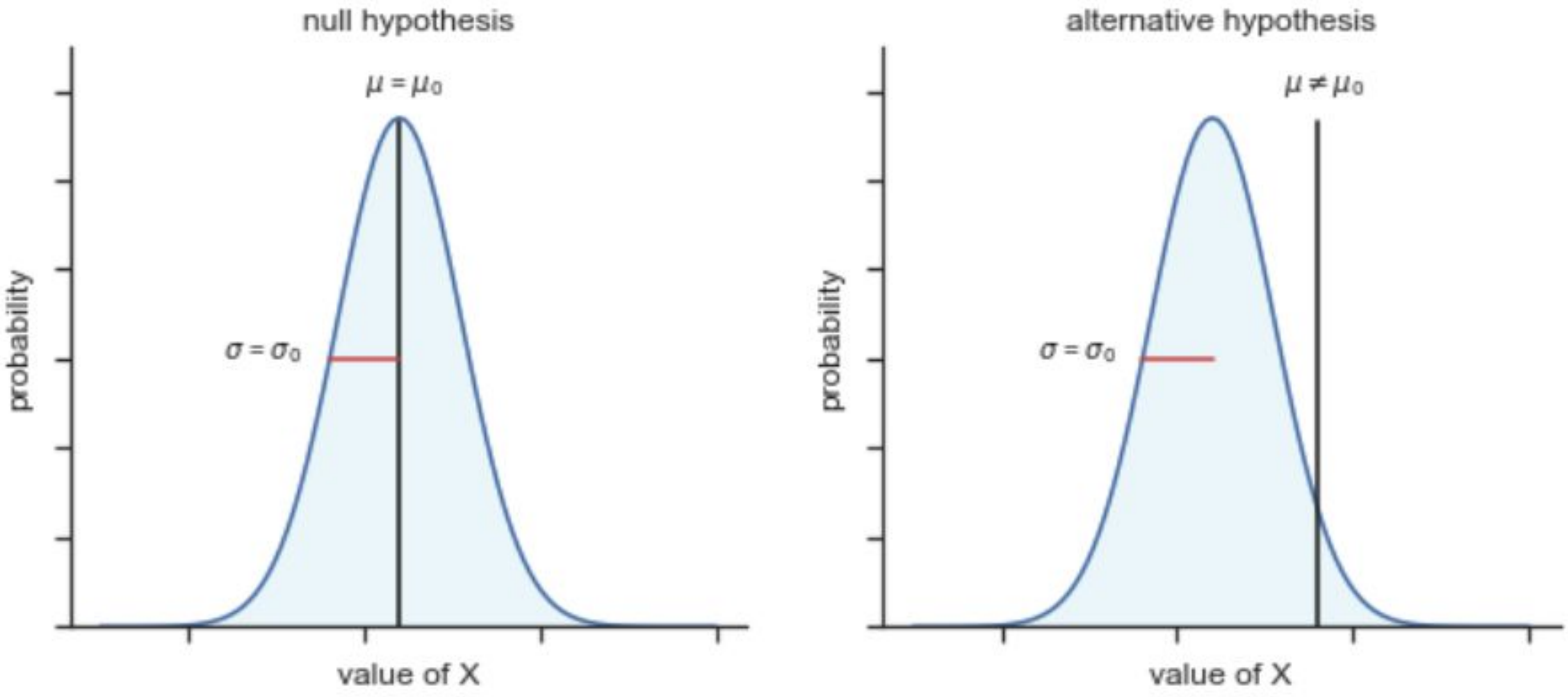
Remember, variance matters




0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
−3
−2
−1
0
1
2
3
index


1
0.8
0.6
0.4
0.2
0
−3
−2
−1
0
1
2
3
index


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
$\mu = \mu_0$
$\sigma = \sigma_0$
probability
value of X
alternative hypothesis
$\mu \neq \mu_0$
$\sigma = \sigma_0$
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

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

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

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

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

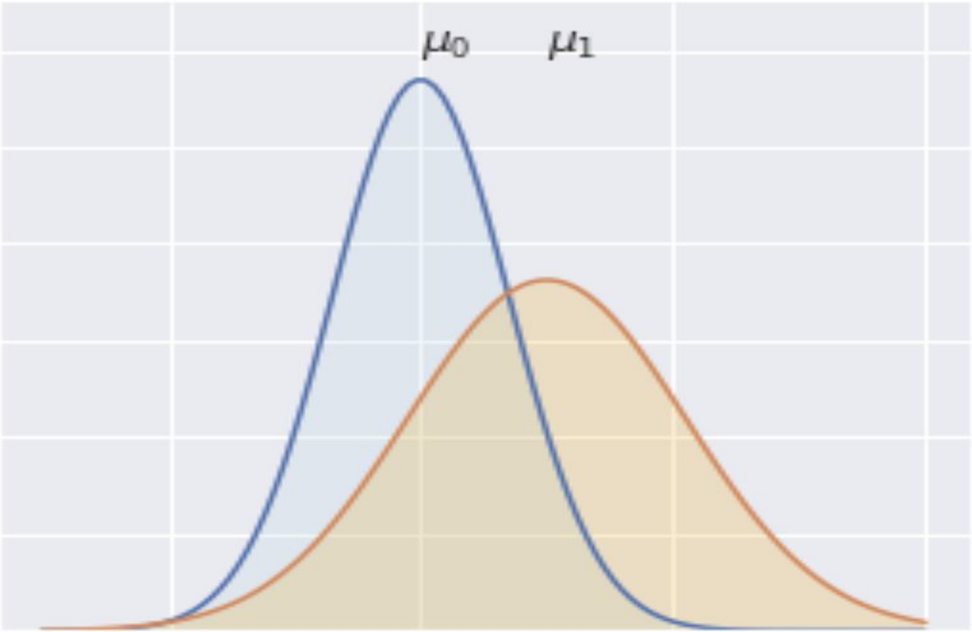
where

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

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: 24
  - 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

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

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

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