



Econ 2250: Stats for Econ

Fall 2022

[Source for pic stats above.](#)

- **Today**
 - **Review of variance**
 - **Proof**
 - **Review of linear regression**
 - **Examples of linear regression**

Variance

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

Variance

Theorem. Computational formula for the variance:

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

Proof:

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

From the example above

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

Sum of demean_sq = 11.1 + 32.1 + 5.4 = 48.6

48.6/3 = 16.222

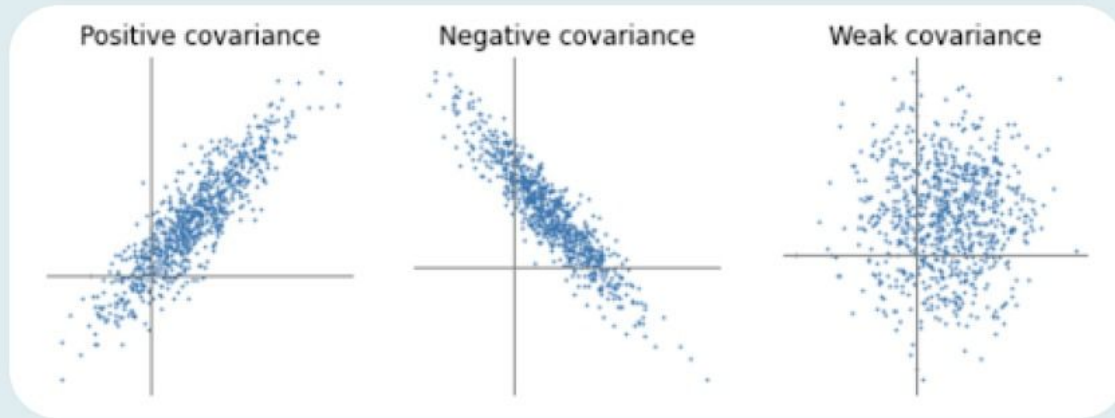
Above we showed $\text{Var}(X) = E(X^2) - \mu_x^2$

$E(X^2) = \frac{1}{3} * 9 + \frac{1}{3} * 144 + \frac{1}{3} * 16 = 56.33$

$\mu_x^2 = (6.333)^2 = 40.111$

$E(X^2) - \mu_x^2 = 56.333 - 40.111 = 16.222$

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



Covariance

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

Covariance

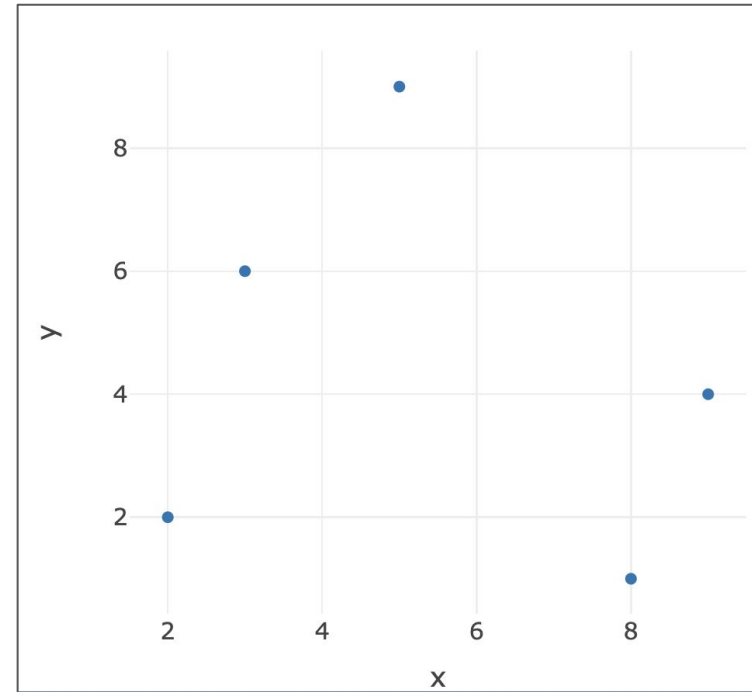
$$\begin{aligned} Cov(x, y) &= E[(X - E(X))(Y - E(Y))] \\ &= \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n - 1} \end{aligned}$$

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

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

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

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

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

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

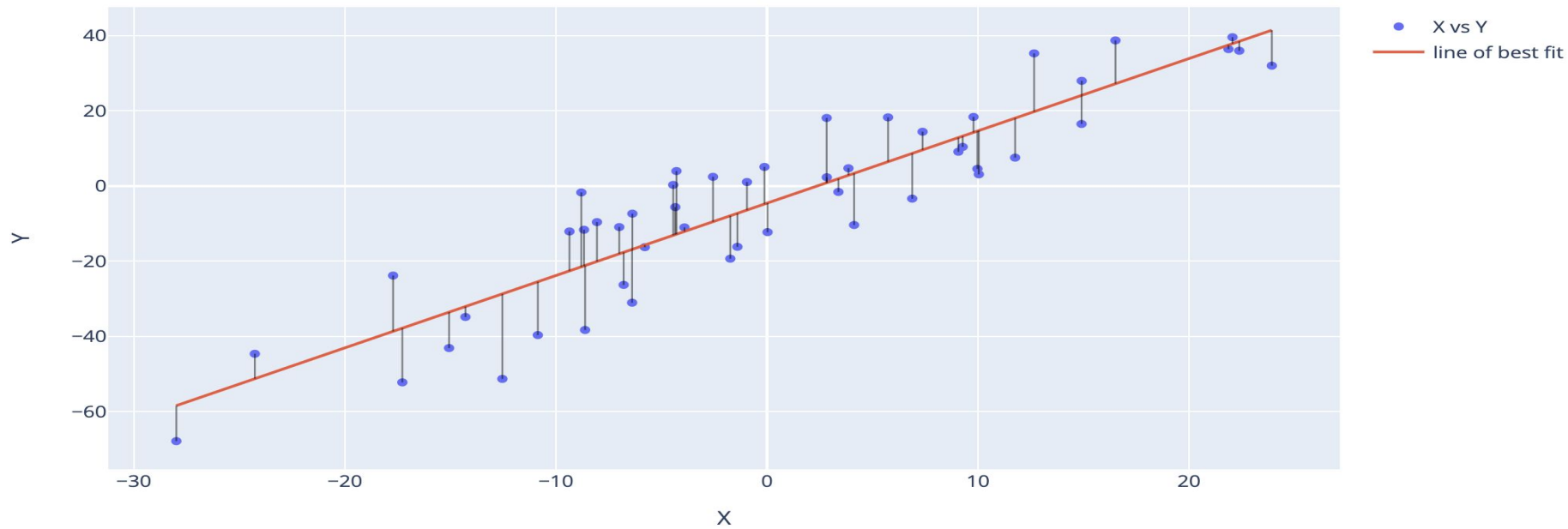
$$\hat{y}_i = \hat{a} + \hat{b} * x_i$$

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

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

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

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



Walk through example

x	y	demean_x	demean_x_sq	demean_y	demean_y_sq	demean_x*demean_y
3	6	-2.4	5.76	1.6	2.56	-3.84
5	9	-0.4	0.16	4.6	21.16	-1.84
2	2	-3.4	11.56	-2.4	5.76	8.16
8	1	2.6	6.76	-3.4	11.56	-8.84
9	4	3.6	12.96	-0.4	0.16	-1.44
			37.2		41.2	-7.8
						sum
						-1.95
						sum/(n-1)

mean_y	4.4
mean_x	5.4

yhat	y-yhat
b0 + b1*x	error
4.90	1.10
4.48	4.52
5.11	-3.11
3.85	-2.85
3.65	0.35

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

End of class form



<https://forms.gle/My9wHi2QFKNLedGC7>