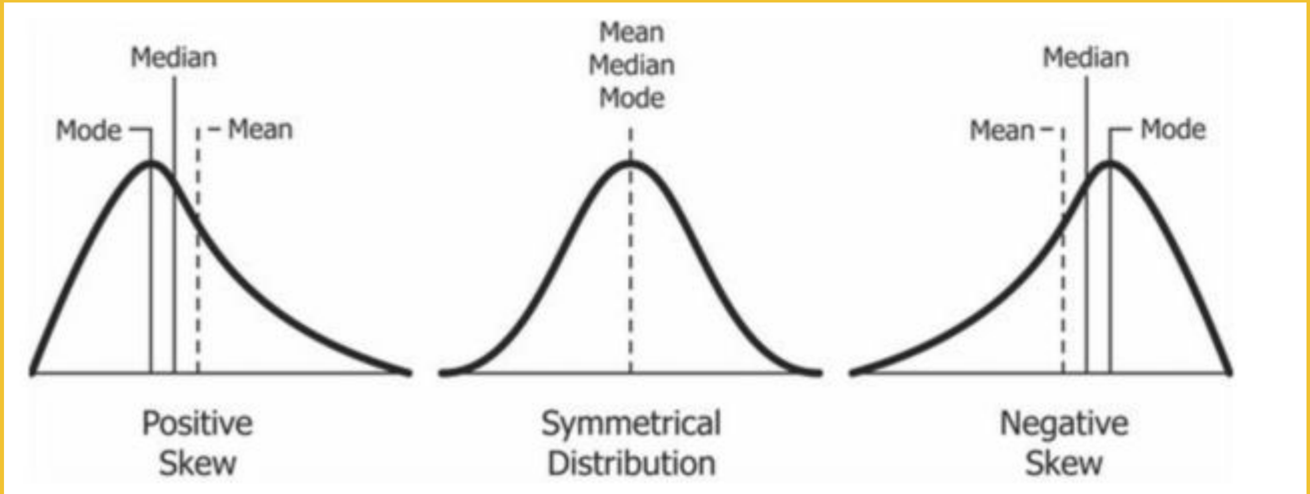


Econ 2250: Stats for Econ

Fall 2022

[Source for pic stats above.]

Announcements

- Homework 6 is due on Tuesday

Resources:

- https://www.probabilitycourse.com/chapter3/3_2_2_expectation.php
- https://mixtape.scunning.com/02-probability_and_regression#variance

What we will do today?

- Deep dive on summary operator
- Deep dive on Expected value
- Revisit Variance
- Revisit Covariance
- Introduce Correlation

Summary Operator

$$\Sigma X = x_1 + x_2 + \ldots + x_n$$

Summary

$$\sum_{i=1}^{n} x_i \equiv x_1 + x_2 + \ldots + x_n$$

Summary Operator Properties

1.) $\sum_{i=1}^{n} c = nc$

2.) $\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i$

3.) For any constant a and b: $\sum_{i=1}^{n}(ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{j=1}^{n} y_i$

Gotchas! Be Careful

$$\sum_{i}^{n} \frac{x_i}{y_i} \neq \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}$$

$$\sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i\right)^2$$

Summary Operator Property 1:

$$\sum_{i=1}^{n} c = nc$$

$$\Sigma_1^3 10 = 10 + 10 + 10 = 30 = 3 * 10$$

```
sum_x = 0
for i in range(3):
  sum_x = sum_x + 10
  print( sum_x)

10
20
30
```

x	cumsum
10	10
10	20
10	30

Summary Operator Property 2:

$$\sum_{i=1}^{n} c x_i = c \sum_{i=1}^{n} x_i$$

$$\Sigma_i^n x_i * c = (3 * 10) + (5 * 10) + (2 * 10)$$
$$= 10 * (3 + 5 + 2) = c * \Sigma_i^n x_i$$

```python
x = [3,5,2]
c = 10
sum_x = 0
for i in range(3):
  sum_x = sum_x + x[i] * c
  print( sum_x)

sum_x == sum(x)*c
```

```
30
80
100
True
```

Summary Operator Property 3:

For any constant a and b: $\sum_{i=1}^{n}(ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{j=1}^{n} y_i$

$$\Sigma_i^n (a * x_i + b * y_i) =$$
$$(a * x_1 + b * y_1) + (a * x_2 + b * y_2) + (a * x_3 + b * y_3) =$$
$$a * x_1 + b * y_1 + a * x_2 + b * y_2 + a * x_3 + b * y_3 =$$
$$a(x_1 + x_2 + x_3) + b(y_1 + y_2 + y_3) =$$
$$a\Sigma_i^n x_i + b\Sigma_i^n y_i$$

Show that…using x = (1,2,3)

$$\sum_{i}^{n} \frac{x_i}{y_i} \neq \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}$$

And

$$\sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i \right)^2$$

Expected Value Operator

$$E(x) = \Sigma\, x_i * Pr(x_i)$$

Expected Value Operator Property 1: $E(c) = c$

C = 10

E(c) = c * p(c) = c* 1 = c

While this is kind of obvious, it will come in handy in lots of proofs.

Expected Value Operator Property 2:

$$E(aX + b) = E(aX) + E(b) = aE(X) + b.$$

x = [3,6,2]

p(x) = ⅓

a = 5

b = 4

E(aX + b) = E(aX) + E(b)

$= ax_1*p(x_1) + ax_2*p(x_2) + ax_3*p(x_3) + b = a(x_1*p(x_1) + x_2*p(x_2) + x_3*p(x_3)) + b$

= a(E(x)) + b = 5(3*⅓ + 6*⅓ + 2*⅓) + 4 = 22⅓

An important extension of this linearity is that

$$E(W + H) = E(W) + E(H)$$

Variance

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

Variance is a measure of the spread of the data

We get the central tendency using the expected value E(X), and to get a measure of the spread of the data we take the expectation of the squared deviations

Expected value of X: $E[X] = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k$

which is the average for equally weighted data $E(X) = \frac{\sum x}{n} = \mu_x$

To get a deviation we subtract off the mean

$$\text{deviation of x} = X - \mu_x = X - E(X)$$

And square this so that it does sum to zero

$$\text{squared deviation of x} = (X - \mu_x)^2 = (X - E(X))^2$$

Example: demean squared

x= [3,12,4]

E(x) = mu = xbar = 3*⅓ + 12*⅓ + 4*⅓= (3+12+4)/3=6.3

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

Expectation is our best guess of what something will equal, so take the expectation of the squared deviation

$$E[(X - \mu_x)^2] = \sum (x_i - \mu_x)^2 * P(x_i)$$

if $P(x_i)$ is $\frac{1}{n}$ for all $i = 1, 2, \ldots, n$

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

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 = 16.2

But, notice our deviations (-3.3, 5.7, -2.3), 16.2 is an awful absolute value estimate. That is because we squared the errors, and x is in levels (not squared).

Standard Deviation = square root of σ^2, sqrt(16.2) = 4.02

Variance Overview

$$V(X) \equiv \sigma^2 = E[(X - E(X))^2]$$

Population model:

$$V(X) = \sigma^2 = E[(X - E(X))^2]$$
$$= E[(X - \mu_x)^2] = \sum (x_i - \mu_x)^2 * P(x_i)$$

if $P(x_i)$ is $\frac{1}{n}$ for all $i = 1, 2, \ldots, n$

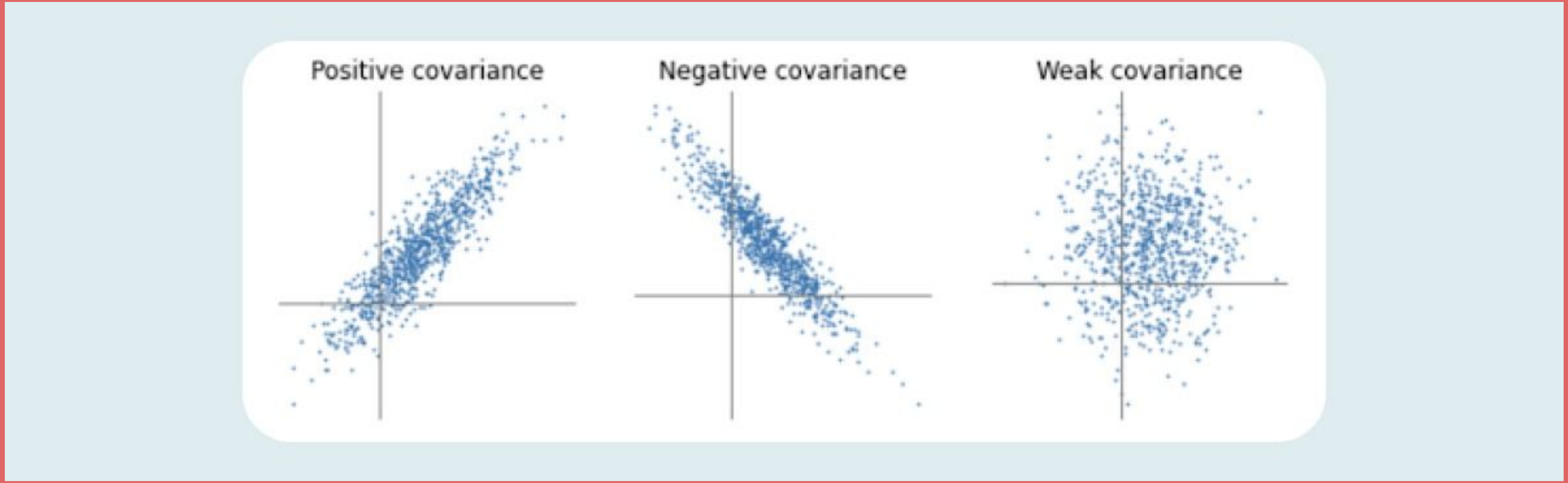
$$V(X) = \sum (x_i - \mu_x)^2 * \frac{1}{n} = \frac{1}{n} \sum (x_i - \mu_x)^2$$

bring squared values back the units of x

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

Nice correlation app

https://shiny.rit.albany.edu/stat/rectangles/



Covariance

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

image

Covariance

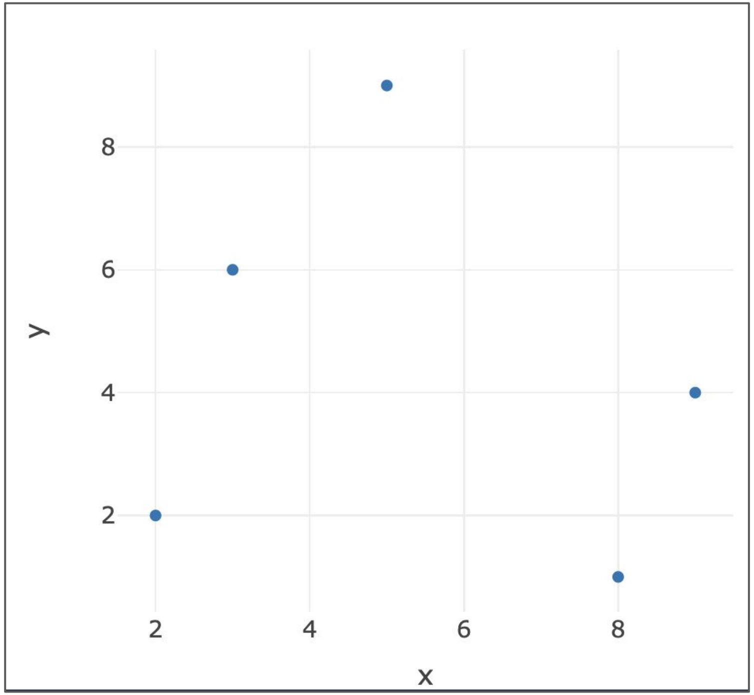
$$
\begin{aligned}
Cov(x, y) &= E[(X - E(X))(Y - E(Y))] \\
&= \frac{\sum (x_i - \mu_x)(y_i - \mu_y)}{n - 1}
\end{aligned}
$$

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



Correlation

$$\rho_{X,Y} = \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}, \quad \text{if } \sigma_X \sigma_Y > 0$$

$$r_{xy} \quad \stackrel{\text{def}}{=} \quad \frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{(n-1)s_x s_y} = \frac{\sum\limits_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i-\bar{x})^2 \sum\limits_{i=1}^{n}(y_i-\bar{y})^2}}$$

Example

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

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

numerator	**-1.95**	-1.95	**-0.22**	**correlation**
denom	sqrt(**37.2** + **41.2**)	8.85		

End of class form

https://forms.gle/kgT2w9wPZo3vJcjA8