Lecture 10

Topics:

I) More facts on $cov(X,Y)$ and $var(X)$

II) z-scores

$Cov(X,Y)$ & $var(X)$ continued

Fact: $var(aX+b) = a^2 var(X)$

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

This tells you how to re-compute E & var after changing units.

It is basically a consequence of MF1 & MF2:

$$var(aX+b) \overset{MF2}{=} a^2 var(X) + b^2 \overbrace{var(1)}^{=0} + 2ab \underbrace{cov(X,1)}_{=0}$$

$$= a^2 var(X)$$

$$E(aX+b) \overset{MF2}{=} aE(X) + b\underbrace{E(1)}_{=1}$$

$$= aE(X)+b$$

e.g. Randomly choose a student in class. Let X be their height in inches. Suppose

$E(X) = 65.2$ inches

$sd(X) = 5.1$ inches

Let $Y = 2.54X =$ height in cm.

Find $E(Y)$ & $sd(Y)$.

$$E(Y) = E(2.54X)$$

$$\overset{MF1}{=} 2.54\,E(X)$$

$$= 2.54(65.2) = 165.608 \text{ cm}$$

$$var(Y) \overset{MF2}{=} var(2.54X)$$

$$= (2.54)^2 var(X)$$

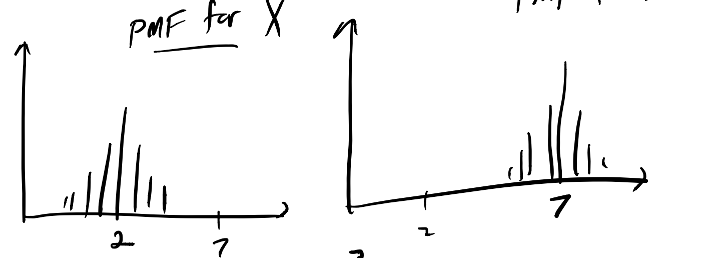
$$= (2.54)^2(5.1)^2 = 167.8 \text{ cm}^2$$

$$\therefore sd(Y) = \sqrt{167.8} \text{ cm.}$$

Note: It should make sense that

$$var(X+5) = var(X)$$

since $X+5$ is just a shifted version of X. So the variability doesn't change



same variability but different expected value.

Fact: $var(X+Y) = var(X)+var(Y)$ when X & Y are indep

$E(X+Y) = E(X)+E(Y)$ always.

Super important! Just a simple consequence of MF1 & MF2 when $a=1$, $b=1$ & $cov(X,Y)=0$.

Says: For indep R.V.s, var & E add!!!

Here is a simple consequence of this

Fact: If $X_1, X_2, \dots, X_n$ are indep where each X_i satisfies

$$E(X_i) = \mu$$

$$var(X_i) = \sigma^2$$

Then
$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n) = \mu + \cdots + \mu = n\mu$$

$$var(X_1 + \cdots + X_n) = var(X_1) + \cdots + var(X_n) = \sigma^2 + \cdots + \sigma^2 = n\sigma^2$$

e.g. Your friend just fliped a coin 100 times & got 75 heads. Is this unusual?

Let
$$X_1 = \begin{cases} 1 & \text{if } 1^{st} \text{ flip is heads} \\ 0 & \text{if } 1^{st} \text{ flip is tails} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if } 2^{nd} \text{ flip is heads} \\ 0 & \text{if } 2^{nd} \text{ flip is tails} \end{cases}$$

$\vdots$

$$X_{100} = \begin{cases} 1 & \text{if } 100^{th} \text{ flip is heads} \\ 0 & \text{if } 100^{th} \text{ flip is tails} \end{cases}$$

Your friend observed

$$X_1 + X_2 + \cdots + X_{100} = 75$$

Lets Find $E(X_1 + \cdots + X_{100})$ & $sd(X_1 + \cdots + X_{100})$ to see what we would expect.

For each $i = 1, \dots, 100$

x	$P_{X_i}(x)$	$xP_{X_i}(x)$	$x^2 P_{X_i}(x)$
0	1/2	0	0
1	1/2	1/2	1/2
	1	1/2	1/2

$\therefore$ $E(X_i) = \frac{1}{2}$

$$var(X_i) = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$

$\therefore$
$$E(X_1 + \cdots + X_{100}) = \underbrace{E(X_1)}_{1/2} + \cdots + \underbrace{E(X_{100})}_{1/2} = \frac{100}{2} = 50$$

$$var(X_1 + \cdots + X_{100}) = \underbrace{var(X_1)}_{1/4} + \cdots + \underbrace{var(X_{100})}_{1/4} = \frac{100}{4} = 25$$

$\therefore$ We would predict $X_1 + \cdots + X_{100}$ to be about

$$50 \pm 5$$

best prediction

$= \sqrt{25}$ = typical deviation from prediction

Your friend got 75 for $X_1 + \cdots + X_{100}$ which is $50 + 25$ ... i.e. 5-times the typical deviation from what we would have predicted.

$\therefore$ either your friend is lying or he has a trick coin!!

Z-scores

Another very important concept. Useful for identifying unlikely values of a random variable.

In the last e.g. we had a R.V. $Y = X_1 + \cdots + X_{100}$ which had

$$E(Y) = 50$$

$$sd(Y) = 5$$

we said "for Y to be 75 it would need to 5-times the typical deviation from what we would predict".

This is the z-score of 75 for Y.

This is E(Y)

This is sd(Y)

To find the "5-times" part we implicity did the following

$$\frac{75 - E(Y)}{sd(Y)} = 5$$

obs

subtract what we expect

divide by the typical deviation

Note rearranging

$$75 = \underbrace{E(Y)}_{\text{prediction}} + \underset{\text{5-times}}{5}\, \underbrace{sd(Y)}_{\text{typical dev.}}$$

Z-scores are themselfs R.V.s in the sense that the Z-score for Y is

$$Z = \frac{Y - E(Y)}{sd(Y)} = \text{The number of typical deviations } Y \text{ is from } E(Y).$$

is a R.V.

we can then analyze $E(Z)$ & $sd(Z)$ using MF1 & MF2.

$$E(Z) = E\left(\frac{Y - E(Y)}{sd(Y)}\right)$$

$$= E(aY + b) \quad \text{where } a = \frac{1}{sd(Y)}, \quad b = -\frac{E(Y)}{sd(Y)}$$

$$\overset{MF1}{=} aE(Y) + b$$

$$= \frac{E(Y)}{sd(Y)} - \frac{E(Y)}{sd(Y)}$$

$$= 0$$

also with the same def of a & b

$$var(Z) = var(aY + b)$$

$$\overset{MF2}{=} a^2 var(Y)$$

$$= \left(\frac{1}{sd(Y)}\right)^2 var(Y)$$

$$= \frac{1}{var(Y)} var(Y)$$

$$= 1$$

Therefore Z-scores always have expected value 0 with typical deviations ± 1!

z-scores give a unit free indication if a observation is unusual or not. 7

8

A z-score within ± 2 is Normal
A z-score $\approx$ 3 is kinda unlikely
A z-score $\approx$ 4 is rare
A z-score $\approx$ 5 is extremely rare.

e.g. Choose a random person from Cali.

Let X = height in inches.

Suppose $E(X) = 67$ & $sd(X) = 8$

Suppose X is observed to be 69.2

How far is the observed X from $E(X)$ in unit of

- inches : $X - E(X) = 69.2 - 67 = 2.2$
- ft : $\frac{X - E(X)}{12} = \frac{2.2}{12} = 0.183$
- standard deviation:

$$\frac{X - E(X)}{sd(X)} = \frac{2.2}{8} = \underbrace{0.275}_{\text{z-score}}$$