

LUBS2575

STATISTICS AND ECONOMETRICS

SEMESTER 1: STATISTICS

TOPIC 2

EXPECTED VALUES

- Expected value of a RV: $E(X)$
- Variance of a RV: $\text{var}(X)$
- Linear transformation of a RV: $Y = a + bX$

Expected Values

Discrete Random Variables

- ♦ Discrete RV: X
- ♦ Its probability distribution: $f(x)$

Expected value of X :

$$E(X) = \sum_x x f(x) \quad \text{for} \quad x = x_1, x_2, \dots, x_n \quad \text{or} \\ x = \dots, x_{i-1}, x_i, x_{i+1}, \dots$$

- summation of
- RV $X \times$ probability distribution of X
- over the whole set of possible values of X

Average:

$$\bar{X} = \sum_x x \frac{f}{n}$$



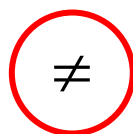
calculated from

relative

frequencies: $\frac{f}{n}$



SAMPLE



Expected value:

$$E(X) = \sum_x x f(x)$$



calculated from

probability

distribution: $f(x)$



POPULATION

← **but related** →

SAMPLE

← **but related** →

POPULATION



in the 'long run'
($n \rightarrow \infty$)

$$\frac{f}{n}$$



$$f(x)$$



in the 'long run'
($n \rightarrow \infty$)

$$\bar{X}$$



$$E(X)$$

sample mean
(or **average**)

population mean
(or **expected value**)

Other expectations:

$$E\left(X^2\right)=\sum_x x^2 f(x) \quad \text{for} \quad x = x_1, x_2, \dots, x_n$$

$$E[g(X)]=\sum_x g(x) f(x) \quad \text{for} \quad x = x_1, x_2, \dots, x_n$$

Example

X : number of copies of a magazine sold per day by a newsagent

Probability distribution of X :

$$f(x) = \frac{x}{6}, \quad x = 1, 2, 3$$

$$x = 1 \rightarrow f(x) = \frac{1}{6}$$

$$x = 2 \rightarrow f(x) = \frac{2}{6}$$

$$x = 3 \rightarrow f(x) = \frac{3}{6}$$

Expected daily sales:

$$E(X) = \sum_{x=1}^3 x f(x) = 1\left(\frac{1}{6}\right) + 2\left(\frac{2}{6}\right) + 3\left(\frac{3}{6}\right) = \frac{1}{6}[1(1) + 2(2) + 3(3)] =$$

$$= \frac{1}{6}(1 + 4 + 9) = \frac{14}{6} = \frac{7}{3} = 2.333 = 2.\bar{3}$$

"2.3 repeating"

ie, the expected (mean) daily sales of the magazine are $\frac{7}{3}$ (or 2.333) copies

Aside

Rounding method: “*round to nearest*”

Important for MCQ!

2.49	2
−2.49	−2
2.50	3
−2.50	−3
2.51	3
−2.51	−3

Rounding method used by **Excel**

Expected value operations

Derived from properties of the Σ operator:

(i) $E(a) = a$ where a is a constant

(ii) $E(aX) = aE(X)$

(iii) $E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$

Also:

$$E[g_1(X) \ominus g_2(X)] = E[g_1(X)] \ominus E[g_2(X)]$$

Continuous Random Variables

- ◆ Continuous RV: X
- ◆ Its *pdf*: $f(x)$

Expected value of X :

$$E(X) = \int_x x f(x) dx \quad \text{for} \quad a < x < b \quad \text{or} \\ -\infty < x < +\infty$$

- integral of
- RV $X \times$ *pdf* of X
- over the whole range of possible values of X

Other expectations:

$$E\left(X^2\right)=\int_x x^2 f(x)dx \quad \text{for} \quad a < x < b$$

$$E[g(X)]=\int_x g(x) f(x)dx \quad \text{for} \quad a < x < b$$

Expected value operations

Derived from properties of the \int operator:

(i) $E(a) = a$ where a is a constant

(ii) $E(aX) = aE(X)$

(iii) $E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$

Also:

$$E[g_1(X) \ominus g_2(X)] = E[g_1(X)] \ominus E[g_2(X)]$$

Example

Example given in Topic 1:

X : daily sales of petrol (in '000 litres) in a petrol station

Probability density function (*pdf*) of X :

$$f(x) = \frac{x}{2}, \quad 0 < x < 2$$

Expected daily sales:

$$\begin{aligned} E(X) &= \int_{x=0}^2 x f(x) dx = \int_{x=0}^2 x \left(\frac{x}{2} \right) dx = \int_{x=0}^2 \frac{x^2}{2} dx = \frac{1}{2} \int_{x=0}^2 x^2 dx = \\ &= \frac{1}{2} \left[\frac{x^3}{3} \right]_0^2 = \frac{1}{6} \left[x^3 \right]_0^2 = \frac{1}{6} [8 - 0] = \frac{4}{3} = 1.333 = 1.\bar{3} \end{aligned}$$

ie, the expected (mean) daily sales of petrol are $\frac{4}{3}$ (or 1.333)
thousand litres, ie 1,333 litres

Variance of X

$\text{var}(X)$ is defined as

mean squared deviation of X around $E(X)$ (its mean)

Also called

Second moment of X measured from $E(X)$ (its mean)

In terms of **expectations**:

$$\begin{aligned} E\left\{[X - E(X)]^2\right\} &= E\left\{X^2 - 2XE(X) + [E(X)]^2\right\} = \\ &= E(X^2) - E[2XE(X)] + E\{[E(X)]^2\} = \\ &= E(X^2) - E[2XE(X)] + [E(X)]^2 = \\ &= E(X^2) - 2E(X)E(X) + [E(X)]^2 = \\ &= E(X^2) - 2[E(X)]^2 + [E(X)]^2 = \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

Sample variance:

$$s^2 = \left(\frac{n}{n-1} \right) \sum_x (x - \bar{X})^2 \frac{f}{n}$$



calculated from

relative

frequencies: $\frac{f}{n}$



SAMPLE

\neq



Population variance:

$$\text{var}(X) = E\{[X - E(X)]^2\}$$



calculated from

probability

distribution: $f(x)$



POPULATION

← **but related** →

SAMPLE

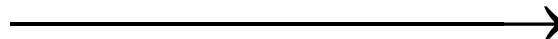
← **but related** →

POPULATION



in the 'long run'
($n \rightarrow \infty$)

$$\frac{f}{n}$$



$$f(x)$$



in the 'long run'
($n \rightarrow \infty$)

$$s^2$$



$$\text{var}(X)$$

sample variance

population variance

Example

X : daily sales of petrol (in '000 litres) in a petrol station

Probability density function (*pdf*) of X :

$$f(x) = \frac{x}{2}, \quad 0 < x < 2$$

It was found that

$$E(X) = \frac{4}{3}$$

To calculate

$$\text{var}(X) = \boxed{E(X^2)} - [\boxed{E(X)}]^{\textcircled{2}}$$

we first find $\boxed{E(X^2)}$:

$$E(\textcolor{blue}{X}^2) = \int_{x=0}^2 \textcolor{blue}{x}^2 f(x) dx = \int_{x=0}^2 x^2 \left(\frac{x}{2}\right) dx = \int_{x=0}^2 \frac{x^3}{2} dx = \frac{1}{2} \int_{x=0}^2 x^3 dx =$$

$$= \frac{1}{2} \left[\frac{x^4}{4} \right]_0^2 = \frac{1}{8} [x^4]_0^2 = \frac{1}{8} (16 - 0) = \boxed{2}$$

Therefore

$$\begin{aligned}\text{var}(X) &= E(X^2) - [E(X)]^2 = \\ &= 2 - \left(\frac{4}{3}\right)^2 = 2 - \frac{16}{9} = \frac{2}{9} = 0.222 = 0.\bar{2}\end{aligned}$$

Thus the standard deviation of X is

$$SD(X) = \sqrt{\text{var}(X)} = \sqrt{\frac{2}{9}} = 0.471$$

Linear transformations of X

General expression for a linear transformation of X :

$$Y = a + bX$$

where a and b are constants

If we know $E(X)$ and $\text{var}(X)$ \rightarrow we can find $E(Y)$ and $\text{var}(Y)$
without knowing $f(y)$

Expected value of Y :

$$E(Y) = E(a + bX) =$$

$$= E(a) + E(bX) =$$

$$= \boxed{a + bE(X)}$$

knowing $E(X)$ \rightarrow derive $E(Y)$ without knowing $f(y)$

Variance of Y :

$$\begin{aligned}\text{var}(Y) &= E\{[Y - E(Y)]^2\} = \\&= E\{[(a + bX) - E(a + bX)]^2\} = \\&= E\{[(a + bX) - E(a + bX)]^2\} = \\&= E\{[(a + bX) - (E(a) + E(bX))]\}^2\} = \\&= E\{[(a + bX) - (a + bE(X))]\}^2\} = \\&= E\{[\cancel{a} + bX - \cancel{a} - bE(X)]^2\} =\end{aligned}$$

$$= E\{[bX - bE(X)]^2\} =$$

$$= b^2 \underbrace{E\{[X - E(X)]^2\}}_{\text{var}(X)} =$$

$$= b^2 \text{var}(X)$$

knowing $\text{var}(X)$ \rightarrow derive $\text{var}(Y)$ without knowing $f(y)$

Also:

$$Y = a \ominus bX$$

$$\text{var}(Y) = b^2 \text{var}(X) \quad \text{same as above}$$

Standard deviation of Y :

$$\begin{aligned} SD(Y) &= \sqrt{\text{var}(Y)} = \sqrt{b^2 \text{var}(X)} = b\sqrt{\text{var}(X)} = \\ &= \boxed{b SD(X)} \end{aligned}$$

knowing $SD(X)$ \rightarrow derive $SD(Y)$ without knowing $f(y)$

NB: These results on

- variance
- linear transformations

apply to both **discrete** and **continuous** RVs

Example

X : number of copies of a magazine sold per day by a newsagent

The newsagent's profit per copy is 9p and his overhead cost of stocking the magazine is 15p per day

Daily profit (Y) is related to sales (X) by

$$Y = -15 + 9X$$

It was found earlier that $E(X) = \frac{7}{3}$

It can be found that $\text{var}(X) = \frac{5}{9}$ (try this at home)

$$Y = \overbrace{-15}^a + \overbrace{9X}^b$$

The expected daily profit can be found:

$$\begin{aligned} E(Y) &= a + bE(X) \\ &= -15 + 9E(X) = -15 + 9\left(\frac{7}{3}\right) = -15 + 3(7) = -15 + 21 = 6 \end{aligned}$$

NB: We found the expected daily profit $E(Y)$ without knowing $f(y)$, which would be needed to obtain $E(Y) = \sum_y y f(y)$ via the standard definition

$$Y = \overbrace{-15}^a + \overbrace{9X}^b$$

The variance of daily profit can be found:

$$\begin{aligned}\text{var}(Y) &= b^2 \text{var}(X) \\ &= 9^2 \text{var}(X) = 9^2 \left(\frac{5}{9} \right) = 9(5) = 45\end{aligned}$$

NB: We found the variance of daily profit $\text{var}(Y)$ **without knowing $f(y)$** , which would be needed to obtain

$$\text{var}(Y) = E(Y^2) - [E(Y)]^2 \quad \text{via the standard definition}$$