Leeds University
Business School

UNIVERSITY OF LEEDS

LUBS2575

STATISTICS AND ECONOMETRICS

SEMESTER 1: STATISTICS

TOPIC 2

EXPECTED VALUES

- Expected value of a RV: $E(X)$
- Variance of a RV: $\operatorname{var}(X)$
- 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:

$$E(X) = \sum_x x\, f(x) \qquad \text{for} \qquad x = x_1, x_2, \ldots, x_n \qquad \text{or}$$

$$x = \ldots, x_{i-1}, x_i, x_{i+1}, \ldots$$

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

Average:

$$\overline{X} = \sum_x x \, \frac{f}{n}$$

$\uparrow$

calculated from

relative frequencies: $\dfrac{f}{n}$

$\uparrow$

SAMPLE

$\ne$

$\leftarrow$ **but related** $\rightarrow$

Expected value:

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

$\uparrow$

calculated from

probability distribution: $f(x)$

$\uparrow$

POPULATION

$$
\begin{array}{ccc}
\textbf{SAMPLE} & \leftarrow \ \textbf{but related} \ \rightarrow & \textbf{POPULATION} \\
\downarrow & & \downarrow \\
\dfrac{f}{n} & \xrightarrow[]{\text{in the `long run'} \ (n \to \infty)} & f(x) \\
\downarrow & & \downarrow \\
\overline{X} & \xrightarrow[]{\text{in the `long run'} \ (n \to \infty)} & E(X) \\
\textbf{sample mean} & & \textbf{population mean} \\
(\text{or } \textbf{average}) & & (\text{or } \textbf{expected value})
\end{array}
$$

Other expectations:

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

$$E\left[g(X)\right] = \sum_x g(x)\, f(x) \qquad \text{for} \qquad x = x_1, x_2, \ldots, x_n$$

Example

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

Probability distribution of X:

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

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

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

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

Expected daily sales:

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

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

"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:

$$E[g_1(X) - g_2(X)] = E[g_1(X)] - E[g_2(X)]$$

Continuous Random Variables

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

Expected value of X:

$$E(X) = \int_x x\, f(x)\, dx \qquad \text{for} \qquad a < x < b \qquad \text{or}$$

$$-\infty < x < +\infty$$

- integral of
- RV X × *pdf* of X
- over the whole range of possible values of X

Other expectations:

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

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

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:

$$E[g_1(X) - g_2(X)] = E[g_1(X)] - E[g_2(X)]$$

Example

Example given in Topic 1:

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

Probability density function (*pdf*) of X:

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

Expected daily sales:

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

$$= \frac{1}{2}\left[\frac{x^3}{3}\right]_0^2 = \frac{1}{6}\left[x^3\right]_0^2 = \frac{1}{6}\left[8-0\right] = \frac{4}{3} = 1.333 = 1.\overline{3}$$

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

Variance of *X*

$\mathrm{var}(X)$ 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**:

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

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

$$= E\left(X^2\right) - E\left[2XE(X)\right] + [E(X)]^2 =$$

$$= E\left(X^2\right) - 2E(X)E(X) + [E(X)]^2 =$$

$$= E\left(X^2\right) - 2[E(X)]^2 + [E(X)]^2 =$$

$$= \boxed{E\left(X^2\right) - [E(X)]^2}$$

Sample variance: $\neq$ **Population variance**:

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

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

$\uparrow$ $\uparrow$

calculated from calculated from

relative frequencies: $\dfrac{f}{n}$

probability distribution: $f(x)$

$\uparrow$ $\uparrow$

SAMPLE $\leftarrow$ **but related** $\rightarrow$ **POPULATION**

$$
\begin{array}{ccc}
\textbf{SAMPLE} & \leftarrow \ \textbf{but related} \ \rightarrow & \textbf{POPULATION} \\
\downarrow & & \downarrow \\
\dfrac{f}{n} & \xrightarrow[]{\text{in the ‘long run’ } (n \to \infty)} & f(x) \\
\downarrow & & \downarrow \\
s^2 & \xrightarrow[]{\text{in the ‘long run’ } (n \to \infty)} & \mathrm{var}(X) \\
\textbf{sample variance} & & \textbf{population variance}
\end{array}
$$

Example

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

Probability density function (*pdf*) of X:

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

It was found that $\boxed{E(X) = \frac{4}{3}}$

To calculate

$$\operatorname{var}(X) = \boxed{E\left(X^2\right)} - [\boxed{E(X)}]^{2}$$

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

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

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

Therefore

$$\operatorname{var}(X) = \boxed{E\left(X^2\right)} - \left[\,\boxed{E(X)}\,\right]^{2} =$$

$$= \boxed{2} - \left(\boxed{\frac{4}{3}}\right)^{2} = 2 - \frac{16}{9} = \frac{2}{9} = 0.222 = 0.\overline{2}$$

Thus the standard deviation of X is

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

Linear transformations of X

General expression for a linear transformation of X :

$$\boxed{Y = a + bX}$$

where a and b are constants

If we know $E(X)$ and $\mathrm{var}(X)$ $\rightarrow$ we can find $E(Y)$ and $\mathrm{var}(Y)$

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:

$$\mathrm{var}(Y) = E\left\{[Y - E(Y)]^2\right\} =$$

$$= E\left\{[\overbrace{(a + bX)} - E\overbrace{(a + bX)}]^2\right\} =$$

$$= E\left\{[(a + bX) - E(a + bX)]^2\right\} =$$

$$= E\left\{[(a + bX) - (E(a) + E(bX))]^2\right\} =$$

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

$$= E\left\{[\cancel{a} + bX - \cancel{a} - bE(X)]^2\right\} =$$

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

$$= b^{2}\, \underbrace{E\left\{\left[X - E(X)\right]^{2}\right\}}_{\operatorname{var}(X)} =$$

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

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

Also:

$$Y = a - bX$$

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

Standard deviation of Y:

$$SD(Y) = \sqrt{\operatorname{var}(Y)} = \sqrt{b^2 \operatorname{var}(X)} = b\sqrt{\operatorname{var}(X)} =$$

$$= \boxed{b\,SD(X)}$$

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 $\operatorname{var}(X) = \frac{5}{9}$ (try this at home)

$$Y = \underbrace{-15}_{a} + \underbrace{9}_{b} X$$

The expected daily profit can be found:

$$E(Y) = \quad a + bE(X)$$

$$= -15 + 9E(X) = -15 + 9\left(\frac{7}{3}\right) = -15 + 3(7) = -15 + 21 = 6$$

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

$$Y = \underbrace{-15}_{a} + \underbrace{9}_{b} X$$

The variance of daily profit can be found:

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

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

$\operatorname{var}(Y) = E\left(Y^2\right) - [E(Y)]^2$ via the standard definition