Topic 2: Review of Probability and Statistics

- In the discussion below we make use of the **summation operator** and its properties.
- Suppose that x is a variable that can assume the values $x_1, x_2, \ldots.., x_n$. We define

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

- The summation operator can be shown to have the following properties:

P1 If c is a constant then

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

For example,

$$\sum_{i=1}^{3} c = c + c + c = 3c.$$

P2

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

For example,

$$\begin{aligned}\sum_{i=1}^{n} 4x_i &= 4x_1 + 4x_2 + ... + 4x_n \\ &= 4(x_1 + x_2 + ... + x_n) \\ &= 4\sum_{i=1}^{n} x_i .\end{aligned}$$

P3 If x is a variable that can assume the values $x_1, x_2, ..., x_n$ and y is a variable that can assume the values $y_1, y_2, ..., y_n$ and c and c_2 are constants then

$$\begin{aligned}\sum_{i=1}^{n}(c_1 x_i + c_2 y_i) &= \sum_{i=1}^{n} c_1 x_i + \sum_{i=1}^{n} c_2 y_i \\ &= c_1 \sum_{i=1}^{n} x_i + c_2 \sum_{i=1}^{n} y_i .\end{aligned}$$

P4

$$\begin{aligned}
\sum_{i=1}^{n}(c_1 x_i + c_2 y_i)^2 &= \sum_{i=1}^{n}[c_1^2 x_i^2 + c_2^2 y_i^2 + 2c_1 c_2 x_i y_i] \\
&= \sum_{i=1}^{n} c_1^2 x_i^2 + \sum_{i=1}^{n} c_2^2 y_i^2 + \sum_{i=1}^{n} 2c_1 c_2 x_i y_i \\
&= c_1^2 \sum_{i=1}^{n} x_i^2 + c_2^2 \sum_{i=1}^{n} y_i^2 + 2c_1 c_2 \sum_{i=1}^{n} x_i y_i .
\end{aligned}$$

- We can use the properties of the summation operator to show that the sum of the deviations from the sample mean is always equal to zero. That is,

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0,$$

where

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

- Using the properties of the summation operator we have

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i - \overline{x}) &= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \overline{x} \\
&= \sum_{i=1}^{n} x_i - n\overline{x} \\
&= \sum_{i=1}^{n} x_i - n\sum_{i=1}^{n} \frac{1}{n} x_i \\
&= \sum_{i=1}^{n} x_i - \frac{n}{n}\sum_{i=1}^{n} x_i \\
&= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i \\
&= 0.
\end{aligned}
$$

- A **random experiment** is a chance mechanism with the following properties:
  - All possible outcomes are known in advance.
  - In a particular trial the outcome is not known in advance.
  - In principle, the experiment can be repeated under identical conditions.
- Consider the experiment of tossing a fair coin twice. The set of possible outcomes of the experiment, which is called the **sample space** for the experiment, and which we will denote by S, is:

$$S = \{(h, h), (h, t), (t, t), (t, h)\}. \tag{1}$$

- Notice that S is known in advance, but in a particular trial, we don't know which of these four possible outcomes will occur.

- An **event** is any subset of S. For example, the event of obtaining two heads in two tosses of a fair coin is the subset of S given by the set

$$E_1 = \{(h, h)\}.$$

- The event of obtaining one head and one tail in two tosses of a fair coin is the subset of S given by

$$E_2 = \{(h, t), (t, h)\}.$$

- An event is said to occur when any of the elements outcomes that define the event occur.
- For example, the event E_2 is said to occur if the outcome of the experiment is either

$$(h, t) \text{ or } (t, h).$$

- Given a random experiment and its associated sample space, we can attach numbers called probabilities to the elements of the sample space and to events defined on the sample space.
- These numbers must satisfy the five laws stated below.
- Suppose that a random experiment has m possible outcomes and let p_i denote the probability of the *ith* outcome. Then:
  - $$0 \leq p_i \leq 1. \qquad \text{L1}$$
  - $$\sum_{i=1}^{m} p_i = p_1 + p_2 + ... + p_m = 1. \qquad \text{L2}$$
  - If A and B are two events then

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B). \qquad \text{L3}$$

- If A and B are mutually exclusively events (cannot both occur) then

$$P(A \text{ or } B) = P(A) + P(B). \qquad \text{L4}$$

- If A and B are two events than

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}. \qquad \text{(L5)}$$

- A **random variable**, X, is a variable associated with a random experiment which assumes a set of countable values, each with a specified probability.
- For example, X denote the number of heads obtained when a fair coin is tossed twice.
- The set of values that X can assume, which we denote by S_X is given by

$$S_X = \{0, 1, 2\}.$$

- Let

$$p_0 \equiv P(X = 0), p_1 \equiv P(X = 1), p_2 \equiv P(X = 2).$$

Intuitively, it is clear that

$$p_0 = \frac{1}{4}, p_1 = \frac{2}{4} = \frac{1}{2}, p_2 = \frac{1}{4}. \tag{2}$$

- Note that L1 and L2 above are satisfied since

$$0 \leq p_i \leq 1, i = 1, 2, 3$$

and

$$\sum_{i=1}^{3} p_i = 1.$$

- Because

$$S_X = \{0, 1, 2\}.$$

is a countable set (you can count the elements in the set), we call X a **discrete random variable**.

- More generally, a random variable X which can assume the countable set of values,

$$S_X = \{x_1, x_2, ..., x_m\},$$

with respective probabilities $p_1, p_2, ..., p_m,$ is a discrete random variable.

- Note that we use X to denote the random variable and we use x_i to denote the *ith* value of the random variable X.
- The **probability density function** (pdf) of a discrete random variable X is a function

$$f : S_X \to [0, 1] \tag{3}$$

with the following properties:

$$f(x_i) = P(X = x_i) = p_i \in [0, 1], i = 1, 2, ...m \tag{4}$$

and

$$\sum_{i=1}^{m} p_i = 1. \tag{5}$$

- In some experiments the set of values that a random variable can take on is not a countable set.
- Let X denote the height of a randomly selected male from the population of males in Australia.
- Since there is an infinite number of values that X could assume, the set of possible outcomes is no longer countable and we can no longer assign a probability to X assuming a specific value.
- Because the set of possible outcomes for X is no longer countable, we call X a **continuous random variable**.
- When X is a continuous random variable, we partition the (uncountable) set of possible outcomes into intervals of real numbers and we assign probabilities to X falling in these intervals.

- If X is a continuous random variable, the pdf of X is defined to be a function $f(x)$ with the following properties:

$$\int_{-\infty}^{a} f(x)dx = P(X \leq a), \tag{6}$$

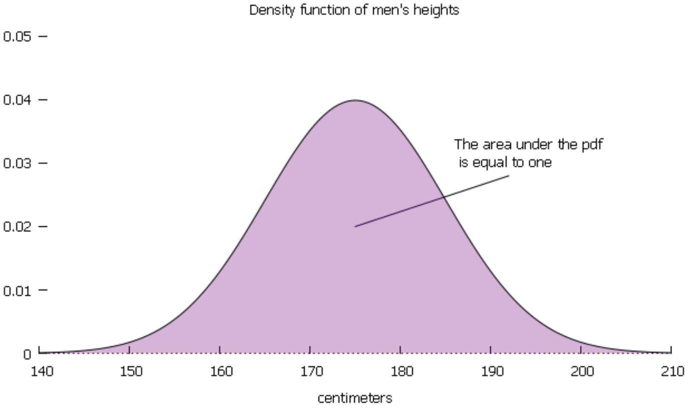
$$\int_{a}^{b} f(x)dx = P(a \leq X \leq b), \tag{7}$$

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{8}$$

- If X can only assume values in the interval [a,b], then (8) reduces to

$$\int_{-\infty}^{\infty} f(x)dx = \int_{a}^{b} f(x)dx = 1. \tag{9}$$

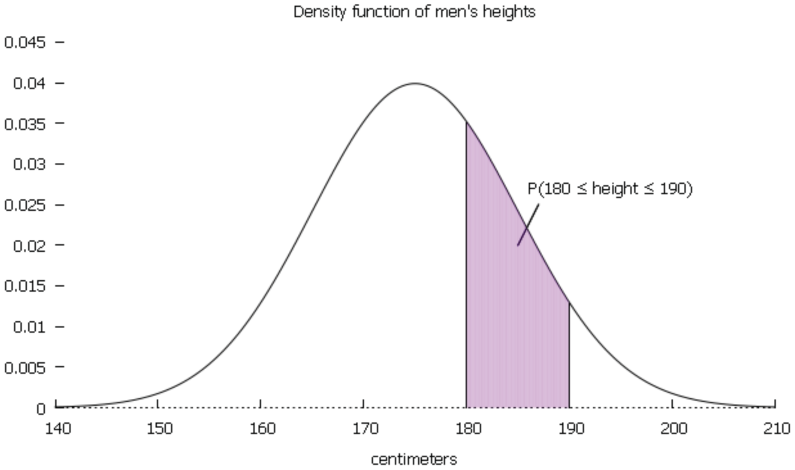
- The probability density function of men's heights is shown in Figure 1 below.



Density function of men's heights
0.05
0.04
0.03
0.02
0.01
0
140
150
160
170
180
190
200
210
centimeters
The area under the pdf is equal to one

Figure: 1

- Note that in the case of a discrete random variable we assign probabilities to specific real numbers and in the case of continuous random variables we assign probabilities to intervals of real numbers.
- This fact is illustrated in Figure 2 below.



Density function of men's heights
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
P(180 ≤ height ≤ 190)
140
150
160
170
180
190
200
210
centimeters

Figure: 2

- Let X be a discrete random variable which can take on the values $(x_1, x_{2}, \ldots\ldots, x_n)$ with probabilities $f(x_1)$, $f(x_2)$, ..., $f(x_n)$ respectively. The **mean** or **expected value** of X, which we denote by $E(X)$, is defined as

$$E(X) = \sum_{i=1}^{n} x_i f(x_i). \tag{10}$$

- The expected value of a discrete random variable is a measure of the "average value" of the random variable.
- If X is a continuous random variable with probability density function (pdf) $f(x)$, then

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx. \tag{11}$$

- Strictly speaking, we should denote the pdf of X as $f_X(x)$. However, to economize on notation we shall denote it as $f(x)$.

- Notice that $E(X)$ is a constant, not a random variable.
- For any set of random variables X, Y and Z, the following rules can be shown to apply to the expectations operator:

R1 The expectation of a sum of random variables is the sum of their expectations (provided that these expectations exist). That is,

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

R2 For any constant k

$$E(k) = k.$$

R3 For any constants a and b,

$$\begin{aligned} E(a+bX) &= E(a)+bE(X) \\ &= a+bE(X). \end{aligned}$$

For example,

$$E(4+10X) = 4+10E(X).$$

- Because it satisfies R1 and R3, the expectations operator is called a **linear operator**. That is, the expectations operator "goes through" linear transformations of a random variable and linear combinations of several random variables.
- However, it does not go through **non-linear transformations** of a random variable or non-linear combinations of several random variables.

- For example

$$
\begin{aligned}
E(X^2) &\neq [E(X)]^2, \\
E[\log(X)] &\neq \log[E(X)], \\
\text{In general, } E(XY) &\neq E(X)E(Y).
\end{aligned}
$$

- The **variance** of the random variable X, which we denote by $Var(X)$, is defined as

$$\begin{aligned} Var(X) &= E\{[X - E(X)]^2\} \\ &= E[X^2 - 2XE(X) + E(X)^2] \\ &= E(X^2) - E[2XE(X)] + E[E(X)^2] \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 \\ &= E(X^2) - 2E(X)^2 + E(X)^2 \\ &= E(X^2) - [E(X)]^2. \end{aligned} \tag{12}$$

- Equation (12) states that $Var(X)$ "is equal to the expectation of X squared minus the square of the expectation of X".
- The notation σ_X^2 is also used to denote the variance of the random variable X.

- Loosely speaking, $Var(X)$ measures how tightly clustered the values of X are around the mean of X.
- A disadvantage of $Var(X)$ as a measure of the dispersion of X is that it is difficult to interpret.
- An alternative measure of the dispersion of the random variable X, called the standard deviation of X, which we denote by σ_x or $sd(X)$, is defined as
$$\sigma_x = \sqrt{Var(X)}.$$
- Note: We should really use the notation σ_X, since we are referring to the standard deviation of the random variable X. However, it is common in econometrics to use the notation σ_x.
- An advantage of σ_x as a measure of dispersion is that, unlike $Var(X)$, it always has the same units as X.

- For example, if X is measured in dollars, then σ_x will also be measured in dollars.

- We are often interested in whether or not two random variables "move together" and, if they do move together, how strong is the covariation.
- One measure of covariation between two random variables, X and Y, is called the **covariance** between X and Y, which we denote by $Cov(X, Y)$ and which is defined as

$$\begin{aligned} Cov(X, Y) &= E\{[X - E(X)][Y - E(Y)]\} \\ &= E[XY - E(X)Y - XE(Y) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y). \end{aligned} \tag{13}$$

- The symbol σ_{xy} is often used to denote $Cov(X, Y)$.

- Equation (13) states that the $Cov(X, Y)$ is equal to "the expectation of their product minus the product of their expectations".
- Note that if follows from (13) that in the special case in which

$$E(X) = 0 \text{ and/or } E(Y) = 0,$$

the formula for the covariance reduces to

$$Cov(XY) = E(XY). \tag{14}$$

- Since

$$Cov(X, Y) = E\{[X - E(X)][Y - E(Y)]\}, \tag{15}$$

it follows that:

- If
$$Cov(X, Y) > 0$$
then, **on average**,
$$X > E(X) \Rightarrow Y > E(Y)$$
and
$$X < E(X) \Rightarrow Y < E(Y).$$
- If
$$Cov(X, Y) < 0$$
then, **on average**,
$$X > E(X) \Rightarrow Y < E(Y)$$
and
$$X < E(X) \Rightarrow Y > E(Y).$$

- It can be shown that In the special case in which X and Y are independently distributed random variables,

$$Cov(X, Y) = 0.$$

- Note that while

$$\text{independence of } X \text{ and } Y \Rightarrow Cov(X, Y) = 0,$$

the converse of this proposition is not true. That is,

$$Cov(X, Y) = 0 \nRightarrow \text{independence of } X \text{ and } Y.$$

- A major limitation of the covariance as a measure of association is that its magnitude is sensitive to the units in which X and Y are measured. For example,

$$Cov(aX, bY) = abCov(X, Y).$$

  That is, if we scale both X and Y by a, the covariance is scaled by a^2.
- For example,

$$Cov(100X, 100Y) = (100)^2 Cov(X, Y).$$

- Consequently, only the sign of $Cov(X, Y)$ is informative (can be interpreted).

- A superior measure of association between X and Y is the **correlation** between X and Y, which we denote by $Corr(X, Y)$, and which is defined as

$$Corr(X, Y) = \frac{Cov(X, Y)}{sd(X)sd(Y)}.$$

- The symbol ρ_{xy} is often used to denote $Corr(X, Y)$.
- As a measure of association $Corr(X, Y)$ has two attractive properties:

  P1 It can be proved that

$$-1 \leq Corr(X, Y) \leq 1.$$

  This property makes the correlation coefficient ρ_{xy} easy to interpret.

- If
$$Corr(X, Y) = 1$$
there is an exact positive linear relationship between X and Y. That is, if we plot the values of X and Y they will form a positively sloped straight line in (X, Y) space.
- The closer the $Corr(X, Y)$ is to 1, the stronger the positive linear association between X and Y.
- If
$$Corr(X, Y) = -1$$
there is an exact negative linear relationship between X and Y. That is, if we plot the values of X and Y they will form a negatively sloped straight line in (X, Y) space.
- The closer the $Corr(X, Y)$ is to -1, the stronger the negative linear association between X and Y.

P2 The magnitude of the $Corr(X, Y)$ is independent of the units in which X and Y are measured. That is,

$$Corr(aX, bY) = Corr(X, Y).$$

For example,

$$Corr(100X, 100Y) = Corr(X, Y).$$

- Now that we have defined $Cov(X, Y)$, we are in a position to discuss the properties of the variance of a random variable and of sums of random variables.

P1

$$Var(k) = 0.$$

P2

$$Var(kX) = k^2 Var(X).$$

P3

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y).$$

P4

$$Var(aX - bY) = a^2 Var(X) + b^2 Var(Y) - 2abCov(X, Y).$$

P5 In the special case in which X and Y are independently distributed random variables,

$$Cov(X, Y) = 0$$

and P3 and P4 respectively reduce to

$$\begin{aligned} Var(aX + bY) &= a^2 Var(X) + b^2 Var(Y), \\ Var(aX - bY) &= a^2 Var(X) + b^2 Var(Y). \end{aligned}$$

- Given two random variables X and Y, it is important to distinguish between the **population parameters** associated with these random variables and the **sample statistics** associated with a sample of n observations drawn from the probability distributions of these random variables.
- Given a sample of n observations on the random variable X, we define the following sample statistics:

$$\text{sample mean} \quad : \quad \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

$$\text{sample variance} \quad : \quad \widehat{\sigma}_x^2 = \frac{1}{n-1}\sum (x_i - \overline{x})^2,$$

$$\text{sample standard deviation:} \quad \widehat{\sigma}_x = \sqrt{\widehat{\sigma}_x^2} = \sqrt{\frac{1}{n-1}\sum (x_i - \overline{x})^2}.$$

- Given a sample of n observations on the random variables X and Y we define the following sample statistics:

$$\text{sample covariance} \quad : \quad \widehat{\sigma}_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}),$$

$$\text{sample correlation} \quad : \quad \widehat{\rho}_{xy} = \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x \widehat{\sigma}_y}.$$

- The most important population parameters associated with two random variables, X and Y, and their sample analogues are summarized in the following table.

pop mean: $E(X) = \sum_{i=1}^{n} x_i f(x_i)$ or $\int_{-\infty}^{\infty} x f(x) dx$
sample mean: $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$
pop var: $\sigma_x^2 = E\{[X - E(X)]^2\}$
sample var: $\widehat{\sigma}_x^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$
pop sd: $\sigma_x = \sqrt{\sigma_x^2}$
sample sd: $\widehat{\sigma}_x = \sqrt{\widehat{\sigma}_x^2}$
pop cov: $\sigma_{xy} = E\{[X - E(X)][Y - E(Y)]\}$
sample cov: $\widehat{\sigma}_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$
pop corr:$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$
sample corr:$\widehat{\rho}_{xy} = \frac{\widehat{\sigma}_{xy}}{\widehat{\sigma}_x \widehat{\sigma}_y}$

- Population parameters cannot be observed, so their values are unknown.
- Once we have collected a sample of observations on the variables of interest, their sample statistics can be computed.
- It is important to realize that populations parameters are not random variables. They are fixed numbers the values of which are usually known.
- Sample statistics on the other hand are random variables in the sense that their values vary from sample to sample and are not known before we collect our sample.
- Sample statistics are often used as estimators of their populations analogues.
- For example, the sample mean is often used as an estimator of the population mean.

- The **joint probability density function** of two random variables X and Y is a function
$$f : S \to [0, 1],$$
where S is a set consisting of all possible combinations of values that X and Y can take on.
- For example, let the random variable Y denote the number bathrooms and the random variable X denote the the number of bedrooms in a randomly selected apartment in Melbourne.
- Assume that Y can assume the values 1 and 2, and X can assume the values 1,2,3.

- The joint probability density function (strictly speaking the joint **probability mass function**, since X and Y are both discrete random variables) is reported in Table 1 below.

Table 1			
$Y \downarrow, X \rightarrow$	1	2	3
1	0.40	0.24	0.04
2	0.00	0.16	0.16

- For example, it follows from Table 1 that

$$\begin{aligned} P(X &= 1, Y = 1) = 0.40, \\ P(X &= 2, Y = 1) = 0.24. \end{aligned}$$

- Notice that

$$0 \leq P(X = x_i, Y = y_j) \leq 1) \text{ for all i and j},$$

  and that the joint probabilities sum to 1.
- We can use the joint pdf in Table 1 to derive both the **marginal pdf** of X and the marginal pdf of Y.
- Since

$$P(X = x_i) = \sum_{j=1}^{2} P(X = x_i, Y = y_j),$$

  summing the elements in

Table 1			
$Y \downarrow, X \rightarrow$	1	2	3
1	0.40	0.24	0.04
2	0.00	0.16	0.16

by column we obtain

$f_X(x) =$

Table 2	
X	$P(X = x)$
1	0.40
2	0.40
3	0.20

.

- Since

$$P(Y = y_j) = \sum_{i=1}^{3} P(X = x_i, Y = y_j),$$

summing the elements in

Table 1			
$Y \downarrow, X \rightarrow$	1	2	3
1	0.40	0.24	0.04
2	0.00	0.16	0.16

by row we obtain

$f_Y(y) =$

Table 3	
Y	$P(Y = y)$
1	0.68
2	0.32

- Note that for each marginal pdf each probability lies between zero and one, and the marginal probabilities sum to 1.
- Given the joint pdf of X and Y and the marginal pdf of X, we can derive the **conditional pdf** of Y.
- Recall that

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}. \quad \text{(L5)}$$

- Using L5 and the information in

Table 1			
$Y \downarrow, X \rightarrow$	1	2	3
1	0.40	0.24	0.04
2	0.00	0.16	0.16

, $f_X(x) =$

Table 2	
X	$P(X = x)$
1	0.40
2	0.40
3	0.20

we obtain

$$\begin{aligned} P(y &= 1|x=1) = \frac{P(y=1, x=1)}{P(x=1)} = \frac{0.40}{0.40} = 1, \\ P(y &= 2|x=1) = \frac{P(y=2, x=1)}{P(x=1)} = \frac{0.00}{0.40} = 0. \end{aligned} \tag{1}$$

- The **conditional density** of Y is summarized in Table 4 below.

$f_Y(y|x=1) =$

Table 4	
$y\|x=1$	$P(Y=y\|x=1)$
1	1.00
2	0.00

.

- The conditional pdf in Table 4 specifies the probability with which Y assumes each value in its domain, given that the random variable X has taken on the value 1.

- It is left as an exercise to use the information in Table 1 and Table 2 together with L5 to show that the pdf of X conditional on $Y = 2$ is given by Table 5 below

$f_Y(y|x=2) =$

Table 5	
$y\|x=2$	$P(Y=y\|x=2)$
1	0.60
2	0.40

,

and the pdf of X conditional on $Y = 3$ is given by Table 6 below

$f_Y(y|x=3) =$

Table 6	
$y\|x=3$	$P(Y=y\|x=3)$
1	0.20
2	0.80

.

- Notice that each of the conditional pdfs in Table 4, Table 5 and Table 6 satisfies the laws of probability. In particular, for each conditional pdf, all probabilities are between zero and one and they sum to one.
- From Table 6 we can deduce that if we randomly select an apartment in Melbourne from the population of apartments which have 3 bedrooms, the probability that the apartment we randomly select will have 1 bathroom is 0.20 and the probability that it will have 2 bathrooms is 0.80.
- Note that both of the above statements are **conditional probability statements**. They respectively tell us the probability that a randomly selected apartment will have 1 bathroom, conditional on the selected apartment having 3 bedrooms, and the probability that a randomly selected apartment will have 2 bathrooms, conditional on the selected apartment having 3 bedrooms.

- For each of the conditional pdfs in Table 4, Table 5 and Table 6 there is an associated **conditional mean**.
- From

Table 4: $f_Y(y\|x=1)$	
$y\|x=1$	$P(Y=y\|x=1)$
1	1.00
2	0.00

we obtain

$$E(Y|x=1) = 1x1.00 + 2x0 = 1.00 \tag{2}$$

- From

Table 5: $f_Y(y \mid x=2)$	
$y \mid x=2$	$P(Y=y \mid x=2)$
1	0.60
2	0.40

we obtain

$$E(Y \mid x=2) = 1x0.60 + 2x0.40 = 1.40. \tag{3}$$

- From

Table 6: $f_Y(y \mid x=3)$	
$y \mid x=3$	$P(Y=y \mid x=3)$
1	0.20
2	0.80

we obtain

$$E(Y \mid x=3) = 1x0.2 + 2x0.8 = 1.80. \tag{4}$$

- Equation (4) states that the average number of bathrooms in the population of apartments which have 3 bedrooms is 1.8.
- From the previous example, we note that the conditional mean of Y need not be a value that Y can actually assume. It is not possible for an apartment to have 1.8 bathrooms!
- Notice that while

$$E(Y|X = x)$$

is a fixed number, its magnitude changes as the value of X changes. Combining (2), (3) and (4) we can see from Table 7 below that $E(Y|X = x)$ is a function of X.

Table 7	
X	$E(Y\|X = x)$
1	1.00
2	1.40
3	1.80

As the value of X changes, so does the conditional mean of Y.

- From Table 7 we can deduce that:
  - In the population of apartments which have 1 bedroom, the average number of bathrooms is 1.
  - In the population of apartments which have 2 bedrooms, the average number of bathrooms is 1.4.
  - In the population of apartments which have 3 bedrooms, the average number of bathrooms is 1.8.