2.1 Boxplots (P.15-16)

A popular graphical representation (in the form of a box) of the following information from a sample is known as a boxplot:

- Quartiles $Q_1$, $Q_2$ and $Q_3$,
- Smallest and largest observations within (LT,UT),
- Outliers, if exist.

**Diagram:** Draw a rectangular box from $Q_1$ to $Q_3$ on a suitable scale (width of this rectangle is not relevant/significant). Now mark $Q_2$ within this rectangular box using * or a dotted line. Then draw two whiskers to the smallest and the largest observations within (LT,UT) from the above rectangle. Finally, if there are any outliers out of (LT,UT), then mark each point by o.

**Example:** Consider the previous sample of 13 observations from an experiment:

$$4 \ 6 \ 6 \ 7 \ 7 \ 9 \ 10 \ 11 \ 13 \ 15 \ 22 \ 24 \ 30$$

Draw a boxplot for this sample following the steps:

(i) Draw a rectangle (horizontal or vertical) of arbitrary width from $Q_1$ to $Q_3$.

(ii) Draw a dotted-line across the rectangle at $Q_2$.

**Solution:** From the previous example, we have calculated:

$Q_1 = 7; \ Q_2 = 10; \ Q_3 = 15.$ Thus

$\text{IQR} = 8; \ LT = -5; \ UT = 27.$

Observations within (-5,27) are considered as “legitimate”. Clearly, 30 is outside (-5,27) and is considered as abnormal or an outlier.

The following boxplot summarises the above information. The outlier (30) is marked as o.

**Boxplot in R:**

R can be used to draw a boxplot. Let $x$ be the vector containing the data.

```r
> x=c(4,6,6,7,7,9,10,11,13,15,22,24,30)
> boxplot(x)
```
Notes:

- Length of a whisker in R is (by default) chosen to be $1.5 \times \text{IQR}$.

- A boxplot gives an alternative, simple visual display of the corresponding data set. This can be used to identify the shape as indicated below:
  - Symmetrical: left and right tails are similar
  - Left skewed: boxplot is stretched to the left.
  - Right skewed: boxplot is stretched to the right.

Now we look at a number of additional statistical summaries from a data set as they are important in further studies.

2.2 Measures of Location and Spread (P.9-11)

Measures of Location

We have seen that the mean and median of a sample measure its center.

A Useful Notation

Suppose that we have $n$ observations (or a sample) from an experiment. Let $x_1$ be the first observation; $x_2$ be the second observation etc and $x_n$ be the $n^{th}$ (the last) observation.

For example, suppose that we have a sample of five observations \{4, 7, 9, 5, 3\}. For this sample, the first observed values is 4 and therefore we write $x_1 = 4$ to identify it. Similarly, $x_2 = 7$, $x_3 = 9$, $x_4 = 5$, $x_5 = 3$.

Summation Notation:

The sum of these $n$ observations $x_1, x_2, \ldots, x_n$ is abbreviated by the sigma notation as follows:

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

Note: Many calculators use this notation. Please check your calculator now.

Example: Consider the previous sample of five observations. Write down $\sum_{i=1}^{5} x_i$ and evaluate it.

Solution:

$$\sum_{i=1}^{5} x_i = x_1 + x_2 + x_3 + x_4 + x_5 = \text{___________}$$
Example: Evaluate the following expressions for four values $(3, 4, 5, 1)$: $\sum_{i=1}^{4} x_i$ and $\sum_{i=1}^{4} x_i^2$.

Solution:
\[
\sum_{i=1}^{4} x_i = x_1 + x_2 + x_3 + x_4 = 3 + 4 + 5 + 1 = 13
\]
\[
\sum_{i=1}^{4} x_i^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 = 3^2 + 4^2 + 5^2 + 1^2 = 51
\]

2.2.1 The Sample Mean, p9

For $n$ observations $x_1, x_2, \ldots, x_n$, the sample mean is denoted by $\bar{x}$ (called $x$ bar) and is given by

\[
\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i.
\]

Example:

The mean of the sample of 4 values from a previous example is

\[
\bar{x} = \frac{3 + 4 + 5 + 1}{4} = \frac{13}{4} = 3.25
\]

Exercise: Look at your calculator now. Change the mode of your calculator to ‘stat’ or ‘sd’ or as per calculator instructions. Check the above answer using your calculator.

Note: The mean is very sensitive to outliers in a sample. However, the median (the center of an ordered sample) is not affected by any outliers. Therefore, in a sample with outliers it is better to use the median as a measure of the “centre” than the mean.

Examples:

1. The median house price is often mentioned in real estate as there are varying prices in many suburbs. For example, many houses in a suburb may be around $800000 while 2 or 3 houses are above $1500000. In such a situation the median is a good indicator of price than the mean.

2. Many people in a certain suburb may earn around $60000 pa while there may be a few executives earn over $250000 pa. Clearly, the median is a good indicator of salary than the mean.

In such cases median gives a good indication of the center.

Use of R

R can be used to find the mean and median of a sample. Practice this example.

> x=c(3,4,5,1)
> mean(x)
> 3.25
> median(x)
> 3.5

Exercise: Find the median, mean and mode for the data set:
13.3, 10.7, 11.0, 11.1, 12.9, 11.8, 11.9, 12.2, 10.8, 12.2, 11.6, 11.8
Solution: Enter the data (just as is) in R:

```r
> x=c(13.3,10.7,11.0,11.1,12.9,11.8,11.9,12.2,10.8,12.2,11.6,11.8)
> mean(x)
> median(x)
```

by hand: Order the data to find the median:

10.7, 10.8, 11.0, 11.1, 11.6, 11.8, 11.8, 11.9, 12.2, 12.2, 12.9, 13.3

Ans: mean = 11.775; median = 11.8; mode = 11.8 and 12.2

In this case, the mode is not unique. Such datasets are also called bimodal.

Exercise: Check the mean of this sample using your calculator (now). Remember to changing the mode to stat.

2.2.2 Sample Variance and Standard Deviation, p12

In order to motivate this concept, consider the following two sets of observations:

- Sample 1: 2, 5, 15, 20, 38
- Sample 2: 12, 13, 15, 19, 21

It is easy to verify that both sets have the same mean at $\bar{x} = 16$. However, the two samples visually appear radically different. This difference lies in the greater spread or variability or dispersion in the first dataset than the second. Therefore, we need a universal measure to study an indication of the amount of variation that a sample exhibits.

Note: A data set with many values far away from the sample mean is called highly volatile. If most of the values are around its mean, then it is called less volatile sample.

The most popular measure of the spread (or volatility) is known as the sample variance. This is defined as the average squared deviation of observations from the sample mean.

The Sample Variance

The difference between an observation and the sample mean is known as the ‘deviation of the observation’ from the sample mean. For example, in sample 1 the deviations from its mean are: $2 - 16 = -14$, $5 - 16 = -11$, $15 - 16 = -1$, $20 - 16 = 4$, $38 - 16 = 22$.

Although the sum of squared deviations divided by 5 is the average, it can be shown that the sum of squared deviations divided by 4 gives a good measure of the sample variance.

Example: For the above sample 1:

$$\text{the variance} = \frac{(-14)^2 + (-11)^2 + (-1)^2 + 4^2 + 22^2}{4} = \frac{818}{4} = 204.5.$$ 

Similarly for the sample 2, the variance is 15. As we have noticed from the data, the above calculations indicate that the sample 1 has more variability than in sample 2.

Note: In general, for a set of $n$ observations $x_1, x_2, \ldots, x_n$, the sample variance is denoted by $s^2$ and is given by

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$
Calculation of the Sample Variance

It is easy show that the following equivalent calculation formula is useful in practice:

\[
 s^2 = \frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left( \sum_{i=1}^{n} x_i \right)^2 \right].
\]

**exercise:** show that

\[
 s^2 = \frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i^2 - n \bar{x}^2 \right].
\]

**A Good News:** You do not need to memorize this formula. It will provided in all examinations. This formula sheet is available from the course web site.

**Note:** Clearly, the above value of \( s^2 \) is in squared units. For example, if the measurements are in cm, then \( s^2 \) is in cm\(^2\).

**Example:** Find the mean and variance of the sample:

\[55, 48, 59, 64, 65, 57, 58, 41, 57, 59, 64, 62\]

**Solution:** \( n = 12 \). First calculate

\[\sum_{i=1}^{12} x_i = \] \\
\[\sum_{i=1}^{12} x_i^2 = \]

**Mean:** \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \) \\
**Variance:**

\[
 s^2 = \frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} \right] = 
\]

**Standard Deviation of a Sample**

Since the sample variance is in squared units, its square root will provide value in original units. Therefore, the square root of \( s^2 \) is known as the sample standard deviation.

**Example:** Find the standard deviation of the above sample.

**Solution:** Simply take the square root of the variance. Thus, the Standard Deviation is:

\[ s = \]

**Use of your calculator:** Many scientific calculators and computer packages (including R) can be used to find the variance and standard deviation of a given dataset. Look at your calculator now:

- Change the mode of your calculator to STAT (or similar depending on your calculator).
- Look for buttons \( \bar{x} \), \( s^2 \) or \( \sigma^2 \). Many calculators have \( s^2_{n-1} \) or \( \sigma^2_{n-1} \) button to find the sample sd.
Note: It can be shown that after a change in origin of a data set, the variance and standard deviation remain the same. If the sample points change in scale by a factor $c$, then the variance changes by a factor of $c^2$ and the sd changes by a factor of $c$.

Exercise: Consider the data set (twice the previous set):

110, 96, 118, 128, 130, 114, 116, 82, 114, 118, 128, 124.

Find the mean, variance and sd of this data set. Show that the mean is twice the previous mean and variance is four times previous variance. What can you say about the sd?

Ans: (approx) 114.84, 194.52, 13.95.

2.2.3 The Coefficient of Variation

The coefficient of variation, denoted CV, is the ratio of the standard deviation to the mean.

For a dataset with $\bar{x} \neq 0$, this is defined as:

$$CV = \frac{s}{\bar{x}}.$$

This ratio of the standard deviation to the mean is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from each other.

Example: The CV for the previous dataset is

$$CV = \frac{s}{\bar{x}} = \frac{13.95}{114.84} = 0.1214449$$

or the s.d. accounts for 12% of the mean.

Note: It is clear that the CV is dimensionless as it is a proportion. That is, the CV is not affected by multiplicative changes of the data. For example, the first and second data sets have the same CV. Therefore, the CV is a useful measure for comparing the dispersions of two or more variables that are measured on different scales. For example, lengths in m and and cm.

The next section considers the corresponding statistics for grouped data.

2.3 Grouped Data (P.16-17)

Recall that large datasets can be summarised with a suitable frequency distribution table with $k$ groups or intervals or bins like this:

<table>
<thead>
<tr>
<th>Group/Class interval</th>
<th>Class center</th>
<th>Frequency</th>
<th>Relative frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1 &lt; x \leq y_2$</td>
<td>$u_1 = (y_1 + y_2)/2$</td>
<td>$f_1$</td>
<td>$f_1/n$</td>
</tr>
<tr>
<td>$y_2 &lt; x \leq y_3$</td>
<td>$u_2 = (y_2 + y_3)/2$</td>
<td>$f_2$</td>
<td>$f_2/n$</td>
</tr>
<tr>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
<td>$\vdots$</td>
</tr>
<tr>
<td>$y_k &lt; x \leq y_{k+1}$</td>
<td>$u_k = (y_k + y_{k+1})/2$</td>
<td>$f_k$</td>
<td>$f_k/n$</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>TOTAL</strong></td>
<td><strong>n</strong></td>
<td>1.000</td>
</tr>
</tbody>
</table>

Now we look the problem of calculating the mean and variance from such a frequency table.
2.3.1 The mean of Grouped Data

Suppose that we only have the information provided by a grouped frequency table for a data set. That is, we only have access to the published report and not the original data set. Let \( k \) be the number of bins (groups or intervals) and \( u_1, u_2, \ldots, u_k \) be the centres of each interval with corresponding frequencies \( f_1, f_2, \ldots, f_k \). Then an approximate sample mean is given by

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{k} f_i u_i.
\]

**Example:** Consider the data on weight in pounds (recorded to the nearest pound) of 35 female students from week 1.

Females:
140 120 130 138 121 125 116 145 150 112 125 130 120 130 131 120 118 125 135 122 115 102 115 150 110 116 108 95 125 133 110 150 108

We have the frequency distribution from last week:

<table>
<thead>
<tr>
<th>CLASS INTERVAL</th>
<th>CLASS CENTER</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>94-104</td>
<td>99</td>
<td>2</td>
</tr>
<tr>
<td>104-114</td>
<td>109</td>
<td>5</td>
</tr>
<tr>
<td>114-124</td>
<td>119</td>
<td>11</td>
</tr>
<tr>
<td>124-134</td>
<td>129</td>
<td>10</td>
</tr>
<tr>
<td>134-144</td>
<td>139</td>
<td>3</td>
</tr>
<tr>
<td>144-154</td>
<td>149</td>
<td>4</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>35</td>
</tr>
</tbody>
</table>

Find the grouped mean.

**Solution:** \( n = 35 \) (the number of values) and we have

\[
\sum_{i=1}^{6} f_i u_i = \ldots
\]

This gives \( \bar{x} = \frac{1}{n} \sum_{i=1}^{6} f_i u_i = \ldots \)

**Exercise:** Find the exact mean of the data and compare it to the above approximation.

**Answer:** Using the complete data, check with your calculator and R, sum of all 35 values = 4333 and hence the exact mean, \( \bar{x} = \ldots \).

**Note:** The grouped mean and the exact mean are close to each other.

2.3.2 The Variance of Grouped Data

Following the previous argument, for a set of data from a frequency table, the grouped sample variance is:

\[
s^2 = \frac{1}{n-1} \sum_{j=1}^{k} f_j (u_j - \bar{x})^2.
\]

**Calculation formula:** The following equivalent form can be used in practice:

\[
s^2 = \frac{1}{n-1} \left[ \sum_{j=1}^{k} f_j u_j^2 - \frac{1}{n} (\sum_{j=1}^{k} f_j u_j)^2 \right].
\]
Some authors use the following form:

\[ s^2 = \frac{1}{n-1} \left[ \sum_{j=1}^{k} f_j u_j^2 - n(\bar{x}^2) \right]. \]

**Example:** Find the sample variance from the previous frequency distribution table of 35 female students.

**Solution:**

\[
\sum_{i=1}^{6} f_i u_i^2 = 
\Rightarrow s^2 = 
\Rightarrow s = 
\]

**Example:** Find the exact sample sd and compare with the grouped sd=13.35581.

**Solution:** Check with your calculator and R the following:

\[
\sum x = 4333; \quad \sum x^2 = 542505, \quad s^2 = 178.8118 \text{ and sd}=13.37205.
\]

Notice that grouped variances and sd are close to exact variances and sd.

**Exercise:** Using the following frequency table for 57 male students from week1, compute the grouped mean and sd using your calculator and R. Compare them with exact values.

<table>
<thead>
<tr>
<th>CLASS INTERVAL</th>
<th>CLASS CENTER</th>
<th>FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>122-136</td>
<td>129</td>
<td>6</td>
</tr>
<tr>
<td>136-150</td>
<td>143</td>
<td>17</td>
</tr>
<tr>
<td>150-164</td>
<td>157</td>
<td>17</td>
</tr>
<tr>
<td>164-178</td>
<td>171</td>
<td>7</td>
</tr>
<tr>
<td>178-192</td>
<td>185</td>
<td>8</td>
</tr>
<tr>
<td>192-206</td>
<td>199</td>
<td>1</td>
</tr>
<tr>
<td>206-220</td>
<td>213</td>
<td>1</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>57</td>
</tr>
</tbody>
</table>

**Answer:** Grouped mean=157.2456 and


Exact mean=\(9021/57 = 158.2632\) and

Exact variance=347.3045, sd=18.63611.

**Additional worked example:**

Consider the two samples:

Sample 1, x: 1.76, 1.45, 1.03, 1.53, 2.34, 1.96, 1.79, 1.21
Sample 2, y: 0.49, 0.85, 1.00, 1.54, 1.01, 0.75, 2.11, 0.92

For each samples:

1. Calculate the mean and the standard deviation,
2. Find \(Q_1, Q_2, Q_3, LT\) and \(UT\),
3. Find CV,
4. Draw both boxplots on the same page.
Ans: 1. We have \( n = 8 \) is even and
\[
\sum_{i=1}^{8} x_i = 13.07, \quad \sum_{i=1}^{8} x_i^2 = 22.5873, \quad \sum_{i=1}^{8} y_i = 8.67, \quad \sum_{i=1}^{8} y_i^2 = 11.2153
\]
Sample 1: mean=1.63, sd \( s_x = 0.42 \)
Sample 2: mean= 1.08, sd= 0.51
2. In ascending order:
Sample 1 x: 1.03, 1.21, 1.45, 1.53, 1.76, 1.79, 1.96, 2.34
Sample 2 y: 0.49, 0.75, 0.85, 0.92, 1.00, 1.01, 1.54, 2.11

Sample 1: \( Q_1 = 1.330; \ Q_2 = 1.645; \ Q_3 = 1.875 \), IQR = \( Q_3 - Q_1 = 1.875 - 1.330 = 0.545 \)
LT = \( Q_1 - 1.5 \times IQR = 1.330 - 1.5(0.545) = 0.5125 \)
UT = \( Q_3 + 1.5 \times IQR = 1.875 + 1.5(0.545) = 2.6925 \). There is no outlier.

Sample 2: \( Q_1 = 0.80; \ Q_2 = 0.96; \ Q_3 = 1.28 \), IQR = \( Q_3 - Q_1 = 1.28 - 0.80 = 0.48 \)
LT = \( Q_1 - 1.5 \times IQR = 0.80 - 1.5(0.48) = 0.08 \)
UT = \( Q_3 + 1.5 \times IQR = 1.28 + 1.5(0.48) = 2.00 \)
Since the max = 2.11 lies outside (LT,UT) = (0.08,2.00).
3. CVs are 0.258 and 0.472 respectively.

R commands:
\[
x = c(1.03, 1.21, 1.45, 1.53, 1.76, 1.79, 1.96, 2.34)
y = c(0.49, 0.75, 0.85, 0.92, 1.00, 1.01, 1.54, 2.11)
\]
mean(x)
sd(x)
sort(x)
median(x)
sd(x)/mean(x) #cv
fivenum(x)
boxplot(x,y) #2 boxplots side by side
where x and y are vectors of measurements.
In order to develop further concepts and applications of biostatistics, it is necessary to understand the basic theory of probability. Now we look at this topic in a simple way.

3 An Introduction to Probability: Theory and Applications, P29

Most of you have covered the initial part of this topic in yr 7-12. Therefore, this chapter considers the following:

- A review of basic terminology,
- Theory of sets and Venn diagrams,
- Probability axioms and counting methods,
- Conditional probability and independence.

Preliminaries

- The word *fair* or *unbiased* is regularly used in many life science situations. This means that all possible outcomes of an experiment have the same chance to occur.
- Any experiment to collect information is called a *random experiment*, if we are not certain or cannot predict of its outcome(s).

It is clear that in a random experiment, one cannot state (before the experiment) what a particular outcome will be. However, we can make a list of all possible outcomes.

For example:

1. In 1, we observe one of these numbers: 1 or 2 or 3 or 4 or 5 or 6.
2. In 2, we observe one colour from: red or black or white.

Now we provide the following definition for later reference:

**Definition:** The collection (or the set) of all possible outcomes of a random experiment is called the *sample space*. This is denoted by $S$ or $\Omega$ and be written within curly brackets as $S = \{ \cdots \}$.

For example,

1. in the experiment 1, the sample space is $S = \{1, 2, 3, 4, 5, 6\}$.
2. in the experiment 2, the sample space is $S = \{R, B, W\}$.
The following terminology will be useful in many applications:

**Definition:** An event of a random experiment is a collection of outcomes with specific or interested features.

**Example:** Revisit the example 1: List the event $A$ of observing a number less than 3.

**Ans:** $A = \{1, 2\}$.

**Example:** A card is selected at random from a box containing 10 cards with numbers 1 to 10. Write-down the sample space. List the following events:

- $A$: observing even numbers,
- $B$: observing numbers divisible by 4.

**Ans:** $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ and hence $A = \{2, 4, 6, 8, 10\}; B = \{4, 8\}$.

### 3.1 Probability of equally likely outcomes/events

First consider the concept of equally likely outcomes.

**Equally Likely Outcomes:** The outcomes of a random experiment (or in a sample space) are called equally likely if all of them have the same chance of occurrence.

In a historical note, the probability was considered as the chance of an event to occur which expresses the strength of one’s belief. Therefore, this was known as *subjective probability*. However, this was later developed with a number of common concepts including equally likely outcomes. Therefore, we have the following definition:

**Definition:** The probability of an event $A$ is the relative frequency of its set of outcomes over an indefinitely large number of repeated trials under identical conditions. This is denoted by $P(A)$ or $Pr(A)$.

### Calculating Probabilities

Suppose we have a random experiment, which has exactly $n$ equally likely outcomes. Let $A$ be an event of interest within this sample space containing $m$ number of simple outcomes. Then the probability assigned to $A$, $P(A)$ is given by:

$$P(A) = \frac{m}{n}.$$ 

**Examples:**

1. Throw a fair six-sided die. There are 6 equally likely possible outcomes each with probability $\frac{1}{6}$. The sample space, $S$ of this experiment is $S = \{1, 2, 3, 4, 5, 6\}$.

   If $A$ denotes the event of observing an even number, then

   $\text{Prob(an even number)} = P(A) = \frac{1}{2}$.

2. Toss a fair coin 3 times. There are 8 possible equally likely outcomes each with probability $\frac{1}{8}$. The sample space, $S$ of this experiment is

   $S = \{\text{HHH, HHT, HTH, THH, TTH, HTT, THT, TTT}\}$.
• Let $A$ be the event of observing exactly two heads in this experiment. Then $A = \{HHT, HTH, THH\}$ and the probability of observing exactly two heads is $P(A) = \ldots$.

• Let $B$ be the event of observing at least one head. Then the event is $B = \{HHH, HHT, HTH, THH, TTH, HTT, THT\}$. Hence, the probability of observing at least one head is $P(B) = \ldots$.

### 3.2 Probability using tree diagrams, p33

Probability Trees or Tree Diagrams can be used to visualize independent events and to calculate simple probabilities related to random experiments.

**Example:** Draw a suitable tree diagram for the experiment of tossing a fair coin two times independently. Hence list the sample space.

**Solution (i):**

Tree diagram for the distribution of gender of three children

\[
\begin{align*}
\text{Tree diagram for the distribution of gender of three children} \\
B & \quad 0.6 \quad B \quad P(BBB) = 0.6 \times 0.6 \times 0.6 \\
0.6 & \quad B \quad P(BBG) = 0.6 \times 0.6 \times 0.4 \\
0.4 & \quad G \quad P(BGB) = 0.4 \times 0.6 \times 0.6 \\
G & \quad 0.6 \quad G \quad P(BGG) = 0.6 \times 0.4 \times 0.4 \\
0.6 & \quad B \quad P(GBB) = 0.4 \times 0.6 \times 0.4 \\
0.4 & \quad G \quad P(GBG) = 0.4 \times 0.6 \times 0.4 \\
0.6 & \quad B \quad P(GGB) = 0.4 \times 0.4 \times 0.6 \\
G & \quad 0.6 \quad G \quad P(GGG) = 0.4 \times 0.4 \times 0.4
\end{align*}
\]

**Solution (ii):**

\[
\begin{align*}
\text{(a) } P(\text{at most 1 boy}) & = \ldots \\
& = \ldots \\
\text{(b) } P(\text{at least 1 boy}) & = \ldots \\
& = \ldots
\end{align*}
\]

**Exercise:** Draw a tree diagram for the experiment of tossing a fair coin three times independently.