Tutorial Questions

Submit solutions to ALL Tutorial questions by 2:30 pm, Friday 3 August. Also attach your completed Laboratory Sheet 1 to your solutions. Make sure you attach a cover sheet (available online) to your solutions.

1. (a) To plan production for Father’s day, a producer of shirts wants to know which shirt size is most popular among men. Which measure of central tendency would be the most useful: the mean, the median or the mode? Explain.

Solution:
The mode is the most useful statistic, as it is the most frequently occurring shirt size.

(b) A dataset contains 100 observations. Which of the following statements are true and which are false. Give reasons for your answer.
   i. If you add 10 to each observation, the mean increases by 10.
   ii. If you add 10 to each observation, the standard deviation increases by 10.
   iii. If you multiply each observation by 3, the mean is multiplied by 3.
   iv. If you multiply each observation by 3, the standard deviation is multiplied by 3.
   v. If you change the sign of each observation, the sign of the mean is changed.
   vi. If you change the sign of each observation, the sign of the standard deviation is unchanged.

Solution:
   i. True
   ii. False
   iii. True
   iv. True
   v. True
   vi. True

(c) In a large corporation, a very small group of employees have extremely high salaries, whereas the majority of employees receive much lower salaries. You were the bargaining agent for the union seeking a salary increase. Write a submission (half a page) to the employer arguing your case for a salary increase for the lower paid workers. Think in terms of the statistics you would use, the key aspects of the salary data here, and how you would argue your case.

If you were the employer, which statistic would you use to counter the union and why?

Solution:
In this situation the data is right skewed, so the median is less than the mean.

As a union representative you would use the median (or the mode if one exists and is small), since the median is not sensitive to extreme data values. Thus the small group of large observations will not affect the median - the larger number of smaller observations will determine the median.

As an employer, you would use the mean. The small group of larger observations cause the mean to be greater than the median.

2. Of three events $A$, $B$, and $C$, suppose $A$ and $B$ are independent and $B$ and $C$ are disjoint. Their probabilities are $P(A) = 0.5$, $P(B) = 0.3$ and $P(C) = 0.1$. Calculate the probabilities of the following events:

(a) Both $B$ and $C$ occur.
(b) At least one of $A$ and $B$ occurs.
(c) $B$ does not occur.
(d) All three events occur.

Solution:

If $A$ and $B$ are independent, then $P(A \cap B) = P(A)P(B)$.
If $B$ and $C$ are disjoint (mutually exclusive), then $P(B \cap C) = 0$.

(a) $P(B \cap C) = 0$.
(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.3 - (0.5)(0.3) = 0.65$.
(c) $P(\overline{B}) = 1 - P(B) = 1 - 0.3 = 0.7$.
(d) $P(A \cap B \cap C) = P(B \cap C)P(A \mid (B \cap C)) = 0$.
   [OR, $P(A \cap B \cap C) = 0$ since $B$ and $C$ are disjoint.]

3. Three machines $A$, $B$ and $C$ fail independently of each other with probabilities $1/6$, $1/4$ and $1/3$ respectively. Define the event $A$ as “Machine A fails”, and define events $B$ and $C$ similarly.

(a) Find the probability that exactly one of the machines fails. [Hint: one of the three ways of achieving this is $A \cap \overline{B} \cap \overline{C}$.]
(b) Given that only one of them fails, what is the probability that it was $A$?
(c) In order to fulfil an order, it is required that Machines $A$ and at least one of Machines $B$ and $C$ are operational. What is the probability that the order will be fulfilled?

Solution:

Let $F$ be the event “exactly one machine fails”.

(a) $P(F) = P[(A \cap \overline{B} \cap \overline{C}) \cup (\overline{A} \cap B \cap \overline{C}) \cup (\overline{A} \cap \overline{B} \cap C)]$, (mutually exclusive).
Since $A$, $B$, and $C$ are independent,

$$P(A \cap \overline{B} \cap \overline{C}) = P(A)P(\overline{B})P(\overline{C}) = (1/6)(3/4)(2/3) = 6/72,$$

and similarly for the others. Hence $P(F) = (6 + 10 + 15)/72 = 31/72$.

(b) $P(A \mid F) = P(A \cap F)/P(F) = P(A \cap \overline{B} \cap \overline{C})/P(F) = 6/72/31/72 = 6/31$.

(c) We need

$$P[\overline{A} \cap (\overline{B} \cup \overline{C})] = P(\overline{A})P(\overline{B} \cup \overline{C}) (\text{independence})$$
$$= P(\overline{A})[P(\overline{B}) + P(\overline{C}) - P(\overline{B} \cap \overline{C})]$$
$$= P(\overline{A})[P(\overline{B}) + P(\overline{C}) - P(\overline{B})P(\overline{C})]$$
$$= \frac{5}{6} \left( \frac{3}{4} + \frac{2}{3} - \frac{6}{12} \right)$$
$$= \frac{55}{72}.$$
Practice Questions

1. Black et al. 4.48

Solution:

Defining our events:
A: Company A has treated the lawn.
G: Company B has treated the lawn.
H: A lawn is rated as very healthy in a month’s time after a service.

Using Bayes Rule (see pg.142 of Black et al.), we can compute the revised probabilities.

\[ P(A|H) = \frac{P(H|T)P(T)}{P(H)} \]

where \( P(H) = P(H|T)P(T) + P(H|G)P(G) \). Therefore, the probability that Company A treated the lawn given that the lawn is healthy is \( P(A|H) = 0.794 \).

Similarly, for Company B, we find that \( P(B|H) = 0.206 = 1 - P(A|H) \). Note that this problem can also be done using tree diagrams, and that is an acceptable solution.

2. For five data values \( x_i, i = 1, 2, \ldots, 5 \), the following calculations were made:

\[ \sum_{i=1}^{5} x_i = 10, \quad \sum_{i=1}^{5} x_i^2 = 29. \]

Use these to calculate \( \bar{x}, s^2 \) and \( s \).

Solution:

\[ \bar{x} = \frac{10}{5} = 2; \quad s^2 = \frac{29 - 5(2^2)}{5 - 1} = \frac{29 - 20}{4} = 2.25; \quad s = 1.5. \]

3. Suppose a sample space contains 6 outcomes denoted by \( e_1, e_2, \ldots, e_6 \). Let \( A, B \) and \( C \) be the events

\[ A = \{e_1, e_2, e_3\}, \quad B = \{e_3, e_4, e_5\}, \quad \text{and} \quad C = \{e_5, e_6\}, \]

and suppose probabilities are assigned as follows:

\[ P(A) = 0.4, \quad P(B) = 0.5, \quad P(C) = 0.5, \quad P(A \cap B) = 0.2, \quad P(B \cap C) = 0.2. \]

(a) By using a Venn diagram, verify that such an assignment of probabilities is possible.

(b) Find the probabilities of the events \( A \cup B, A \cup B \cup C \) and \( C \cap \bar{B} \).

(c) The event \( A \cap B \) can be described in words as “both the event A and the event B occur”. Describe similarly in words, the three events of part (b).

Solution:

(a) Without showing the Venn diagram here, it can be deduced that \( P(\{e_3\}) = 0.2 \) and \( P(\{e_5\}) = 0.2 \), so that \( P(\{e_4\}) = 0.1 \) and \( P(\{e_6\}) = 0.3 \), and finally that \( P(\{e_1, e_2\}) = 0.2 \).

It is not possible, from the given information, to determine unique values for \( P(\{e_1\}) \) and \( P(\{e_2\}) \). Any allocation for which \( P(\{e_1\}) + P(\{e_2\}) = 0.2 \) would be acceptable. However, the remaining questions can be answered knowing only \( P(\{e_1, e_2\}) = 0.2 \).

(b) [These questions can be done directly from the Venn diagram, or by using Probability Rules. The methods used here are only suggestions. You might like to try other methods.]

\[
P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.5 - 0.2 = 0.7, \]

[or: \( P(A \cup B) = P(\{e_1, e_2, e_3, e_4, e_5\}) = 1 - P(\{e_6\}) = 1 - 0.3 = 0.7 \)]

\[
P(A \cup B \cup C) = P(\{e_1, e_2, e_3, e_4, e_5, e_6\}) = 1, \quad P(C \cap \bar{B}) = P(\{e_6\}) = 0.3.
\]
(c) $A \cup B$: at least one of the events $A$ or $B$ occur,

$A \cup B \cup C$: at least one of the events $A$ or $B$ or $C$ occur,

$C \cap \overline{B}$: the event $C$ occurs, and the event $B$ does not occur.

4. (a) What is wrong with the following assignment of probabilities?

$$P(A) = 0.6, \quad P(B) = 0.3, \quad P(A \cap B) = 0.4.$$  

(b) What is wrong with the following assignment of probabilities?

$$P(A) = 0.6, \quad P(B) = 0.7, \quad P(A \cup B) = 0.5.$$  

(c) We are given $P(A) = 0.4$, $P(B) = 0.5$ and $P(A \cap B) = 0.2$. Find $P(A \cup B)$.

(d) Black et al. 4.19, 4.27 and 4.32.

Solution:

(a) The assignment of probabilities is wrong because $P(A \cap B)$ cannot be bigger than $P(B)$.

(b) The assignment of probabilities is wrong because $P(A \cup B)$ cannot be smaller than $P(A)$ or $P(B)$.

(c) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.4 + 0.5 - 0.2 = 0.7$.

(d) Black et al. 4.19 Let $A$ be the percentage of households that have cable television and $B$ be of the percentage of households that have two or more television sets.

(a) $P(A \cup B) = ?$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = 0.67 + 0.74 - 0.55 = 0.86.$$  

(b) Here, we are asked about "only" events $A$ or $B$ occurring by themselves i.e. $P(A \cap B)$ is not accounted for. Therefore, we have 0.12(0.67 - 0.55) + 0.19(0.74 - 0.55) = 0.31. Thus 31% of households have cable or two or more television sets but not both.

(c) Here, we are asked to find an event which is not a part of $A$ or $B$. i.e. $P(A \cup B) = 0.14$

(d) The special law of addition says that if $X$ and $Y$ are mutually exclusive, then $P(X \cup Y) = P(X) + P(Y)$. This does not hold in this case since, $P(A \cap B) \neq 0$. In other words, having cable and two or more television sets are not mutually exclusive.

Black et al. 4.19 Black et al. 4.27 Black et al. 4.32

5. Of 250 employees of a company, a total of 130 are full-time employees. The remainder are part-time employees. There are 150 males working for this company, 85 of whom are full-time employees.

(a) What is the probability that an employee chosen at random

i. is a part-time employee?

ii. is female and a full-time employee?

iii. is a full-time employee, given that the employee is female?

iv. is a female, given that the employee is full-time?

(b) Are the events "employee chosen at random is female" and "employee chosen at random is full-time" statistically independent?

Solution:

The given information can be summarised and completed in a table.

<table>
<thead>
<tr>
<th></th>
<th>Full-time</th>
<th>Part-time</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>85</td>
<td>65</td>
<td>150</td>
</tr>
<tr>
<td>Female</td>
<td>45</td>
<td>55</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>130</td>
<td>120</td>
<td>250</td>
</tr>
</tbody>
</table>

(a) i. From the table, $P(\text{Part-time})= 120/250 = 0.48$

ii. From the table, $P(\text{Female and Full-time})= 45/250 = 0.18$. 

4
iii. Directly from the table, \( P(\text{Full-time} \mid \text{Female}) = \frac{45}{100} = 0.45. \)

[OR, using the definition of conditional probability, \( P(\text{Full-time} \mid \text{Female}) = \frac{P(\text{Female AND Full-time})}{P(\text{Female})} = \frac{(45/250) / (100/250)}{100/250} = 0.45 \).

iv. Directly from the table, \( P(\text{Female} \mid \text{Full-time}) = \frac{45}{130} = 0.35. \)

[OR, \( P(\text{Female} \mid \text{Full-time}) = \frac{P(\text{Female AND Full-time})}{P(\text{Full-time})} = \frac{(45/250) / (130/250)}{130/250} = 0.35 \).

(b) The two events are not independent, since \( P(\text{Female}) = 0.4 \neq P(\text{Female} \mid \text{Full-time}) = 0.35 \).

6. The table below shows the profits for the year from the sale of pairs of shoes. Calculate the overall average profit per pair sold (to the nearest cent).

<table>
<thead>
<tr>
<th>Type of shoe</th>
<th>Profit per pair ($)</th>
<th>Pairs sold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imported</td>
<td>10</td>
<td>6000</td>
</tr>
<tr>
<td>Locally made</td>
<td>8</td>
<td>2500</td>
</tr>
<tr>
<td>In-store made</td>
<td>16</td>
<td>1450</td>
</tr>
</tbody>
</table>

Solution:

Mean profit ($) per pair sold

\[
= \frac{10(6000) + 8(2500) + 16(1450)}{6000 + 2500 + 1450} = $10.37 \text{ (to nearest cent).}
\]

[Note, as a check, that this result lies between the smallest and largest profit values, 8 and 16 respectively.]

7. (a) The sample mean and variance for a random sample of numbers \( x_1, x_2, \ldots, x_n \) are given by

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{and} \quad s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]

respectively. If a linear transformation \( u_i = ax_i + b, \ i = 1, 2, \ldots, n, \) is applied to these data, show that

i. \( \bar{u} = a\bar{x} + b \)

ii. \( s_u^2 = a^2 s_x^2. \)

[Hints: For part i., define \( \bar{u} \) in terms of \( \sum u_i \), then substitute \( ax_i + b \) for \( u_i \). For part ii., first use part i. to show that, for each \( i, u_i - \bar{u} = a(x_i - \bar{x}) \).]

Solution:

i.

\[
\bar{u} = \frac{1}{n} \sum_{i=1}^{n} (ax_i + b) = \frac{1}{n} (a \sum_{i=1}^{n} x_i) + \frac{1}{n} nb = a\bar{x} + b.
\]

ii. For each \( i, u_i - \bar{u} = ax_i + b - (a\bar{x} + b) = a(x_i - \bar{x}), \) so

\[
s_u^2 = \frac{1}{n-1} \sum_{i=1}^{n} (u_i - \bar{u})^2 = \frac{1}{n-1} \sum_{i=1}^{n} a^2(x_i - \bar{x})^2 = a^2 s_x^2.
\]

(b) Suppose \( x_1 = -3, x_2 = 0, x_3 = 2, x_4 = 1 \) and \( x_5 = 6. \)

i. Calculate \( \bar{x} \) and \( s_x. \)

ii. Given that

\[
u_i = 3x_i - 2, \ i = 1, 2, \ldots, 5,
\]

determine the values of the sample mean \( \bar{v} \) and standard deviation \( s_v \) of the set of numbers \( u_1, u_2, u_3, u_4, \) and \( u_5. \)

iii. Write down a linear transformation that would convert \( x_i \) to \( v_i, \ i = 1, 2, \ldots, 5, \) such that \( \bar{v} = 0 \) and \( s_v = 1. \)

Solution:
i. \( \bar{x} = 6/5 = 1.2 \); \( s_x = \sqrt{\frac{50 - 5(1.2)^2}{5 - 1}} = \sqrt{10.7} = 3.27 \) (2dp).

ii. \( \bar{u} = 3\bar{x} - 2 = 3(1.2) - 2 = 1.6 \); \( s_u = 3s_x = 3\sqrt{10.7} = 9.81 \).

iii. Let \( v_i = \frac{x_i - \bar{x}}{s_x} = \frac{x_i - 1.2}{\sqrt{10.7}}, \; i = 1, 2, \ldots, 5 \).

8. The annual income of 100 people was entered into a computer file. The largest income was $96,000 but this was typed as $960,000 by mistake.

(a) Suppose the sample mean based on the incorrect income of $960,000 is $46,000. What is the mean using the correct income of $96,000?
(b) Suppose the sample median based on the incorrect income of $960,000 is $32,000. What is the median using the correct income of $96,000?

Solution:

In this question we need the result that \( \sum_{i=1}^{N} x_i = nx_\overline{\text{r}} \).

(a) The sum of the data is \( 100 \times 46,000 = 4,600,000 \). This sum includes the incorrect data value of $960,000. Thus the corrected sum of the data is $4,600,000-$960,000+$96,000=$3,736,000, so the correct mean is $37,360.

(b) The median remains unchanged, as the both the correct and incorrect data value lie on the same side of the median.

9. The following data show the interest rates, on a specific type of account, charged by a financial institution for each month from January through October 1999.

\[
\begin{align*}
6.05, & \quad 6.3, \quad 7.05, \quad 7.3, \quad 6.8, \\
6.2, & \quad 5.9, \quad 6.4, \quad 6.9, \quad 6.9.
\end{align*}
\]

Find the sample median, the first and third quartiles.

Solution:

The data arranged in ascending order:

\[
\begin{align*}
5.9, & \quad 6.05, \quad 6.2, \quad 6.3, \quad 6.4, \quad 6.8, \quad 6.9, \quad 6.9, \quad 7.05, \quad 7.3
\end{align*}
\]

Median = \( \frac{6.4 + 6.8}{2} = 6.6 \)

First quartile = 6.2

(note that at least 25% of data values are smaller than or equal to 6.2, and at least 75% of data values are greater than or equal to 6.2.)

Third quartile = 6.9

(note that at least 75% of data values are smaller than or equal to 6.9, and at least 25% of data values are greater than or equal to 6.9.)