Please answer all questions. Note that the questions are not worth equal marks.

Question 1 contains short answer questions worth a total of 65 marks. Questions 2-7 are longer and are worth a total of 95 marks.

Show all relevant working.

PLEASE NOTE

Examination candidates may only bring authorised materials into the examination room. If a supervisor finds, during the examination, that you have unauthorised material, in whatever form, in the vicinity of your desk or on your person, whether in the examination room or the toilets or en route to/from the toilets, the matter will be reported to the head of school and disciplinary action will normally be taken against you. This action may result in your being deprived of any credit for this examination or even, in some cases, for the whole unit. This will apply regardless of whether the material has been used at the time it is found.

Therefore, any candidate who has brought any unauthorised material whatsoever into the examination room should declare it to the supervisor immediately. Candidates who are uncertain whether any material is authorised should ask the supervisor for clarification.
1. [65 marks]

(a) A market researcher obtains data on salaries of residents in a particular suburb. Discuss in not more than half a page the exploratory data analysis that should be performed, and the purpose of such an analysis. [6 marks]

(b) If \( P(A) = 0.5, \ P(B \mid A) = 0.3 \) and \( P(A \cup B) = 0.8 \), find \( P(B) \). [3 marks]

(c) A particular type of printer cartridge is produced by two companies, Alamo and Jersey. Alamo produces 65% of the cartridges, of which eight percent are defective. Twelve percent of the cartridges produced by Jersey are defective. Let \( J \) denote the event that a randomly selected cartridge is produced by Jersey, and \( D \) the event that the cartridge is defective. A customer purchases a cartridge and finds it to be defective. What is the probability that the cartridge was produced by Jersey? [5 marks]

(d) A business analyst assesses that the joint probability mass function of demand \( (D) \) and price \( (P) \) for a product for the next quarter is as shown in the table below.

<table>
<thead>
<tr>
<th>( D )</th>
<th>( 2000 )</th>
<th>( 2500 )</th>
<th>( 3000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$30</td>
<td>0.0</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>$35</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>$40</td>
<td>0.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

(i) Find the expected demand in the next quarter? [2 marks]

(ii) Find the expected demand if the price is fixed at $35. [4 marks]

(iii) What is the expected revenue from sales of this product in the next quarter? [4 marks]

(e) The delay in hours for delivering a consignment is a continuous random variable \( X \) with probability density function

\[
f_X(x) = \begin{cases} 
\frac{1}{2} + x & \text{if } 0 < x < 1, \\
0 & \text{otherwise.}
\end{cases}
\]

What is the probability that the delay is more than half an hour? [4 marks]

(f) A company manufactures bolts and washers. The diameters \( B \) of bolts and \( W \) of washers are independently normally distributed, with respective means
1(f) (Continued)

10 mm and 11 mm, and respective variances 0.15 mm$^2$ and 0.1 mm$^2$. Quality specifications require the diameter of a bolt to be less than that of a washer (or else the bolt will not pass through the washer) and the diameter of the washer to be no more than 2 mm greater than that of the bolt (or else the washer does not fit snugly). What is the probability that a bolt and a washer that are randomly selected meet the quality specifications? [5 marks]

[HINT: Find the distribution of $D = W - B$, and state the required probability in terms of $D$.]

(g) In a nursery the probability that a seedling survives to the selling age of two weeks is 0.8. If the nursery plants 2,500 seeds and they all germinate, use a suitable approximation to estimate the probability that at least 2,050 of these survive to the selling age? [5 marks]

(h) A 95% confidence interval for a population proportion is $(0.35, 0.40)$, based on a sample size of 200.

(i) Find the value of the observed sample proportion. [1 mark]

(ii) Find a 99% confidence interval for the population proportion. [4 marks]

(i) Discuss the three factors that affect the width of a confidence interval for a population mean, and the effect of each. [6 marks]

(j) A test of hypothesis for a population mean fails to reject the null hypothesis at the 5% level of significance. What is the decision at the 1% level of significance? Justify your answer. [2 marks]

(k) A partially filled ANOVA table is given below.

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>627.12</td>
<td>B</td>
<td>A</td>
<td>20.1</td>
</tr>
<tr>
<td>Within</td>
<td>D</td>
<td>C</td>
<td>10.4</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>E</td>
<td>40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Find the values of the missing entries. [5 marks]
1 (Continued)

(l) The following contingency table represents the frequency count of types of transportation used by the publishing and the computer hardware industries over a given week.

<table>
<thead>
<tr>
<th></th>
<th>Air</th>
<th>Train</th>
<th>Truck</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Publishing</td>
<td>32</td>
<td>12</td>
<td>41</td>
<td>85</td>
</tr>
<tr>
<td>Computer Hardware</td>
<td>5</td>
<td>6</td>
<td>24</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>37</td>
<td>18</td>
<td>65</td>
<td>120</td>
</tr>
</tbody>
</table>

A chi-square test of independence is performed in Excel.

(i) State the null and alternative hypotheses being tested here. [2 marks]

(ii) If the null hypothesis is true, find the expected frequency of Publishing companies using Air as transport. [2 marks]

(iii) What is the distribution of the test statistic assuming the null hypothesis is true (include the value of any parameters)? [2 marks]

(iv) The hypothesis test is performed in Excel, giving a p-value of 0.0401. Determine using this information if the mode of transport is independent of industry. Justify your answer. [3 marks]
2. [15 marks] In a manufacturing process a machine is set to punch a hole of diameter 1.84 cm in a strip of metal. The strip of metal is then creased and sent to the next phase where a metal rod is slipped through the hole. It is important that the hole be punched to the specified diameter. To test punching accuracy, a technician has taken a random sample of 12 punched holes and measured their diameters, giving a sample mean and sample standard deviation of 1.85 cm and 0.0245 cm respectively.

Let $\mu$ denote the mean diameter of the holes.

(a) State the null and alternative hypotheses. [2 marks]

(b) Give a formula for the test statistic and state its distribution under the null hypothesis, specifying the value of any parameters. State any assumption(s) required for this distribution to be valid. [3 marks]

(c) Find the observed value of the test statistic. [1 mark]

(d) Test the hypotheses at the 5% level of significance. Write a short statement reporting your conclusion to the production manager. [5 marks]

(e) Find a 95% confidence interval for the mean hole diameter. [2 marks]

(f) How might your analysis change if the sample size was 120? Give reasons for your answer. [2 marks]

3. [10 marks] The vice president of a marketing company brought to the attention of sales managers that most of the company’s 10 sales representatives contacted clients and maintained client relationships in a disorganised and haphazard way. The 10 sales representatives were given a three day training programme. Data on the number of client visits for each sales representative over a month before and after the training programme showed an average increase of 1.222 with a sample standard deviation of 1.0929.

Has the the number of visits to clients increased after the training programme? Use a 2.5% level of significance, and state any assumptions required in the analysis. [10 marks]
4. **[15 marks]** A company that makes computer keyboards has specifications that are designed to ensure that no more than 3% of the production is defective. The company has recently been receiving more customer complaints than usual. To investigate if the proportion of defective keyboards has increased, a quality control officer takes a random sample of 500 keyboards and finds 20 of these to be defective.

Let $p$ denote the proportion of keyboards that are defective.

(a) State the null and alternative hypotheses that the officer is testing. 

(b) Give an expression for the test statistic and state its distribution under the null hypothesis. 

(c) Find the observed value of the test statistic. 

(d) Conduct the hypothesis test at the 2.5% level of significance, and state your conclusion. 

(e) Find a 95% confidence interval for the proportion of defective items produced by the company. 

(f) Verify any assumptions required in the above analysis. 

5. **[15 marks]** Few things are more frustrating than needing information and having to wait for it. The time spent on hold waiting for technical support to resolve software problems is one of the largest drains on productivity experienced by small businesses. Before a company will consider switching from its current word processing software (Microsoft Word) to a competitor (WordPerfect) the company will have to be convinced that it will spend less time on hold. The company decides to perform a test by placing 10 calls to each technical support line and recording the amount of time (in minutes) that the caller is on hold. The summary statistics are given below.

<table>
<thead>
<tr>
<th></th>
<th>MS Word</th>
<th>WordPerfect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>$n_1 = 10$</td>
<td>$n_2 = 10$</td>
</tr>
<tr>
<td>Sample mean</td>
<td>$\bar{x}_1 = 11.2$</td>
<td>$\bar{x}_2 = 9.7$</td>
</tr>
<tr>
<td>Sample standard deviation</td>
<td>$s_1 = 3.4$</td>
<td>$s_2 = 1.6$</td>
</tr>
</tbody>
</table>

**QUESTION 5 CONTINUES OVER THE PAGE**
5 (Continued)

[Information: The test statistic for a hypothesis test for a difference of two population means based on two independent samples is]

\[ T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}, \]

where

\[ S_p^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2} \]

is the pooled variance.]

Let \( \mu_1 \) and \( \mu_2 \) be the mean hold time for MS Word and WordPerfect respectively.

(a) State the relevant null and alternative hypotheses. [2 marks]

(b) State any assumptions required in the analysis. [2 marks]

(c) Find the observed value of the test statistic. [3 marks]

(d) Perform the hypothesis test using a significance level of 2.5%. [3 marks]

(e) Write a short statement with a recommendation to the manager based on your analysis. [2 marks]

(f) Find a 95% confidence interval for the difference in mean hold times for the two call lines. [3 marks]

6. [20 marks] National Airlines recently introduced a daily early-bird flight between Hong Kong and Singapore. The vice president of marketing for National Airlines wants to compare National’s average passenger load on this new flight and that of each of two major competitors, Competitor 1 and Competitor 2. Ten early morning flights with equal seating capacity were selected at random from each of the three airlines and the number of empty seats on each were recorded. An ANOVA was performed on the data and the results, along with some diagnostics, are given below.

Let \( \mu_N \), \( \mu_1 \) and \( \mu_2 \) denote the mean number of empty seats on National Airlines, Competitor 1 and Competitor 2 flights respectively.
(a) State the hypotheses that are being tested here. [2 marks]

(b) State the assumptions of the analysis and use the output to check if they hold. [6 marks]

(c) Perform a test of the hypotheses and state your conclusion. [3 marks]

(d) Find 95% confidence intervals for the difference in the mean number of empty seats between National Airlines and each of Competitor 1 and Competitor 2. Comment on the implications of these intervals. [6 marks]

(e) Write a short statement to the vice president informing her of the results of your analysis. [3 marks]
7. [20 marks] Data on 150 houses was collected to explore the relationship between selling price (Price, the dependent or Y variable) and the appraised value (Value, the independent or X variable) in $,000 ($ thousands). A linear regression analysis was performed in Excel, and some of the output is given below.

(a) State the linear regression model and explain each term in the model. [4 marks]

QUESTION 7 CONTINUES OVER THE PAGE
7 (Continued)

(b) State the estimated equation of regression between Price and Value. [2 marks]

(c) If the appraised value of a house increases by $1,000, what is the expected increase in selling price? [1 mark]

(d) State and check the assumptions of the linear regression model using the output provided. [8 marks]

(e) Is there a significant positive linear relationship between Price and Value? As part of your answer state the relevant hypotheses and test them using the output provided. [4 marks]

(f) State a 95% confidence interval for the slope parameter. [1 mark]