Answering the questions

- All answers to be written in the answer book.
- Write in blue or black biro. Pencil scripts will not be accepted.
- Number your answers corresponding to the question numbers and their parts.
- Explain what the answers are, e.g. $P(X > 5.2) = 0.56$.
- Where the answer requires a number of steps, the setting out should be neat and easily readable.
- You may use a calculator or indicate where you have approximated.
Question 1 [2 marks]
What is the 80\% \text{‘ile} of the following ranked sample?

\[1\] 50 60 70 75 79 89 90 100 100 100
\[11\] 112 114 122 126 130 150 176 340 476

Question 2 [2 marks]

Figure 1: Box-plot for 2 treatments

Rewrite the following table in your answer book and complete it using the information in Figure 1.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Range</th>
<th>25%\text{‘ile}</th>
<th>50%\text{‘ile}</th>
<th>75%\text{‘ile}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Question 3 [4 marks]
Fifty-two \% of students at a university are female. Five \% of students are studying Computer Science. Two \% of the students are women studying Computer Science. If a student is selected at random, find the conditional probability that

(a) the student is female, given that the student is studying Computer Science, [2 marks]

(b) the student is studying Computer Science, given that the student is female. [2 marks]
Question 4  [4 marks]
The cumulative distribution function of a normal distribution is plotted in Figure 2.

(a) Determine the range of the central 68% of the data from the cdf. [2 marks]

(b) Estimate the mean of these data. [1 mark]

(c) Estimate the standard deviation of these data. [1 mark]

Question 5  [2 marks]
The cumulative distribution function of rainfall is plotted in Figure 3.

Estimate (approximately) the probability of between 50 and 80 mm of rain.
**Question 6**

The numbers of applicants accepted and rejected by 6 departments in a university are listed in Table 1.

Table 1: Acceptances and rejections for departments A-F

<table>
<thead>
<tr>
<th>Department</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accepted</td>
<td>175</td>
<td>393</td>
<td>512</td>
<td>138</td>
<td>53</td>
<td>122</td>
</tr>
<tr>
<td>Rejected</td>
<td>225</td>
<td>207</td>
<td>288</td>
<td>262</td>
<td>147</td>
<td>378</td>
</tr>
</tbody>
</table>

Rewrite the table so that it readily shows differences amongst departments with respect to acceptances.

**Question 7**

Table 2 is a 2-way table of probabilities for Rank × Age category.

Table 2: Probabilities of Academic rank and Age

<table>
<thead>
<tr>
<th>Rank</th>
<th>Full Professor $R_1$</th>
<th>Associate Professor $R_2$</th>
<th>Senior Lecturer $R_3$</th>
<th>Lecturer $R_4$</th>
<th>$P(A_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$ (Under 30)</td>
<td>0.002</td>
<td>0.003</td>
<td>0.049</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>$A_2$ 30-39</td>
<td>0.045</td>
<td>0.146</td>
<td>0.140</td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td>$A_3$ 40-49</td>
<td>0.134</td>
<td>0.107</td>
<td>0.052</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>$A_4$ 50-59</td>
<td>0.125</td>
<td>0.058</td>
<td>0.031</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>$A_5$ (Over 60)</td>
<td>0.064</td>
<td>0.013</td>
<td>0.003</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>$P(R_j)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Write the marginal probabilities, carefully labelled.
Question 8

Figure 4: Binomial(20,0.6) probability and distribution functions

Determine (approximately) from Figure 4 the probabilities:-

(a) \( P(X = 10) \) [1 mark]
(b) \( P(X \leq 10) \) [1 mark]
(c) \( P(X > 15) \) [1 mark]

Question 9

Figure 5: Poisson (\( \lambda = 7.2 \)) probability and distribution function

Determine (approximately) from Figure 5 the probabilities:-

(a) \( P(X = 10) \) [1 mark]
(b) \( P(X > 6) \) [1 mark]
(c) \( P(6 < X \leq 8) \) [1 mark]
Question 10 [3 marks]
The mean and standard deviation of a population, indicated by records in a large data base are \( \mu = 60 \) and \( \sigma = 10 \). A new small sample is, 48 34 28 66 87 58
Rewrite these numbers in the exam book and alongside each denote as True (T) if it is likely to have been sampled from the same population represented by the data base, False (F) otherwise. You may refer to Figure 9 in the reference section on page 11.

Question 11 [4 marks]
A random sample of 25 was taken from a \( N(40, 15^2) \) distribution. What is the probability that the average for this sample is between 37 and 43? You may refer to Figure 10 in the reference section on page 11.

Question 12 [2 marks]
The mean and standard deviation of ages (in years) of a class of buildings are 30 and 5 respectively. What is the distribution of the mean of a sample of size \( n = 50 \)?

Question 13 [3 marks]
In a study of 100 patients, 50 received a treatment and 50 a control. Eighteen of the control group showed no improvement and 8 of the treatment group showed no improvement. What is the 95% confidence interval for the difference in sample proportions? Use:
\[
\sqrt{0.36 \times 0.64 \times \frac{1}{50} + 0.16 \times 0.84 \times \frac{1}{50}} = 0.085
\]

Question 14 [3 marks]
A new procedure was compared with a conventional one by measuring the length of time (days) to effect a successful result. The sample size, mean and standard deviation for the new treatment are \( n_1 = 8, \bar{X}_1 = 3.5, s_1 = 1.5 \) respectively and for the conventional treatment are \( n_2 = 10, \bar{X}_2 = 8.0, s_2 = 3.5 \). The pooled variance is \( S_p^2 = 3.23 \).
Test the hypothesis that the new procedure is better.
Use the result that the new procedure is better.
\[
\sqrt{3.23 \times \left( \frac{1}{8} + \frac{1}{10} \right)} = 0.85 \text{ and that } P(t_{10} > 2.58) = 0.01.
\]
Question 15 [4 marks]
The diastolic blood pressure of 10 hypertensive patients was measured before and after treatment. The summary statistics are listed in Table 3.

Table 3: Means, s.e.’s of Before, After and Change of blood pressure

<table>
<thead>
<tr>
<th></th>
<th>Before</th>
<th>After</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}$</td>
<td>82.7</td>
<td>97</td>
<td>14.3</td>
</tr>
<tr>
<td>s.e.</td>
<td>1.3</td>
<td>2.3</td>
<td>1.25</td>
</tr>
</tbody>
</table>

(a) Write the formula for the t-test and calculate the T statistic (approximately) for this case.

(b) Use Figure 11 in the reference section on page 12 to infer whether or not the change is a result of treatment?

Question 16 [6 marks]
The responses of a survey on opinions on an issue are shown in Table 4.

Table 4: Frequencies of responses to a Y/N/U question

<table>
<thead>
<tr>
<th></th>
<th>Supports</th>
<th>Opposes</th>
<th>Undecided</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>80</td>
<td>15</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>70</td>
<td>25</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Rewrite the table in your exam book and compute the row, column and grand totals. Leave room for extra cell entries for (b).

(b) Enter into each cell the expected frequencies under the model of independence between Gender and Position.

(c) Use the R calculations to infer whether opinion is dependent upon Gender.

```r
> source("Otab.r")
-- Otab --
S O U
M 80 15 5
F 70 25 5

Pearson’s Chi-squared test
data: Otab
X-squared = 3.2, df = 2, p-value = 0.2053
```
Question 17 [5 marks]

A sample of size 100 was taken from a population and the data analyst proposed a distribution whose cumulative distribution function is plotted in Figure 6.

The observed bin frequencies of the sample are:

<table>
<thead>
<tr>
<th>bins</th>
<th>(0-5)</th>
<th>(5-10)</th>
<th>(10-15)</th>
<th>(15-20)</th>
<th>(20-25)</th>
<th>&gt; 25</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>28</td>
<td>34</td>
<td>15</td>
<td>11</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

(a) Make a table of observed frequencies and the expected frequencies from the gamma model for each of the bin widths. [2 marks]

(b) State the degrees of freedom for the Chi square test. [1 mark]

(c) The calculation of the $\chi^2$ goodness of fit test ($\chi^2 = \sum (O - E)^2 / E$) was done in R,

```
> source("chisq.r")
sum((O-E)^2/E) = 14
```

What is the approximate probability of getting a value as least as large as (ie $\geq$) the observed Chi square if the data genuinely had the proposed distribution. Use Figure 12 in the reference section on page 12. [1 mark]

(d) Explain whether the proposed model could be regarded as a reliable approximation for these data. [1 mark]
Question 18  [7 marks]

The data in Table 5 are heights of wheat plants (in inches) at maturity measured from an experiment with 3 varieties.

Table 5: Heights in inches of wheat plants from 3 varieties

<table>
<thead>
<tr>
<th>Variety</th>
<th>Heights</th>
<th>Mean (rounded)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>17 18 17 16 18</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>21 19 18 19 20</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>18 21 20 21 19</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 6 is the AOV to compare varieties.

Table 6: Analysis of variance for comparing heights from 3 varieties

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variety</td>
<td>2</td>
<td>19.60</td>
<td>9.80</td>
<td>7.95</td>
<td>0.0063</td>
</tr>
<tr>
<td>Residuals</td>
<td>12</td>
<td>14.80</td>
<td>1.23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) State the conclusion you draw from the AOV about the mean heights.  [1 mark]

(b) Calculate the standard error of the means, approximate to 1 decimal place.  [1 mark]

(c) Calculate the 95% Confidence Interval for each mean.
\[(t_{12}(0.025) = -2.2, \ t_{12}(0.975) = 2.2)\]  [3 marks]

(d) Draw a plot which provides a visual comparison of the means and their confidence intervals.  [2 marks]
Question 19

A statistical model is fitted to data relating car stopping distance (feet) to speed (miles per hour) as depicted in Figure 7. The points are observations \((x_i, y_i)\) and the smooth line represents the systematic component. That is,

\[ y_i = f(x_i) + \epsilon_i \quad \epsilon_i \sim N(0, \sigma^2) . \]

(a) What is the estimate (approximately) of the most likely stopping distance when the speed is 15 m.p.h.?

[1 mark]

(b) The distribution function of the random component when \(x = 15\) is plotted in Figure 8. What are the estimates (approximately) of the 5% and 95%'iles of error?

[2 marks]

(c) What is the 90% confidence interval (approximately) for the predicted stopping distance at speed 15 m.p.h.?

[2 marks]

Figure 7: Stopping distance ~ speed

Figure 8: Distribution function of errors
Figure 9: Density of the standard normal showing $\frac{1}{2}\%$ tail regions

Figure 10: Distribution function of the standard normal
Figure 11: Distribution function of $t$ with 9 degrees of freedom

Figure 12: Distribution functions for $\chi^2$ 5,6,7 df