1. The following data represent the balances (in hundreds of dollars) from a sample of saving accounts:

   27  32  23  35  25
   43  51  28  31  60

   (a) Construct a stem-and-leaf plot for the above sample data. [2 marks]

   (b) Find the median, the first quartile, and the third quartile, of the data. [3 marks]
2. Is anything wrong with the following assignment of probabilities:

\[ P(A^c \cup B) = 0.6 \quad P(B) = 0.3 \quad P(A \mid B^c) = 0.7. \]

Justify your answer. [4 marks]

**Hint:** Recall that \((A^c \cup B)^c = A \cap B^c\).

3. An accounting firm has advertised the availability of its report on recent changes to the tax act. The first 200 callers requesting a copy of the report are classified in the following table according to the medium by which the caller became aware of the report and the caller’s primary interest.

<table>
<thead>
<tr>
<th>PRIMARY INTEREST</th>
<th>MEDIUM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Radio</td>
</tr>
<tr>
<td>Personal Tax</td>
<td>34</td>
</tr>
<tr>
<td>Corporate Tax</td>
<td>36</td>
</tr>
</tbody>
</table>

One caller is selected at random and two events are defined as follows:

A: The caller is primarily interested in corporate tax
B: The caller became aware of the report through the newspaper

(a) Find

(i) \( P(A \cap B) \) [2 marks]

(ii) \( P(A \cup B) \) [2 marks]

(iii) \( P(A \cap B \mid A \cup B) \). [2 marks]

(b) Are the events \( A \) and \( B \) independent? Justify your answer. [3 marks]
4. (a) Let $X$ be a binomial random variable with $n = 20$ and $p = 0.4$, $(X \sim Bin(20, 0.4))$. Find $P(5 \leq X \leq 12)$. [2 marks]

(b) Let $Y$ be normally distributed random variable with mean $\mu = 6$ and variance $\sigma^2 = 9$, $(Y \sim N(6, 9))$. Find $P(Y \geq 8)$. [2 marks]

(c) It is known that 80% of the households in a community would pay to have cable TV. A salesperson visits a random sample of 50 households from that community. Find the probability that less than 35 of these households want cable TV. (Use an appropriate approximation, if necessary) [4 marks]

(d) Suppose the mean and standard deviation of $X$, the yearly cost of employee medical insurance, are $3500$ and $400$, respectively. An official obtains a random sample of 200 employees and calculates the sample mean $\overline{X}$. What is the (approximate) probability that $\overline{X}$ is between 3450 and 3520? [4 marks]

5. The table below gives the joint probability mass function (p.m.f.) of the random variables $X$ and $Y$.

<table>
<thead>
<tr>
<th>$y$</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.12</td>
<td>0.24</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>0.18</td>
<td>0.36</td>
<td>0.06</td>
</tr>
</tbody>
</table>

(a) Find

(i) The marginal p.m.f.’s $p_X(x)$ and $p_Y(y)$, [2 marks]

(ii) $E(X)$ and $E(Y)$, [2 marks]

(iii) $Var(X)$ and $Var(Y)$, [2 marks]

(b) Are $X$ and $Y$ independent random variables? Justify your answer. [3 marks]

(c) Find $Cov(X + Y, X - Y)$. [2 marks]
6. Let $X$ be a continuous random variable with the probability density function $f(x)$ given by:

$$f(x) = \begin{cases} 
  x + 1 & \text{if } -1 \leq x < 0 \\
  x & \text{if } 0 \leq x \leq 1 \\
  0 & \text{otherwise}
\end{cases}$$

(a) Sketch the graph of the probability density function. [1 mark]

(b) Find $P(X < -0.5)$. [2 marks]

(c) Find $P(-0.5 < X < 0.3)$. [3 marks]

7. The mean monthly sales of insurance agents in a large company is $72000. In an attempt to improve sales, a new training program has been devised. Ten agents are randomly selected to participate in the program. At its completion, the sales of the agents in the next month are recorded as follows (in thousands of dollars)

$$63 \ 87 \ 95 \ 75 \ 83 \ 78 \ 69 \ 79 \ 103 \ 98.$$ 

The sample mean and sample standard deviation are $\bar{x} = 83$, and $s = 12.85$, respectively. Assume that the sales are normally distributed.

(a) State the relevant null and alternative hypotheses, in terms of a relevant parameter, in order to answer the following question. Is there enough evidence to suggest that the new training program is successful? [3 marks]

(b) Write down a formula for the relevant test statistic. What is its distribution given that the null hypothesis is true? [2 marks]

(c) Carry out the relevant test specifying the critical region (use $\alpha = 0.01$) and state clearly the conclusion from the test. [3 marks]

(d) From the above sample, find a 95% confidence interval for the mean monthly sales, first writing down the relevant formula for such an interval. [3 marks]
8. A supplier of tomato seeds claims that at least 95% of her tomato seeds will eventually produce tomato plants. A farmer takes a random sample of 400 seeds and plants them. Eventually, only 368 of these seeds produced tomato plants.

(a) State the relevant null and alternative hypotheses, in terms of a relevant parameter, in order to answer the following question. Is there enough evidence to suggest that the claim of the supplier is not justified? [3 marks]

(b) Write down a formula for the relevant test statistic. What is its (approximate) distribution given that the null hypothesis is true? [2 marks]

(c) Carry out the relevant test specifying the critical region (use $\alpha = 0.01$) and state clearly the conclusion from the test. [3 marks]

(d) From the above sample, find a 99% confidence interval for the proportion of seeds that eventually produce tomato plants, first writing down the relevant formula for such an interval. [3 marks]

9. A statistician wants to compare the relative success of two department store chains. She decides to measure the sales per square metre. She takes a random sample of five stores from chain 1 and independently a random sample of five stores from chain 2. The gross sales per square metre are shown below

Chain 1: 65.50 72.00 103.00 93.50 82.60
Chain 2: 82.50 63.50 68.00 70.00 66.50.

Assume that the population distributions for both samples are normal with the same variance. Note that the corresponding sample means and sample variances are:

$\bar{x}_1 = 83.32$, $s_1^2 = 234.29$, $\bar{x}_2 = 70.10$, $s_2^2 = 53.67$.

(a) State the relevant null and alternative hypotheses, in terms of relevant parameters in order to answer the following question. Is there enough evidence to suggest that the two chains are not equally successful? [2 marks]
9(a) (Continued)

(b) Write down a formula for the relevant test statistic. What is its distribution given that the null hypothesis is true? [3 marks]

(c) Carry out the relevant test specifying the critical region (use $\alpha = 0.05$) and state clearly the conclusion from the test. [3 marks]

10. A major computer manufacturer has received numerous complaints concerning the short life of its disk drives. In his search for a better disk drive, the manufacturer finds three new products. He decides to test these three plus his current disk drive to determine if differences in lifetimes exist among the products. He takes a random sample of five disk drives of each type and links each drive with a computer. The number of weeks until the drive breaks down is recorded and shown in the table below:

<table>
<thead>
<tr>
<th>Type 1</th>
<th>78</th>
<th>92</th>
<th>101</th>
<th>105</th>
<th>98</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 2</td>
<td>125</td>
<td>110</td>
<td>116</td>
<td>88</td>
<td>128</td>
</tr>
<tr>
<td>Type 3</td>
<td>143</td>
<td>125</td>
<td>133</td>
<td>108</td>
<td>121</td>
</tr>
<tr>
<td>Current Type</td>
<td>101</td>
<td>96</td>
<td>88</td>
<td>125</td>
<td>128</td>
</tr>
</tbody>
</table>

The question of interest is: Is there any difference among the mean lifetimes of the four types of disk drive? To answer this question:

(a) State the relevant assumptions concerning the distributions of the populations from which the four samples were drawn. [2 marks]

(b) State the relevant null and alternative hypotheses in terms of the population means, $\mu_1, \mu_2, \mu_3, \mu_4$, corresponding to the four types of disk drive. [1 mark]

(c) Write down a detailed (in terms of the sample means and variances) formula for the relevant test statistic and state its distribution given the null hypothesis is true. Find its observed value. [3 marks]

(d) Carry out the test, at a significance level of 0.05, and state clearly your conclusion. [2 marks]

**Note:** The corresponding sample means and sample variances for the four populations are:

\[ \bar{x}_1 = 94.8, \ \bar{x}_2 = 113.4, \ \bar{x}_3 = 126.0, \ \bar{x}_4 = 107.6, \]

SEE OVER
$s_1^2 = 110.7$, $s_2^2 = 252.8$, $s_3^2 = 171.9$, $s_4^2 = 320.4$. 

SEE OVER
11. An educational economist wants to establish the relationship between an individual’s income and education. He takes a random sample of 10 individuals and asks for their income (in thousands of dollars) and education (in years). The results are shown below. Also the usual assumptions on the underlying model for simple linear regression are made.

<table>
<thead>
<tr>
<th>Education ($x_i$)</th>
<th>11</th>
<th>12</th>
<th>11</th>
<th>15</th>
<th>8</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>17</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income ($y_i$)</td>
<td>25</td>
<td>33</td>
<td>22</td>
<td>41</td>
<td>18</td>
<td>28</td>
<td>32</td>
<td>24</td>
<td>53</td>
<td>26</td>
</tr>
</tbody>
</table>

Excel output for the regression analysis is shown below.

Using the Excel output answer the following questions.

(a) Is there evidence of a relationship between education and income? Justify your answer. (Use $\alpha = 0.05$.) [2 marks]

(b) What is the least squares estimate ($b_0$) of the intercept parameter $\beta_0$? [1 mark]

(c) Assume you are going to carry out a test where the alternative hypothesis is $H_1: \beta_1 > 0$ ($\beta_1$ denotes the slope parameter). What is the distribution of the relevant test statistic? Find its observed value from the above output. [2 marks]

(d) Predict the income of an individual whose education equals 9 years. Is this prediction reliable? Explain. [2 marks]
11. (Continued)

(e) Write down a 95% confidence interval for the slope parameter $\beta_1$. [2 marks]

12. The records of an investment banking firm show that, historically, 60% of its clients were primarily interested in the stock market, 30% in the bonds market, and 10% in the futures market. A recent sample of 50 clients showed that 25 were primarily interested in stocks, 19 in bonds, and 6 in futures. Carry out a relevant goodness-of-fit test in order to answer the following question. Is there enough evidence to suggest, at the level of significance 0.01, that there has been a change in the distribution of the primary interest of clients? [6 marks]