Sample Exam Solution Outlines

**IMPORTANT NOTE:** This document contains an outline of the solutions to the sample exam. Students are expected to answer questions more completely. For many of the discussion questions, the suggestions given here are only some of the possible correct answers. Students are not expected to list every single possible correct answer!

**Section A**

**Question 1**
(a) Population – employees of the airline.
   Sampling frame – a list of union members who are employees of the airline.
(b) Not all employees will be union members. There may be bias since those who join the union might not be typical of all employees (eg. they may be more likely to join if they are concerned about losing their job).
(c) (i) Categorical, nominal  (ii) Numerical, continuous, ratio

**Question 2**
(a) (i) Stem-and-leaf plot of weekly sales figures for the deli (in hundreds of dollars)

```
2 7
3 8 9
4 2 2 6 7 7 8 8 9
5 0 1 4 8
6 0 9
7 2
```

(ii) Only 6 of the 18 values are greater than $5000. The owner’s statement is unlikely to be true.
(b) Some possible answers:
• Results are distorted. People will interpret the area of each person, not just the height – the higher readership numbers are over-emphasised.
• The numbers at the head of each person are not explained and no y-axis label is given.
• The title is not an objective description.
Some possible improvements:
• Keep the bar widths equal (and probably avoid the use of the people!).
• Include a label for the y-axis.
• Make the title descriptive not interpretive (for example, “Buzz readership numbers over recent years”).

**Question 3**
(a) The midpoints of the classes \( m \) are 20, 30, 40, etc.
\[
\sum mf = 20(9) + 30(18) + 40(32) + 50(27) + 60(24) + 70(8) = 5350
\]
\[
\sum m^2 f = (20)^2 (9) + (30)^2 (18) + (40)^2 (32) + (50)^2 (27) + (60)^2 (24) + (70)^2 (8) = 264100
\]
\[
n = \sum f = 9 + 18 + 32 + 27 + 24 + 8 = 118
\]
\[
\bar{X} = \frac{\sum mf}{n} = \frac{5350}{118} = 45.3390 \text{ (4 dec. pl.)}
\]
\[
S = \sqrt{\frac{\sum m^2 f - \left( \frac{\sum mf}{n} \right)^2}{n-1}} = \sqrt{\frac{264100 - \left( \frac{5350}{118} \right)^2}{117}} = 13.5673 \text{ (4 dec. pl.)}
\]
Correlation does not imply causation. The two variables are highly correlated, but this does not imply that one causes the other. So spending more on advertising will not necessarily force an increase in sales.

Question 4
(a) (i) Contingency table:

<table>
<thead>
<tr>
<th>Married to a smoker?</th>
<th>Lung cancer?</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>15</td>
<td>35</td>
</tr>
<tr>
<td>No</td>
<td>5</td>
<td>145</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
<td>180</td>
</tr>
</tbody>
</table>

(ii) \( P(\text{lung cancer}|\text{married to a smoker}) = \frac{15}{50} = 0.3 \)

(iii) \( P(\text{lung cancer}|\text{not married to a smoker}) = \frac{5}{150} = 0.0333 \) (4 dec. pl.)

(b) \( \binom{18}{3} = \frac{18!}{15!3!} = 816 \) ways.

Section B

Question 5
(a) (i) \( E(X) = \mu_X = \sum X_i P(X_i) = -3000(0.2) + 4000(0.5) + 12000(0.3) = \$5000 \)

\( E(Y) = \mu_Y = \sum Y_i P(Y_i) = 1000(0.2) + 5000(0.5) + 10000(0.3) = \$5700 \)

(ii) \( \sigma_{XY} = \sum (X_i - E(X))(Y_i - E(Y))P(X_i,Y_i) \)

\( = (-3000-5000)(1000-5700)(0.2) + (4000-5000)(5000-5700)(0.5) \)

\( + (12000-5000)(10000-5700)(0.3) \)

\( = 16,900,000 \)

Since the covariance is positive, there is a positive relationship between the two stocks (i.e. as one increases, so does the other).

(b) Poisson distribution, \( P(X) = \frac{e^{-\lambda} \lambda^x}{X!} \), with \( \lambda = 3 \).

\( P(X < 2) = P(X = 0) + P(X = 1) = \frac{e^{-3}3^0}{0!} + \frac{e^{-3}3^1}{1!} = 0.1991 \) (4 dec. pl.)

Question 6
(a) \( \mu = 300 \), \( \sigma = 54 \). Looking for \( P(250 < X < 320) \). Use \( Z = \frac{X - \mu}{\sigma} \).

\( Z_1 = \frac{250 - 300}{54} = -0.93 \), \( Z_2 = \frac{320 - 300}{54} = 0.37 \), so

\( P(250 < X < 320) = P(-0.93 < Z < 0.37) = 0.6443 - 0.1762 = 0.4681 \)

So the probability that a randomly selected photocopier will have an annual expenditure of between \$250 and \$320 is 0.4681.
The distribution of the data is not symmetric. It is skewed to the right (see box-and-whisker and five number summary). The slight curve in the normal probability plot also indicates the data is skewed to the right. The data is unlikely to be normally distributed.

**Question 7**

\[ \mu = 16, \sigma = 0.75 \text{ (from 45 seconds)}. \]

(a) \[ n = 35, Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{15 - 16}{0.75/\sqrt{35}} = -7.89 \text{ (2 dec. pl.)}, \]

\[ P(\bar{X} < 15) = P(Z < -7.89) = 0. \text{ So the probability that the average time to run the first five kilometres will be less than 15 minutes is 0.} \]

(b) We are told that the population is normally distributed. Therefore, even though the sample size is less than 30, we can say that the sample mean will be normally distributed and we could answer part (a).

**Question 8**

\[ p_s = \frac{340}{400} = 0.85, \text{ } n = 400, \text{ for 99% confidence we have } Z = \pm 2.58. \]

\[ p_s \pm Z \sqrt{\frac{p_s(1-p_s)}{n}} = 0.85 \pm 2.58 \sqrt{\frac{(0.85)(0.15)}{400}} = 0.85 \pm 0.0461 = 0.8039 \text{ to } 0.8961 \text{ (4 dec. pl.)} \]

We can say with 99% confidence that the true proportion of licensed drivers who would pass the written driving test is between 0.8039 and 0.8961.

**Section C**

**Question 9**

(a) (i) \[ \bar{X} = 26, S = 4, n = 50, \alpha = 0.05 \]

\[ H_0 : \mu \leq 20 \]

\[ H_1 : \mu > 20 \]

\( \sigma \) is unknown so use \( t \)-test with \( df = n - 1 = 49 \Rightarrow t_{0.0.0.05} = 1.6766 \)

Rejection rule: Reject \( H_0 \) if \( t > 1.6766 \), otherwise do not reject \( H_0 \).

\[ t = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{26 - 20}{4/\sqrt{50}} = 10.6066 \text{ (4 dec. pl.)} \]

Since 10.6066 > 1.6766, reject the null hypothesis and conclude that there is sufficient evidence to believe that there has been an increase in the number of incorrectly reported patients.

(ii) Some possible ethical considerations:

- Is the sample representative of the population? Probably not since the sample is of a “suspected group”, although this needs a definition.
- Is there still a hidden number of non-reported patients? Doctors who are already being dishonest are unlikely to give full information…
- How practically significant is the difference between 20 and 26 patients given that there are likely to be 1000’s of doctors with 1000’s (or even 10,000’s) of patients each?

(b) We reject the null hypothesis when the \( p \)-value is less than the significance level. In this case, \( 0.00096 < 0.01 \) and so we reject the null hypothesis.
**Question 10**

$H_0$ : there is no relationship between satisfaction with service and length of time as a client

$H_1$ : there is a relationship between satisfaction with service and length of time as a client

$df = (r-1)(c-1) = (3-1)(2-1) = 2$

Given $\alpha = 0.01$, the critical value is $\chi^2_{0.01} = 9.210$ and so the rejection rule is: Reject $H_0$ if $\chi^2 > 9.210$, otherwise do not reject $H_0$.

<table>
<thead>
<tr>
<th>$f_o$</th>
<th>$f_e$</th>
<th>$f_o - f_e$</th>
<th>$(f_o - f_e)^2$</th>
<th>$(f_o - f_e)^2 / f_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>80</td>
<td>24</td>
<td>56</td>
<td>3136</td>
<td>130.6667</td>
</tr>
<tr>
<td>20</td>
<td>76</td>
<td>-56</td>
<td>3136</td>
<td>41.2632</td>
</tr>
<tr>
<td>10</td>
<td>33.6</td>
<td>-23.6</td>
<td>556.96</td>
<td>16.5762</td>
</tr>
<tr>
<td>130</td>
<td>106.4</td>
<td>23.6</td>
<td>556.96</td>
<td>5.2346</td>
</tr>
<tr>
<td>30</td>
<td>62.4</td>
<td>-32.4</td>
<td>1049.76</td>
<td>16.8231</td>
</tr>
<tr>
<td>230</td>
<td>197.6</td>
<td>32.4</td>
<td>1049.76</td>
<td>5.3126</td>
</tr>
</tbody>
</table>

Total: 215.8764

Since $215.8764 > 9.210$, we reject the null hypothesis and conclude that there is sufficient evidence to believe there is a relationship between satisfaction with service and length of time as a client.

Comparing observed and expected cell counts, we see that in general more new clients are dissatisfied and more existing clients are satisfied than would be expected if there were no relationship. This makes intuitive sense since happy clients are more likely to stay with the firm.

**Question 11**

(a) The data appears to roughly follow a straight line and so the simple linear regression model may be appropriate.

(b) $SS_{XY} = \frac{\sum_{i=1}^{n} X_i Y_i}{n} - \left(\frac{\sum_{i=1}^{n} X_i}{n}\right) \left(\frac{\sum_{i=1}^{n} Y_i}{n}\right) = 2905000 - \frac{(3000)(9300)}{10} = 115000$
\[
SSX = \sum_{i=1}^{n} X_i^2 - \left( \frac{\sum_{i=1}^{n} X_i}{n} \right)^2 = 945000 - \left( \frac{3000}{10} \right)^2 = 45000
\]

\[
b_1 = \frac{SSYY}{SSX} = \frac{115000}{45000} = 2.55555556
\]

\[
\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{3000}{10} = 300, \quad \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{9300}{10} = 930
\]

\[
b_0 = \bar{Y} - b_1 \bar{X} = 930 - \left( \frac{115000}{45000} \right)(300) = 163.33333333
\]

So the fitted linear regression model (with coefficients rounded to two decimal places) is:

\[
\hat{Y} = 163.33 + 2.56X_i
\]

(c) \( b_0 = 163.33333333 \) tells us that there is a base profit of $163,333.33 earned no matter how many books are sold. \( b_1 = 2.55555556 \) tells us that for each additional 1000 copies sold, there is an increase in profit of $2,555.56.

(d) 320000 copies is \( X = 320 \). So \( \hat{Y} = 163.33 + 2.56(320) = 981.1111... \) (using all the decimal places in the working) and so we can predict a profit of $981,111.11 when 320,000 copies are sold.

**Question 12**

(a) Moving average calculations are:

\[
MA(3) = \frac{0.98 + 1.02 + 1.06}{3} = 1.02, \quad MA(3) = \frac{1.02 + 1.06 + 1.07}{3} = 1.05,
\]

\[
MA(3) = \frac{1.06 + 1.07 + 1.22}{3} = 1.1167 \quad (4 \text{ dec. pl.})
\]

<table>
<thead>
<tr>
<th>Year</th>
<th>Share price ($)</th>
<th>MA(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1955</td>
<td>0.98</td>
<td></td>
</tr>
<tr>
<td>1956</td>
<td>1.02</td>
<td>1.02</td>
</tr>
<tr>
<td>1957</td>
<td>1.06</td>
<td>1.05</td>
</tr>
<tr>
<td>1958</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td>1959</td>
<td>1.22</td>
<td>1.1167</td>
</tr>
</tbody>
</table>
(b) Exponential smoothing calculations are:

\[ E_1 = Y_1 = 0.98, \quad E_2 = WY_2 + (1-W)E_1 = 0.5(1.02) + 0.5(0.98) = 1 \]

\[ E_3 = WY_3 + (1-W)E_2 = 0.5(1.06) + 0.5(1) = 1.03 \]

\[ E_4 = WY_4 + (1-W)E_3 = 0.5(1.07) + 0.5(1.03) = 1.05 \]

\[ E_5 = WY_5 + (1-W)E_4 = 0.5(1.22) + 0.5(1.05) = 1.135 \]

<table>
<thead>
<tr>
<th>Year</th>
<th>Share price ($)</th>
<th>Smoothed values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1955</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>1956</td>
<td>1.02</td>
<td>1.03</td>
</tr>
<tr>
<td>1957</td>
<td>1.06</td>
<td>1.03</td>
</tr>
<tr>
<td>1958</td>
<td>1.07</td>
<td>1.05</td>
</tr>
<tr>
<td>1959</td>
<td>1.22</td>
<td>1.135</td>
</tr>
</tbody>
</table>

(c) Using the formula \( I_i = \frac{P_i}{P_{base}} \times 100 \) the index values are:

\[ I_1 = \frac{0.98}{0.98} \times 100 = 100, \quad I_2 = \frac{1.02}{0.98} \times 100 = 104.08, \quad I_3 = \frac{1.06}{0.98} \times 100 = 108.16 \]

\[ I_4 = \frac{1.07}{0.98} \times 100 = 109.18, \quad I_5 = \frac{1.22}{0.98} \times 100 = 124.49 \]

<table>
<thead>
<tr>
<th>Year</th>
<th>Share price ($)</th>
<th>Price Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1955</td>
<td>0.98</td>
<td>100</td>
</tr>
<tr>
<td>1956</td>
<td>1.02</td>
<td>104.08</td>
</tr>
<tr>
<td>1957</td>
<td>1.06</td>
<td>108.16</td>
</tr>
<tr>
<td>1958</td>
<td>1.07</td>
<td>109.18</td>
</tr>
<tr>
<td>1959</td>
<td>1.22</td>
<td>124.49</td>
</tr>
</tbody>
</table>

(d) The share price in 1959 is 24.49% greater than in 1955.