EUROPEAN UNIVERSITY OF LEFKE  
FACULTY OF ECONOMICS AND ADMINISTRATIVE SCIENCES  
STAT 204 STATISTICS II SPRING 13M14 FINAL EXAM QUESTIONS  

Date: 09. 06. 2014  
Instructor: Prof. Dr. Hüseyin Öğuz  
Duration: 14:30-16:00  
Room#: AS205&207  
Student Registration No:___________________  
Student NameMSurname:__________________________________  

Important Note: Your own calculator and appendix tables/guidelines are allowable to use during exam with forbidding of their exchange. Please use first the rear side of this exam paper for your solutions, then the blank sheets attached.

1. (35 p) Two different independent random samples of consumers were asked about satisfaction with their computer system each in a slightly different way. The options available for answer were slightly different in the two cases. When asked how satisfied they were with their computer system, 138 of the first group of 240 sample numbers opted for “very satisfied”. When the second group was asked how dissatisfied they were with their computer system, 128 of the second group of 240 sample numbers opted for “very satisfied”. Test, at the 5% significance level against the obvious one-sided alternative, the null hypothesis that the two population proportions are equal.

2. (35 p) A large consumer goods company has been studying the effect of advertising on total profits. As part of this study, data on advertising expenditures and total sales were collected for a five-month period and are as follows:

<table>
<thead>
<tr>
<th>Advertising Expenditures (X, 10^3$)</th>
<th>10</th>
<th>15</th>
<th>7</th>
<th>12</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly Sales (Y, 10^3$)</td>
<td>100</td>
<td>200</td>
<td>80</td>
<td>120</td>
<td>150</td>
</tr>
</tbody>
</table>

a) Prepare a scatter plot of the data  
b) Compute covariance and correlation coefficient  
c) Compute and interpret  
d) How much total sales as (10^3$) would you estimate if the advertising expenditure is 13 (10^3$)

3. (35 p) How do customers first hear about a new product? A random sample of 200 users of a new product was surveyed to determine the answer to this question. Other demographic data such as age were also collected as follows:

<table>
<thead>
<tr>
<th>Age</th>
<th>Method of learning about product</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 21</td>
<td>Friend</td>
</tr>
<tr>
<td>21-35</td>
<td>30</td>
</tr>
<tr>
<td>35+</td>
<td>18</td>
</tr>
</tbody>
</table>

Is there an association between the consumer’s age and the method by which the customer heard about the new product at the 0.5% significance level?

Recall:

$$S_X = \sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \over n-1} ; S_Y = \sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \over n-1} ; S_{XY} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \over n-1 \ ;$$

$$r_{XY} = S_{XY} \over S_X S_Y ; b_1 = \frac{\text{Cov}(X,Y)}{S_X^2} ; b_0 = \bar{Y} - b_1 \bar{X} ; \bar{X} = \sum_{i=1}^{n}X_i \over n ; \bar{Y} = \sum_{i=1}^{n}Y_i \over n \ ;$$

$$\chi^2 = \sum \sum \left( O_{ij} - E_{ij} \right)^2 \over E_{ij} ; E_{ij} = R_i C_j \over n$$
1. Let $X = $ When asked how satisfied they were; $Y = $ When asked how dissatisfied they were
First identify the null and alternative hypothesis

$$H_0: \hat{P}_x - \hat{P}_y = 0$$
$$H_1: \hat{P}_x - \hat{P}_y > 0$$

To test $H_0: \hat{P}_x - \hat{P}_y = 0$ against the alternative $H_1: \hat{P}_x - \hat{P}_y > 0$ the decision rule is to reject $H_0$ and accept $H_1$ if

$$z > z_{a} \quad (upper - tail\ test, \ From\ Appendix\ Guidelines\ Diagram\ E - 4)$$

$$z = \frac{(\hat{P}_x - \hat{P}_y) - (P_x - P_y)}{\sqrt{\hat{P}_o(1 - \hat{P}_o)\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}} > z_a \quad (Guidelines\ E - 4)$$

$$\hat{P}_o = \frac{n_x\hat{P}_x + n_y\hat{P}_y}{n_x + n_y}$$

$$\hat{P}_x = \frac{138}{240} = 0.575, \hat{P}_y = \frac{128}{240} = 0.533, n_x = 240, n_y = 240$$

$$\hat{P}_o = \frac{138 + 128}{240 + 240} = 0.554 \quad (common\ proportion)$$

Now substitute these values into the formula to find the test statistic, rounding to three decimal places.

$$z = \frac{(0.575 - 0.533) - 0}{\sqrt{0.554(1 - 0.554)\left(\frac{1}{240} + \frac{1}{240}\right)}} = 0.926$$

$$\alpha = 5\%, z_{0.05} = 1.645 \quad (Read\ from\ Appendix\ Table\ 1,p.815)$$

Compare the test statistic to $z_{0.05}$

- If $z > z_{0.05}$, Reject $H_0$
- $0.926 < 1.645$

Do Not Reject $H_0$ at the 5% significance level: $P_x - P_y = 0$

Conclusion: There is sufficient evidence that the two population proportions are equal.

2.a) Scatter plot-Advertising Expenditures ($10^3$ $\$ $) vs. Monthly Sales ($10^3$ $\$ $)
b) Use tabulated method to compute covariance and correlation coefficient as follows:

<table>
<thead>
<tr>
<th>$i$</th>
<th>$X_i$</th>
<th>$Y_i$</th>
<th>$(X_i - \bar{X})$</th>
<th>$(X_i - \bar{X})^2$</th>
<th>$(Y_i - \bar{Y})$</th>
<th>$(Y_i - \bar{Y})^2$</th>
<th>$(X_i - \bar{X})(Y_i - \bar{Y})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>100</td>
<td>-1.6</td>
<td>2.56</td>
<td>-30</td>
<td>900</td>
<td>48</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>200</td>
<td>3.4</td>
<td>11.56</td>
<td>70</td>
<td>4900</td>
<td>238</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>80</td>
<td>-4.6</td>
<td>21.16</td>
<td>-50</td>
<td>2500</td>
<td>230</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>120</td>
<td>0.4</td>
<td>0.16</td>
<td>-10</td>
<td>100</td>
<td>-4</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>150</td>
<td>2.4</td>
<td>5.76</td>
<td>20</td>
<td>400</td>
<td>48</td>
</tr>
<tr>
<td>$\Sigma$</td>
<td>58</td>
<td>650</td>
<td>41.2</td>
<td>8800</td>
<td>560</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{58}{5} = 11.6; \quad \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{650}{5} = 130
\]

\[
S_{XY} = \text{Cov}(X, Y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n - 1} = \frac{140}{5 - 1} = 3.2094; S_{\bar{X}}^2 = 10.3
\]

\[
S_X = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n - 1}} = \sqrt{\frac{41.2}{5 - 1}} = 3.2094; S_{\bar{X}}^2 = 10.3
\]

\[
S_Y = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n - 1}} = \sqrt{\frac{8800}{5 - 1}} = 46.9042; S_{\bar{Y}}^2 = 2200
\]

\[
r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\text{Cov}(X, Y)}{(3.2094)(46.9042)} = 0.93002
\]

c)

\[
b_1 = \frac{\text{Cov}(X, Y)}{S_X^2} = \frac{140}{10.3} = 13.5922
\]

Interpretation of $b_1$ : For a one dollar increase in the advertising expenditures, we estimate that the monthly sales would increase by 13.59 dollars (advertising has a positive effect on sales).

\[
b_0 = \bar{Y} - b_1 \bar{X} = 130 - (13.5922)(11.6) = -27.67
\]

Interpretation of $b_0$ : If the advertising expenditures were $0$, we would expect to have monthly sales -$27.67$. Interpret with caution-note that we are extrapolating the results beyond the observed data.

d) Prediction (Estimation) from the least-square linear correlation ($Value of r_{XY} = 0.93002$ shows good correlation of monthly sales with the advertising expenditures positively. \[
\hat{Y}_{n+1} = b_0 + b_1 X_{n+1} = -27.67 + 13.5922(13) = 149.03 (\text{$/\text{month}$})
\]

3. Complete the contingency table (3x2 matrix) given to calculate expected counts ($E_{ij}$) from the corresponding formula written below observed counts ($O_{ij}$) as boldface and italic with the row and column total as follows:

<table>
<thead>
<tr>
<th>Age</th>
<th>Method of learning about product</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;21</td>
<td>Friend</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Advertisement</td>
<td>27</td>
</tr>
<tr>
<td>21-35</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>48.6</strong></td>
</tr>
<tr>
<td>35+</td>
<td></td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>32.4</strong></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>108</td>
</tr>
</tbody>
</table>
\( H_0: \) No association exists between the method of learning and age of the respondent. \\
\( H_1: \) Association exists between the method of learning and age of the respondent.

\[ E_{ij} = \frac{R_i C_j}{n} \rightarrow E_{11} = \frac{R_1 C_1}{n} = \frac{(50)(108)}{200} = 27 \]

\[ E_{12} = \frac{R_1 C_2}{n} = \frac{(50)(92)}{200} = 23 \]

\[ E_{21} = \frac{R_2 C_1}{n} = \frac{(90)(108)}{200} = 48.6 \]

\[ E_{22} = \frac{R_2 C_2}{n} = \frac{(90)(92)}{200} = 41.4 \]

\[ E_{31} = \frac{R_3 C_1}{n} = \frac{(60)(108)}{200} = 32.4 \]

\[ E_{32} = \frac{R_3 C_2}{n} = \frac{(60)(92)}{200} = 27.6 \]

\[
\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}
\]

\[
= \frac{(O_{11} - E_{11})^2}{E_{11}} + \frac{(O_{12} - E_{12})^2}{E_{12}} + \frac{(O_{21} - E_{21})^2}{E_{21}} + \frac{(O_{22} - E_{22})^2}{E_{22}}
\]

\[
+ \frac{(O_{31} - E_{31})^2}{E_{31}} + \frac{(O_{32} - E_{32})^2}{E_{32}}
\]

\[
\chi^2 = \frac{(30 - 27)^2}{27} + \frac{(20 - 23)^2}{23} + \frac{(60 - 48.6)^2}{48.6} + \frac{(30 - 41.4)^2}{41.4} + \frac{(18 - 32.4)^2}{32.4}
\]

\[
+ \frac{(42 - 27.6)^2}{27.6} = 0.333 + 0.391 + 2.674 + 3.139 + 6.4 + 7.513
\]

\[
= 20.45
\]

Decision rule:

reject \( H_0 \) if \( \chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} > \chi^2_{(r-1)(c-1), \alpha} \)

\[
\chi^2 = 20.45 > \chi^2_{(3-1)(2-1), 0.005}
\]

\[
\chi^2_{2, 0.005} = 10.60 \text{ (Read from Appendix Table 7)}
\]

Conclusion: Reject \( H_0 \) of no association at the 0.5% significance level (\( p - value = P(\chi^2_{v=2} > 20.45) = 0.000 \)