FACULTY OF ECONOMICS AND ADMINISTRATIVE SCIENCES OF EUL
STAT 204 STATISTICS II SPRING 14-15 MAKEUP EXAM QUESTIONS
Date: 16. 06. 2015  Instructor: Prof. Dr. Hüseyin Oğuz
Duration: 09:00-10:30  Room#: AS005
Student Registration No:
Student Name-Surname:
Important Note: Your own calculator and appendix tables/guidelines are allowable to use during exam with the prohibition of their exchanges.

1. (30 p) In a study comparing banks in Country A and Country B, a sample of 145 matched pairs of banks was formed. Each pair contained one bank from Country A and one from from Country B. The pairings were made in such a way that the two members were as similar as possible in regard to such factors as size and age. The ratio of total loans outstanding to total assets was calculated for each of the banks. For this ratio, the sample mean difference (Country A-Country B) was 0.0518, and the sample standard deviation of the differences was 0.3055. Test, against a two-sided alternative, the null hypothesis that the two population means are equal at a 5% significance level.

2. (35 p) A corporation administers an aptitude test to all new sales representatives. Management is interested in the extent to which this test is able to predict sales representatives’ eventual success. The accompanying table records average weekly sales (in thousands of dollars) and aptitude test scores for a random sample of eight representatives as follows:

<table>
<thead>
<tr>
<th>Weekly sales ($10^3$)</th>
<th>Test score (Y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>55</td>
</tr>
<tr>
<td>12</td>
<td>60</td>
</tr>
<tr>
<td>28</td>
<td>85</td>
</tr>
<tr>
<td>24</td>
<td>75</td>
</tr>
<tr>
<td>18</td>
<td>80</td>
</tr>
<tr>
<td>16</td>
<td>85</td>
</tr>
<tr>
<td>15</td>
<td>65</td>
</tr>
<tr>
<td>12</td>
<td>60</td>
</tr>
</tbody>
</table>

a) Prepare a scatter plot of the data  
b) Compute covariance and correlation coefficient  
c) Compute and interpret  
d) Estimate the weekly sales if the test score were 70.

3. (35 p) A study was conducted to determine if there is a difference in the investing preferences of mid-level managers working in the public and private sectors in a city. A random sample of 320 public sector employees and 380 private sector employees was taken. The sample participants were then asked about their retirement investment decisions and classified as being either ‘aggressive’, if they invested only in stocks or stock mutual funds, or ‘balanced’, if they invested in some combination of stocks, bonds, cash, and other. The following results were found:

<table>
<thead>
<tr>
<th>Sector</th>
<th>Investing Preference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aggressive</td>
</tr>
<tr>
<td>Public</td>
<td>164</td>
</tr>
<tr>
<td>Private</td>
<td>236</td>
</tr>
</tbody>
</table>

a) State the hypothesis of interest and conduct the appropriate hypothesis test to determine there is a relationship between employment sector and investing preference. Use a level of significance of 0.01.

b) State the conclusion of the test conducted in part (a).
1. Let $A =$ Country A; $B =$ Country B. First identify the null and alternative hypothesis
   
   $H_0: \mu_A - \mu_B = 0$
   
   $H_1: \mu_A - \mu_B \neq 0$

   To test $H_0: \mu_A - \mu_B = 0$ against the alternative $H_1: \mu_A - \mu_B \neq 0$ the decision rule is to reject $H_0$ and accept $H_1$ if

   \[
   t > t_{a/2} \text{ or } t < -t_{a/2} \text{(two - tail test, Paired Samples, From Appendix Guidelines Diagram E - 3)}
   \]

   \[
   t = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\frac{s_d}{\sqrt{n}}} > t_{a/2} \text{ (Guidelines E - 3)} \text{ or }
   \]

   \[
   t = \frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\frac{s_d}{\sqrt{n}}} < -t_{a/2} \text{ (Guidelines E - 3)}
   \]

   Now substitute the values given in the question into the formula to find the test statistic as follows:

   \[
   t = \frac{(0.0518) - (0)}{0.3055} = 2.04
   \]

   \[
   \alpha = 5\%, t_{144,0.025} = t_{144,0.05/2} = t_{144,0.025} = 1.96
   \]

   (Read from Appendix Table 8, p. 844

   for $\nu \to \infty$ t value approaches to z value for large sample sizes)

   Compare the test statistic to $t_{\infty,0.025}$

   If $t > t_{\nu,a/2}$, Reject $H_0$

   $2.04 > 1.96$

   Since the test statistic, 2.04, is higher than the critical value of 1.96, **Reject the null hypothesis**.

   Conclusion: There is a difference between the population means of ratios of the paired banks in Country A and B.

2.a) Scatter plot-Weekly Sales vs. Test Scores by using Excel as follows:

   **Regression Plot**

   \[
   \text{WklySales} = -11.5046 + 0.401835 \text{TestScores}
   \]

   \[
   S = 4.27924 \quad \text{R-Sq} = 60.0 \% \quad \text{R-Sq(adj)} = 53.4 \%
   \]
b) Use tabulated method to compute covariance and correlation coefficient as follows:

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$X_i$</td>
<td>$Y_i$</td>
<td>$(X_i - \bar{X})$</td>
<td>$(Y_i - \bar{Y})$</td>
<td>$(X_i - \bar{X})(Y_i - \bar{Y})$</td>
</tr>
<tr>
<td>1</td>
<td>55</td>
<td>10</td>
<td>-15.625</td>
<td>-6.875</td>
<td>107.4219</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>12</td>
<td>-10.625</td>
<td>-0.475</td>
<td>76.1265</td>
</tr>
<tr>
<td>3</td>
<td>85</td>
<td>28</td>
<td>14.375</td>
<td>11.125</td>
<td>164.375</td>
</tr>
<tr>
<td>4</td>
<td>75</td>
<td>24</td>
<td>4.375</td>
<td>7.125</td>
<td>31.0938</td>
</tr>
<tr>
<td>5</td>
<td>80</td>
<td>18</td>
<td>9.375</td>
<td>1.125</td>
<td>10.625</td>
</tr>
<tr>
<td>6</td>
<td>85</td>
<td>16</td>
<td>14.375</td>
<td>0.875</td>
<td>12.625</td>
</tr>
<tr>
<td>7</td>
<td>65</td>
<td>15</td>
<td>-5.625</td>
<td>-1.875</td>
<td>10.625</td>
</tr>
<tr>
<td>8</td>
<td>60</td>
<td>12</td>
<td>-10.625</td>
<td>-4.875</td>
<td>51.0938</td>
</tr>
<tr>
<td>(\Sigma)</td>
<td>565</td>
<td>135</td>
<td>(1021.875)</td>
<td>(274.875)</td>
<td>(410.625)</td>
</tr>
</tbody>
</table>

\[
\bar{X} = \frac{\sum_{i=1}^{n}X_i}{n} = \frac{565}{8} = 70.625; \quad \bar{Y} = \frac{\sum_{i=1}^{n}Y_i}{n} = \frac{135}{8} = 16.875
\]

\[
S_{XY} = Cov(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1} = \frac{1021.875}{8-1} = 120.823; \quad S_{XY} = 145.9821
\]

\[
S_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{1021.875}{8-1}} = 12.0823; \quad S_X^2 = 145.9821
\]

\[
S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}} = \sqrt{\frac{274.875}{8-1}} = 6.2664; \quad S_Y^2 = 39.2679
\]

\[
r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{Cov(X, Y)}{S_X S_Y} = \frac{58.6607}{(120.823)(6.2664)} = 0.775
\]

c) \[
b_1 = \frac{Cov(X, Y)}{S_X^2} = \frac{58.6607}{145.9821} = 0.401835
\]

Interpretation of $b_1$: For a one unit increase in the test score, we estimate or predict that there will be an increase of $401.84 in weekly sales.

$B_0 = Y - b_1 \bar{X} = 16.875 - (0.401835)(70.625) = -11.5046$

Interpretation of $B_0$: If the test score were 0, we would expect to have weekly sales of $-11504.6$. Interpret with caution—note that we are extrapolating the results beyond the observed data.

d) Prediction (Estimation) from the least-square linear correlation ($Value of r_{XY} = 0.810$) shows a positive moderately good relationship between the test score of sales representative and the weekly sales.

$\bar{Y}_{n+1} = b_0 + b_1 X_{n+1} = -11.5046 + 0.401835(70) = 16.62385 ($16623.85$)

3. a) \[
H_0: \text{Investing preference is independent of employment sector.}
\]

H1: There is a relationship between the employment sector and investing preference. In other words, investing preference is not independent of employment sector. 

b) Complete the contingency table (2x2 matrix) given to calculate expected counts (\(E_{ij}\)) from the corresponding formula below. These values were written under the observed counts (\(O_{ij}\)) as boldface and italic with the row and column total as follows:

\[
E_{ij} = \frac{R_i C_j}{n} \rightarrow E_{11} = \frac{R_1 C_1}{n} = \frac{(320)(400)}{700} = 182.857
\]
\[
E_{12} = \frac{R_1 C_2}{n} = \frac{(320)(300)}{700} = 137.143
\]
\[
E_{21} = \frac{R_2 C_1}{n} = \frac{(380)(400)}{700} = 217.143
\]
\[
E_{22} = \frac{R_2 C_2}{n} = \frac{(380)(300)}{700} = 162.857
\]

<table>
<thead>
<tr>
<th>Sector</th>
<th>Investing Preference</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aggressive</td>
<td>Balanced</td>
</tr>
<tr>
<td>Public</td>
<td>164</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td><strong>182.857</strong></td>
<td><strong>137.143</strong></td>
</tr>
<tr>
<td>Private</td>
<td>236</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td><strong>217.143</strong></td>
<td><strong>162.857</strong></td>
</tr>
<tr>
<td>Total</td>
<td>400</td>
<td>300</td>
</tr>
</tbody>
</table>

The calculated value of the chi-square test statistic is:

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

\[
\chi^2 = \frac{(164 - 182.857)^2}{182.857} + \frac{(156 - 137.143)^2}{137.143} + \frac{(236 - 217.143)^2}{217.143} + \frac{(144 - 162.857)^2}{162.857} = 1.9446 + 2.5928 + 1.6376 + 2.1834 = 8.3584
\]

Decision rule:

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

\[
\chi^2 = 8.3584 > \chi^2_{(2-1)(2-1), 0.01} = 6.63 \text{ (Read from Appendix Table 7)}
\]

Because the calculated value of the test statistic=8.3584 is greater than the critical value of 6.63, reject the null hypothesis.

b) Conclusion: There is a relationship between the employment sector and investing preference. Investing preference is not independent of employment sector.

\( p - \text{value} = P(\chi^2 > 8.3584) < 0.005 \text{ from Table 7, using Excel p - value = 0.00384} \)