1. (30 p) For a random sample of 125 entrepreneurs, the mean number of job changes was 1.91 and the sample standard deviation was 1.32. For an independent random sample of 86 corporate managers, the mean number of job changes was 0.21 and the sample standard deviation was 0.53. Test the null hypothesis that the population means are equal against the alternative that the mean number of job changes is higher for entrepreneurs than for corporate managers at the 1% significance level.

2. (35 p) Generale Bank agreed to a takeover from the Dutch-Belgian bank and insurance concern, named Fortis Group. This acquisition would propel Fortis to the top league of Europe’s largest financial institutions and speed up other cross-border mergers in the European financial services sector, an industry roiled by mounting competition and the launch of euro. Generale Bank’s management advisor is interested in the relationship between the market capitalization of Europe’s main bank insurance concerns and their total assets. The following data are listed as a random sample of these European companies with their market capitalization and total assets in billions of dollars:

<table>
<thead>
<tr>
<th>European Bank Insurance Firm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deutsche Bank</td>
</tr>
<tr>
<td>Credit Suisse</td>
</tr>
<tr>
<td>ING Group</td>
</tr>
<tr>
<td>Benelux Bank</td>
</tr>
<tr>
<td>Generale Bank</td>
</tr>
<tr>
<td>CGER-ASLK</td>
</tr>
<tr>
<td>Market Capitalization (Y) ($10^3)</td>
</tr>
<tr>
<td>Total Assets (X) ($10^3)</td>
</tr>
</tbody>
</table>

a) Prepare a scatter plot of the data
b) Compute covariance and correlation coefficient
c) Compute and interpret $b_1$ and $b_0$
d) Estimate the market capitalization if the total assets were 250 billion dollars.

3. (35 p) A contingency analysis table has been constructed from data obtained in a phone survey of customers in a market area in which respondents were asked to indicate whether they owned a domestic or foreign car and whether they were a member of a union or not. The following contingency table is provided:

<table>
<thead>
<tr>
<th>Car</th>
<th>Member of Union</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>155</td>
</tr>
<tr>
<td>Foreign</td>
<td>40</td>
</tr>
</tbody>
</table>

a) Use the chi-square approach to test whether the type of car owned (domestic or foreign) is independent of union membership. Test using an $\alpha = 0.05$ level.
b) State the conclusion of the test conducted in part (a).
STAT 204 STATISTICS II SPRING 14-15 RESIT EXAM SOLUTIONS

1. Let $1 = \text{entrepreneurs}; 2 = \text{corporate managers}$

First identify the null and alternative hypothesis

\[
H_0: \mu_1 - \mu_2 = 0 \\
H_1: \mu_1 - \mu_2 > 0
\]

To test $H_0: \mu_1 - \mu_2 = 0$ against the alternative $H_1: \mu_1 - \mu_2 > 0$ the decision rule is to reject $H_0$ and accept $H_1$ if

\[
t > t_{n_1+n_2-2,\alpha} \quad (upper - tail \ test, From \ Appendix \ Guidelines \ Diagram \ E - 2)
\]

\[
t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (Guidelines \ E - 2)
\]

\[
s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}; \text{ pooled standard deviation}
\]

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

\[
s_p = \sqrt{\frac{(125 - 1)(1.32)^2 + (86 - 1)(0.53)^2}{125 + 86 - 2}} = 1.07; t = \frac{(1.91 - 0.21) - 0}{1.07 \sqrt{\frac{1}{125} + \frac{1}{86}}} = 11.34
\]

\[
\alpha = 0.01, t_{209,0.01} = t_{\infty,0.01} = 2.326
\]

(For $v = \infty$ degree of freedom, $t$ value approaches to $z$ value of 2.33, Read from Appendix Table 8, p. 844)

Compare the test statistic to $t_{\infty,0.01}$

\[
11.34 > 2.326
\]

Reject $H_0$ at the 1% significance level: $\mu_x - \mu_y > 0$

Conclusion: There is sufficient evidence that the true mean number of job changes is higher for entrepreneurs than for corporate managers at all common levels of significance.

2.a) Scatter plot-Market Capitalization vs. Total Asset in billions of dollars as follows:
b) Use tabulated method to compute covariance and correlation coefficient as follows:

\[
\begin{array}{ccccccc}
 i & X_i & Y_i & (X_i - \bar{X}) & (Y_i - \bar{Y}) & (X_i - \bar{X})^2 & (Y_i - \bar{Y})^2 & (X_i - \bar{X})(Y_i - \bar{Y}) \\
1 & 585.0 & 46.3 & 264.667 & 70048.6209 & 9.3 & 86.49 & 2461.4031 \\
2 & 370.0 & 59.1 & 49.667 & 2466.8109 & 22.1 & 488.41 & 1097.6407 \\
3 & 309.4 & 49.4 & -10.933 & 119.5305 & 12.4 & 153.76 & -135.5692 \\
4 & 168.0 & 23.8 & -152.333 & 23205.3429 & -13.2 & 174.24 & 2010.7956 \\
5 & 160.8 & 9.8 & -159.533 & 25450.7781 & -27.2 & 739.84 & 4339.2976 \\
6 & 328.8 & 33.6 & 8.467 & 71.6901 & -3.4 & 11.56 & -28.7878 \\
\end{array}
\]

\[
\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{1922}{6} = 320.333; \quad \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{222}{6} = 37
\]

\[
S_{XY} = Cov(X,Y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} = \frac{121362.77}{6-1} = 1948.956
\]

\[
S_X = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{121362.77}{6-1}} = 155.7965; S_X^2 = 24272.554
\]

\[
S_Y = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1}} = \sqrt{\frac{1654.3}{6-1}} = 18.1896; S_Y^2 = 330.86
\]

\[
r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{1948.956}{(155.7965)(18.1896)} = 0.688
\]

c) For a one (or one billion) dollar increase in the total asset, we estimate or predict that there will be an increase of $0.0803 (or $80.3 million) in market capitalization.

\[
b_1 = \frac{\text{Cov}(X,Y)}{S_X^2} = \frac{1948.956}{24272.554} = 0.0803
\]

Interpretation of $b_0$: If the total asset were $0, we would expect to have market capitalization of $11.278. 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 $r_{XY} = 0.688$) shows a positive moderately good relationship between the total asset and the market capitalization.

\[
\hat{Y}_{n+1} = b_0 + b_1 X_{n+1} = 11.278 + 0.0803(250) = 31.353 \text{ billion dollars}
\]

3. a)

$H_0$: Type of car owned is independent of union membership.

$H_1$: There is a relationship between the union membership and type of car owned.

in other words, type of car owned is not independent of union membership

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:

\[
E_{ij} = \frac{R_i C_j}{n} \rightarrow E_{11} = \frac{R_1 C_1}{n} = \frac{(625)(195)}{990} = 123.1061
\]

\[
E_{12} = \frac{R_1 C_2}{n} = \frac{990}{(25)(795)} = 501.8939
\]

\[
E_{21} = \frac{R_2 C_1}{n} = \frac{71.8939}{990}
\]

\[
E_{22} = \frac{R_2 C_2}{n} = \frac{(365)(795)}{990} = 293.1061
\]
The calculated value of the chi-square test statistic is:

\[
\chi^2 = \sum_{r=1}^{2} \sum_{c=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(155 - E_{11})^2}{E_{11}} + \frac{(123.1061 - E_{12})^2}{E_{12}} + \frac{(123.1061 - E_{21})^2}{E_{21}} + \frac{(123.1061 - E_{22})^2}{E_{22}}
\]

\[
\chi^2 = \frac{(155 - 123.1061)^2}{123.1061} + \frac{(470 - 501.8939)^2}{501.8939} + \frac{(40 - 71.8939)^2}{71.8939} + \frac{(325 - 293.1061)^2}{293.1061}
\]

\[
= 8.2630 + 2.0268 + 14.1489 + 3.4705 = 27.9092
\]

Decision rule:

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

\[
\chi^2 = 27.9092 > \chi^2_{(2-1)(2-1),0.05}
\]

\[
\chi^2_{1,0.05} = 3.84 \text{ (Read from Appendix Table 7)}
\]

Since the calculated value of 27.9092 > 3.84, reject the null hypothesis.

b) Conclusion: There is a relationship between the type of the car owned and the union membership. Type of the car owned is not independent of union membership.

\( p - value = P(\chi^2_{v=1} > 27.9092) < 0.005 \text{ from Table 7,} \)

\( p - value = 1.27144 \times 10^{-7} \text{ using Excel CHIDIST function} \)

Small p-values provide strong evidence to reject the null hypothesis.