AMS 572.01 Midterm Exam # 2  ▪ Fall, 2003

Name:                  SSN:

Instructions.
This is a close book and close notes exam. Please provide detailed solutions. Anyone who cheats on the exam shall receive a course grade of F.

1. A genetics experiment on characteristics of tomato plants provided the following data on the number of offspring expressing four phenotypes. Please test at the significance level of 0.05 that the four phenotypes will appear in the proportion 9:3:3:10.

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Tall, cut-leaf</th>
<th>Dwarf, cut-leaf</th>
<th>Tall, potato-leaf</th>
<th>Dwarf, potato-leaf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>875</td>
<td>275</td>
<td>325</td>
<td>1025</td>
</tr>
</tbody>
</table>

(Solution) \( n = 2500 \), \( p_{1,0} = \frac{9}{25} \), \( p_{2,0} = \frac{3}{25} \), \( p_{3,0} = \frac{3}{25} \), \( p_{4,0} = \frac{10}{25} \),
\( e_i = n \cdot p_{i,0} \) \( (i = 1, 2, 3, 4) \)
Therefore \( e_1 = 900, e_2 = 300, e_3 = 300, e_4 = 1000 \).
The hypotheses are
\[
H_0 : p_1 = \frac{9}{25}, p_2 = \frac{3}{25}, p_3 = \frac{3}{25}, \text{and } p_4 = \frac{10}{25} \\
H_a : \text{At least, one of the above is not true.}
\]
Using the Chi-square goodness-of-fit test, the test statistic value is
\[
\chi_0^2 = \sum_{i=1}^{4} \frac{(n_i - e_i)^2}{e_i} = \frac{(875 - 900)^2}{900} + \frac{(275 - 300)^2}{300} + \frac{(325 - 300)^2}{300} + \frac{(1025 - 1000)^2}{1000} = 5.486.
\]
The critical value of this test is
\[
\chi_{4, 0.05}^2 = \chi_{5, 0.05}^2 = 7.815.
\]
Since \( \chi_0^2 = 5.486 < 7.815 \), we fail to reject \( H_0 \) at the significance level of 0.05 and claim that the four phenotypes will appear in the proportion 9:3:3:10.

2. An experiment with 5 subjects is conducted to determine the relationship between the percentage of a certain drug in the blood stream and the length of time it takes to react to a certain stimulus. The results are shown in the following table.

<table>
<thead>
<tr>
<th>Subject</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount of drug (%)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Reaction time (seconds)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>
(a) At $\alpha = 0.05$, is there a significant linear relationship between these two variables? What is the p-value of your test?

(Solution) Let $x_i$ ($i = 1, \ldots, 5$) be the amount of drug and $y_i$ ($i = 1, \ldots, 5$) be the corresponding reaction time.

The hypotheses are $H_0 : \beta_1 = 0$ v.s. $H_a : \beta_1 \neq 0$.

\[
S_{XX} = \sum_{i=1}^{5} x_i^2 - \frac{1}{n} (\sum_{i=1}^{5} x_i)^2 = 10.
\]
\[
S_{XY} = \sum_{i=1}^{5} x_i y_i - \frac{1}{n} (\sum_{i=1}^{5} x_i)(\sum_{i=1}^{5} y_i) = 7.
\]
\[
S_{YY} = \sum_{i=1}^{5} y_i^2 - \frac{1}{n} (\sum_{i=1}^{5} y_i)^2 = 6.
\]

\[
R^2 = r^2 = \frac{S_{XY}^2}{S_{XX}S_{YY}} = 0.8167 \quad \text{and} \quad SSE = S_{YY}(1 - R^2) = 1.0998.
\]

\[
s_e = \sqrt{\frac{SSE}{n - 2}} = \sqrt{\frac{1.0998}{3}} = 0.6055 \quad \text{and} \quad \hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = 0.7
\]

Therefore the test statistic is

\[
t_0 = \frac{\hat{\beta}_1 - 0}{s_e/\sqrt{S_{XX}}} = \frac{0.7}{0.6055/\sqrt{10}} = 3.656.
\]

Since $|t_0| > t_{n-2, \alpha/2} = t_{3,0.025} = 3.182$, we reject $H_0$. Therefore there is a significant linear relationship.

Since $t_{3,0.025} = 3.182 < |t_0| = 3.656 < t_{3,0.01} = 4.541$,

\[
: \quad 0.02 < \text{p-value} < 0.05
\]

(b) What is the sample Pearson correlation coefficient between these two variables? At $\alpha = 0.05$, does the sample correlation indicate a significant positive linear relationship between the variables? What is the p-value of your test?

(Solution)

The sample pearson correlation

\[
r = \frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}} = 0.904.
\]

The hypotheses are $H_0 : \rho = 0$ v.s. $H_a : \rho > 0$ and the test statistic

\[
t_0 = \frac{r - 0}{\sqrt{1-r^2/\sqrt{n-2}}} = \frac{0.904}{\sqrt{1-0.904^2/3}} = 3.662.
\]
Since $|t_0| > t_{3.0,0.05} = t_{3.0.05} = 2.353$, we can reject $H_0$. Therefore there is a significant positive linear relationship between the variables. From the fact that $t_{3.0.025} = 3.182 < t_0 = 3.66 < t_{3.0.01} = 4.541$,

\[ 0.01 < p\text{-value} < 0.025 \]

(c) What is the expected reaction time when the percentage of drug is 4.5?

(Solution)

\[
\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 2 - 0.7 \cdot 3 = -0.1. 
\]

When $x = 4.5$, \( \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = -0.1 + 0.7 \times 4.5 = 3.05 \text{ (sec)} \).

3. A consumer group aimed to compare the sugar content of 3 brands of cereals. A sample of 10 was taken from each brand. A summary of the sugar content (grams per serving) is as follows:

<table>
<thead>
<tr>
<th>Cereal Brand</th>
<th>Sugar Content (mean, std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy Nuts</td>
<td>6.2, 2.9</td>
</tr>
<tr>
<td>Hearty Pops</td>
<td>7.3, 3.2</td>
</tr>
<tr>
<td>Sweet Oats</td>
<td>10.3, 3.4</td>
</tr>
</tbody>
</table>

(a) Test at $\alpha = 0.05$ whether there are significant differences in sugar content between the three brands.

(Solution)

This is a one-way ANOVA. The hypotheses are

\[ H_0 : \mu_1 = \mu_2 = \mu_3. \]
\[ H_a : \text{At least, one of the above is not true.} \]

$t = 3$, $n_i = n = 10$ ($i = 1, 2, 3$), $n_T = 30$, $\bar{y}_1 = 6.2$, $\bar{y}_2 = 7.3$, $\bar{y}_3 = 10.3$, $\bar{y} = 7.93$.

\[
SST = \frac{n \sum_{i=1}^{3} (\bar{y}_i - \bar{y})^2}{t - 1} = \frac{10}{2} [ (6.2 - 7.93)^2 + (7.3 - 7.93)^2 + (10.3 - 7.93)^2] = 45.03.
\]
\[
SSW = \frac{s^2}{n_T - t} = \frac{1}{3} (2.9^2 + 3.2^2 + 3.4^2) = 10.07.
\]

\[ F_0 = \frac{45.03}{10.07} = 4.47. \]

Since $F_0 = 4.47 > F_{1,10-1,0.05} = F_{2,0.05} = 3.35$, we reject $H_0$ and conclude that there are significant differences in sugar content.
(b) Please enumerate assumptions necessary for the above test.

*Solution* Normality assumption, equal population variances, independent samples.

4. (Extra credit). Prove that the test for no Pearson correlation and the test for zero slope in a simple linear regression are, indeed, equivalent.

*Solution*

\[
\frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{S_{XY}}{S_{XX} S_{XY}} \sqrt{\frac{1-r^2}{n-2}} = \frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}} \sqrt{\frac{1-r^2}{n-2}} = \frac{S_{XY}}{S_{XX}} \sqrt{S_{XX}} \sqrt{\frac{1}{n-2}} = \frac{\hat{\beta}_1 \sqrt{S_{XX}}}{s_\epsilon} = \frac{\hat{\beta}_1}{s_\epsilon / \sqrt{S_{XX}}}
\]