Math 251 — Review for Final Test (Autumn 2007)

Notes. The final test will be comprehensive with approximately 3/4 of the test covering material that was previously tested. Approximately 1/4 of the test will cover the material that has not yet been tested. This includes: hypotheses tests on mean when $\sigma$ is unknown (9.2); hypotheses tests on proportions (9.3); correlation coefficients (10.1) and linear regression (10.2); and goodness of fit (11.2). Notice that we did not cover (9.5), so it will not be on the test.

The questions here focus on the new material. See your previous tests and reviews for an overview of previously tested material.

1. For healthy adults, the mean blood pH is $\mu = 7.4$. It is suspected a new arthritis drug changes blood pH (either higher or lower). Test this at a 5% significance level. To do this, a sample of 31 patients on the new drug were tested, and the sample had a mean $\bar{x} = 8.1$ and a standard deviation $s = 1.9$.

2. It was reported that the average life span of people in Hawaii is 77 years. Conduct an hypothesis test to see if the life span of people in Honolulu is less than 77 years. Use a level of significance of $\alpha = .05$. For this test, a random sample of 20 obituary notices of people in Honolulu had a mean $\bar{x} = 71.4$ years, and $s \approx 20.65$.

3. (a) If you did a right-tailed hypothesis test on a mean, and found that $t = 1.9641$ with $d.f. = 16$, what $P$-value would you report.

(b) Same question, but with a two-tailed test.

(c) Same question, but with a left-tailed test.

(d) Same question, for left-tailed test, but $t = -1.9641$.

4. Nationally about 28% of the population believes that NAFTA benefits America. A random sample of 48 interstate truck drivers showed that 19 believe NAFTA benefits America. Conduct an hypothesis test to determine whether the population proportion of interstate truckers who believe NAFTA benefits America is higher than 28%. Test at a level of significance of $\alpha = .05$.

(a) State the null and alternative hypotheses. Is this a right-tailed, left-tailed or two-tailed test?

(b) Find the $P$-value for the test, and report your conclusion for the test.

(c) Would you reject the null hypothesis at a level of significance of $\alpha = .04$?

(d) Would you reject the null hypothesis at a level of significance of $\alpha = .01$?

(e) Do you think the proportion of interstate truckers who believe NAFTA benefits America is higher than 28%? Explain your answer.
5. Draw scatter plots that exhibit: (a) no linear correlation; (b) moderate to good positive linear correlation; (c) moderate to good negative linear correlation (d) perfect positive linear correlation (e) perfect negative linear correlation.

(f) In general, does the correlation coefficient $r$ measure a cause and effect relationship between the variables $x$ and $y$?

6. Approximately what value would you expect for the correlation coefficient when a scatter plot exhibits: (a) no linear correlation; (b) good positive linear correlation; (c) good negative linear correlation (d) perfect positive linear correlation (e) perfect negative linear correlation?

(f) What are the possible values that a correlation coefficient can take?

7. The following sample data concerns the number of years a student studied German in school versus their score on a proficiency test.

<table>
<thead>
<tr>
<th>Years ($x$)</th>
<th>3</th>
<th>4</th>
<th>4</th>
<th>2</th>
<th>5</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>3</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Score ($y$)</td>
<td>57</td>
<td>78</td>
<td>72</td>
<td>58</td>
<td>89</td>
<td>63</td>
<td>73</td>
<td>84</td>
<td>75</td>
<td>48</td>
</tr>
</tbody>
</table>

The sums are: $\sum x = 35$ \hspace{1em} $\sum y = 697$ \hspace{1em} $\sum x^2 = 133$ \hspace{1em} $\sum y^2 = 50085$ \hspace{1em} $\sum xy = 2554$

(a) Find the equation of the least squares line for this data.

(b) Use your line from (a) to predict the score on the proficiency test of a person who had 3.5 years of German.

(c) Use the regression line in (a) to predict the number of years of German required to achieve a proficiency score of 75.

(d) Compute the correlation coefficient $r$ for this data. What does this coefficient suggest about a linear relationship between number of years German was studied in school and test scores for this sample? That is, determine whether it is a good fit, and whether it indicates a positive or negative linear relationship.

(e) Compute the coefficient of determination, and interpret what it means.
8. The following data is for investigating the relation between salary in thousands \( (x) \) and average number of absences per year \( (y) \) for employees of a company.

<table>
<thead>
<tr>
<th>Salary (( x ))</th>
<th>20</th>
<th>23</th>
<th>28</th>
<th>30</th>
<th>33</th>
<th>35</th>
<th>37</th>
<th>40</th>
<th>42</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absences (( y ))</td>
<td>2.4</td>
<td>2.2</td>
<td>1.9</td>
<td>2.1</td>
<td>1.5</td>
<td>1.4</td>
<td>1.3</td>
<td>0.5</td>
<td>0.4</td>
</tr>
</tbody>
</table>

For this data: \( \sum x = 288 \), \( \sum x^2 = 9660 \), \( \sum y = 13.7 \), \( \sum y^2 = 24.93 \), \( \sum xy = 398.2 \).

(a) Find the equation of the least squares regression line.

(b) Do the data appear to be positively or negatively correlated? Explain.

(c) Use the regression line to find: (i) How many absences per year would be expected from an employee that makes $38,000 per year? (ii) The expected salary for someone with 2 absences per year?

(d) Given that the correlation coefficient is \( -0.945 \), determine how well the data fits the line of best fit. Explain your answer.

9. (a) A local radio station claims that 15 percent of all people in Riverside say it is their favorite station, 65 percent of all people in Riverside listen to it occasionally, while 20 percent never listen to it. Suppose you surveyed 200 randomly selected people in Riverside and found that of those 200 people, 20 claimed it was their favorite station, 131 said they listen to it occasionally, while 49 never listen to it. Conduct an hypothesis test at a level of significance of \( 0.05 \) to determine whether the station’s claim concerning the distribution of listeners is correct.

(b) Explain why the sample size was sufficiently large for this test to be valid.

10. Suppose that in an attempt to determine whether a die is fair, you tossed it 120 times with the following results:

<table>
<thead>
<tr>
<th>Outcome</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>24</td>
<td>18</td>
<td>15</td>
<td>26</td>
<td>21</td>
<td>16</td>
</tr>
</tbody>
</table>

Does this provide sufficient evidence at a level of significance of \( \alpha = 0.05 \) to conclude that the die is not fair? Be sure to find the test statistic, P-value and state the conclusion.