Homework 9

In this homework you will review some previous material, one and two sample tests on proportions and the chi-squared tests.

(1) What is the purpose of an ANOVA?

(A) To test if the means of two populations are the same against the alternative that they are different.
(B) To test if the means of more than two populations are same against the alternative that at least one is different.
(C) To test if the variances of more than two populations are the same against the alternative that at least one is different.

(2) A meteorologist wants to understand whether the mean temperature in College Station is the different to the mean temperature in Snook (which is about 20 miles away from College Station). In order to test this hypothesis she measures the temperature at 10am in College Station and Snook between April 6th and April 15th (10 readings from both towns).

(a) What test would you recommend she does to test equality of the mean temperatures against the alternative that they are different?

(A) A two sample independent t-test.
(B) A paired t-test.
(C) A Wilcoxon sign rank test.
(D) A Wilcoxon sum rank test.
(E) ANOVA.

(b) Considering how the data was collected, what fundamental assumption (that all the above tests require) is likely to have been violated.

(3) In a few weeks time local elections will take place in Sutown. Mr. Big is contesting against Dr. Small. In a recent poll of 700 people, it was found that 380 were in favour of Dr. Small and 320 were in favour of Mr. Big.

(i) Test the hypothesis that Dr. Small will win the election (this will be the alternative hypothesis) (state precisely the null and alternative and do the test at the 5% level).

(ii) How many people need to be in the sample such that the 95% confidence interval for the proportion who will vote for Dr. Small will have length 3% (look at lecture 28)?
(4) Castaneda v Partida is an important court case in which statistical methods were used as part of the legal argument.

In Castaneda the plaintiffs alleged that the method for selecting juries in a county in Texas as biased against Mexican Americans. For the period of time of issue, there were 181,535 persons eligible for jury duty, of whom 143,611 were Mexican Americans. Of the 870 people selected for jury duty, 339 were Mexican American.

(a) We can do a one-sample proportion test, by using \( p = \frac{143,611}{181,535} = 0.79 \) as the overall proportion of mexicans in the county and comparing the proportion of Mexican american on jury to this.

<table>
<thead>
<tr>
<th>Hypothesis test results:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p ) : proportion of successes for population</td>
</tr>
<tr>
<td>( H_0 : p = 0.79 )</td>
</tr>
<tr>
<td>( H_A : p &lt; 0.79 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proportion</th>
<th>Count</th>
<th>Total</th>
<th>Sample Prop.</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p )</td>
<td>339</td>
<td>870</td>
<td>0.38965517</td>
<td>0.013809042</td>
</tr>
</tbody>
</table>

Figure 1:

(i) Explain why we are testing \( H_0 : p = 0.79 \) against \( H_A : p < 0.79 \)?

(ii) Using the output in Figure 1 calculate the z-transform and the the corresponding p-value.

(iii) Is there any evidence to support the view that Mexican Americans are being underrepresented in jury duty?

(b) An alternative method is to do a two-sided test and compare proportions of Mexican Americans who did jury duty with the proportion of non-Mexican American who did jury duty.

<table>
<thead>
<tr>
<th>Hypothesis test results:</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_1 ) : proportion of successes for population 1</td>
</tr>
<tr>
<td>( \gamma_2 ) : proportion of successes for population 2</td>
</tr>
<tr>
<td>( \gamma_1 - \gamma_2 ) : difference in proportions</td>
</tr>
<tr>
<td>( H_0 : \gamma_1 - \gamma_2 = 0 )</td>
</tr>
<tr>
<td>( H_A : \gamma_1 - \gamma_2 &gt; 0 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Difference</th>
<th>Count1</th>
<th>Total1</th>
<th>Count2</th>
<th>Total2</th>
<th>Simple Diff.</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_1 - \gamma_2 )</td>
<td>531</td>
<td>37924</td>
<td>339</td>
<td>143611</td>
<td>0.011641144</td>
<td>3.987177E-4</td>
</tr>
</tbody>
</table>

Figure 2:

(i) What hypothesis are you testing using this method?

(ii) Using the output in Figure 2 calculate the z-transform and the the corresponding p-value.
(iii) Is there any evidence to support the view that Mexican Americans are being underrepresented in jury duty?

(5) Three instructors gave STAT 651 exams last week. The number of As and Bs are recorded below. Students want to know whether the instructor has an influence on the grade a student obtains.

<table>
<thead>
<tr>
<th>Instructor 1</th>
<th>Instructor 2</th>
<th>Instructor 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>20</td>
<td>40</td>
</tr>
<tr>
<td>B</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>170</td>
<td></td>
</tr>
</tbody>
</table>

(a) State the null and alternative hypothesis to investigate.
(b) State the test you would do and why?
(c) Do the test at the 5% level, and report your findings.

(6) A gaming magazine is conducting a survey on the number of different gaming consoles an individual owns. There are currently three different gaming consoles on the market (Nintendo, Playstation and Xbox) and it thought that the probability an individual should own any one of these consoles is 0.4 (ie. the probability of owning a Nintendo is 40% and not owning a Nintendo is 60%, the probability of owning a Playstation is 40% and not owning a Playstation is 60% etc).

(a) Assuming that ownership of one system is completely independent of ownership of another system calculate the probability that an individual owns none, one, two or all three systems.

<table>
<thead>
<tr>
<th></th>
<th>No consoles</th>
<th>1 console</th>
<th>2 consoles</th>
<th>3 consoles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) The gaming magazine interviews 220 people, and asks them how many consoles they own (if any). The results are given below. Based on the above survey data, is there evidence to suggest that the probabilities derived in part are not the correct probabilities? State the test you will do, do the test at the 5% level and give the conclusions of the test.