I. T or F. (1 point each)
1. The \( \chi^2 \) -distribution is symmetric. ___F____
2. The \( \chi^2 \) may be negative, zero, or positive_____F____
3. The chi-square distribution is skewed to the right.______T____
4. The observed frequency of a cell should not be allowed to be smaller than 5 when a chi-square test is being conducted.____ F____
5. If the value of the chi-square statistic is less than the critical value, the null hypothesis must be rejected at a predetermined level of significance._____F____
6. If the data points form a straight horizontal or vertical line, there is strong correlation.___F____
7. If two variables are not linearly correlated then they are not related._____F____
8. If the value of the coefficient of linear correlation, \( r \), is near -1 for two variables, then the variables are not related._____F____
9. A set of data for which \( r = -1 \) or \( r = +1 \) will be such that the residuals equals zero._____T____
10. If your confidence level for two different proportion is \(-.11 < p_1 - p_2 < .13 \) than there is evidence that there is no difference between the two proportions.____T____

II. Multiple Choices. (2 points each)
1. In bivariate data, where both responses are quantitative order pairs \((x, y)\), what name do we give to the variable \( x \)?
   a. Attribute variable   b. Dependent variable   **c. Independent variable**   d. Output variable

2. For which of the following situations is it appropriate to use a scatter diagram?
   a. Presenting two qualitative variables
   b. Presenting one qualitative and one quantitative variables
   **c. Presenting two quantitative variables.**
   d. All of the above.

3. Which of the following would be the alternative hypothesis that would be used to test the claim that the mean IQ of individuals in population A is significantly different from the mean IQ of individuals in population B, assume independent sampling?
   
   \[ A. H_a : \mu_A - \mu_B = 0 \]
   \[ B. H_a : \mu_A - \mu_B > 0 \]
   \[ C. H_a : \mu_A - \mu_B < 0 \]
   **D. H_a : \mu_A - \mu_B \neq 0**

4. What is your conclusion for a right tail chi-square test with CV= 17.34 and \( \chi^2 \) test statistics of 2.5 ?
   A. Reject the null hypothesis
   **B. Fail to reject the null hypothesis**
   C. Unable to reject or fail to reject the null hypothesis
   D. None of the above is correct.

5. When computing \( \chi^2 \) we see that:
   A. large values of \( \chi^2 \) indicate agreement between the two sets of frequencies.
   **B. large values of \( \chi^2 \) indicate disagreement between the two sets of frequencies.**
   C. \( \chi^2 \) uses only continuous variables
   D. \( \chi^2 \) uses both continuous and categorical variables.
Note: Problems III-X are all worth 10 points each.

III. **Assume that the samples are independent and that they have been randomly selected.**

In a random sample of 360 women, 63% favored stricter gun control laws. In a random sample of 220 men, 58% favored stricter gun control laws. Test the claim that the proportion of women favoring stricter gun control is greater than the proportion of men favoring stricter gun control. Use a significance level of 0.05.

a) Set up hypothesis.

b) Find test statistics. You can do this manually or using calculator, either one you choose you must show work. For calculator state function you use, output you use, name the values that you use (example : s=2.5, this is the sample standard deviation)

c) State decision and WHY!

D) Interpretation.

**SOLUTION**

\[ n_1 = 360 \text{ women} \quad p_1 = 0.63 \text{ women favored stricter gun control} \]

\[ n_2 = 220 \text{ men} \quad p_2 = 0.58 \text{ men favored stricter gun control} \]

Claim: \[ p_1 > p_2 \quad \alpha = 0.05 \]

a) \[ H_o : p_1 = p_2 \]

\[ H_a : p_1 > p_2 \]

Requirement is satisfied \( np \geq 5 \) and \( nq \geq 5 \) hence we can use \( z \)-distribution

Using TI-83 calculator and 2-PropZTest rounding my \( x_1 \) value to 226 and \( x_2 \) to 127 a more conservative way. (If you round up i.e. \( x_1 = 227, x_2 = 128 \) you will get similar answer)

Output from calculator:

b) Test - statistics of \( z = 1.2 \) (rounded to two decimal places)

\[ p\text{-value} = 0.1133 \]

Because \( p\text{-value} \) of 0.1133 is more than 0.05 we fail to reject \( H_0 \) that men and women favor gun control equally!

d) There is not enough evidence to support the claim that proportion of women favoring gun control is greater than the proportion of men favoring gun control.
IV. Construct the indicated confidence interval for the difference between population proportions \( p_1 - p_2 \). Assume that the samples are independent and that they have been randomly selected.

In a random sample of 300 women, 48% favored stricter gun control legislation. In a random sample of 200 men, 29% favored stricter gun control legislation. Construct a 90% confidence interval for the difference between the population proportions \( p_1 - p_2 \).

a) If you are using a calculator, state function you are using, and output. If you are doing this manually, show work.

- If you are using a calculator the interpretation is worth 7 points. If you are doing it manually where you construct the confidence interval interpretation is 4 points.

SOLUTION

\( n_1 = 300 \) women \( p_1 = \) women 0.48 favored stricter gun control
\( n_2 = 200 \) men \( p_2 = \) men 0.29 favored stricter gun control

Requirement is satisfied \( np \geq 5 \) and \( nq \geq 5 \) hence we can use \( z \)-distribution

Using TI-83 calculator and 2-PropZInt > with 90% CI

Output: 90% CI is (0.1190, 0.2610)

Interpretation:

\( 0.119 < (p_1 - p_2) < 0.261 \)

Zero is not included in the interval.

We are 90% confident that the proportion of women who favored gun control differs from the proportion of men who favored gun control from 11.9% to 26.1%
V. Test the indicated claim about the means of two populations. Assume that the two samples are independent and that they have been randomly selected.
A researcher wishes to determine whether people with high blood pressure can reduce their blood pressure by following a particular diet. Use the sample data below to test the claim that the treatment population mean $\mu_1$ is smaller than the control population mean $\mu_2$. Test the claim using a significance level of 0.01.

<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_1 = 85$</td>
<td>$n_2 = 75$</td>
</tr>
<tr>
<td>$\bar{x}_1 = 189.1$</td>
<td>$\bar{x}_2 = 203.7$</td>
</tr>
<tr>
<td>$s_1 = 38.7$</td>
<td>$s_2 = 39.2$</td>
</tr>
</tbody>
</table>

a) Set up hypothesis.
b) Find test statistics. You can do this manually or using calculator, either one you choose you must show work. For calculator state function you use, output you use, name the values that you use (example: $s=2.5$, this is the sample standard deviation)
c) State decision and WHY!
d) Interpretation.

Claim: Treatment population mean is smaller than Control population mean
$\mu_1 =$ Treatment population mean
$\mu_2 =$ Control population mean

$a) H_0 : \mu_1 = \mu_2$
$H_a : \mu_1 < \mu_2$

$b$) Using TI-83 calculator $> 2$-SampTTest $>$ Stats $>$ ...put in Treatment and Control stats... $>$ pooled : NO

Note: We are using $t$-test because we don't know sigma

Output:
b) Test - statistic $t$-value = $-2.37$ (rounded to two decimal places)
$p = 0.0096$
$df = 155.01$ (rounded to two decimal places)

c) Decision: $p$-value 0.0096 is less than alpha 0.01, Reject $H_0$ equal means!
d) Interpretation: There is enough evidence to support the claim that Treatment population mean is smaller than Control population mean. This means that diet to reduce blood pressure is effective.
VI. Construct the indicated confidence interval for the difference between the two population means. Assume that the two samples are independent and that they have been randomly selected.

A researcher wishes to determine whether people with high blood pressure can reduce their blood pressure by following a particular diet. Use the sample data below to construct a 99% confidence interval for \( \mu_1 - \mu_2 \) where \( \mu_1 \) and \( \mu_2 \) represent the mean for the treatment group and the control group respectively.

<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 = 85 )</td>
<td>( n_2 = 75 )</td>
</tr>
<tr>
<td>( \bar{x}_1 = 189.1 )</td>
<td>( \bar{x}_2 = 203.7 )</td>
</tr>
<tr>
<td>( s_1 = 38.7 )</td>
<td>( s_2 = 39.2 )</td>
</tr>
</tbody>
</table>

\begin{align*}
\text{a) State the function you use in your calculator, why} \\
\text{that function is used, show output of your results.} \\
\text{b) State conclusion and interpretation. Using} \\
\text{calculator interpretation is 7 points, doing manually} \\
\text{interpretation is 4 points.}
\end{align*}

SOLUTION

\( \mu_1 = \) Treatment group

\( \mu_2 = \) Control group

Using a TI 83 calculator > 2-SampleTint > Pooled: NO

Output: 99% CI (-30.7, 1.45); df:155.01

\(-30.7 < (\mu_1 - \mu_2) < 1.45 \)

Zero is included in the interval.

Interpretation: We are 99% confident that the mean of Treatment group doesn’t differ from Control group because the limit of -30.7 and 1.45 contain the difference between means. Since zero is included in this interval confidence interval suggests that population’s means are equal.

VII. Perform the indicated goodness-of-fit test.

A company manager wishes to test a union leader's claim that absences occur on the different week days with the same frequencies. Test this claim at the 0.05 level of significance if the following sample data have been compiled.

<table>
<thead>
<tr>
<th>Day</th>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thur</th>
<th>Fri</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absences</td>
<td>37</td>
<td>15</td>
<td>12</td>
<td>23</td>
<td>43</td>
</tr>
</tbody>
</table>

\begin{align*}
\text{a) Set up hypothesis.} \\
\text{b) State the function you use in your calculator, why} \\
\text{that function is used, show output of your results,} \\
\text{label your test statistic.} \\
\text{c) State conclusion of your decision, why and} \\
\text{interpretation.}
\end{align*}

SOLUTION
$p =$ days of week

a) $H_0 : p_1 = p_2 = p_3 = p_4 = p_5$

$H_a : $ At least one is different

b) Using a TI-83 calculator > program goodness of fit

Observed values are : Mon = 37, Tue = 15, ...., Fri = 43

Expected values are $\frac{n}{k} \frac{130}{5} = 26;$

Output : $\chi^2$ - test statistics $= 28.31$, $p$ - value $\approx 0$, df $= 4$

c) $p$ - value of approx 0 is less than 0.05, Reject $H_0$ equality of proportions!

Interpret : There IS enough evidence to reject the claim that absences occur on different week days with the same frequencies.

VIII. Find the value of the linear correlation coefficient $r$ and test for significance at a 0.05 level. Then construct a regression equation and predict the growth for a temperature 54 and 95, comment on your predicted values.

The paired data below consist of the temperatures on randomly chosen days and the amount a certain kind of plant grew (in millimeters):

<table>
<thead>
<tr>
<th>Temp</th>
<th>62</th>
<th>76</th>
<th>50</th>
<th>51</th>
<th>71</th>
<th>46</th>
<th>51</th>
<th>44</th>
<th>79</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth</td>
<td>36</td>
<td>39</td>
<td>50</td>
<td>13</td>
<td>33</td>
<td>33</td>
<td>17</td>
<td>6</td>
<td>16</td>
</tr>
</tbody>
</table>

SOLUTION
a) $H_0 : \rho = 0$

$H_a : \rho \neq 0$

$x = \text{Temp}, \ y = \text{Growth}$

Using TI-83 calculator > LinRegTTest

Output : $y = a + bx$, $p$ - value $= 0.6141$, $a = 15.21$, $b = 0.21$, $r = 0.1955$

c) $p$ - value of 0.6141 is greater than significance level 0.05, Fail to reject $H_0$!

There is not enough evidence to suggest there is correlation between Temp and Growth

d) Regression equation : $\hat{y} = 15.21 + 0.21x$

e) Because there is no significant correlation between Temp and Growth the best predicted value is Average of Growth $= \bar{y} = 27$. Also, for a Temp of 95 the predicted value would have much error since 95 is outside by 16 degrees (this is a lot) of our Temp range.
IX. Use a $\chi^2$ test to test the claim that in the given contingency table, the row variable and the column variable are independent.

160 students who were majoring in either Math or English were asked a test question, and the researcher recorded whether they answered the question correctly. The sample results are given below. At the 0.10 significance level, test the claim that response and major are independent.

<table>
<thead>
<tr>
<th></th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>27</td>
<td>53</td>
</tr>
<tr>
<td>English</td>
<td>43</td>
<td>37</td>
</tr>
</tbody>
</table>

SOLUTION

a) $H_0$ : Response and Major are independent

$H_a$ : Response and Major are dependent

b) Using a TI-83 calculator $> \chi^2$ – test of independence

Note : Observed values are above in a 2x2 contingency table

Output : $\chi^2 test - statistics = 6.5, p-value = 0.0108, df = 1$

d) p-value of 0.0108 is less than 0.1 significance level, Reject Ho of independence!

Interpret : There is enough evidence to reject the claim that Response and Major are independent.

There is an association between Major and Responses.

X. **Perform test of Homogeneity.** A researcher wishes to test the effectiveness of a flu vaccination. 150 people are vaccinated, 180 people are vaccinated with a placebo, and 100 people are not vaccinated. The number in each group who later caught the flu was recorded. The results are shown below.

<table>
<thead>
<tr>
<th></th>
<th>Vaccinated</th>
<th>Placebo</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caught the flu</td>
<td>8</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>Did not catch the flu</td>
<td>142</td>
<td>161</td>
<td>79</td>
</tr>
</tbody>
</table>

Use a 0.05 significance level to test the claim that the proportion of people catching the flu is the same in all three groups.

a) Set up hypothesis.

b) Find test statistics. State the function you use in your calculator, why that function is used, show output of your results, label your test statistic.

c) State conclusion of your decision and interpretation.
SOLUTION

a) $H_0$ : The proportion of people catching the flu is the same for all three groups.

$H_a$ : At least one group is different.

b) Using TI - 83 calculator $\chi^2$ - test

Note : The above observed contingency table is a 2x3

Output : $\chi^2$ test statistics = 14.97, p - value $\approx$ 0, df = 2

c) p - value of approx 0 and is less than 0.05 significance level, Reject $H_0$ !

Interpret : There is enough evidence to reject the claim that the proportion of people cathicing the flu is the same for all three groups. There is evidence that vaccinations works against the flu.