NASEJJE MARIAM LUMALA

REG NO ;2023/DSCE/0083/SS

Assignment

What statistical hypothesis test can be applied to two samples of categorical data with the same categories to determine if the frequency distributions are significantly different?

To determine if the frequency distributions of two samples of categorical data with the same categories are significantly different, you can use the Chi-Square test for independence (also known as the Chi-Square test of association). This test is appropriate when you have two categorical variables and you want to assess whether there is a significant association or dependency between them.

Here's how you can apply the Chi-Square test for independence to compare two samples of categorical data:

1. \*\*Set Up Your Hypotheses\*\*:

- Null Hypothesis (H0): There is no significant association between the two categorical variables, and the observed frequencies are consistent with what would be expected by chance.

- Alternative Hypothesis (Ha):

There is a significant association between the two categorical variables, and the observed frequencies are not consistent with what would be expected by chance.

2. \*\*Create a Contingency Table\*\*:

Construct a contingency table (also known as a cross-tabulation or contingency table) that shows the observed frequencies of each category for the two variables. It will look something like this:

```

| Category A | Category B | Category C | ... | Total

--------------------------------------------------------------

Sample 1 | Obs\_11 | Obs\_12 | Obs\_13 | ... | Total\_1

Sample 2 | Obs\_21 | Obs\_22 | Obs\_23 | ... | Total\_2

--------------------------------------------------------------

Total | Total A | Total\_B | Total\_C | ... | Grand Total

```

3. \*\*Calculate the Chi-Square Statistic\*\*:

The Chi-Square statistic is calculated using the formula:

\[

\chi^2 = \sum \frac{(O\_{ij} - E\_{ij})^2}{E\_{ij}}

\]

Where:

- \(O\_{ij}\) is the observed frequency in cell (i, j).

- \(E\_{ij}\) is the expected frequency in cell (i, j), which is calculated as \(\frac{Total\_{i} \times Total\_{j}}{Grand Total}\).

4. \*\*Determine Degrees of Freedom (df)\*\*:

The degrees of freedom for the Chi-Square test for independence is calculated as \((R - 1) \times (C - 1)\), where \(R\) is the number of rows and \(C\) is the number of columns in the contingency table.

5. \*\*Find the Critical Value or P-Value\*\*:

You can use a Chi-Square distribution table or a statistical software package to find the critical value for your chosen significance level (e.g., 0.05) and degrees of freedom, or you can calculate the p-value directly.

6. \*\*Make a Decision\*\*:

Compare the calculated Chi-Square statistic to the critical value or p-value. If the calculated Chi-Square statistic is greater than the critical value (or if the p-value is less than your chosen significance level), you reject the null hypothesis, indicating that there is a significant association between the two categorical variables.

In summary, the Chi-Square test for independence is a suitable statistical test for comparing the frequency distributions of two categorical samples to determine if they are significantly different. It assesses whether there is an association between the two categorical variables, and if the association is significant, it implies that the distributions are different in a statistically meaningful way.