Department of Information & Communication Engineering

Assignment Name: Three-Way Contingency Tables and Large Sample Test of Significance

Submitted By:

SAMIUL ISLAM TONMOY

Roll: 2206042

Session 2021-22

Dept. of Information and Communication Engineering

Submitted To :

Dr. Md. Sarwar Hossain

Associate Professor

Dept. of Information and Communication Engineering

Assignment Name :Three-Way Contingency Tables and Large Sample Test of Significance.

Introduction: In the field of statistics, analyzing relationships between variables is crucial for understanding patterns, associations, and causality. Contingency tables are widely used to examine associations between categorical variables. While two-way tables are common, sometimes we need to explore interactions among three categorical variables simultaneously—this is where three-way contingency tables come in. Furthermore, when working with large datasets, we use large sample tests of significance to make reliable statistical inferences. This assignment explores both concepts in depth.

Three-Way Contingency Tables

Definition:

A three-way contingency table is a data table that displays the frequency distribution of three categorical variables. Each cell in the table represents a combination of values across the three variables.

Structure:

Let's denote the three variables as:

- **A** with *r* categories
- **B** with *c* categories

- **C** with k categories

A three-way table then contains $r \times c \times k$ cells.

These can be visualized as a series of $r \times c$ two-way tables, one for each level of variable C.

Example:

Suppose a researcher studies the association between:

- Gender (Male, Female)
- Smoking Status (Smoker, Non-Smoker)
- Age Group (Teen, Adult, Senior**)**

A three-way table would display how gender and smoking status relate within each age group.

Gender / Smoking	**Smoker**	**Non-Smoker**
Male	a_1	b_1
Female	a_2	b_2

Purposes and Uses

- Understanding conditional associations (e.g., does gender affect smoking status *within* a certain age group?)
- Testing for interaction effects
- Performing log-linear modeling (advanced statistical modeling of multi-way tables).

Analysis of Three-Way Contingency Tables

Types of Associations

1. **Marginal Association**: The association between two variables, ignoring the third.
   Example: Gender and Smoking Status, ignoring Age.

2. **Conditional Association**: The association between two variables within levels of the third variable.
   Example: Gender and Smoking within each Age Group.

3. **Homogeneous Association**: If the conditional association between two variables is the same across all levels of the third variable.

   Large Sample Test of Significance:

   Definition

   Large sample tests are statistical hypothesis tests that are reliable when the sample size is large (typically $n > 30$). These tests help determine whether the observed relationships or differences are statistically significant or due to random chance.

Chi-Square Test of Independence

This is the most commonly used test for categorical data in large samples. It is used to examine whether two (or more) variables are independent.

Formula:

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

O: Observed frequency

E: Expected frequency (calculated under the assumption of independence).

Steps:

1. **State Hypotheses:**
    - H_0: The variables are independent
    - H_1: The variables are not independent
2. **Calculate Expected Frequencies** for each cell:

$$E_{ijk} = \frac{(Row\ Total) \times (Column\ Total)}{Grand\ Total}$$

Compute the Chi-Square Statistic

Find degrees of freedom:

- For a three-way table:

$$df = (r-1)(c-1)(k-1)$$

Compare the calculated value with a critical value from the chi-square distribution table, or compute the **p-value**.

Decision Rule:

- If $\chi^2_{calc} > \chi^2_{critical}$

or p-value < significance level (e.g., 0.05), reject H_0.

Conditions for Using Chi-Square Test

- Data must be in frequency format
- Observations must be independent
- Expected frequency in each cell should be at least 5

Applications and Real-World Examples

- **Sociology**: Examining how education, income level, and ethnicity relate to voting behavior
- **Medical Research**: Studying how age, treatment type, and disease stage influence recovery rates
- **Marketing**: Investigating the relationship between age group, product preference, and region.

Conclusion:

Three-way contingency tables are an essential tool in multivariate categorical data analysis. They provide insights not only into marginal relationships but also into deeper conditional and interaction effects. When used with large samples, statistical tests like the chi-square test allow researchers to draw meaningful inferences and test hypotheses rigorously. Understanding and applying these techniques enables statisticians and analysts to uncover complex patterns and relationships in data.