

# ALY 6015: INTERMEDIATE ANALYTICS

Assignment 3: CHI-SQUARE and ANOVA Testing

**Submitted to**

# Prof. Fatemeh Ahmadi Abkenari

**Submitted by** Abhilash Dikshit Mrityunjay Gupta Sidddharth Alashi Smit Parmar

*College of Professional Studies Northeastern University - Vancouver Canada*

**Title - CHI-SQUARE and ANOVA Testing**

**Mrityunjay Gupta, Siddharth Alashi, Abhilash Dikshit, Smit Parmar**

***College of Professional Studies Northeastern University Vancouver, Canada***

## Abstract

## Introduction

1. **Problem#1: Blood Types**   
   A medical researcher wishes to see if hospital patients in a large hospital have the same blood type   
   distribution as those in the general population. The distribution for the general population is as  
   follows: type A= 20%; type B= 28%; type O= 36%; and type AB = 16%. He selects a random sample   
   of 50 patients and finds the following: 12 have type A blood, 8 have type B, 24 have type O, and   
   6 have type AB blood.   
   At α = 0.10, can it be concluded that the distribution is the same as that of the general population?

**Answer:**

To determine whether the blood type distribution of the hospital patients is the same as the general population, we can perform a chi-squared goodness of fit test.

The null hypothesis is that the distribution of blood types in the hospital patients is the same as the general population, while the alternative hypothesis is that the distribution is different.

The expected frequency for each blood type in the sample of 50 patients can be calculated based on the general population distribution:

* Type A: 20% of 50 patients = 10 patients
* Type B: 28% of 50 patients = 14 patients
* Type O: 36% of 50 patients = 18 patients
* Type AB: 16% of 50 patients = 8 patients

The chi-squared statistic can then be calculated as the sum of the squared differences between the observed and expected frequencies, divided by the expected frequencies:

$\chi^2 = \sum\_{i=1}^4 \frac{(O\_i - E\_i)^2}{E\_i} = \frac{(12 - 10)^2}{10} + \frac{(8 - 14)^2}{14} + \frac{(24 - 18)^2}{18} + \frac{(6 - 8)^2}{8} = 1.6 + 4.0 + 1.2 + 0.75 = 7.55$

Finally, the p-value can be calculated based on the chi-squared distribution with 3 degrees of freedom (since there are 4 categories minus 1).

If the p-value is less than 0.10 (the significance level), we can reject the null hypothesis and conclude that the distribution of blood types in the hospital patients is different from the general population.

In this case, the p-value can be calculated using a chi-squared distribution table, or by using statistical software. If the p-value is greater than 0.10, we fail to reject the null hypothesis and conclude that there is not enough evidence to say that the distribution of blood types in the hospital patients is different from the general population.

**Problem#2: Women in the Military**   
This table lists the numbers of officers and enlisted personnel for women in the military. At α =   
0.05, is there sufficient evidence to conclude that a relationship exists between rank and branch of   
the Armed Forces?

|  |  |  |
| --- | --- | --- |
| Action | Officers | Enlisted |
| Army | 10,791 | 62,491 |
| Navy | 7,816 | 42,750 |
| Marine Corps | 932 | 9,525 |
| Air Force | 11,819 | 54,344 |

**Answer:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Branch** | **Rank** | **Observed** | **Expected** |
| Army | Officers | 10,791 | (10,791 + 62,491) / 2 = 36,641 |
| Enlisted | 62,491 | (10,791 + 62,491) / 2 = 36,641 |
| Navy | Officers | 7,816 | (7,816 + 42,750) / 2 = 25,283 |
| Enlisted | 42,750 | (7,816 + 42,750) / 2 = 25,283 |
| Marine Corps | Officers | 932 | (932 + 9,525) / 2 = 5,229 |
| Enlisted | 9,525 | (932 + 9,525) / 2 = 5,229 |
| Air Force | Officers | 11,819 | (11,819 + 54,344) / 2 = 33,081 |
| Enlisted | 54,344 | (11,819 + 54,344) / 2 = 33,081 |

To determine if there is a relationship between rank (officers vs enlisted personnel) and branch of the Armed Forces, you can perform a chi-squared contingency test. The null hypothesis is that there is no relationship between rank and branch, while the alternative hypothesis is that there is a relationship.

The expected frequency for each combination of rank and branch can be calculated based on the overall distribution of rank and branch:

The chi-squared statistic can then be calculated as the sum of the squared differences between the observed and expected frequencies, divided by the expected frequencies:

$\chi^2 = \sum\_{i=1}^4 \sum\_{j=1}^2 \frac{(O\_{ij} - E\_{ij})^2}{E\_{ij}}$

Finally, the p-value can be calculated based on the chi-squared distribution with (2-1) \* (4-1) = 3 degrees of freedom. If the p-value is less than 0.05 (the significance level), we can reject the null hypothesis and conclude that there is a relationship between rank and branch of the Armed Forces.

In this case, the p-value can be calculated using a chi-squared distribution table, or by using statistical software. If the p-value is greater than 0.05, we fail to reject the null hypothesis and conclude that there is not enough evidence to say that there is a relationship between rank and branch of the Armed Forces.

1. Problem#3: Sodium Contents of Food   
   The amount of sodium (in milligrams) in one serving for a random sample of three different kinds   
   of foods is listed. At the 0.05 level of significance, is there sufficient evidence to conclude that a   
   difference in mean sodium amounts exists among condiments, cereals, and desserts?

Condiments Cereals Desserts   
270 260 100   
130 220 180   
230 290 250   
180 290 250   
80 200 300   
70 320 360   
200 140 300   
 160   
Source: The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter

Problem#4: Sales for Leading Companies   
The sales in millions of dollars for a year of a sample of leading companies are shown. At α = 0.01,   
is there a significant difference in the means?   
Cereal Chocolate Candy Coffee   
578 311 261   
320 106 185   
264 109 302   
249 125 689   
237 173   
Source: Information Resources, Inc.

Problem#5: Increasing Plant Growth   
A gardening company is testing new ways to improve plant growth. Twelve plants are randomly   
selected and exposed to a combination of two factors, a “Grow-light” in two different strengths   
and a plant food supplement with different mineral supplements. After a number of days, the plants   
are measured for growth, and the results (in inches) are put into the appropriate boxes. Can an   
interaction between the two factors be concluded? Is there a difference in mean growth with respect   
to light? With respect to plant food? Use α = 0.05.   
   
Grow-light 1 Grow-light 2   
Plant food A 9.2, 9.4, 8.9 8.5, 9.2, 8.9   
Plant food B 7.1, 7.2, 8.5 5.5, 5.8, 7.6