# statistics Assignment 4

## Problem statement 1:

Is gender independent of education level? A random sample of 395 people were surveyed & each person was asked to report the highest education level they obtained. The data that resulted from the survey is summarized in following tabel.

| | High school | Bachelors | Masters | ph.d | Total |
|---|---|---|---|---|---|
| Female | 60 | 54 | 46 | 41 | 201 |
| Male | 40 | 44 | 53 | 57 | 194 |
| Total | 100 | 98 | 99 | 98 | 395 |

Question: Are gender & education level dependent at 5% level of significance? In other words. given the data collected above is there a relationship between the gender of an individual & the level of education that they have obtained?

→ expected counts:

| | High school | Bachelors | Masters | ph.d | Total |
|---|---|---|---|---|---|
| Female | 50.886 | 49.868 | 50.377 | 49.868 | 201 |
| Male | 49.114 | 48.132 | 48.623 | 48.132 | 194 |
| Total | 100 | 98 | 99 | 98 | 395 |

$$\chi^2 = \frac{(60-50.886)^2}{50.886} + \dots + \frac{(54-48.132)^2}{48.132}$$

$$= 8.006$$

The critical value of $\chi^2$ with 3 degree of freedom is 7.815. since $8.006 > 7.815$ we reject null hypothesis & conclude that education level depends on gender at 5% level of significance

problem statement 2.

Using the following data perform a oneway analysis of variance using $\alpha = 0.5$ write up the results in APA format.

[ Group 1: 51, 45, 33, 45, 67 ]
[ Group 2: 23, 43, 23, 43, 65 ]
[ Group 3: 56, 76, 74, 87, 56 ]

| | Group 1 | Group 2 | Group 3 |
|---|---|---|---|
| | 51 | 23 | 56 |
| | 45 | 43 | 76 |
| | 33 | 23 | 74 |
| | 45 | 43 | 87 |
| | 67 | 65 | 56 |
| Mean | 48.2 | 35.4 | 69.8 |

Group 1

| | value | mean | deviations | sq. deviations |
|---|---|---|---|---|
| | 51 | 48.2 | 2.8 | 7.84 |
| | 45 | 48.2 | -3.2 | 10.24 |
| | 33 | 48.2 | -15.2 | 231.04 |
| | 45 | 48.2 | -3.2 | 10.24 |
| | 61 | 48.2 | 18.8 | 353.44 |
| SS | | | | 612.8 |

Group 2

| | value | mean | deviations | sq. deviations |
|---|---|---|---|---|
| | 23 | 35.4 | -12.4 | 153.76 |
| | 43 | 35.4 | 7.6 | 57.76 |
| | 23 | 35.4 | -12.4 | 153.76 |
| | 43 | 35.4 | 7.6 | 57.76 |
| | 45 | 35.4 | 9.6 | 92.16 |
| SS | | | | 515.2 |

Group 3

| | value | mean | deviations | sq. deviations |
|---|---|---|---|---|
| | 56 | 69.8 | -13.8 | 190.44 |
| | 76 | 69.8 | 6.2 | 38.44 |
| | 74 | 69.8 | 4.2 | 17.64 |
| | 87 | 69.8 | 17.2 | 295.84 |
| | 56 | 69.8 | -13.8 | 190.44 |
| SS | | | | 732.8 |

Mean square [Error or within group] = 155.07
degree of freedom [Error or within group] = 12
sum of squares [Within Group] = 1800.8

Problem statement 3

calculate F Test for given 10, 20, 30, 40, 50 & 5, 10, 15, 20, 25. for 10, 20, 30, 40, 50

Ans ⇒ Variance for 1st set

(N) = (10, 20, 30, 40, 50)
(N) = 5
Mean = $(x_1 + x_2 + x_3 - \cdots + x_n) / N$
Mean = $\frac{150}{5} = 30$

$SD = \text{sqrt}\left(1/(N-1) * ((x_1 - xm)^2 + (x_2 - m)^2 \cdots\right.$

$= \text{sqrt}\left(1/(5-1)\left((10-30)^2 + (20-30)^2 + (30-30)^2 + (40-30)^2 + (50-30)^2\right)\right.$
$= \text{sqrt}\left(1/4(-20)^2 + (-10)^2 + (0)^2 + (10)^2 + (20)^2\right)$
$= \text{sqrt}\left(1/4\,(400) + (100) + (0) + (100) + (400)\right)$
$= \text{sqrt}\,(250)$
$= 15.8114$

Variance $= SD^2$

Variance $= 15.8114^2$

Variance $= 250$

Variance of second set

for 5, 10, 15, 20, 25

$(N) = (5, 10, 15, 20, 25)$

$(N) = 5$

Mean $= (x_1 + x_2 + x_3 + N)/N$

$\frac{75}{5} = 15$

$SD = \text{sqrt}\left(1/(N-1)^* (x_1 - x_m)^2 + (x_2 - x_m)^2 - \cdots\right)$

$= \text{sqrt}\left(1/(5-1)(5-15)^2 + (10-15)^2 + (15-15)^2 + (20-15)^2 + (25-15)^2\right)$

$= \text{sqrt}\left(1/4\ (-10)^2 + (-5)^2 + (0)^2 + (5)^2 + (10)^2\right)$

$= \text{sqrt}\left(1/4\ (100) + (25) + (0) + (25) + (100)\right)$

$= \text{sqrt}\ (62.5)$

$= 7.905 f$

Variance $= SD^2$

$= (7.905 f)^2$

$= 62.5$

F- Test

F- Test = (Variance 1st) / (Variance 2nd)

$= 250/62.5$

$= 4$