**CHI-SQUARE**

Chi-Square can be used for three purposes namely

1. As a test of independence

2. To test the Homogeneity

3. To test the goodness of fit

**PURPOSE:**

Chi-square which is available in cross tab is used to test whether there is a significant association between two variables.

**ASSUMPTION:**

* The two variables that are to be applied in chi-square analysis must be of type category.
* The total number of observation used in this test must be large i.e., n>=30.
* Independence - The observations are always assumed to be independent of each other. This means chi-square cannot be used to test correlated data (like: matched pairs, panel data).
* It is frequency based test

**HYPOTHESIS:**

Null hypothesis: H0- There is no significant association between the two variables.

Alternate hypothesis: H1- There is significant association between the two variables.

**The two variables considered here are total income and total saving.**

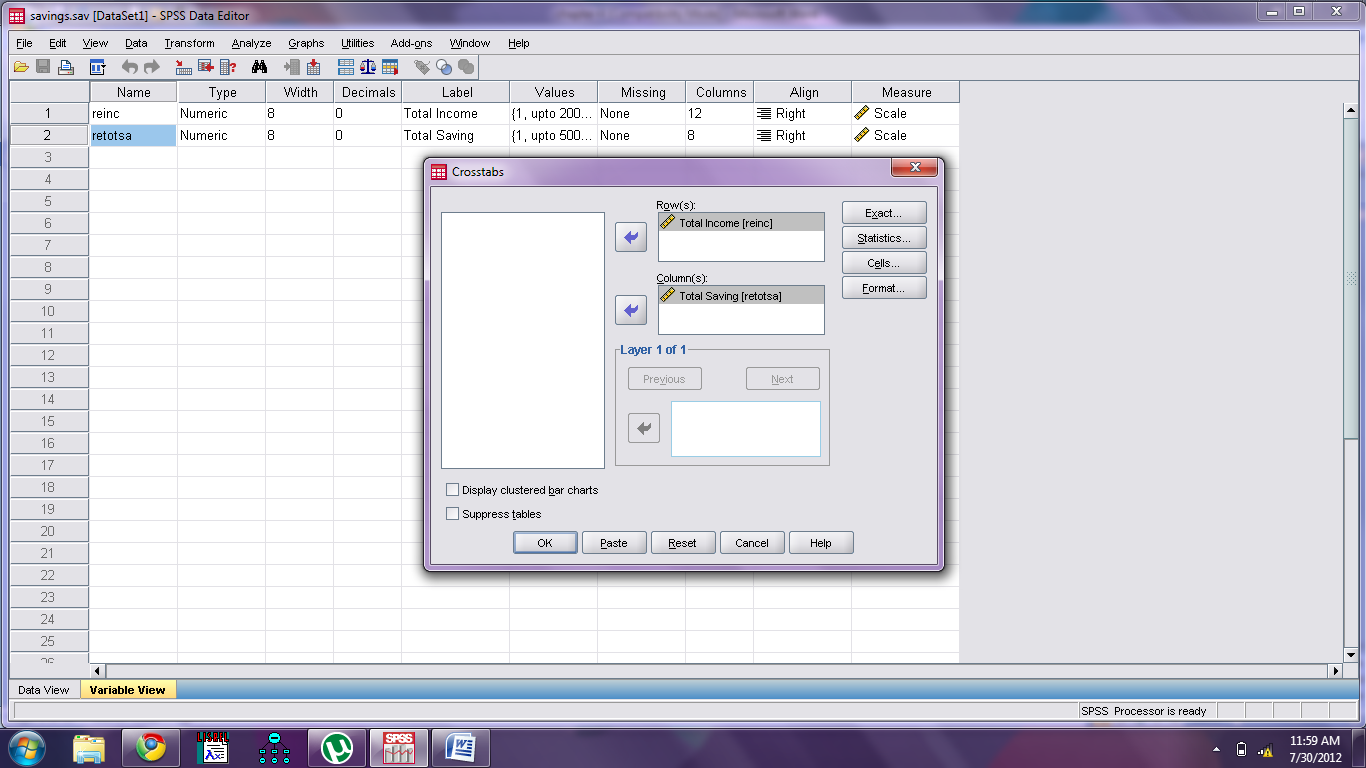
**PROCEDURE:**

STEP 1: Open the database, say SAVINGS.

STEP 2: To run a cross tab analysis, from the menu choose:

Analyze----->Descriptive Statistics------>cross tabs

STEP 3: The cross tab dialog box appears as shown



* select total income as the row variable
* select total saving as the column variable
* click statistics
* Select chi-square
* Click continue

STEP 4: Press ok in the cross tabs dialogue box table showing Cross Tab between total income and total saving would be generated.

| **Total Income \* Total Saving Cross tabulation** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Count |  |  |  |  |  |  |  |
|  |  | Total Saving | | | | | Total |
|  |  | upto 5000 | 5001-10000 | 10001-15000 | 15001-20000 | 20001 and above |
| Total Income | upto 20000 | 67 | 42 | 13 | 2 | 0 | 124 |
| 20001-30000 | 89 | 80 | 42 | 6 | 2 | 219 |
| 30001-40000 | 25 | 32 | 26 | 15 | 5 | 103 |
| 40001-50000 | 8 | 21 | 10 | 11 | 4 | 54 |
| 50001 and above | 7 | 9 | 19 | 7 | 10 | 52 |
| Total | | 196 | 184 | 110 | 41 | 21 | 552 |

| **Total Income \* Total Saving Cross tabulation** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Total Saving | | | | | Total |
|  |  |  | upto 5000 | 5001-10000 | 10001-15000 | 15001-20000 | 20001 and above |
| Total Income | upto 20000 | Count | 67 | 42 | 13 | 2 | 0 | 124 |
| Expected Count | 44.0 | 41.3 | 24.7 | 9.2 | 4.7 | 124.0 |
| % within Total Saving | 34.2% | 22.8% | 11.8% | 4.9% | .0% | 22.5% |
| 20001-30000 | Count | 89 | 80 | 42 | 6 | 2 | 219 |
| Expected Count | 77.8 | 73.0 | 43.6 | 16.3 | 8.3 | 219.0 |
| % within Total Saving | 45.4% | 43.5% | 38.2% | 14.6% | 9.5% | 39.7% |
| 30001-40000 | Count | 25 | 32 | 26 | 15 | 5 | 103 |
| Expected Count | 36.6 | 34.3 | 20.5 | 7.7 | 3.9 | 103.0 |
| % within Total Saving | 12.8% | 17.4% | 23.6% | 36.6% | 23.8% | 18.7% |
| 40001-50000 | Count | 8 | 21 | 10 | 11 | 4 | 54 |
| Expected Count | 19.2 | 18.0 | 10.8 | 4.0 | 2.1 | 54.0 |
| % within Total Saving | 4.1% | 11.4% | 9.1% | 26.8% | 19.0% | 9.8% |
| 50001 and above | Count | 7 | 9 | 19 | 7 | 10 | 52 |
| Expected Count | 18.5 | 17.3 | 10.4 | 3.9 | 2.0 | 52.0 |
| % within Total Saving | 3.6% | 4.9% | 17.3% | 17.1% | 47.6% | 9.4% |
| Total | | Count | 196 | 184 | 110 | 41 | 21 | 552 |
| Expected Count | 196.0 | 184.0 | 110.0 | 41.0 | 21.0 | 552.0 |
| % within Total Saving | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |

**Result Generated By SPSS**

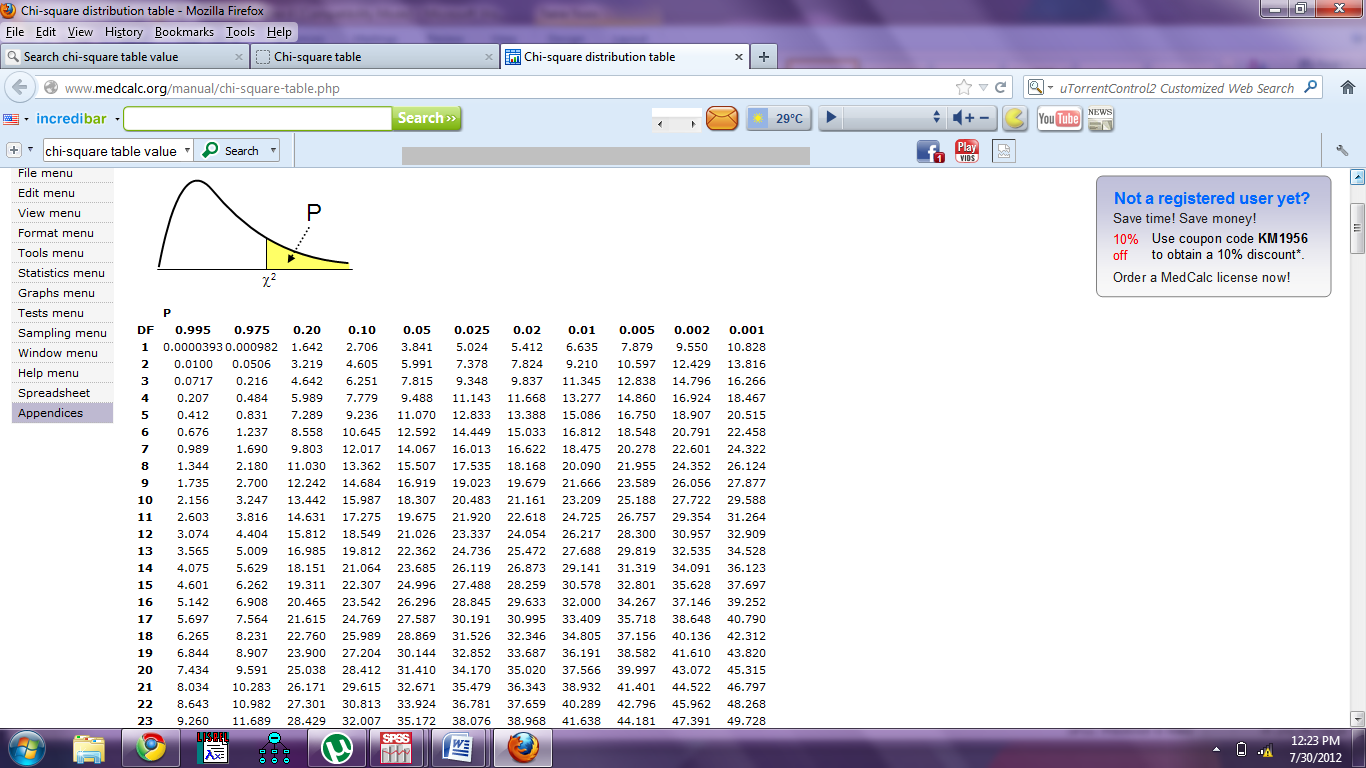
| **Chi-Square Tests** | | | |
| --- | --- | --- | --- |
|  | Value | df | Sig. (2-sided) |
| Pearson Chi-Square | 1.287E2a | 16 | .000 |
| Likelihood Ratio | 120.928 | 16 | .000 |
| Linear-by-Linear Association | 100.919 | 1 | .000 |
| N of Valid Cases | 552 |  |  |

The calculated value is 1.287 E2 which means 128.7 and it significant at this level of significance 0.000 at degrees of freedom 16.

In SPSS, if the significant value is less than 0.05 then reject null hypothesis and accept alternate hypothesis. In the above obtained result the significant value is less than 0.05 so, reject null hypothesis.

Hence there is significant association between total income and total saving.

Calculated value (128.7) is greater than table value (39. 253) at degrees of freedom 16 in the 0.000 level of significance in the chi-square table. So we can reject the null hypothesis and accept the alternative hypothesis.



Savings and Income

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **reinc \* retotsa Cross tabulation** | | | | | | | |
|  | | retotsa | | | | | Total |
| upto 5000 | 5001-10000 | 10001-15000 | 15001-20000 | 20001 and above |
| reinc | upto 20000 | 67 | 42 | 13 | 2 | 0 | 124 |
| 20001-30000 | 89 | 80 | 42 | 6 | 2 | 219 |
| 30001-40000 | 25 | 32 | 26 | 15 | 5 | 103 |
| 40001-50000 | 8 | 21 | 10 | 11 | 4 | 54 |
| 50001 and above | 7 | 9 | 19 | 7 | 10 | 52 |
| Total | | 196 | 184 | 110 | 41 | 21 | 552 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Chi-Square Tests** | | | |
|  | Value | Df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square | 128.683a | 16 | .000 |
| Likelihood Ratio | 120.928 | 16 | .000 |
| Linear-by-Linear Association | 100.919 | 1 | .000 |
| N of Valid Cases | 552 |  |  |
| a. 6 cells (24.0%) have expected count less than 5. The minimum expected count is 1.98. | | | |

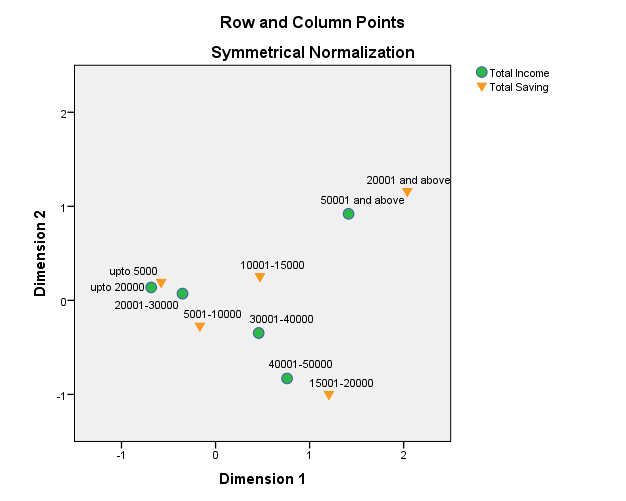
Post-Hoc –Column %

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **reinc \* retotsa Cross tabulation** | | | | | | | | |
|  | | | retotsa | | | | | Total |
| upto 5000 | 5001-10000 | 10001-15000 | 15001-20000 | 20001 and above |
| reinc | upto 20000 | Count | 67 | 42 | 13 | 2 | 0 | 124 |
| % within retotsa | 34.2% | 22.8% | 11.8% | 4.9% | 0.0% | 22.5% |
| 20001-30000 | Count | 89 | 80 | 42 | 6 | 2 | 219 |
| % within retotsa | 45.4% | 43.5% | 38.2% | 14.6% | 9.5% | 39.7% |
| 30001-40000 | Count | 25 | 32 | 26 | 15 | 5 | 103 |
| % within retotsa | 12.8% | 17.4% | 23.6% | 36.6% | 23.8% | 18.7% |
| 40001-50000 | Count | 8 | 21 | 10 | 11 | 4 | 54 |
| % within retotsa | 4.1% | 11.4% | 9.1% | 26.8% | 19.0% | 9.8% |
| 50001 and above | Count | 7 | 9 | 19 | 7 | 10 | 52 |
| % within retotsa | 3.6% | 4.9% | 17.3% | 17.1% | 47.6% | 9.4% |
| Total | | Count | 196 | 184 | 110 | 41 | 21 | 552 |
| % within retotsa | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |

Post-hoc Adjusted Residual

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **reinc \* retotsa Cross tabulation** | | | | | | | | |
|  | | | retotsa | | | | | Total |
| upto 5000 | 5001-10000 | 10001-15000 | 15001-20000 | 20001 and above |
| reinc | upto 20000 | Count | 67 | 42 | 13 | 2 | 0 | 124 |
| Adjusted Residual | 4.9 | .1 | -3.0 | -2.8 | -2.5 |  |
| 20001-30000 | Count | 89 | 80 | 42 | 6 | 2 | 219 |
| Adjusted Residual | 2.0 | 1.3 | -.4 | -3.4 | -2.9 |  |
| 30001-40000 | Count | 25 | 32 | 26 | 15 | 5 | 103 |
| Adjusted Residual | -2.6 | -.5 | 1.5 | 3.1 | .6 |  |
| 40001-50000 | Count | 8 | 21 | 10 | 11 | 4 | 54 |
| Adjusted Residual | -3.3 | .9 | -.3 | 3.8 | 1.5 |  |
| 50001 and above | Count | 7 | 9 | 19 | 7 | 10 | 52 |
| Adjusted Residual | -3.5 | -2.6 | 3.2 | 1.7 | 6.1 |  |
| Total | | Count | 196 | 184 | 110 | 41 | 21 | 552 |

Correspondence Analysis



Strength of Association

|  |  |  |  |
| --- | --- | --- | --- |
| **Symmetric Measures** | | | |
|  | | Value | Approximate Significance |
| Nominal by Nominal | Phi | .483 | .000 |
| Cramer's V | .241 | .000 |
| N of Valid Cases | | 552 |  |

Saving Motives and Perception

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **mocl3 \* perclus Cross tabulation** | | | | | |
| Count | | | | | |
|  | | perclus | | | Total |
| High Positive Cluster | Positive Cluster | Negative Cluster |
| mocl3 | Highly Motivated | 74 | 79 | 53 | 206 |
| Weakly Motivated | 25 | 63 | 49 | 137 |
| Self Centred | 49 | 108 | 52 | 209 |
| Total | | 148 | 250 | 154 | 552 |

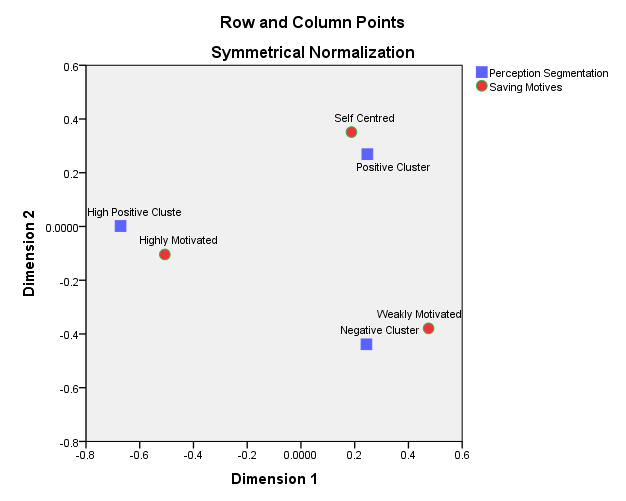
|  |  |  |  |
| --- | --- | --- | --- |
| **Chi-Square Tests** | | | |
|  | Value | Df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square | 19.165a | 4 | .001 |
| Likelihood Ratio | 18.885 | 4 | .001 |
| Linear-by-Linear Association | 2.526 | 1 | .112 |
| N of Valid Cases | 552 |  |  |
| a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 36.73. | | | |

Column %

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **mocl3 \* perclus Crosstabulation** | | | | | | |
|  | | | perclus | | | Total |
| High Positive Cluster | Positive Cluster | Negative Cluster |
| mocl3 | Highly Motivated | Count | 74 | 79 | 53 | 206 |
| % within perclus | 50.0% | 31.6% | 34.4% | 37.3% |
| Weakly Motivated | Count | 25 | 63 | 49 | 137 |
| % within perclus | 16.9% | 25.2% | 31.8% | 24.8% |
| Self Centred | Count | 49 | 108 | 52 | 209 |
| % within perclus | 33.1% | 43.2% | 33.8% | 37.9% |
| Total | | Count | 148 | 250 | 154 | 552 |
| % within perclus | 100.0% | 100.0% | 100.0% | 100.0% |

Post-hoc Adjusted Residual

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **mocl3 \* perclus Cross tabulation** | | | | | | |
|  | | | perclus | | | Total |
| High Positive Cluster | Positive Cluster | Negative Cluster |
| mocl3 | Highly Motivated | Count | 74 | 79 | 53 | 206 |
| Adjusted Residual | 3.7 | -2.5 | -.9 |  |
| Weakly Motivated | Count | 25 | 63 | 49 | 137 |
| Adjusted Residual | -2.6 | .2 | 2.4 |  |
| Self Centred | Count | 49 | 108 | 52 | 209 |
| Adjusted Residual | -1.4 | 2.4 | -1.2 |  |
| Total | | Count | 148 | 250 | 154 | 552 |



Strength of Assciation

|  |  |  |  |
| --- | --- | --- | --- |
| **Symmetric Measures** | | | |
|  | | Value | Approximate Significance |
| Nominal by Nominal | Phi | .186 | .001 |
| Cramer's V | .132 | .001 |
| N of Valid Cases | | 552 |  |

2X2 Matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **purchase \* gender Crosstabulation** | | | | |
| Count | | | | |
|  | | gender | | Total |
| Male | Female |
| purchase | No | 59 | 118 | 177 |
| Yes | 150 | 255 | 405 |
| Total | | 209 | 373 | 582 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Chi-Square Tests** | | | | | |
|  | Value | Df | Asymptotic Significance (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
| Pearson Chi-Square | .734a | 1 | .392 |  |  |
| Continuity Correctionb | .582 | 1 | .446 |  |  |
| Likelihood Ratio | .739 | 1 | .390 |  |  |
| Fisher's Exact Test |  |  |  | .400 | .223 |
| Linear-by-Linear Association | .733 | 1 | .392 |  |  |
| N of Valid Cases | 582 |  |  |  |  |
| a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 63.56. | | | | | |
| b. Computed only for a 2x2 table | | | | | |

Assumptions

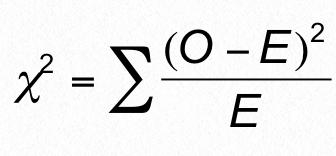
1. The expected should not be less than 5

|  |  |  |  |
| --- | --- | --- | --- |
| **Chi-Square Tests** | | | |
|  | Value | Df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square | 128.683a | 16 | .000 |
| Likelihood Ratio | 120.928 | 16 | .000 |
| Linear-by-Linear Association | 100.919 | 1 | .000 |
| N of Valid Cases | 552 |  |  |
| a. 6 cells (24.0%) have expected count less than 5. The minimum expected count is 1.98. | | | |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **reinc \* retotsa Crosstabulation** | | | | | | | | |
|  | | | retotsa | | | | | Total |
| upto 5000 | 5001-10000 | 10001-15000 | 15001-20000 | 20001 and above |
| Reinc | upto 20000 | Count | 67 | 42 | 13 | 2 | 0 | 124 |
| Expected Count | 44.0 | 41.3 | 24.7 | 9.2 | 4.7 | 124.0 |
| 20001-30000 | Count | 89 | 80 | 42 | 6 | 2 | 219 |
| Expected Count | 77.8 | 73.0 | 43.6 | 16.3 | 8.3 | 219.0 |
| 30001-40000 | Count | 25 | 32 | 26 | 15 | 5 | 103 |
| Expected Count | 36.6 | 34.3 | 20.5 | 7.7 | 3.9 | 103.0 |
| 40001-50000 | Count | 8 | 21 | 10 | 11 | 4 | 54 |
| Expected Count | 19.2 | 18.0 | 10.8 | 4.0 | 2.1 | 54.0 |
| 50001 and above | Count | 7 | 9 | 19 | 7 | 10 | 52 |
| Expected Count | 18.5 | 17.3 | 10.4 | 3.9 | 2.0 | 52.0 |
| Total | | Count | 196 | 184 | 110 | 41 | 21 | 552 |
| Expected Count | 196.0 | 184.0 | 110.0 | 41.0 | 21.0 | 552.0 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | B1 | B2 | B3 | B4 | Total |
| G1 | 25 |  |  |  |  |
| G2 | 25 |  |  |  |  |
| G3 | 25 |  |  |  |  |
| G4 | 25 |  |  |  |  |
|  | 100 |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | B1 | B2 | B3 | B4 | Total |
| G1 | 70 |  |  |  |  |
| G2 | 15 |  |  |  |  |
| G3 | 5 |  |  |  |  |
| G4 | 10 |  |  |  |  |
| Total | 100 |  |  |  |  |



Expected value

|  |  |  |  |
| --- | --- | --- | --- |
|  | Upto 5000 | 5000-10000 | Total |
| Upto 20000 |  |  | 150 |
| 20000-30000 |  |  | 50 |
| Total | 100 | 100 | 200 |

150 x 100 = 75

200

E = Rows Total x Column Total

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

Total Total Or Grand Total

Degrees of Freedom

|  |  |  |  |
| --- | --- | --- | --- |
| Total Income\Total Savings | Upto 5000 | 5000-10000 | Total |
| Upto 20000 |  |  | 150 |
| 20000-30000 |  |  | 50 |
| Total | 100 | 100 | 200 |

|  |  |  |  |
| --- | --- | --- | --- |
| Total Income\Total Savings | Upto 5000 | 5000-10000 | Total |
| Upto 20000 | Any value 80 | No Freedom only 70 | 150 |
| 20000-30000 | No Freedom only 20 | NO Freedom Only 30 | 50 |
| Total | 100 | 100 | 200 |

So only one freedom

Dof = (r-1) \* (c-1)

2-1 \* 2-1

1\*1 =1