Practical work 6 : The Chi-square test

The Chi-square test

• The Chi-square test is used to test the existence of a relationship between two discrete (qualitative) variables.

• The procedure for the test is as follows:

1. Formulation of hypotheses:

   H0: "There is no relationship between the variables X and Y to be tested."

   H1: "There is a relationship between the variables X and Y to be tested."

2. Determination of the observed Chi-square value (Chi-2Obs) from the studied table.
3. Determination of the number of degrees of freedom (z) of the studied table, and setting the significance level alpha for rejecting H0.
4. Determination of the Chi-square value, Chi-2(z, alpha), which represents the Chi-square value from a contingency table with z degrees of freedom. This value is obtained from a Chi-square test table found in appendices of statistical manuals.
5. Conducting the test:

   Accept H0 if: Chi-2Obs is less than or equal to Chi-2(z, alpha).

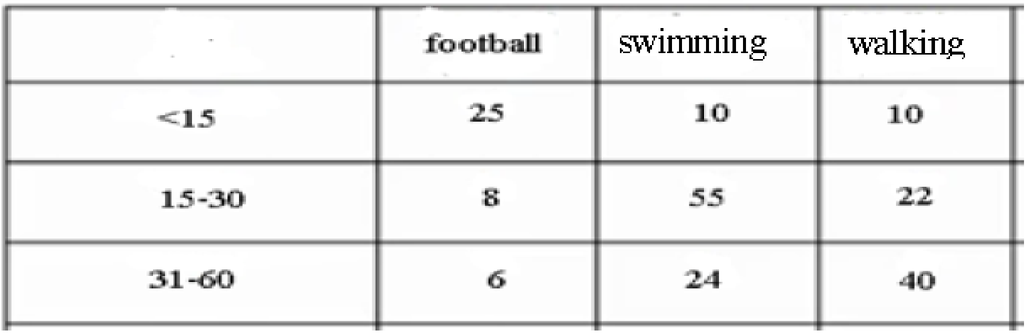
   Reject H0 if: Chi-2Obs is greater than Chi-2(z, alpha).

   And we accept the alternate hypothesis H1 ("there is a relationship of dependence between X and Y") with a risk of error alpha.

   Example

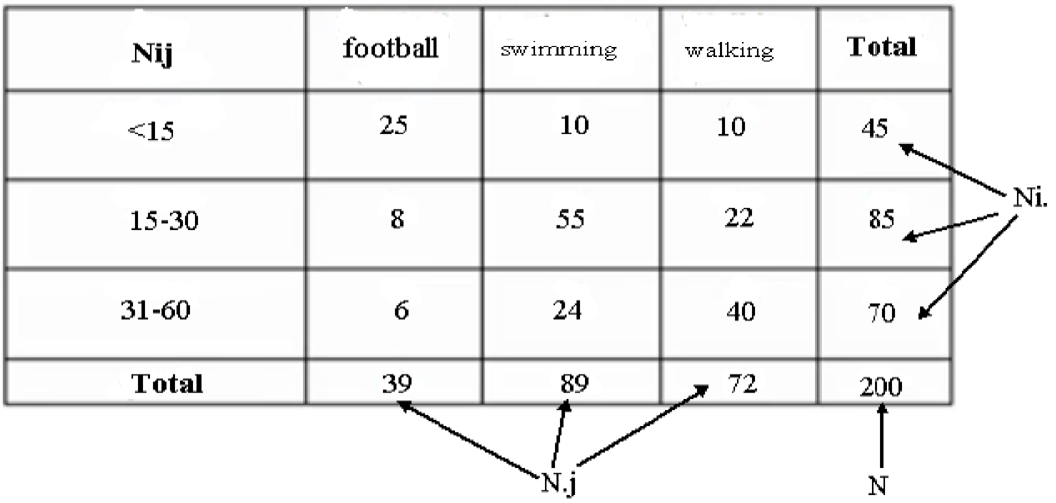
   Consider a sample of 200 individuals based on age and their preferred sports program.

   Is age independent of the preferred sports program of the studied sample?



	football	swimming	walking
<15	25	10	10
15-30	8	55	22
31-60	6	24	40

1. The creation of the contingency table

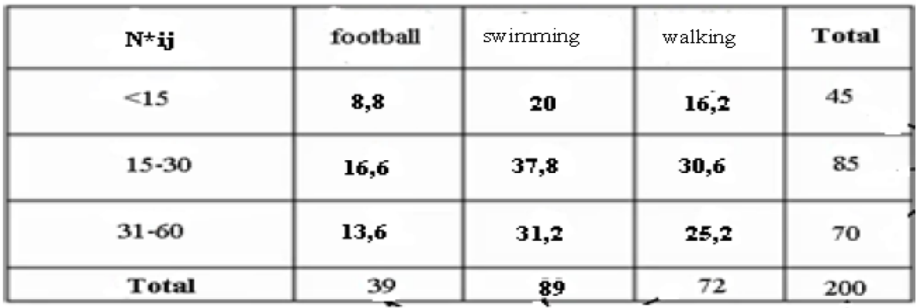


Nij	football	swimming	walking	Total
<15	25	10	10	45
15-30	8	55	22	85
31-60	6	24	40	70
Total	39	89	72	200

Nij: Frequency of the cell corresponding to the i-th row and j-th column of the table, meaning the number of individuals having the i-th attribute of X and the j-th attribute of Y.

Ni.: Sum of the i-th row, indicating the number of individuals having the i-th attribute of X. N.j: Sum of the j-th column, indicating the number of individuals having the j-th attribute of Y. N: Total number of individuals studied.

2. Calculation of theoretical frequencies (Nij*)
   N*ij = (Ni. * N .j) / N

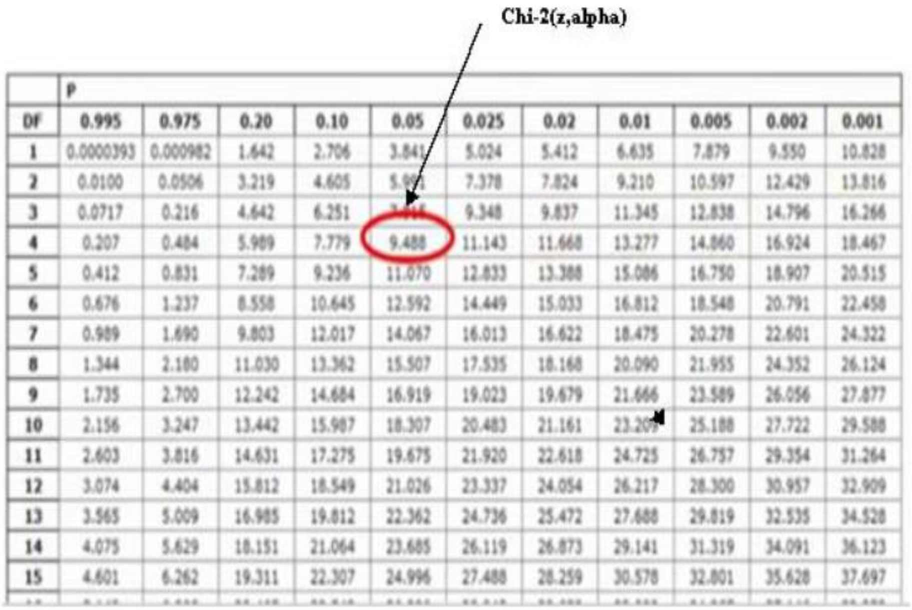


N*ij	football	swimming	walking	Total
<15	**8,8**	**20**	**16,2**	45
15-30	**16,6**	**37,8**	**30,6**	85
31-60	**13,6**	**31,2**	**25,2**	70
Total	39	89	72	200

$$CH-2_{obs} = \sum_{i} \sum_{j} (N_{ij} - N^*_{ij})^2 / N^*_{ij}$$

CH-2$_{obs}$=(25-8,8) *(25-8,8)/8,8+(10-20)*(10-20)/20+….(40-25,2)*(40-25,2)/25,2=**66,6**

3. Determination of the number of degrees of freedom z: z = (k-1)*(p-1) = (3-1)*(3-1)



Chi-2(z,alpha)

	P										
DF	0.995	0.975	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697

Chi-2Obs is greater than Chi-2(z, alpha), i.e., 66.6 > 9.488. We reject H0 and accept H1; thus, age and preferred sport are dependent with a 5% error rate.