The chi-square test

Here is how to do a chi square test of whether two categorical variables are independent.

1. Create a contingency table that shows how many scores fall in each pair of categories.

observed contingency table

	grade A	**grade B**
year 1	10	20
year 2	18	5

2. Find the marginal probabilities in the observed contingency table.

observed contingency table

	grade A	**grade B**	
year 1	10	20	30/53 = 0.566
year 2	18	5	23/53 = 0.434
	28/53 = 0.528	25/53 = 0.472	

3. Find the joint probabilities in the expected contingency table according to the null hypothesis, which says that the two categorical variables are independent.

expected contingency table

	grade A	**grade B**	
year 1	0.566*0.528 = 0.299	0.566*0.472 = 0.267	0.566
year 2	0.434*0.528 = 0.229	0.434*0.472 = 0.205	0.434
	0.528	0.472	

4. Multiply the joint probabilities by the sample size, to find the expected counts according to the null hypothesis.

expected contingency table

	grade A	**grade B**	
year 1	0.299*53 = 15.8	0.267*53 = 14.2	0.566
year 2	0.229*53 = 12.1	0.205*53 = 10.9	0.434
	0.528	0.472	

5. Calculate the chi-square statistic to summarize the difference between the observed and expected contingency tables.

$chi^2 = (10-15.8)^2/15.8 + (20-14.2)^2/14.2 + (18-12.1)^2/12.1 + (5-10.9)^2/10.9 = 10.6$

6. Calculate the number of degrees of freedom: df = (R-1)(C-1).

df = (2-1)(2-1) = 1

7. Use Appendix E to see whether the chi-square statistic is large enough to reject the null hypothesis.

Appendix E shows that at the alpha=0.05 level with one degree of freedom, we need a chi square statistic of 3.84 or greater to reject the null hypothesis. Our chi square statistic is 10.6, so we reject the null hypothesis that year of study and grade level are independent.