

Math 1180:Lab14

Kyle Gaffney

April 16th, 2014

1 Independence

Earlier we calculated independence by comparing the joint probabilities to the product of the marginal probabilities. Now instead of being given the probability distributions we instead only have data at our disposal? Can we still determine whether the events are independent?

Suppose that we have the following data set collected from a dart tournament:

Scores	P1 Hit	P1 Miss
P2 Hit	6	29
P2 Miss	27	38

Now we sum all of the cases like we did for the marginal probabilities.

Scores	P1 Hit	P1 Miss	Probs
P2 Hit	6	29	$\Pr(\text{P2 Hit}) = \frac{6+29}{100} = 0.35$
P2 Miss	27	38	$\Pr(\text{P2 Miss}) = \frac{27+38}{100} = 0.65$
Probs	$\Pr(\text{P1 Hit}) = \frac{6+27}{100} = 0.33$	$\Pr(\text{P1 Miss}) = \frac{29+38}{100} = 0.67$	1

Now with these observed marginal probabilities we get the following expected data table if the scores were independent:

Scores	P1 Hit	P1 Miss
P2 Hit	$0.33 \cdot 0.35 \cdot 100 = 11.55$	$0.67 \cdot 0.35 = 23.45$
P2 Miss	$0.33 \cdot 0.65 = 21.45$	$0.67 \cdot 0.65 = 43.55$

Now we wish to compare what the experimental marginal probabilities should have given us if the problem is independent to what we actually observed experimentally. To do this we compute the χ^2 statistic.

NOTE: The degree of freedom for a matrix or table of size (R)x(C) where R is the number of rows and C is the number of columns is degree of freedom = $(R-1) \cdot (C-1)$

$$\chi^2 = \frac{(6 - 11.55)^2}{11.55} + \frac{(29 - 23.45)^2}{23.45} + \frac{(27 - 21.45)^2}{21.45} + \frac{(38 - 43.55)^2}{43.55} = 6.12$$

For 1 degree of freedom the Critical value for $p=0.05$ is 3.841 and for $p=0.01$ is 6.635 So our result of $\chi^2 = 6.12$ is very significant that we do not have independence, or in other words the players effect one another.

2 Assignment for the Week

Suppose that you collect the following data set:

	0 lice	1 louse	2 lice
0 mites	10	10	15
1 mite	20	5	10
2 mites	10	10	10

1. Calculate the marginal probabilities for the experimental data.
2. Construct the expected data table if the number of lice and mites are independent.
3. Compare using the χ^2 test the 2 data sets (the collected data and the expected data). How many degrees of freedom are there? What is the χ^2 value? What is its significance level? What conclusion do you draw about the relationship between mites and lice?
4. If this data table looked familiar it is the expected value data of what we would generate using the joint probability distribution from Lab 7 (Covariance and Correlation). How does your conclusion from the χ^2 test compare to the results you came to using covariance and correlation?