



9

FIRST STEPS IN TESTING RELATIONSHIPS - GOODNESS OF FIT AND INDEPENDENCE TESTS

© Nick Lee and Mike Peters 2016.

QUESTION 1.

In the box below state in simple terms the purpose of a goodness of fit test for a multinomial population.

A goodness of fit test is used to compare the observed values with the expected results assuming the null hypothesis is true.

QUESTION 2.

The company for which you work has recently bought a new machine to fill bags of flour. On the machine which has been replaced, the standard deviation was 25 grams. A random sample of 20 bags from the new machine gave a standard deviation of 32 grams. You have to decide if this variability is due to chance or if the new machine needs to be recalibrated.

Complete the following:

$H_0: \sigma = 0.25$ and the variability is due to chance.

$H_1: \sigma > 0.25$ and the machine needs calibrating.

The value of χ^2 is 32.8.

The number of degrees of freedom is 19.

At $\alpha = 0.05$ the sample value of $\chi^2 = 30.1$, therefore the null hypothesis would be rejected.

At $\alpha = 0.01$ the sample value of $\chi^2 = 36.2$, therefore the null hypothesis would be rejected.

The variability is therefore probably due to a fault and the machine should be recalibrated.

QUESTION 3.

You have been asked to analyse the email requirements of a business. You conducted a survey of 100 employees and found that 25 send 5 or less emails per week, 35 send more than 5 but less than 20 per week, and 40 send more than 20 per week. The national statistics are: 30% send 5 or less per week, 40% send more than 5 but less than 20 per week, and 30% send more than 20 per week.

As part of your analysis you want to compare the business you are analysing with the national figures.

Your first step is to construct the null and alternative hypotheses.

H_0 : No difference between the business mean and the national mean.

H_1 : There is a difference between the means.

Now complete the table of actual and expected frequencies.

Expected	Actual
30	25
40	35
30	40

Using an α level of 0.05 decide if the null or alternative hypothesis is supported. At the stated α level the null hypothesis is supported.

My conclusion is that the business is performing as expected compared to the national figures.

QUESTION 4.

Verify your conclusion using Excel and SPSS.

The p -value is 0.091 and therefore the alternative hypothesis is rejected.

QUESTION 5.

The table below shows the result of a survey of politicians from 3 UK parties on their attitudes towards a change in VAT (Value Added Tax): Conservatives, Liberal Democrats and Labour.

VAT policy	Conservative	Liberal Democrat	Labour
Cut VAT	25	55	20
Raise VAT	60	30	10

The null hypothesis is H_0 : The two variables are independent.

The alternative hypothesis is H_1 : The two variables are dependent.

The expected value for each cell is greater than 5, therefore the χ^2 distribution can be used to test for independence.

The number of degrees of freedom is calculated by using $(r - 1)(c - 1)$.

The number of degrees of freedom for this analysis is 2.

The calculated χ^2 value is 25.098.

The critical α^2 value for a value of 0.05 is 5.991.

Therefore, since the calculated value is greater, the null hypothesis is rejected.

QUESTION 6.

Repeat the above using Excel and SPSS.

The calculated p -value is 0.0000035. Therefore the attitudes of the politicians are dependent on their political party affiliation.

MINI PROJECT

A company has commissioned you to investigate if there is a relationship between age and shopping online. The purpose of the study is for the company to decide how much they should invest in developing an online shopping portal. The products they manufacture are sold in a number of shops and tend to be purchased by people over 40.

The table below shows the results of a survey carried out in 5 of their shops. Using this data produce a report which analyses the data and based on this analysis, make a recommendation to the company as to the whether they should potentially invest in an on line shopping portal.

You should also point out in your report the limitations of your analysis and make a recommendation as to how they could obtain a more detailed analysis of their potential online customer base.

Age group	Online purchase			
	None	1-5	6-10	10+
below 20	25	50	75	150
20-40	30	60	70	140
40+	95	85	70	50

And finally...

How would a large chai tea be advertised in a tea shop for mathematicians and statisticians?