Using the Chi-Square Test for Association

You are a quality engineer preparing for your monthly report to the leadership team. One chart you always include is a Pareto diagram on last month's defects. As usual, scratches is the highest bar on the Pareto diagram. During the meeting, someone asks the question: "Is it the same for every shift?" You respond: "Good question."

But you are prepared and have that data as backup. You look at the Pareto diagrams by shift. The heights of the bars look a little different from shift to shift. But are the differences you see "real" or just normal variation in the process?

This is the question that the chi-square test for association addresses. This test compares two categorical variables to see if they are related. In this example, the two categorical variables are shift and defect. This publication introduces the chi-square test for association.

In this publication:

Example Data

Your plant assembles metal housings. Data is collected on 5 defect types by shift: scratches, dents, misalignment, porosity and other. There were 420 defects last month, each classified into one of the 5 defect types and by the shift where the defect occurred. The defect by shift data is shown in Table 1.

Table 1: Observed Defect Counts by Shift

Defect Type	Day	Evening	Night	Total
Scratches	52	47	45	144
Dents	31	29	33	93
Misalignment	14	18	48	80
Porosity	24	20	22	66
Other	12	10	15	37
Total	133	124	163	420

This type of table is called a contingency table. It provides counts for each combination of the two categories, defect type and shift in this example.

It is easy to see from Table 1 that the night shift has the most defects. But it is very possible that the night shift runs more parts. This is quite common, but for this analysis, it does not matter. The question you want to answer is the following:

Is the mix of defect types the same on each shift?

Table 2 shows the data in a different way. It shows each shift's defects as a percentage of that shift's total.

Table 2: Percent of Each Shift's Defects by Type

Defect Type	Day	Evening	Night
Scratches	39.1%	37.9%	27.6%
Dents	23.3%	23.4%	20.2%
Misalignment	10.5%	14.5%	29.4%
Porosity	18.0%	16.1%	13.5%
Other	9.0%	8.1%	9.2%
Total	100%	100%	100%

As expected, due to variation, the numbers are different. Scratches are 39.1% on days but only 27.6% on nights. Dents are about 23% on days and evenings, and 20.2% on nights.

The chi-square test for association will tell us if these numbers are the "same" (just chance variation) or if the differences in the numbers are statistically significant. The chi-square test for association will tell you if your two categorical variables, like shift and defect type are related. If not related, the mix of defect types should be about the same on every shift. If they are related, the mix will not be the same.

Check for Stability

Before combining a month of data into a single contingency table, it is good practice to consider whether the underlying process was reasonably stable during that period. If the process or the way defects are classified changed substantially during the month, combining all the observations can mask important time-related changes and make the results more difficult to interpret. In our example, we will assume that we have checked and the process is stable – in statistical control.

The Hypotheses and Expected Counts

The chi-square test for association is a test of hypothesis – like many other tests. There is a null (H_0) hypothesis and alternative (H_1) hypothesis. These are given by:

- H_0 (null): There is no association between the two categorical variables; the variables are independent.
- H_1: There is an association between the two categorical variables; the variables are not independent.

This test compares what we actually observe with what we would expect to observe if there is no association between the two categorical variables, e.g., if the null hypothesis is true. If there is no association, then each shift should have the same share of each defect type as the overall plant.

Consider the scratches defect. There are 144 total during the month, or about 34.3%. Day shift had 133 defects, so we expect 34.3% (or 45.60) of them to be scratches. The actual data was 52 scratches. Is this close enough to be the "same" as the 45.60 expectation? Our chi-square test for association will tell us.

The formula for the expected count (E) in a cell is given by:

$$E = (\text{Row Total})(\text{Column Total})/(\text{Grand Total})$$

The totals are given in Table 1. For example, for scratches on the day shift, the expected count (E) is:

$$E = (144)(133)/420 = 45.60$$

The expected count for porosity for the evening shift is:

$$E = (66)(124)/420 = 19.49$$

Table 3 shows the expected counts for all the cells when there is no association between the defect type and shift.

Table 3: Expected Counts if There Is No Association

Defect Type	Day	Evening	Night	Total
Scratches	45.60	42.51	55.89	144
Dents	29.45	27.46	36.09	93
Misalignment	25.33	23.62	31.05	80
Porosity	20.90	19.49	25.61	66
Other	11.72	10.92	14.36	37
Total	133	124	163	420

One assumption of the chi-square test is that the expected counts are large enough to use the chi-square approximation. A common guideline is that all expected counts should be at least 5. In this example, the smallest expected count is 10.92, so this requirement is satisfied.

Take a look at Tables 1 and 3. What differences do you see? For example, the night shift had 48 misalignment defects; you would expect 31. Is this difference significant? Let's see.

The Chi-Square Statistic and p-Value

There are differences between the observed and expected counts – even if there is really no association between the two categorical variables. The chi-square statistic measures this difference. It adds up how far each observed count is from its expected count.

To determine chi-square (χ^2) , for each cell, you square the difference between the observed count (O) and the expected count (E) and then divide by the expected count.

$$\chi^2 = \Sigma\,(O - E)^2 / E$$

Each of these values is called the cell's contribution to chi-square. Table 4 shows the results for each cell.

Table 4: Contribution to Chi-Square

Defect Type		Day	Evening	Night
Scratches	Observed	52	47	45
	Expected	45.60	42.51	55.89
	Contribution to χ^2	0.898	0.473	2.120
Dents	Observed	31	29	33
	Expected	29.45	27.46	36.09
	Contribution to χ^2	0.082	0.087	0.265
Misalignment	Observed	14	18	48
	Expected	25.33	23.62	31.05
	Contribution to χ^2	5.070	1.337	9.256
Porosity	Observed	24	20	22
	Expected	20.90	19.49	25.61
	Contribution to χ^2	0.460	0.014	0.510
Other	Observed	12	10	15
	Expected	11.72	10.92	14.36
	Contribution to χ^2	0.007	0.078	0.029

This data analysis is from the SPC for Excel software. You can learn more about SPC for Excel at this link.

Next, you add up all the contributions to χ^2. This gives $\chi^2 = 20.69$.

Remember, we are assuming that the null hypothesis is true. We are assuming that each cell's observed value matches the expected value. If this is true, then $\chi^2 = 0$. Normal variation prevents this from happening. The larger χ^2 is, the more likely there is an association. So, how do we determine if χ^2 is large enough to conclude that there is an association? We will compare the value with the chi-square distribution.

To use the chi-square distribution, we have to determine the degrees of freedom, df. This is given by:

$$\text{df} = (\text{number of rows} - 1) \times (\text{number of columns} - 1)$$

With 5 defect types and 3 shifts,

$$\text{df} = (5 - 1) \times (3 - 1) = 8.$$

You also have to determine the value for alpha. Alpha is the risk of concluding that defect type and shift are related when they really are not. Most of the time, alpha = 0.05. We will calculate a p-value based on this risk.

If the p-value is less than alpha, we reject the null hypothesis and conclude that there is evidence of an association. If the p-value is greater than or equal to alpha, we fail to reject the null hypothesis; there is not sufficient evidence to conclude that an association exists.

You can calculate the value of p using the Excel formula CHISQ.DIST.RT(χ^2, df) .

$$p = \text{CHISQ.DIST.RT}(20.69, 8) = .008$$

The p-value is the probability, assuming there is no association between the two categorical variables, of obtaining a χ^2 statistic at least as large as the one observed. Since the p-value is small (less than 0.05), we reject the null hypothesis and conclude that there is evidence of an association between defect type and shift.

For more on interpreting alpha and p-values, see our SPC Knowledge Base publication Interpretation of Alpha and p-Value.

Which Cells Drive the Result?

So, we have concluded that there is an association between defect type and shift. But what causes this? The chi-square test for association doesn't tell us where the issue is.

Look at Table 4. You are looking for large values of χ^2. Two cells stand out compared to the others. Misalignment on the night shift contributed 9.256 and misalignment on days contributed 5.07. Together they are responsible for 69% of the total contribution to χ^2. The other cells contribute very little.

The contribution to χ^2 is always positive, so it does not show direction. Are the expected counts more than the observed? Or less?

The residuals do give you direction. Table 5 shows the residuals. The residuals are simply the observed minus the expected value (see Table 4).

Table 5: Residuals

Defect Type	Day	Evening	Night
Scratches	6.40	4.49	-10.89
Dents	1.55	1.54	-3.09
Misalignment	-11.33	-5.62	16.95
Porosity	3.10	0.51	-3.61
Other	0.28	-0.92	0.64

Look at the two cells again. Misalignment on days has fewer defects than expected. Misalignment on nights has almost 17 more defects than expected. This is the problem and where you should focus your attention. What is happening on night shift that creates misalignment errors?

Summary

The chi-square test for association answers a common question in quality: are two categories related? In our example, the mix of defect types was not the same across shifts. The cell contributions and residuals showed that the difference came almost entirely from misalignment, which was much higher on night shift and lower on day shift.

Quick Links

Visit our home page

SPC for Excel Software

Download SPC for Excel Demo

SPC Knowledge Base

Ordering Information

Thanks so much for reading our publication. We hope you find it informative and useful. Happy charting and may the data always support your position.

Sincerely,

Dr. Bill McNeese