Notes:

Testing Hypotheses With Categorical Variables - Chi-Square Test

Key Learning Points

1. Describe the importance of completing a Chi-Square test.
2. Explain how to compare tables of categorical variables.
3. Utilize Chi-Square tests in improvement projects.

What is a Chi-square Test?

A Chi-square test is used to analyze data with two-way tables of discrete variables. Unlike the proportion tests, the data can take more than two values and more than two categories can be compared.

The Chi-square test uses data from a sample to estimate if a discrete response depends on the level of a discrete classification variable.

Potential Hypotheses:
H_o: {A} is independent of {B}
H_a: {A} is not independent of {B}

Minitab: Stat > Tables > Chi-square Test for Association

Contingency Table Description

A contingency table is used to analyze data via a two-way classification. In a table, this can be described as a row factor and as a column factor.

It tests the relationship between two variables/factors:

H_o: "{A} is independent of {B}"
H_a: "{A} is NOT independent of {B}"

A test statistic is calculated. This statistic uses a c2 (Chi-Square) distribution to test hypotheses about the frequency of occurrence of some event.

Given a and the degrees of freedom associated with each factor, a critical value is determined.

Are Requests Dependent on Location?

In this example, a project team is trying to determine if the relative frequency of imaging service requests by equipment type is independent of the five hospitals in their health system.

Step 1: State the Practical Problem

In this case, the question is:
Is the relative frequency of base service requests by imaging type independent of the five hospitals?

The Data

	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5
MR Scan	13	5	8	21	43
CT Scan	18	10	36	56	29
x-Ray	16	16	35	51	10

Step 2: Establish the Hypotheses

H_o: Relative frequency of base service requests by equipment type is independent of the hospital

H_a: Relative frequency of base service requests by equipment type is dependent on the hospital

Step 3: Decide on Appropriate Statistical Test

To determine independence of requests for service from the hospital, a Chi-Square Test of Independence should be used.

Step 4: Set the Alpha Level

We choose 95% confidence for our test.
$\alpha = 0.05$

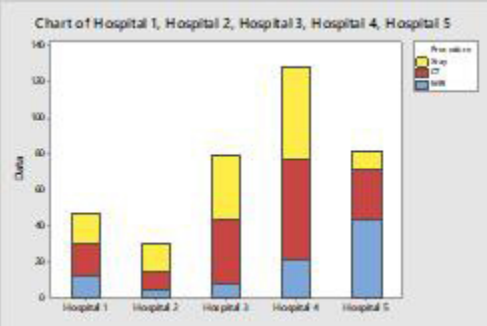
Steps 5&6: Set the Sample Size, and Collect the Data

In this case, the entire population of data will be used. Over the course of one year, in total, there were a combined 367 imaging service requests across all five hospitals for MRI Scans, CT Scans, and X-Rays.

Notes:

Step 7: Use the Appropriate Graphical Tool To Explore the Data

Graph the output counts by hospital for the three base imaging service requests. Does there appear to be a difference?



Step 8: Check Data Assumptions

For this example, there are no data assumptions that need testing.

Step 9: Run the Statistical Test

- **Minitab: Stat > Tables > Chi-Square Test for Association**
  - In the drop-down select Summarized data in a two-way table
  - Select the five columns of hospitals
  - Statistics: Check off Each cell's contribution to Chi-Square
  -

Notes:

Statistical Conclusion

Chi-Square Test for Association: Service, Worksheet columns

Rows: Service Columns: Worksheet columns

	Hospital 1	Hospital 2	Hospital 3	Hospital 4	Hospital 5	All
MR Scan	13	5	8	21	43	90
	11.53	7.60	19.37	31.39	20.11	
	0.1885	0.8907	6.6768	3.4389	26.0579	
CT Scan	18	10	36	56	29	149
	19.08	12.59	32.07	51.97	33.29	
	0.0613	0.5313	0.4807	0.3129	0.5532	
X-Ray	16	16	35	51	10	128
	16.39	10.81	27.55	44.64	28.60	
	0.0094	2.4894	2.0127	0.9052	12.0960	
All	47	31	79	128	82	367

Cell Contents
Count
Expected count
Contribution to Chi-square

Chi-Square Test

	Chi-Square	DF	P-Value
Pearson	56.705	8	0.000
Likelihood Ratio	56.325	8	0.000

Since the P-Value ≤ 0.05, reject the Ho.

Notice that Hospital 5 has the largest Chi Square value (38.707). Was something overlooked?

Step 10: Translate the Statistical Conclusion Into a Practical Conclusion

When comparing all hospitals, the relative frequency of imaging service by type of equipment is dependent on the hospital. Hospital 5 has a different rate of service requests than Hospitals 1-4.

When comparing only Hospitals 1-4, the relative frequency of imaging service requests by type of equipment is not dependent on the hospital.

When Should Chi-Square Tests Be Used?

Use a Chi-square test when comparing a discrete Y with a discrete X in a two-way contingency table.

Pitfalls to Avoid

- Discrete tests require large data sets to detect small differences.
- Each cell should have an expected count of at least five for appropriate results.

Notes: