Chi-Square Analysis

Define Measure Analyze Improve Control

LEAN SIX SIGMA

Bonacorsi Consulting

**Chi-Square Training for Attribute Data**

# What Is A Chi-Square Test?

- **What is a Chi-Square?**
- The probability density curve of a chi-square distribution is asymmetric curve stretching over the positive side of the line and having a long right tail.
- The form of the curve depends on the value of the degrees of freedom.

# Types of Chi-Square Tests?

- **Types of Chi-Square Analysis:**
- Chi-square Test for Association is a (non-parametric, therefore can be used for nominal data) test of statistical significance widely used bivariate tabular association analysis.
  - Typically, the hypothesis is whether or not two different populations are different enough in some characteristic or aspect of their behavior based on two random samples.
  - This test procedure is also known as the Pearson chi-square test.
- Chi-square Goodness-of-fit Test is used to test if an observed distribution conforms to any particular distribution. Calculation of this goodness of fit test is by comparison of observed data with data expected based on the particular distribution.

The numbers in a chi-square test come from the following formula:

$$\chi^2 = \frac{(\#_{observed} - \#_{expected})^2}{\#_{expected}}$$

# The Basics

## When to apply a Chi-Squared Test:

- Chi-Squared test is used to determine if there is a statistically significant difference in the proportions for different groups. To accomplish this, it breaks all outcomes into groups.

## What the Chi-Squared Test does:

- It starts by determining how many defects, for example, would be “expected” in each group involved.
- It does this by assuming that all groups have the same defect rate (which Minitab approximates from the data provided).
- Minitab then compares the expected counts with what was actually observed.
- If the numbers are different by a large enough amount, Chi-Square determines that the groups do not have the same proportion.

# The Basics (Cont.)

**Chi-Square Requirements:**

- Data is typically attribute (discrete). At the very least, all data must be able to be categorized as being in some category or another).
- Expected cell counts should not be low (definitely not less than 1 and preferable not less than 5) as this could lead to a false positive indication that there is a difference when, in fact, none exists.

**Chi-Square Hypotheses:**

- **Ho:** The null hypotheses (P-Value > 0.05) means the populations have the same proportions.
- **Ha:** The alternate hypotheses (P-Value <= 0.05) means the populations do **NOT** have the same proportions.

# Using Minitab – Step 1

Lean Six Sigma

| C1-T | C2 | C3 |
|---|---|---|
| Engineering Change Proposal | Success | Defect |
| A | 40 | 60 |
| B | 27 | 43 |
| C | 33 | 22 |
| D | 16 | 10 |
| E | 40 | 32 |
| F | 21 | 20 |
| G | 20 | 30 |
| H | 10 | 22 |
| I | 42 | 42 |
| J | 40 | 60 |

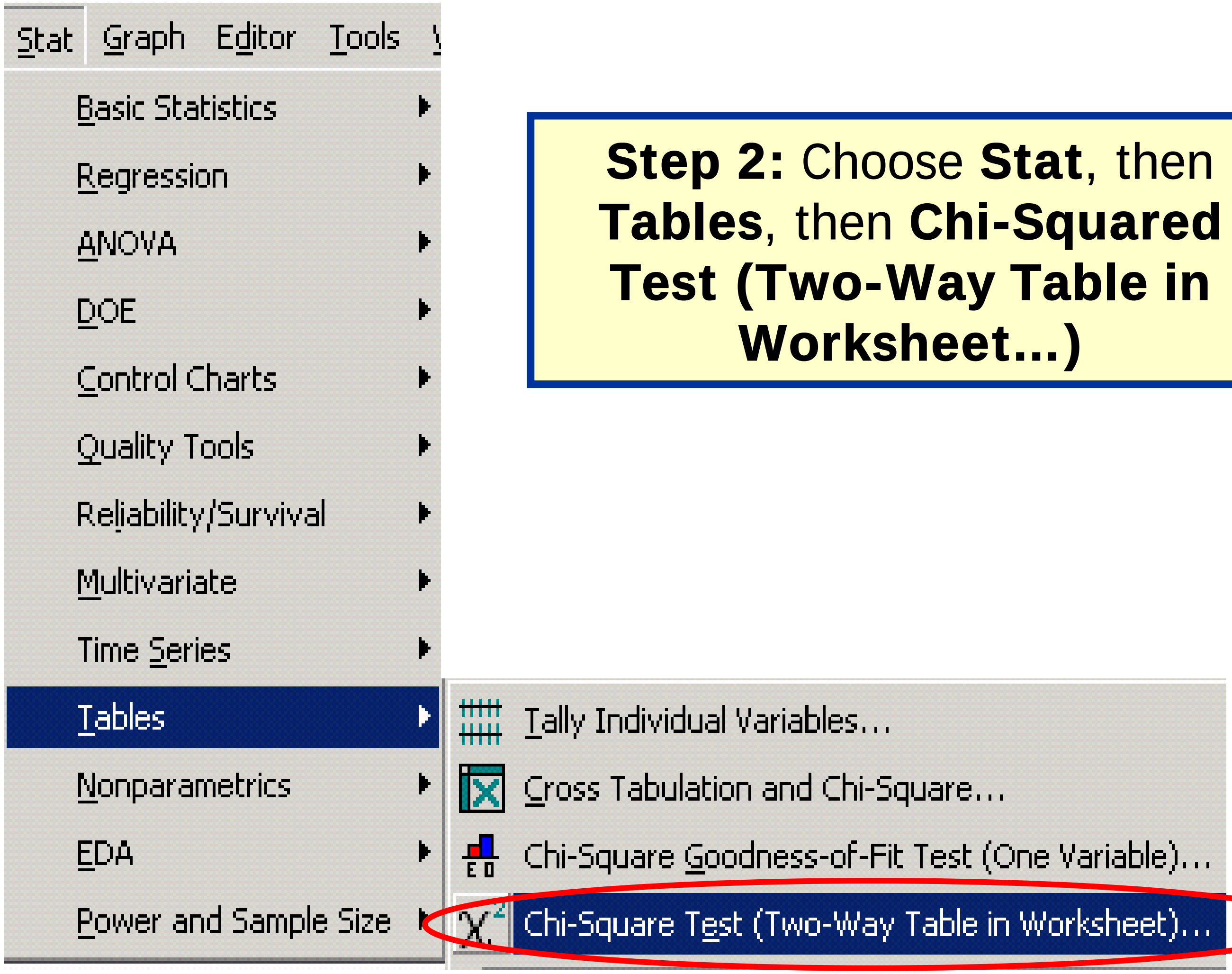
**Step 1:** Enter data into Minitab

**Enter Category Names**

**Enter the Defect Qty for the Data Collection Period**

**Enter the Success Qty for the Data Collection Period**

# Using Minitab – Step 2



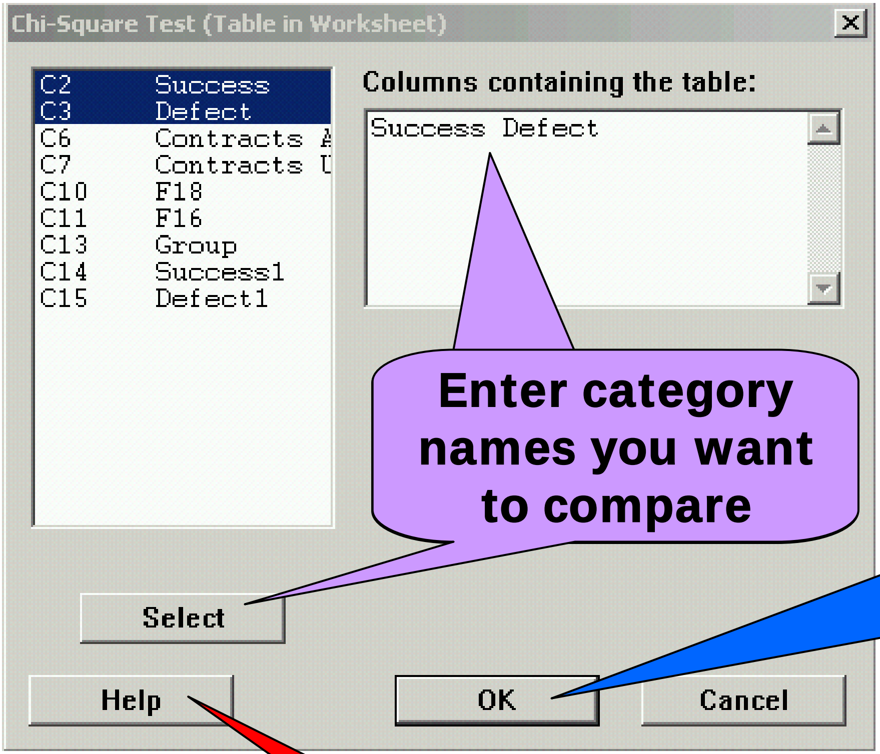

**Step 2:** Choose **Stat**, then **Tables**, then **Chi-Squared Test (Two-Way Table in Worksheet...)**

# Using Minitab – Step 3

**Step 3:** Enter the subgroup data to compare

Determine the subgroups you want to compare for variation.

In this example we examine the variation between success (C2) & Defect (C3) for each Engineering Change Proposal (ECP)




Enter category names you want to compare

Choose Ok to run Chi-Squared Test

Click help for Minitab help on Chi-Square

# The Session Window Results

Lean Six Sigma

| | Success | Defect | Total |
|---|---|---|---|
| 1 | 40 | 60 | 100 |
| | 45.87 | 54.13 | |
| | 0.752 | 0.637 | |
| 2 | 27 | 43 | 70 |
| | 32.11 | 37.89 | |
| | 0.814 | 0.689 | |
| 3 | 33 | 22 | 55 |
| | 25.23 | 29.77 | |
| | 2.393 | 2.028 | |
| 4 | 16 | 10 | 26 |
| | 11.93 | 14.07 | |
| | 1.391 | 1.179 | |
| 5 | 40 | 32 | 72 |
| | 33.03 | 38.97 | |
| | 1.471 | 1.247 | |
| 6 | 21 | 20 | 41 |
| | 18.81 | 22.19 | |
| | 0.255 | 0.217 | |
| 7 | 20 | 30 | 50 |
| | 22.94 | 27.06 | |
| | 0.376 | 0.319 | |
| 8 | 10 | 22 | 32 |
| | 14.68 | 17.32 | |
| | 1.492 | 1.264 | |
| 9 | 42 | 42 | 84 |
| | 38.53 | 45.47 | |
| | 0.312 | 0.264 | |
| 10 | 40 | 60 | 100 |
| | 45.87 | 54.13 | |
| | 0.752 | 0.637 | |
| | 289 | 341 | 630 |

Chi-Sq = 18.489, DF = 9, P-Value = 0.030

**The Chi-Square test calculates using the formula**
**$(40 - 45.87)^2 / 45.87 = 0.752$**

**Observed counts**

**Expected counts**

**Sum of each category (Machine)**

**Machine #5**

**Since the P-Value is <= 0.05, there is a 95% chance that there IS a statistically significant difference, in other words the groups (Machines) involved do NOT have the same defect rate.**

**Sum of all observed counts**

**Degree of Freedom (df = n-1)**

# Chi-Square Statistic

| | Success | Defect | Total |
|---|---|---|---|
| 1 | 40 | 60 | 100 |
| | 45.87 | 54.13 | |
| | 0.752 | 0.637 | |
| 2 | 27 | 43 | 70 |
| | 32.11 | 37.89 | |
| | 0.814 | 0.689 | |
| 3 | 33 | 22 | 55 |
| | 25.23 | 29.77 | |
| | 2.393 | 2.028 | |
| 4 | 16 | 10 | 26 |
| | 11.93 | 14.07 | |
| | 1.391 | 1.179 | |
| 5 | 40 | 32 | 72 |
| | 33.03 | 38.97 | |
| | 1.471 | 1.247 | |
| 6 | 21 | 20 | 41 |
| | 18.81 | 22.19 | |
| | 0.255 | 0.217 | |
| 7 | 20 | 30 | 50 |
| | 22.94 | 27.06 | |
| | 0.376 | 0.319 | |
| 8 | 10 | 22 | 32 |
| | 14.68 | 17.32 | |
| | 1.492 | 1.264 | |
| 9 | 42 | 42 | 84 |
| | 38.53 | 45.47 | |
| | 0.312 | 0.264 | |
| 10 | 40 | 60 | 100 |
| | 45.87 | 54.13 | |
| | 0.752 | 0.637 | |
| Total | 289 | 341 | 630 |

Chi-Sq = 18.489, DF = 9, P-Value = 0.030

**The series of numbers are added together for the overall chi-square statistic. When the value of the Chi-square statistic is high enough, the differences between the groups are concluded to be too large to be mere chance and the groups are determined to be truly different.**

# Errors in the Chi-Square

**Note:** if the expected cell counts are below 5, Minitab will print a warning. The warning is generated because of the fact that with the expected count in the denominator, a small value potentially creates an artificially large chi-square statistic. This is particularly troublesome if more than 20% of the cells have expected counts less than 5 and the contribution to the overall chi-square statistic is considerable.

**Additionally**, if any of the expected cell counts are below 1, Minitab will not even produce a p-value since the chi-square statistic is sure to be artificially inflated. In either of these cases, the **binomial distribution** (Minitab: Stat/ ANOVA/ Analysis of Means) may be able to be used.

**Lastly:** Attribute Gage R&R (AR&R) or Kappa Test is needed with an acceptable level of measurement system error prior to running a Chi-Square Analysis

# Lets See How You Do

**Example test questions for Analyze**

1. Consider a certain operation which produced 100 documents on line 1, 200 documents on line 2, and 300 documents on line 3. Additionally, it is known that group 1 produced 17 defective documents, group 2 produced 8 defective documents, and group 3 produced 11 defective documents. If a chi-square test were done on this data, what would be the expected number of defects seen by group 3 if the null hypothesis were completely true?

a) 24
b) 18
c) 12
d) 6

**Enter data in to Minitab like this:**

| Group | Defective | Acceptable |
|---|---|---|
| (group 1) | 17 | 83 |
| (group 2) | 8 | 192 |
| (group 3) | 11 | 289 |

2. Using the same data set, is this data statistically significant? **Yes** or **No**

**Note:** Answers are located in the note section of this slide

# Tips and Tricks

**Tips:**

- Determine the subgroups and categories to be tested for variation (differences in proportions) as part of your data collection plan.
- Define the operational definitions for success/defect, the stratifications layers (subgroups) and the Cause & Effect diagram (fishbone) to pre-determine where the team believes differences in proportions may exist.
- Continuous (Variable) data can usually be converted into Discrete (Attribute) data by using categories (Example: cycle time (continuous 1 hr, 1.5 hr, 2 hr) can be categorized into Cycle Time Met = 1 where success is cycle time <= 8 hrs or Cycle Time Not Met = 0 where a defect is cycle time > 8 hrs.)

**Tricks**

- An (MSA) Attribute R&R (Kappa Analysis) for discrete data or Gage R&R for continuous (variable) data is used prior to calculating the Chi-Square Test to ensure that the measurement variation < 10% Contribution. If the measurement variation is > 10% then the variation you will see in the Chi-Square Test is not valid as too much of the variation seen is coming from your measurement system (10% MSA error) and not your process variation.


# Appendix A

Backup Slides

# Chi-Square (X2) Distribution

**Machine Example Summary**

```
Chi-Sq =  0.752 +  0.637 +
          0.814 +  0.689 +
          2.393 +  2.028 +
          1.391 +  1.179 +
          1.471 +  1.247 +
          0.255 +  0.217 +
          0.376 +  0.319 +
          1.492 +  1.26[illegible]
          0.312 +  0.264 +
          0.752 +  0.637 = 18.489
DF = 9, P-Value = 0.030
```

**Row for 95% confidence**

**Degree of Freedom (df) = n-1**

**Machine example for Success or Defect per Month where we had n = 10 (Machines) Df = 10 – 1 = 9.**

**To be 95% confident of a difference, the P-Value <= 0.05**

**The sum of the Chi-Squared Test (18.489 would need to be > 16.92 where df = 9) for Ho**

VassarStats: Critical Values of Chi-Square

| | Level of Significance | | | | |
|---|---|---|---|---|---|
| df | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 |
| 1 | 3.84 | 5.02 | 6.63 | 7.88 | 10.83 |
| 2 | 5.99 | 7.38 | 9.21 | 10.60 | 13.82 |
| 3 | 7.81 | 9.35 | 11.34 | 12.84 | 16.27 |
| 4 | 9.49 | 11.14 | 13.28 | 14.86 | 18.47 |
| 5 | 11.07 | 12.83 | 15.09 | 16.75 | 20.51 |
| 6 | 12.59 | 14.45 | 16.81 | 18.55 | 22.46 |
| 7 | 14.07 | 16.01 | 18.48 | 20.28 | 24.32 |
| 8 | 15.51 | 17.53 | 20.09 | 21.95 | 26.12 |
| 9 | 16.92 | 19.02 | 21.67 | 23.59 | 27.88 |
| 10 | 18.31 | 20.48 | 23.21 | 25.19 | 29.59 |
| 11 | 19.68 | 21.92 | 24.73 | 26.76 | 31.26 |
| 1[illegible] | 21.03 | 23.34 | 26.22 | 28.30 | 32.91 |
| [illegible] | 22.36 | 24.74 | 27.69 | 29.82 | 34.53 |
| 14 | 23.68 | 26.12 | 29.14 | 31.32 | 36.12 |
| 15 | 25.00 | 27.49 | 30.58 | 32.80 | 37.70 |
| 16 | 26.30 | 28.85 | 32.00 | 34.27 | 39.25 |
| 17 | 27.59 | 30.19 | 33.41 | 35.72 | 40.79 |
| 18 | 28.87 | 31.53 | 34.81 | 37.16 | 42.31 |
| 19 | 30.14 | 32.85 | 36.19 | 38.58 | 43.82 |
| 20 | 31.41 | 34.17 | 37.57 | 40.00 | 45.31 |
| 21 | 32.67 | 35.48 | 38.93 | 41.40 | 46.80 |
| 22 | 33.92 | 36.78 | 40.29 | 42.80 | 48.27 |
| 23 | 35.17 | 38.08 | 41.64 | 44.18 | 49.73 |
| 24 | 36.42 | 39.36 | 42.98 | 45.56 | 51.18 |
| 25 | 37.65 | 40.65 | 44.31 | 46.93 | 52.62 |

**Chi Squared Distribution table is used by Minitab to calculate the P-Value using the Degree of Freedom (df) and the variation between each category's data sets being compared for differences.**

# Chi-Square (for Two-Way)

**Chi-Square equation for two-way tables:**

$$\chi^2 = \sum_{j=1}^{g} \frac{(f_o \quad f_e)^2}{f_e}$$

**Where:**

$f_0$ = observed frequency
$f_e$ = expected frequency
r = number of rows
c = number of columns
g = number of groups = r * c
degrees of freedom = (r-1) * (c-1)
All $f_e$'s should be > 5 *

* For expected frequencies less than 5 the calculated chi-square value changes dramatically as $f_e$ changes. Therefore the calculated value becomes less reliable and should be interpreted cautiously.

$H_o$: Independent (There is *no* relationship between the populations)
$H_a$: Dependent (There *is* a relationship between the populations)

**Rejection Criteria:**
**Fail to reject $H_o$ when $p \geq 0.05$; Accept $H_a$ when $p < 0.05$**
**Compare a calculated value to a critical values from a $\chi^2$ table.**

# Bonacorsi Consulting


Steven Bonacorsi, President

Lean Six Sigma Master Black Belt

14 Clinton Street

Salem NH 03079

sbonacorsi@comcast.net

603-401-7047