Hypothesis Testing

User Agreement and Copyright Information

air academy ASSOCIATES © 2020

Hypothesis Testing

- In this session, we will discuss Hypothesis Testing for both variables and attribute data.

- Hypothesis Testing Fundamentals

- Variables Data
    - Two Samples Test of Means – t Test
    - Power and Sample Size
    - Two Paired Samples Test of Means – Paired t Test
    - Multiple Samples Test of Means – Analysis of Variance (ANOVA)
    - Two Samples Test of Standard Deviations – F Test

- Attribute Data
    - Test of Proportions
   - Chi – Square Test for Independence

- Review

- A list of supplemental material and additional practice/review questions for this session are provided at the end of this presentation

- You can download the pdf of this presentation, along with any supporting data files, on the site where you are accessing this course

Hypothesis Testing

- Statistical hypothesis tests can be very valuable in both Lean Six Sigma (LSS - DMAIC) and Design for Six Sigma (DFSS – IDOV) methodologies.

- Hypothesis Tests can be used to:
  - Compare two (or more) sets of data
  - Identify the critical causes/variables (Analyze phase of LSS) (Design phase of DFSS)
  - Validate a significant improvement (Improve phase of LSS and Optimize phase of DFSS)

- Hypothesis testing starts with two hypotheses, the null hypothesis (H_0) and the alternative hypothesis (H_1). We assume the null hypothesis is true without any justification. We think the alternative hypothesis is now true, so we collect sample data to see if the alternative hypothesis is significantly supported!

- Example:
$$H_0: \text{Fair}$$
$$H_1: \text{Cheating}$$

Hypothesis Testing

- A method for looking at data and comparing results
  - Method 1 vs. Method 2
  - Option A vs. Option B
  - Before vs. After Project results

- Helps us make good decisions and not get fooled by random variation:
  - "Is a difference we see REAL, or is it just random variation and no real difference exists at all?"

- We set up 2 hypotheses
  - Example:

$$H_0 : \mu_1 = \mu_2 \quad \text{vs.} \quad H_1 : \mu_1 \neq \mu_2$$
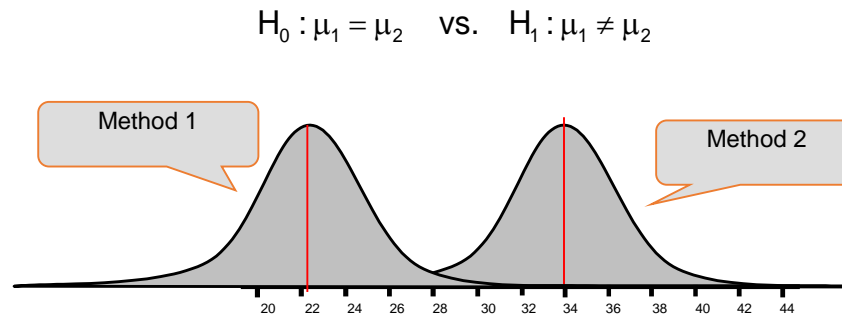
- Based on the sample data we collect to estimate the population, we must decide in favor of either H_0 or H_1. We assume the two population means are equal. Which hypothesis does the evidence support?

Nature of Hypothesis Testing

H_0:   **Defendant is Innocent (assumed to be true)**

H_1:   **Defendant is Guilty (trying to show)**

- Since verdicts are arrived at with less than 100% certainty, either conclusion has some probability of error. Consider the following table.

		True State of Nature	
		H_0	H_1
Conclusion Drawn	H_0	Conclusion is Correct	Conclusion results in a Type II error
	H_1	Conclusion results in a Type I error	Conclusion is Correct

- Type I or II Error Occurs if Conclusion Not Correct
  - The probability of committing a Type I error is defined as α $(0 \leq \alpha \leq 1)$ and is often called the false detection error. In this example, sending an innocent person to jail.
  - The probability of committing a Type II error is β $(0 \leq \beta \leq 1)$ and is often called missed detection error. Power is the complement of β, $(1 - \beta)$. In this example, letting a guilty person go.
  - The most critical decision error is usually a Type I error, but we should be concerned about the Type II error as well. Sample size controls both errors!

2-Sample Hypothesis Test Example

Sensor

Target Acquisition Time
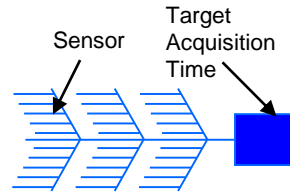
- In a target kill chain process, a team suspects that one of the important contributors to destroying the target is the amount of time it takes a sensor to acquire the target. The team further suspects that there might be a significant difference in the average amount of time it takes two different sensors to acquire the target. The team decides to conduct a test, to see if "sensor type" really is an important factor. In other words, is there a significant difference in the average target acquisition times between the first and second sensors?

- A random sample of 9 data points for each of the two sensors, shown below, was collected to help answer this question.

Sensor Target Time
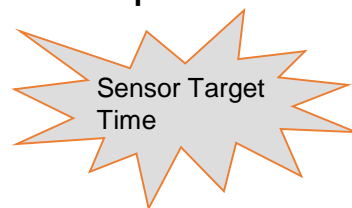
Hypothesis Testing Data Files

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

Target Acquisition Time (in seconds)											
Sensor	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	$\bar{y}$	s
1	2.8	3.6	6.1	4.2	5.2	4.0	6.3	5.5	4.5	4.6889	1.17
2	7.0	4.1	5.7	6.4	7.3	4.7	6.6	5.9	5.1	5.8667	1.08

2-Sample Hypothesis Test Example (cont.)

- The graphical interpretation of the hypotheses to be tested are:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$
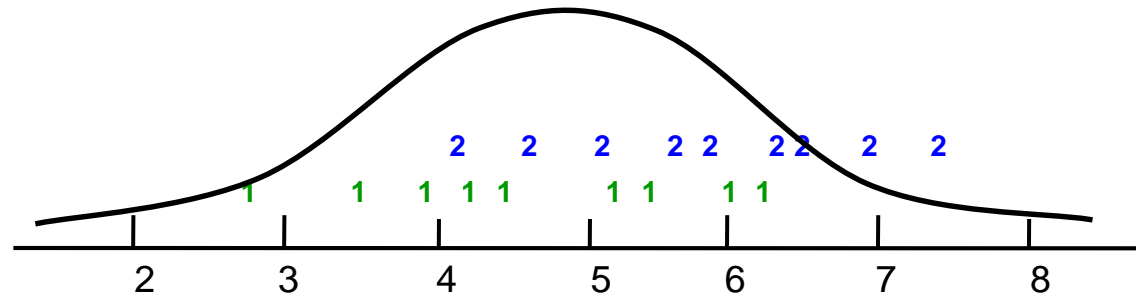
$H_0: \mu_1 = \mu_2$

versus

$H_1: \mu_1 \neq \mu_2$
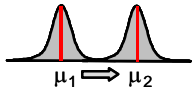
Testing for Differences in Averages (t-test)

- The statistical test for detecting a shift in average is called the t-test. The result of the test is a p-value, which indicates the probability of making a type I error. P-values are derived from the data.

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

- Rule of Thumb:
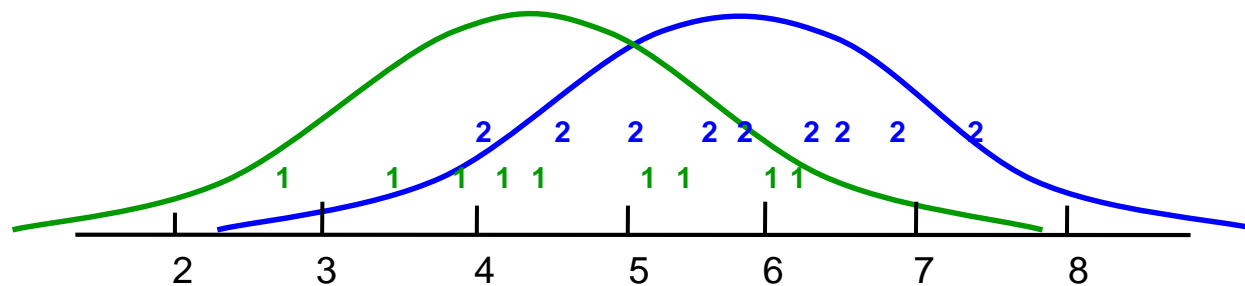
  - If p-value < 0.05 (red), highly significant difference in the averages (H_1).

  - If 0.05 < p-value < 0.10 (blue), moderately significant difference in the averages. Perhaps get more data!

  - If p-value > 0.10 (black), no significant difference in the averages (H_0).
  - (1 – p-value) • 100% is our percent confidence that there is a significant difference in the averages (H_1).

- For video instruction on hypothesis testing, p-value, and conducting a t test in SPC XL, go to: https://airacad.com/our-insights/training-videos/spc-xl/

Testing for Differences in Averages (SPC XL)

SPC XL > Analysis Tools > t Test matrix (Mean)
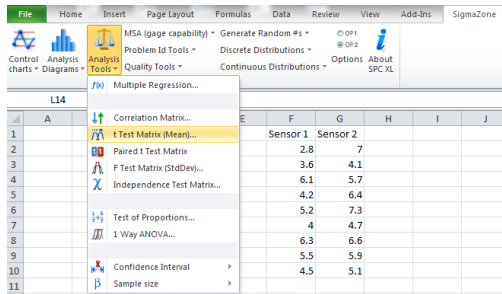
(specify data orientation)
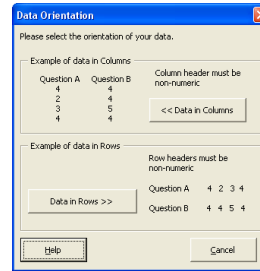
(specify one-sided or two-sided test)
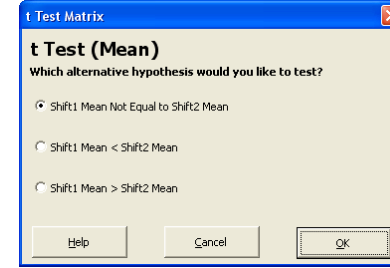
Hypothesis Testing Data Files

t-Test Result

| Hypothesis Tested: | H0: Sensor 1 Mean = Sensor 2 Mean |
| | H1: Sensor 1 Mean not equal to Sensor 2 Mean |

p-value (probability of Type I Error)	0.041
Confidence that Sensor 1 Mean not equal to Sensor 2 Mean	95.9%

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Summary Statistics

	Sensor 1	Sensor 2
Mean	4.6889	5.8667
StDev	1.1731	1.0759
Count	9	9

SPC XL results

- Since the p-value = 0.041 (and red), we can be at least (1 – p-value) • 100% confident—in this case more than 95% confident—that the two population means are different. That is, we are more than 95% confident that H_1 is the correct conclusion. This is said to be a statistically significant result. But this says nothing about how different μ_1 and μ_2 are. Enter Power!!

Managing the α and β Risks

- α risk = P(false detection) = probability of falsely concluding that a factor is important; i.e., we conclude H_1 but H_0 is true (a type I error).
  - P(false detection) = p-value (software calculates this value from the test data).
  - (1 - p-value) x 100% provides the percent confidence in the H_1 conclusion.
  - Rule of Thumb (ROT): p-value < 0.05 (confidence > 95%).

- β risk = P(missed detection) = probability we fail to detect that a factor is important; i.e., H_1 is true but we conclude H_0 (a type II error). *
  - Power = [1 – P(missed detection)] x 100%.
  - Power calculations are good for planning purposes, but we need to know the effect size when calculating power, i.e., how small a change (delta) do we want to be able to detect.
  - Rule of Thumb (ROT): Power ≥ 75%

- Sample size is the means for controlling both types of risk.
  - Rule of Thumb (ROT): p-value < 0.05 (confidence > 95%) and Power ≥ 75%

* We don't really "conclude H_0". Rather, we "fail to reject H_0" because there is not enough evidence to conclude H_1 with high confidence.

Determining Sample Size (for a given α, β, Δ, and σ)

- In the previous example, the team took 9 data points from each sensor, without considering power, and found a statistically significant result. Suppose that, before the test, the team wanted to be able to detect a change in mean as small as 1 second between the two sensors' average target acquisition times. What sample size should they have taken?

- Suppose $\alpha = 0.05$ (the most commonly used value).

- Suppose the team wants to have 80% power in the test. That is, $1 - \beta = 0.80$, the probability of detecting a change in means between the two sensors as small as $\Delta = 1$ second.

- We will need an estimate of σ. Suppose from historical data that the team estimates the standard deviation as 1.1 seconds for both sensors.

SPC XL > Analysis Tools > Sample Size > Hyp Test 2 Means (2 side)

Sample Size Calculation for Dual Sided 2 Sample Test (Mean)	
User defined parameters	
Estimated Standard Deviation (population 1)	1.1
Estimated Standard Deviation (population 2)	1.1
Size of Difference in Means	1
Desired Power of Test	80.00%
Desired Confidence Level	95.00%
Results	
Sample Size	19

beta risk = 20.00%

alpha risk = 5.00%

SPC XL is Copyright (C) 1999-2008 SigmaZone.com and Air Academy Associates, LLC. All Rights Reserved. Unauthorized duplication prohibited by law.

Determining Sample Size (Practice)

- Suppose we want to test two statapult ball materials (rubber, nerf) to see if material type has a significant effect on the average launch distance.

- The estimated launch distance standard deviation is 1 inch.

- The desired confidence level is 95% (i.e., alpha = 0.05) and the team wants to be able to detect a change in the means between the two materials as small as 1.5 inches (change or delta) with 95% power.

- What sample size is required?

- How would this sample size change if the estimated launch distance standard deviation is 1.5 inches?

Hypothesis Test for the Means (Paired t Test)

- Standard two-sample tests assume that the samples are independent. Sometimes, measurements are made on the same element of the population

- Examples:
  - Pre-test score and post-test score for the same individual
  - Weight of a sample before bake and after bake
  - Others?

- Why not just do a regular two-sample hypothesis test?

- Many software packages (including SPC XL) have a paired t test in their analysis tools sections. The output of the paired t test is a p-value just like the independent t test! The same rules for the p-value still apply for the paired t test.

$$H_0 : \quad \Delta = 0$$
$$H_1 : \quad \Delta \neq 0$$

- Rule of Thumb:
  - If p-value < 0.05 (red), highly significant mean difference (non-zero) in the before and after averages (H_1).
  - If 0.05 < p-value < 0.10 (blue), moderately significant mean difference (non-zero) in the before and after averages. Perhaps get more data!
  - If p-value > 0.10 (black), no significant mean difference (zero) in the before and after averages (H_0).
  - (1 – p-value) • 100% is our percent confidence that there is a significant mean difference in the before and after averages (H_1).

Paired t Test Example

- Nine foreign language teachers participated in a short course over the summer to improve their Spanish skills. At the beginning of the period, the teachers took the Modern Language Association's listening test of understanding of spoken Spanish. After four weeks of immersion in Spanish in and out of the classroom, they took the listening test again. Their pre and post test scores are shown below. (the maximum possible score is 36)

- Does the data support a significant differences in scores?

Hypothesis Testing Data Files

Scores –
Paired t test
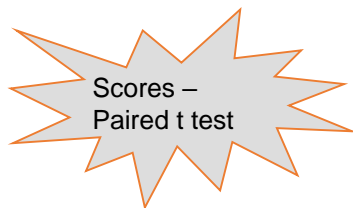
Score Pre-Test	Score Post Test	Gain
32	34	2
31	31	0
29	33	4
10	13	3
30	36	6
33	35	2
22	28	6
32	30	-2
20	26	6

Paired t Test Example SPC XL

- Perform the appropriate hypothesis test on the pre and post test data given on the previous page. What is your conclusion about the difference in test scores?

Paired t-Test Result		
Hypothesis Tested:	H0: Mean of Paired Differences (Pre-test - Post-test) = Zero	
	H1: Mean of Paired Differences (Pre-test - Post-test) not equal to Zero	
p-value (probability of Type I Error)		**0.013**
Confidence that Mean of Paired Differences (Pre-test - Post-test) not equal to Zero		**98.7%**

Summary Statistics			
	Pre-test	**Post-test**	**Paired Differences**
Mean	26.556	29.556	-3.0
StDev	7.7154	7.0198	2.8284
Count	9	9	9

- If a regular 2 sample hypothesis test had inadvertently been performed, what conclusion would be drawn?

| t-Test Result	
Hypothesis Tested:	H0: Pre-test Mean = Post-test Mean
	H1: Pre-test Mean not equal to Post-test Mean
p-value (probability of Type I Error)	**0.401**
Confidence that Pre-test Mean not equal to Post-test Mean	**59.9%**

Summary Statistics		
	Pre-test	**Post-test**
Mean	26.556	29.556
StDev	7.7154	7.0198
Count	9	9

Comparing More than 2 Sample Averages
Analysis of Variance (ANOVA))

- Suppose we have 5 different processing methods each with their own average to compare to each other: Method 1 (μ_1), Method 2 (μ_2), Method 3 (μ_3), Method 4 (μ_4), and Method 5 (μ_5)

- How many two sample t-tests would need to be performed?

- Can you think of any problems with doing that many tests?

- Analysis of Variance test (ANOVA) extends the t-test to several groups for differences in means and gives a statistical significance to the differences.

- Hypothesis:
  H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = ... = \mu_g$ g = number of groups
  H_1: means are not all equal

- Rule of Thumb: The same rules for the p-value still apply for ANOVA test.
  - If p-value < 0.05 (red), highly significant probability that at least one pair averages is different (H_1).
  - If 0.05 < p-value < 0.10 (blue), moderately significant probability that at least one pair averages is different. Perhaps get more data!
  - If p-value > 0.10 (black), no significant difference! All the groups averages are similar (H_0).
  - (1 − p-value) • 100% is our percent confidence that there is at least one pair averages is different (H_1).

Analysis of Variance (ANOVA) Example

Materials - ANOVA
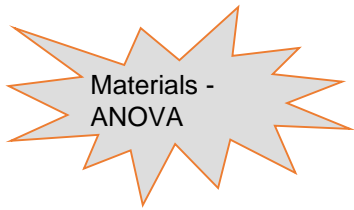
USB ***Hypothesis Testing Data Files***

- Example: Five materials were compared for their ability to improve the tensile strength of a product. Four samples from each of the five materials were tested and the tensile strengths are shown below. The product engineer is interested in whether the average strength is the same for all of the materials, or whether there is at least one pair that is different.

- Using SPC XL, conduct an ANOVA test to see if there is a statistically significant difference in the means among the five materials! The next page has the results!

material 1	material 2	material 3	material 4	material 5
28	32	33	41	41
36	21	33	33	38
37	23	40	31	35
39	28	30	39	34

One Way ANOVA Using SPC XL
(Tensile Strength Data)

SPC XL > Analysis Tools > 1 Way ANOVA

ANOVA Analysis

Source of Variation	SS	df	MS	F	P-value
Between Groups	308.8	4	77.2	3.9122	0.0227
Within Groups	296	15	19.73333		
Total	604.8	19			

Group Summary

Group	Count	Average	Std Dev
material 1	4	35	4.830459
material 2	4	26	4.966555
material 3	4	34	4.242641
material 4	4	36	4.760952
material 5	4	37	3.162278

Testing for Differences in Standard Deviations (F-test)
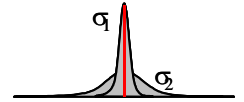
- The statistical test to detect differences in standard deviations is the F-test. The result of the test is a p-value, which indicates the probability of making a type I error. P-values are derived from the data.
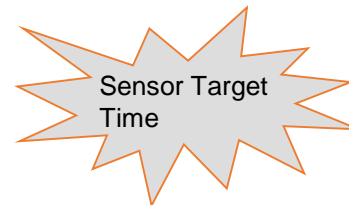
- Rule of Thumb:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

  - If p-value < 0.05 (red), highly significant difference in the standard deviations (H_1).

  - If 0.05 < p-value < 0.10 (blue), moderately significant difference in the standard deviations. Perhaps get more data!

  - If p-value > 0.10 (black), no significant difference in the standard deviations (H_0).
  - (1 – p-value) • 100% is our percent confidence that there is a significant difference in the standard deviations (H_1).

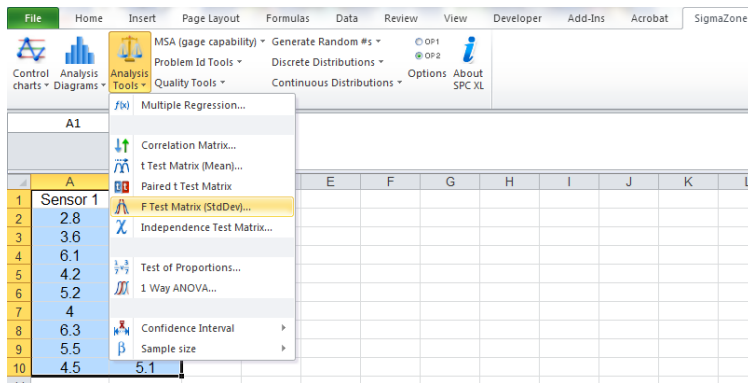- For video instruction on conducting an F test in SPC XL, go to:

https://airacad.com/our-insights/training-videos/spc-xl/

Hypothesis Testing Data Files

Sensor Target Time
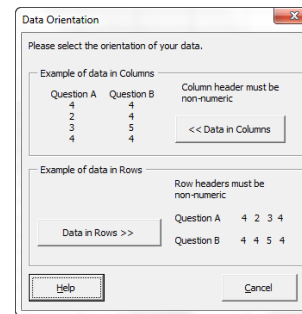
air academy
ASSOCIATES © 2020

18

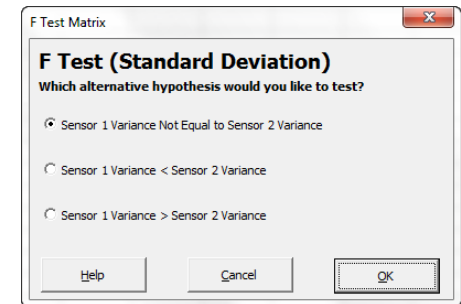# Testing for Differences in Standard Deviations (SPC XL)

SPC XL > Analysis Tools > F-Test matrix (StdDev)

(specify data orientation)

(specify one-sided or two-sided test)

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

F-Test Result

Hypothesis Tested:	H0: Sensor 1 Variance = Sensor 2 Variance
	H1: Sensor 1 Variance not equal to Sensor 2 Variance

p-value (probability of Type I Error)	0.813
Confidence that Sensor 1 Variance not equal to Sensor 2 Variance	18.7%

Summary Statistics

	Sensor 1	Sensor 2
Mean	4.6889	5.8667
StDev	1.1731	1.0759
Count	9	9

SPC XL results

- Since the p-value = 0.813 (and black!), we can be at least (1 – p-value) • 100% confident—in this case 18.7% confident—that the two population standard deviations are different. This is not very strong evidence to support H_1. This is a statistically insignificant result. Our conclusion would be to stay with H_0. The data has failed to reject the null hypothesis. We assumed H_0 was true to start with and the weak evidence requires us to stay with that assumption!

Hypothesis Test Exercise

Hypothesis Testing Data Files

- The data below represents the drying time (in seconds) from samples of two different paints (L, H) that were tested.

Paint	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}	Y_{13}	Y_{14}	$\overline{Y}$	S
L	201	209	215	221	211	213	217	205	218	208	203	214	212	215	211.6	5.8
H	218	225	217	222	223	220	222	216	221	224	224	221	220	219	220.9	2.7

- Use the appropriate hypothesis tests to determine:

1. Is there a significant difference in the average drying time between paint L and paint H? Why or why not?

2. Is there a significant difference in the drying time standard deviation between paint L and paint H? Why or why not?

3. The box plot is a great graphical tool to show, visually, the location and spread differences for two groups. It is a great visual for the group's location and spread differences but it has no statistical significance associated with it! For video instruction on generating a box plot in SPC XL, go to:

https://airacad.com/our-insights/training-videos/spc-xl/

Testing for Differences in Two Proportions (Example)

- A company wants to compare the leakage rates of one of its products produced on two different shifts. Data from the past month is gathered. For shift #1, 11 out of the 65 they produced failed the leak test.   For shift #2, 5 out of 60 failed the leak test. Are the proportion of defective products (i.e., items failing the leak test) being produced by the two shifts significantly different?

Shift #1:

Proportion of defective products =   _____

Shift #2:

Proportion of defective products =   _____

Testing for Differences in Proportions

- This statistical test to detect differences in proportions is called the "Test of proportions", or "Z" test. The result of the test is a p-value, which indicates the probability of making a type I error. P-values are derived from the data.

$$H_0 : \pi_1 = \pi_2$$

- Rule of Thumb:

$$H_1 : \pi_1 \neq \pi_2$$

  - If p-value < 0.05 (red), highly significant difference in the proportions (H_1).

  - If 0.05 < p-value < 0.10 (blue), moderately significant difference in the proportions. Perhaps get more data!

  - If p-value > 0.10 (black), no significant difference in the proportions (H_0).
  - (1 – p-value) • 100% is our percent confidence that there is a significant difference in the proportions (H_1).

- For video instruction on conducting a test of proportions in SPC XL, go to:

https://airacad.com/our-insights/training-videos/spc-xl/

Testing for Differences in Proportions (SPC XL)

SPC XL > Analysis Tools > Test of Proportions

Hypothesis Testing Data Files

$$H_0 : \pi_1 = \pi_2$$

$$H_1 : \pi_1 \neq \pi_2$$

- This test is different in the software than the t and F tests. The software develops a template and the user needs to input the data. The end result is still a p-value. Since the p-value = 0.151 (and black!), we can be at least (1 – p-value) • 100% confident—in this case 84.9% confident—that the two population means are different. This is not very strong evidence to support H_1. This is a statistically insignificant result. Our conclusion would be to stay with H_0. The data has failed to reject the null hypothesis. In this situation we might consider getting more data! What if we wait one more month and the leakage rate stays the same (22 out of 130 and 10 out of 120). What would your conclusion be now?

Hypothesis Testing with Classification Data

- Suppose we can classify data according to two different criteria

- A count is recorded in each <u>cell</u> of the table

	Product A	Product B
East Coast		
Midwest		
West Coast		

	Defect Type A	Defect Type B	Defect Type C
Supplier 1			
Supplier 2			
Supplier 3			
Supplier 4			

- This is called a (r x c) contingency table with r rows and c columns.

- A Chi-square test for independence will help us answer the following question:
  - "Is there a significant relationship between the row variable and the column variable?"

H_0 : Row variable is independent of the column variable (no relationship)

H_1 : Row variable is dependent on the column variable (relationship!)

Chi – Square Test for Independence Example

- A hospital monitors patient satisfaction scores on a 3 point scale:

  (1) highly satisfied, (2) satisfied, and (3) dissatisfied. Data for the past quarter is summarized below, according to which floor the patient was on.

Floor \ Patient Sat Rating	Highly Satisfied	Satisfied	Dissatisfied	Total
4	125	43	12	180
5	136	55	8	199
6	186	75	20	281
7	78	33	10	121
Total	525	206	50	781

- Are "floor" and "patient satisfaction" independent, or is there some relationship between the patient satisfaction scores and floor?

Chi – Square Test for Independence Example (cont.)

- The chi square test statistic is calculated by comparing the observed and actual counts over the entire (r x c) table. A p-value is computed using the Chi-Square distribution.

H_0 : Row variable is independent of the column variable (no relationship)

H_1 : Row variable is dependent on the column variable (relationship!)

- Rule of Thumb:
  - If p-value < 0.05 (red), highly significant relationship between row and column variables (H_1)
  - If 0.05 < p-value < 0.10 (blue), moderately significant relationship. Perhaps get more data!
  - If (p-value) > 0.10 (black), there is no relationship (independent) (H_0)
  - (1 – p-value) • 100% is our percent confidence that there is a significant relationship between the row and column variables (H_1).

- For video instruction on conducting a test of independence in SPC XL, go to:

https://airacad.com/our-insights/training-videos/spc-xl/

Notes:
1. Another consideration might be to test if the proportion of "highly satisfied" is the same for all floors.
2. The Chi-Square test for independence assumes that the sample size is sufficiently large, such that
  - All expected cell frequencies are ≥ 2
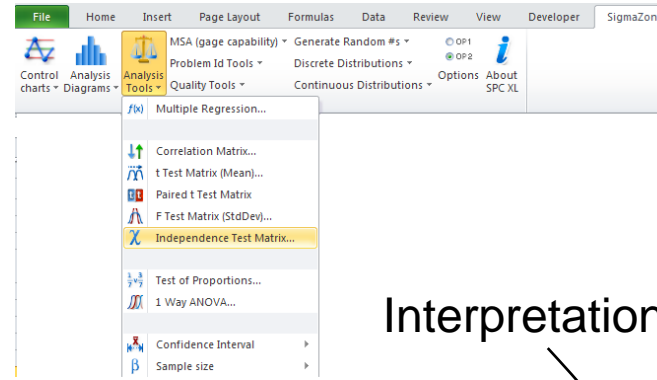 - At least half of the expected cell frequencies are ≥ 5

Chi – Square Test for Independence Example SPC XL

SPC XL > Analysis Tools > Independence Test Matrix

Hypothesis Testing Data Files

	Highly Satisfied	Satisfied	Dissatisfied
Floor 4	125	43	12
Floor 5	136	55	8
Floor 6	186	75	20
Floor 7	78	33	10

Patient Satisfaction

Interpretation?

Independence Test Matrix

		Highly Satisfied	Satisfied	Dissatisfied	Marginals
Floor 4	Actual	125	43	12	180
	Expected	121.0	47.5	11.5	
	Chi	0.1323	0.4223	0.0197	
Floor 5	Actual	136	55	8	199
	Expected	133.8	52.5	12.7	
	Chi	0.0371	0.1201	1.7636	
Floor 6	Actual	186	75	20	281
	Expected	188.9	74.1	18.0	
	Chi	0.0443	0.0105	0.2246	
Floor 7	Actual	78	33	10	121
	Expected	81.3	31.9	7.7	
	Chi	0.1370	0.0369	0.6556	
Marginals		525	206	50	781

p Value	.7301
Chi Statistic	3.6040
df	6

- For this p-value = 0.7301, we would conclude there is no relationship between patient satisfaction ratings and floor!

Hypotheses Test Summary

Question	Hypothesis Test
Are the means of two samples different?	
Are the means of two samples different where individual values in one sample are paired with individual values in the other sample?	
Are there any differences between the means of multiple groups?	
Are the standard deviations of two samples different?	
Are the proportions of two samples different?	
Is there a dependence between two ways to classify data (row versus column)?	

Hypotheses Test Summary

Question	Hypothesis Test
Are the means of two samples different?	t - test (independent)
Are the means of two samples different where individual values in one sample are paired with individual values in the other sample?	Paired t - test
Are there any differences between the means of multiple groups?	ANOVA (Analysis of Variance)
Are the standard deviations of two samples different?	F - test (StdDev)
Are the proportions of two samples different?	Test of Proportions
Is there a dependence between two ways to classify data (row versus column)?	Chi Square test of Independence

Key Takeaways

- As a review techniques, stop the video and summarize the key learnings from this session. When you are finished, continue to the next page.

Key Takeaways

- Hypothesis Testing Basics
  - The secret is the p - value

- Variables Data – use the Box Plot graph to show visually the location and spread
  - Two Samples Test of Means – t Test
  - Power and Sample Size – controls both false detection and missed detection errors
  - Two Paired Samples Test of Means – Paired t Test
  - Multiple Samples Test of Means – Analysis of Variance (ANOVA)
  - Two Samples Test of Standard Deviations – F Test

- Attribute Data
  - Test of Proportions
  - Chi – Square Test for Independence

- Rule of Thumb:
  - If p-value < 0.05 (red), highly significant difference in the averages (H_1).
  - If 0.05 < p-value < 0.10 (blue), moderately significant difference in the averages. Perhaps get more data!
  - If p-value > 0.10 (black), no significant difference in the averages (H_0).
  - $(1 - \text{p-value}) \cdot 100\%$ is our percent confidence that there is a significant difference in the averages (H_1).

Supplemental Material

- Suggested Reading:

  - ***Lean Six Sigma: A Tools Guide*** by Adams, Kiemele, Pollock and Quan (pp. 99 – 101, 129 - 131)

  - ***Basic Statistics – Tools for Continuous Improvement*** by Kiemele, Schmidt and Berdine, 4th edition (chapter 6)

  - ***Knowledge Based Management*** by Kiemele, Pollock and Murrow (pp. xx – xx)

  - ***Design for Six Sigma: The Tool Guide for Practitioners*** by Reagan and Kiemele (section 7.17)

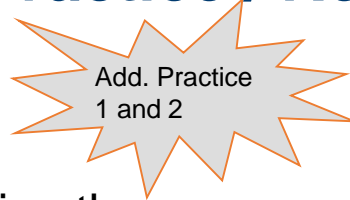 - Air Academy's app: ***Six Sigma Quick Tools***

- SPC XL™ software training tutorials:

  - https://airacad.com/our-insights/training-videos/spc-xl/

- The data files for this session can be downloaded from the site where you are accessing this course.

Additional Practice / Review Questions

Hypothesis Testing Data Files

1. You are interested in comparing the average wait time (in minutes) in an emergency department for two different hospitals. A random sample of wait times over the course of the past week is in the Add. Practice 1 worksheet of the data files. Perform the appropriate graphical and statistical hypothesis tests on this data!

2. A new gas additive was developed. The gas mileages for the same 10 cars before and after the additive are on the Add. Practice 2 worksheet of the data files. Perform the appropriate statistical hypothesis test on this data!

3. A medical facility was studying the effectiveness of adding a 1-800 help/support line for its smoking cessation program. Participants in the study were divided into two groups and followed for one year. One group was used a nicotine patch alone, and the other group used a nicotine patch in addition to being provided 24-hour access to a help/support line. Can we be 95% confident that providing the help/support line increases the chances of remaining smoke-free after 1 year? Why or why not? Results of the study are shown below:

Group 1: Nicotine patch

Sample Size: 1,020 participants

smoke-free after 1 year: 242

Group 2: Nicotine patch plus help/support line

Sample Size: 1,078 participants

smoke-free after 1 year: 298

We can help…
Connect With Us

[Remote Project Coaching](#)

There are times when help outside your organization is needed. When that time comes, benefit from a partner that is experienced, tested, and trusted.

Expert coaching is one of the Top Five Best Practices for generating step change in project execution, as well as enhanced return on investment. We can work remotely with your organization to provide coaching support.

Air Academy Associates
Phone: (719) 531-0777
Email: aaa@airacad.com
https://airacad.com/
https://sixsigmaproductsgroup.com/

There's an app for that!
Six Sigma Quick Tools

© 2020