LeanOhio Green Belt:
Transforming the Public Sector

Week Two

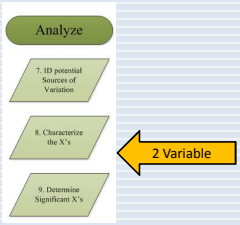
2-Variable Testing

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

DMAIC Flow

- **Analyze Phase:**
  - ID Potential Sources of Variation
  - Characterize the X's
  - Determine Significant X's

Purpose: Determine root causes, estimate population parameters with confidence intervals, and construct hypothesis about the data and test to determine significance.



SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

1 Variable is fine….but limited

- 1 variable is fine if you have a **standard**…but limited in use if you do not
  - How often do you have a known standard?
- Lets move on to 2-Variable testing!!!!
  - We can use observed performance as a standard
    - This can be very important for continuous improvement

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

2 Variable Testing

- Continuous Y data, and Discrete X data
- The X variable has two or more levels (settings), producing two or more sub-groups of Y data
- Assumptions:
  - Normal distribution
  - Stable process
  - Equal variance
  - No bias in the results
- **Continuous Y, Discrete X with 2 or more Levels**

The General Approach – Roadmap!

- Continuous Y, Discrete X with 2 or more Levels
- Check for Shape, Center, Spread, and Stability
  - Will you use a parametric or non-parametric test?
- Perform the appropriate test
  - If parametric testing was used, the next step is to check **residuals** to confirm the results of the test
  - If non-parametric testing was used, proceed with the results, but understand limitations and relative power

What Are Residuals Anyway???

- Hypothesis tests are based on mathematical models
- When we establish a null hypothesis we are identifying a model that is specific to our problem
- In a perfect world, these models would fit our data perfectly; however, in our real, imperfect world there will be some data that doesn't fit the model perfectly
- We can use this imperfect data (AKA **residuals**) to evaluate how well our data fits the assumptions in the model
  - Residuals are a double check on the validity of your test

What Are Residuals Anyway???

- Just like your original data, residuals must also follow these assumptions:
  - Normal distribution
  - Stable process
  - Equal variance
  - No bias in the results
- If all assumptions hold true, use the original test results
- If any of the above are shown to be false…
  - Fall back to conduct non-parametric test
  - proceed with the results, but understand limitations and relative power

Example #1

- Continuous Y, Discrete X with 2-levels

"Everyone knows that Group A is better than Group B…in fact, if only Group B could pick up the pace we would be fine!"

2 Variable Testing Scenario

Continuous Y, Discrete X with 2-Levels
Levels are: Group A and Group B.

- Group A complains that Group B is slower
- Are Group A and Group B statistically different?

- Let's start completing our Hypothesis Form

Hypothesis Form

Hypothesis Testing Form			
What is the Y?		What Type of Data?	
What is the X?		What Type of Data	
How many "levels" does X have?			
Is my data Stable?			
What type of tool would you use?			
Is my data Normal? (Outliers?)			
Comparing Median or Means?			
Ho: (=)			
Ha:			
P value: (0.05)			
Interpret results:			

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Hypothesis Form

Hypothesis Testing Form			
What is the Y?	Time (Seconds)	What Type of Data?	Continuous
What is the X?	Group A and B	What Type of Data	Discrete
How many "levels" does X have?		Two	
Is my data Stable?			
What type of tool would you use?			
Is my data Normal? (Outliers?)			
Comparing Median or Means?			
Ho: (=)			
Ha:			
P value: (0.05)			
Interpret results:			

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Collect the Data

- Remember the importance of Operational Definitions and ensuring reliable data
- If you are using Minitab: All of the variables need to be arranged in columns
  - A single column for the Y variables
  - One column for each of the X variables

* Remember, it might be valuable to collect other data points at this same time, but you have to think about the goal of the project and whether or not additional data collection will be worth the cost

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Let's compare Group A and Group B

(Just an example of what your data might look like)

Call Center Data					
Date	Day	Time	Shift	Region	Calls Answered
3/1/2013	Friday	9:00:00 AM	Shift A	North	6
3/1/2013	Friday	10:00:00 AM	Shift A	South	7
3/1/2013	Friday	11:00:00 AM	Shift A	East	16
3/1/2013	Friday	12:00:00 PM	Shift A	West	13
3/1/2013	Friday	1:00:00 PM	Shift B	North	10
3/1/2013	Friday	2:00:00 PM	Shift B	South	8
3/1/2013	Friday	3:00:00 PM	Shift B	East	9
3/1/2013	Friday	4:00:00 PM	Shift B	West	4
3/4/2013	Monday	9:00:00 AM	Shift A	North	12
3/4/2013	Monday	10:00:00 AM	Shift A	South	14
3/4/2013	Monday	11:00:00 AM	Shift A	East	20
3/4/2013	Monday	12:00:00 PM	Shift A	West	22
3/4/2013	Monday	1:00:00 PM	Shift B	North	23

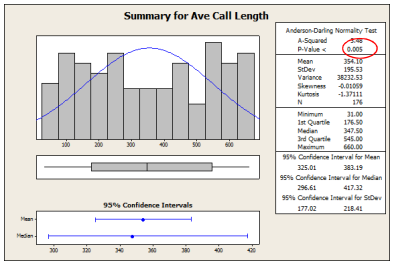
Hypothesis Test in a Hypothesis Test

- Check for Normality: Hypothesis test with-in our hypothesis test:
  - Null Hypothesis (Ho): Data = Normal
  - Alternate Hypothesis (Ha): Data ≠ Normal
- Conduct Anderson Darling Normality Test using Minitab: Stat – Basic Stat – Graphical Summary

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

2 sample t test: Check for Normality

- Test for normality: stat-basic stat- graphical summary:



Compare Groups

- Write our hypothesis test
  - Ho: Group A's mean = to Group B's mean
  - Ha: Group A's mean ≠ to Group B's mean
- Run test
  - If normal data, run 2 sample t-test (means)
  - If non-normal data, run the non-parametric Mann-Whitney Test (medians)
- Analyze results
  - What else might you need to do?

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

2 Sample t test

Group A vs. Group B

Two-Sample T-Test and CI: Ave Call Length, Group

Two-sample T for Ave Call Length

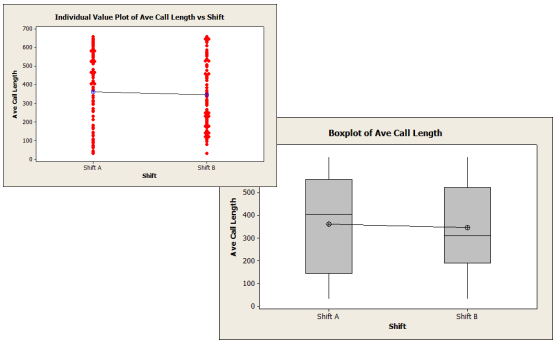
Group	N	Mean	StDev	SE Mean
Group A	88	362	210	22
Group B	88	346	180	19

- Difference = mu (Group A) - mu (Group B)
- Estimate for difference: 16.0
- 95% CI for difference: (-42.3, 74.3)
- T-Test of difference = 0 (vs not =): T-Value = 0.54 P-Value = 0.589 DF = 174
- Both use Pooled StDev = 195.9278

The p is not low so the Ho does not go: The p is 0.598 so the ho is NOT rejected the means the are NOT statistically different.

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Graphical Summary: 2-Sample t test Group A vs. Group B



Think back to cup Stacking Data

- What Hypotheses could we have tested using two-variable?

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

What if X is more than 2 levels?

- If you have more than two levels use the **ANOVA** test.
- **ANOVA: Continuous Y, and Discrete X but >2 levels**
- We were originally concerned with Groups A and B… However, we can extend to the other questions
  - What are some other hypothesis you want to test?
  - BRAINSTORM

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

ANOVA - Statistical Assumptions

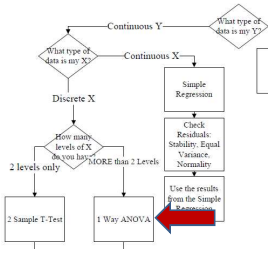
- Just as with the t-tests, ANOVA is built upon some foundation assumptions that must be verified:
  - Normal distribution
  - Stable process
  - Equal variance
  - No bias in the results
- There is a great deal of similarity between the ANOVA methods and the t-test methods
  - In fact, the t-test is really a special case of an ANOVA
  - For the equal variance case, the t-test and the ANOVA will produce exactly the same results

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

ANOVA

- Let's begin with Region.
- Are the regions behaving in a statistically different manner?

ANOVA

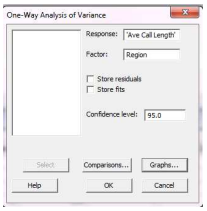


- In the case of a continuous Y, with one discrete X with 3 or more levels, we should be thinking about One-Way ANOVA

Don't Forget The Roadmap

- Check Your Assumptions
  - For this example, let's assume we passed them all
- Develop your hypothesis test
- Ho: North's Mean = East's Mean = West's Mean = South's Mean
- Ha: Mean's ≠
  - At least one mean is statistically different

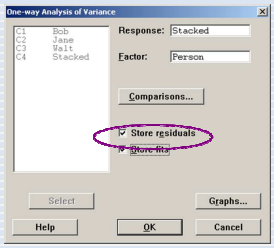
ANOVA - Minitab



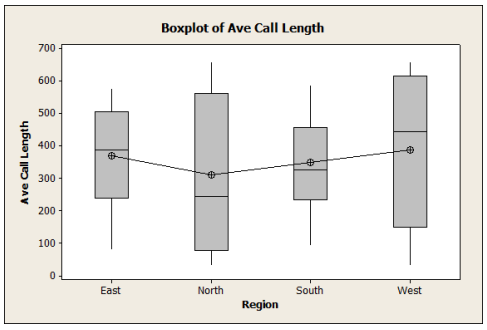

```
Welcome to Minitab, press F1 for help.

One-way ANOVA: Ave Call Length versus Region

Source   DF       SS     MS     F      P
Region    3   138073  46024  1.21  0.308
Error   172  6552620  38097
Total   175  6690692

S = 195.2   R-Sq = 2.06%   R-Sq(adj) = 0.36%

                          Individual 95% CIs For Mean Based on
                          Pooled StDev
Level    N   Mean  StDev  ---------+---------+---------+---------+
East    44  369.2  143.0                (-----------*-----------)
North   44  311.4  246.7  (-----------*-----------)
South   44  349.1  136.0          (-----------*-----------)
West    44  386.7  229.3                   (-----------*-----------)
                          ---------+---------+---------+---------+
                                  300       350       400       450

Pooled StDev = 195.2
```

Minitab Tip

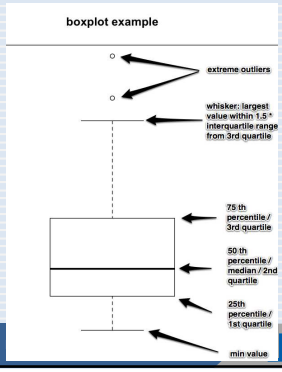
- When running an ANOVA, Minitab will automatically calculate and store residuals
  - The ANOVA is essentially the same as a 2 Sample T-Test and will work for an X with 2 levels
  - You can save some time in calculating residuals by just using the ANOVA instead of the 2 Sample –Test



Graphs for ANOVA



Quick Boxplot Refresher



SIMPLER. FASTER.

LEANOhio

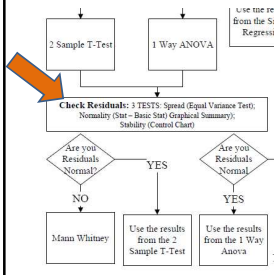
What did the ANOVA Tell Us

- Remember our mantra – if the P is Low the Ho must Go!
  - Is the p value of 0.308 low (below 0.05)?
- The p value of 0.308 is not low which tells us there is NO statistical difference between the regions means. We can see this from the Box plot as well.
  - We're done right!
  - NOPE – we need to check the Residuals.

SIMPLER. FASTER. BETTER. LESS COSTLY.

LEANOhio

One Way ANOVA



- After running the ANOVA, we still have to confirm the assumptions.
- The good news is that this time Minitab will produce the residuals for us…
  - We must remember to set the "Store Residuals" switch if we want Minitab to produce the residuals for us

Residuals Analysis is the Same in ANOVA

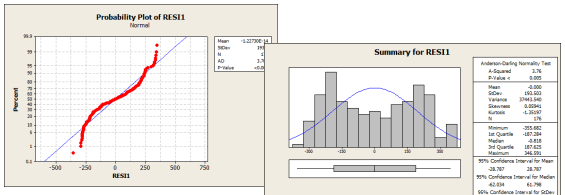
- Letting Minitab calculate the residuals saves some work
- Still need to verify the behavior of the residuals:
  - Normally Distributed: Stat > Basic Stat > Graphical Summary
  - Stable Process: Control Chart
  - Equal Variance: Test for Equal Variance

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Residuals

- We can check the Residuals for normality using the Normality Test or Graphical Summary.

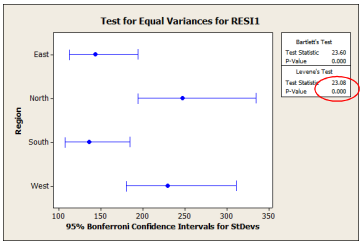
The p value is .005 (what does that tell us?)



Residuals: Test for Equal Variances

Just for fun… let's test for equal variance

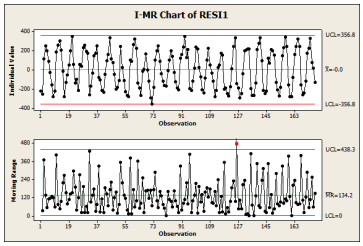
- Stat > ANOVA > Test for Equal Variances
  - What does the P Value tell us?
  - **The Test Fails!**



Failing Tests

- Once you fail any of the three tests – you need to use the non-parametric test. But lets check stability… just for fun!
- Also just for fun… can anyone name all 8 types of waste?
  - Bonus! Which is this an example of?

Using Control Charts to Check Stability: Using Residual DATA!



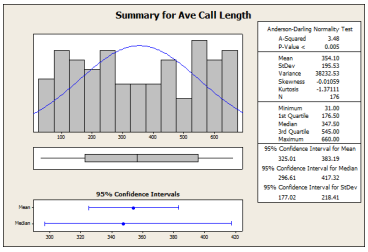
What decision do we make?

Non-parametric equivalent to ANOVA

- **Kruskal-Wallis** and **Mood's Median** tests
- The tests are essentially the same except that if the data has suspect outliers, the Mood's Median test is more robust against their influence
- If there are no outlier conditions of concern, then the Kruskal-Wallis is slightly more powerful than Mood's
- Again, neither nonparametric test is more powerful than the ANOVA

Original Data Set

- Are there any outliers?
- No - Kruskal Wallis
- Yes – Mood's Median



Kruskall-Wallis

Kruskal-Wallis Test: Ave Call Length versus Region

Kruskal-Wallis Test on Ave Call Length

Region	N	Median	Ave Rank	Z
East	44	386.0	91.9	0.51
North	44	244.5	76.9	-1.74
South	44	325.0	87.4	-0.17
West	44	443.0	97.8	1.40
Overall	176		88.5	

H = 3.96 DF = 3 P = 0.266
H = 3.96 DF = 3 P = 0.266 (adjusted for ties)

- Same answer! – no pretty graphs and not as much information.

Questions?

SIMPLER. FASTER. BETTER. LESS COSTLY.

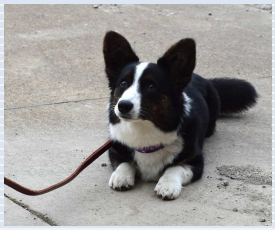
LEANOhio

Now it's your turn – Best Golfer?

- Get in groups:
  - Pick a Team Name
  - How Many Operators?
  - How Many Levels?
    - 2 Ball Types (Golf Ball and Wiffle Ball)
  - Mark the target at 12 feet from the line
  - 10 observations per operator, per ball type
  - ROTATE OPERATORS
  - Measure and record linear distance from center of ball to center of target
  - Measure to the nearest half inch

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Questions?



SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Recap: 2-Variable Testing

- 2 Sample T-Test / Mann-Whitney
  - Continuous Y, Discrete X with 2-levels
- ANOVA / Kruskal-Wallis / Mood's Median
  - Continuous Y, Discrete X with >2-levels

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Case Study: DOP A&A Section

- The Department of Prevention receives Sanctioned Legislative Obligatory Workforce (Slow) Forms – form all of the cabinet agencies. This form formally requests to conduct a project, major initiative or money saving event in their agency.

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Case Study: DOP A&A Section

- The processing of the Slow Form is conducted by the DOP Accuracy Analyzers (A&A) Section.
- The A&A section receives two types of Slow Forms: Renewal and Initials.
- The A&A section receives forms online and by paper.
- The A&A section is divided into four regions: North, South, East, and West.

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Example One: Renewal vs. Initial Apps

- Gus and Gracie Green Belt have had the first team meeting. During that team meeting the group have complained that the Initial applications are taking longer than the Renewal Applications.

	A	B	C	D	E	F	G	H
1	Month	Date	Type of Application	Shift	Region	County	Employee	Days to Process
2	January	Tuesday	Renewal	Shift A	East	Adams County	Murphy	25
3	January	Tuesday	Renewal	Shift A	East	Adams County	Murphy	44
4	January	Thursday	Renewal	Shift A	East	Defiance County	Emma	68
5	April	Wednesday	Initial	Shift A	East	Franklin County	Murphy	35
6	January	Tuesday	Renewal	Shift A	East	Franklin County	Annie	38
7	January	Thursday	Renewal	Shift A	East	Franklin County	Murphy	19
8	January	Thursday	Renewal	Shift A	East	Franklin County	Murphy	51
9	April	Thursday	Initial	Shift A	East	Franklin County	Murphy	36

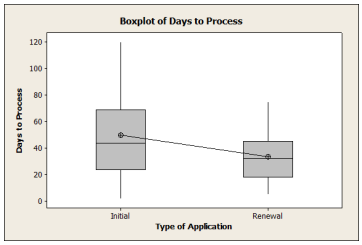
- Complete your Hypothesis Form for this scenario.

Example One: Renewal vs. Initial Apps

Hypothesis Testing Form			
What is the Y?		What Type of Data?	
What is the X?		What Type of Data	
How many "levels" does X have?			
Is my data Stable?			
What type of tool would you use?			
Is my data Normal? (Outliers?)			
Comparing Median or Means?			
Ho: (=)			
Ha:			
P value: (0.05)			
Interpret results:			

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Soooo is there a Statistical Difference?



2-Sample T-Test

P-Value = 0.000

```
Two-sample T for Days to Process

Type of
Application      N  Mean  StDev  SE Mean
Initial       9486  49.6   31.1     0.32
Renewal      37916  33.4   18.7    0.096


Difference = mu (Initial) - mu (Renewal)
Estimate for difference:  16.260
95% CI for difference:  (15.625, 16.934)
T-Test of difference = 0 (vs not =): T-Value = 48.75  P-Value = 0.000  DF =
    11259
```

Individual Value Plot of Days to Process vs Type of Application

Boxplot of Days to Process

Hypothesis Form

Hypothesis Testing Form			
What is the Y?	Days To Complete	What Type of Data?	Continuous
What is the X?	Application Type	What Type of Data	Discrete
How many "levels" does X have?		Two	
Is my data Stable?		Yes	
What type of tool would you use?		2 Sample T-Test	
Comparing Median or Means?		Means	
Is my data Normal (Residuals)?		YES	
Ho: (=)	Mean for Renewals Applications = Mean for Initial Applications.		
Ha:	Mean for Renewals Applications (not equal) Mean for Initial Applications.		
P value: (0.05)	0.000		
Interpret results:	The Means are different – there is a statistical difference between Initial and Renewal Applications		

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Same test if Residual Tests Fail!

P-Value = 0.000

Probability Plot of Days to Process

Mood Median Test: Days to Process versus Type of Application

```
Mood median test for Days to Process
Chi-Square = 910.81    DF = 1    P = 0.000

Type of                                    Individual 95.0% CIs
Application     N<=     N>  Median  Q3-Q1  ---+---------+---------+---------+---
Initial        3454   6032    44.0   45.0                              (--*-)
Renewal       20374  17542    32.0   27.0  (--*
                                          ---+---------+---------+---------+---
                                          32.0      36.0      40.0      44.0

Overall median = 33.0

A 95.0% CI for median(Initial) - median(Renewal): (11.0,14.0)
```

Example Two

- Gus and Gracie Green Belt have had the first team meeting. During that team meeting the group have maintained that the applications received through Online and Paper take roughly the same amount of time.

	A	B	C	D	E	F	G	H
1	Month	Date	Type of Application	Shift	Region	County	Employee	Days to Process
2	January	Tuesday	Renewal	Shift A	East	Adams County	Murphy	25
3	January	Tuesday	Renewal	Shift A	East	Adams County	Murphy	44
4	January	Thursday	Renewal	Shift A	East	Defiance County	Emma	68
5	April	Wednesday	Initial	Shift A	East	Franklin County	Murphy	35
6	January	Tuesday	Renewal	Shift A	East	Franklin County	Annie	38
7	January	Thursday	Renewal	Shift A	East	Franklin County	Murphy	19
8	January	Thursday	Renewal	Shift A	East	Franklin County	Murphy	51
9	April	Thursday	Initial	Shift A	East	Franklin County	Murphy	36

- Complete your Hypothesis Form for this scenario.

Hypothesis Form

Hypothesis Testing Form			
What is the Y?	Time	What Type of Data?	Continuous
What is the X?	App Entry (Paper or Online)	What Type of Data	Discrete
How many "levels" does X have?		Two	
Is my data Stable?		YES	
What type of tool would you use?		1 Sample t Test	
Comparing Median or Means?		Means	
Is my data Normal (Residuals)?			
Ho: (=)			
Ha:			
P value: (0.05)			
Interpret results:			

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Example Two

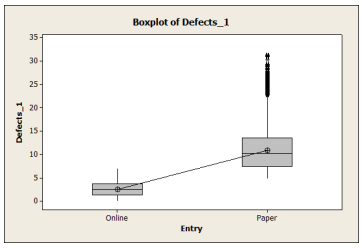
P-Value = 0.000

Two-Sample T-Test and CI: Defects_1, Entry

```
Two-sample T for Defects_1

Entry      N   Mean  StDev  SE Mean
Online  13309  2.45   1.37    0.012
Paper   34093 10.89   4.26    0.023

Difference = mu (Online) - mu (Paper)
Estimate for difference:  -8.4348
95% CI for difference:  (-8.4855, -8.3839)
T-Test of difference = 0 (vs not =): T-Value = -324.65  P-Value = 0.000  DF =
     46261
```

Individual Value Plot of Defects_1 vs Entry

Example Two



Hypothesis Form

Hypothesis Testing Form			
What is the Y?	Time	What Type of Data?	Continuous
What is the X?	App Entry (paper or Online)	What Type of Data	Discrete
How many "levels" does X have?		Two	
Is my data Stable?		YES	
What type of tool would you use?		1 Sample t Test	
Comparing Median or Means?		Means	
Is my data Normal (Residuals)?		YES	
Ho: (=)	Mean for defects on paper = Mean for defects online.		
Ha:	Mean for defects on paper (not equal) Mean for defects online.		
P value: (0.05)	0.000		
Interpret results:	The Means are different – there is a statistical difference between defects committed online and by paper		

SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Example Three: Regions

- Gus and Gracie Green Belt have had the first team meeting. During that team meeting the group have complained that the North Region doesn't pull it's weight.

	A	B	C	D	E	F	G	H
1	Month	Date	Type of Application	Shift	Region	County	Employee	Days to Process
2	January	Tuesday	Renewal	Shift A	East	Adams County	Murphy	25
3	January	Tuesday	Renewal	Shift A	East	Adams County	Murphy	44
4	January	Thursday	Renewal	Shift A	East	Defiance County	Emma	68
5	April	Wednesday	Initial	Shift A	East	Franklin County	Murphy	35
6	January	Tuesday	Renewal	Shift A	East	Franklin County	Annie	38
7	January	Thursday	Renewal	Shift A	East	Franklin County	Murphy	19
8	January	Thursday	Renewal	Shift A	East	Franklin County	Murphy	51
9	April	Thursday	Initial	Shift A	East	Franklin County	Murphy	36

- Complete your Hypothesis Form for this scenario.

Hypothesis Form

Hypothesis Testing Form			
What is the Y?	Time to complete App	What Type of Data?	Continuous
What is the X?	Region	What Type of Data	Discrete
How many "levels" does X have?		Four	
Is my data Stable?		YES	
What type of tool would you use?		Anova	
Comparing Median or Means?		Means	
Is my data Normal (Residuals)?		YES	
Ho: (=)	Mean for North = South = West = East		
Ha:	Mean for North (not equal) Mean for South – East – West (at least one is different.		
P value: (0.05)	0.201		
Interpret results:	The Means are not significantly different!		

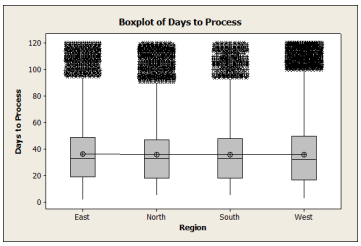
SIMPLER. FASTER. BETTER. LESS COSTLY. LEANOhio

Example Three

One-way ANOVA: Days to Process versus Region

```
Source      DF        SS     MS     F      P
Region       3      2445  815  1.54  0.201
Error    47398  25062457  529
Total    47401  25064903

S = 22.99   R-Sq = 0.01%   R-Sq(adj) = 0.00%

                                  Individual 95% CIs For Mean Based on
                                  Pooled StDev
Level      N   Mean  StDev  ---------+---------+---------+---------+
East   10042  36.34  22.61                (-----------*-----------)
North  12933  35.80  22.26  (-----------*----------)
South   8596  36.03  22.45       (-----------*-----------)
West   15831  35.77  24.09  (--------*---------)
                            ---------+---------+---------+---------+
                                35.70     36.05     36.40     36.75

Pooled StDev = 22.99
```

Regions



Questions?

Thank You!

Joe Rust
Deputy Director
Auditor of State, Ohio Performance team
JCRust@ohioauditor.gov
614-387-6291