# What's the Big Deal?

***Is Agile Software Development Really Different in the DoD Acquisition Environment?***


**Katelyn Barbre**

*Scientist / Cost Analyst*
*Naval Surface Warfare Center, Dahlgren Division*

**Contributing Members**

*Dr. Jonathan Brown*
*Katiana Lemmons*
*Markie Harris*


*14 May, 2019*

Date

**The Leader in Warfare Systems Development and Integration**

# Agenda

- Introduction
  - Purpose
  - Agile vs. Non-Agile Definition
  - Data Selection
  - Data Sources
  - Agile Project Descriptions
  - Non-Agile Project Descriptions
  - Project Data Summary
- Analysis Details
- Summary


Presented at the 2019 ICEAA Professional Development & Training Workshop - www.iceaaonline.com

NSWCDD/PN/19/094

# Purpose

- Explore cost, schedule, and performance metric differences between Department of Defense software acquisition programs using Agile development and those not using Agile development
- Determine if the two sample populations of data are different using nonparametric analysis
- Highlight takeaways and path forward


Presented at the 2019 ICEAA Professional Development & Training Workshop - www.iceaaonline.com

NSWCDD/PN/19/094

# Definition

| Agile | Non-Agile |
| --- | --- |
| Projects using any form of Agile development | Project using any form of development other than Agile |
| Examples:<br>Modified Agile, Scrum Agile | Examples:<br>Waterfall, Spiral |
| Development type collected from Software Resources Data Report (SRDR) | Development type collected from SRDR |

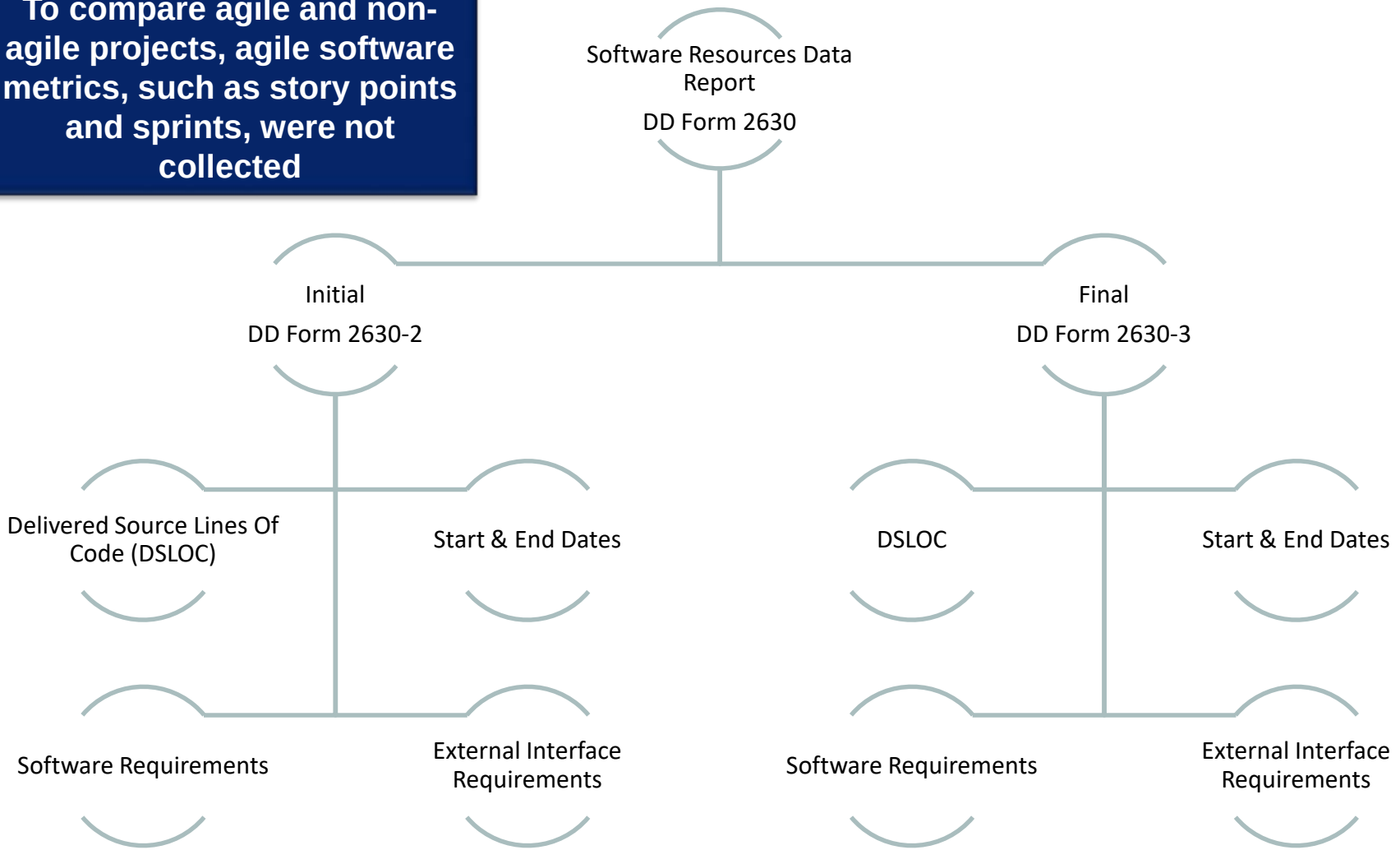
# Data Selection

- Agile data points were chosen using the SRDR Database maintained on the DACIMS site by NAVAIR 4.2[1]
  - Only data points with "Good" or "Good – Allocation" Verification and Validation (V&V) Quality Tags were used from the database
- Non-agile data points were selected using two methods:
  1. By finding analogous systems of the Agile data points and then verifying quality tags
  2. Some data points were randomly selected from the pool of "Good" data points
- An agile data point is not defined by a sprint timeframe but by the overarching initial and final reports

1. OSD CADE. (2018, 05). SRDR Data Compilation. *SRDR Data Compilation as of 20180516*. Retrieved from https://www.osd.cade.mil

# Data Sources

**To compare agile and non-agile projects, agile software metrics, such as story points and sprints, were not collected**



# Data Sources (Cont'd)

**Cost Summary Data Report**

**(CSDR)**

**DD Form 1921**

- Reports costs in contract work breakdown structure
- Initial and final reports
- Reports actual costs of work performed (ACWP) and estimates at completion (EAC)
- ACWP = EAC on the final report

**Functional Cost-Hour and Progress Curve Report**

**(FCHPCR)**

**DD Form 1921-1**

- Reports costs and hours at the level of indenture required in the contract
- Only design, code, test, and integration (DCTI) hours were collected
- Every CSDR submitted has a corresponding FCHPCR
- Initial and final reports


Presented at the 2019 ICEAA Professional Development & Training Workshop - www.iceaaonline.com

NSWCDD/PN/19/094

# Agile Project Descriptions

| Project/Marker | ACAT | ESLOC | Cost BY19$K | Duration (Months) |
|---|---|---|---|---|
| 1 (1) | I | 588,951 | $40,100.4 | 49.7 |
| 2 (2) | I | 321,999 | $6,819.0 | 4.0 |
| 3 (3) | I | 474,410 | $7,773.2 | 21.1 |
| 4 (4) | I | 841,098 | $6,028.5 | 13 |
| 5 (5) | I | 938,931 | $136,666.7 | 1.6 |
| 6 (6) | I | 186,690 | $38,764.9 | 1.4 |
| 7 (7) | I | 90,384 | $9,581.2 | 28.3 |
| 8 (8) | I | 87,866 | $3,262.0 | 24.3 |

Standard effort adjustment factors (EAFs) used to calculate equivalent source lines of code (ESLOC) from DSLOC for comparison

# Non-Agile Project Descriptions

| Project/Marker | ACAT | ESLOC | Cost BY19$K | Duration (Months) |
|---|---|---|---|---|
| A | I | 4,477 | $1,074.3 | 45.0 |
| B | I | 645,156 | $78,060.6 | 117.1 |
| C | I | 661,231 | $40,621.7 | 41.1 |
| D | I | 137,861 | $4,304.9 | 1.3 |
| E | I | 1,076,792 | $140,971.6 | 1.6 |
| F | I | 353,195 | $63,768.9 | 71.5 |
| G | I | 366,949 | $32,718.2 | 62.1 |
| H | I | 735,799 | $106,017.0 | 71.5 |
| I | I | 119,369 | $13,108.9 | 62.1 |
| J | I | 188,560 | $68,111.8 | 71.5 |
| K | I | 92,175 | $11,735.0 | 59.0 |

Standard EAFs used to calculate ESLOC from DSLOC for comparison

# Project Data Summary

| Metric | AGILE | | | | | | | | NON-AGILE | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | A | B | C | D | E | F | G | H | I | J | K |
| ESLOC Per Hour (Initial) | ** | X | X | X | ** | X | X | X | X | ** | X | X | ** | ** | X | ** | X | X | X |
| ESLOC Per Hour (Final) | ** | X | X | X | ** | X | X | X | X | ** | X | X | ** | ** | X | ** | X | X | X |
| Hours Per Requirement (Initial) | ** | X | X | X | ** | X | X | X | X | ** | X | X | ** | ** | X | ** | X | X | X |
| Hours Per Requirement (Final) | ** | X | X | X | ** | X | X | X | X | ** | X | X | ** | ** | X | ** | X | X | X |
| Cost Growth | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| Hours Growth | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| SW Growth | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |
| Requirements Growth | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X |

**Hours were not provided for subcontractor on 1921-1; generic labor rate used to calculate subcontractor hours

# Agenda

- Introduction
- Analysis Details
  - Statistics Definition
  - Hypothesis
  - Initial Predictions
  - Mann-Whitney U Test
  - Hypothesis Testing
  - Final Results Table
- Summary

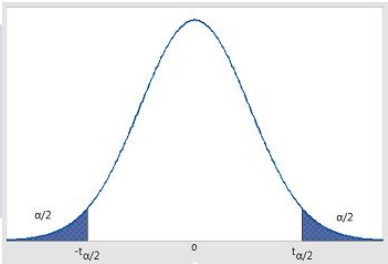

Presented at the 2019 ICEAA Professional Development & Training Workshop - www.iceaaonline.com

NSWCDD/PN/19/094

# Statistics Definitions

**Critical Value**

- Any value that separates the critical region, where we reject the null hypothesis, from the values of the test statistic that do not lead to rejection of the null hypothesis
- In the figure, $-t_{\alpha/2}$ and $t_{\alpha/2}$ are the critical values




**Test Statistic**

- A random variable calculated from sample data and used in hypothesis testing
- If the test statistics falls within the critical region, we reject the null hypothesis

**Significance Level, $\alpha$**

- The probability that the test statistic will fall in the critical region when the null hypothesis is actually true

**Fail to Reject the Null Hypothesis**

- When we fail to reject the null hypothesis, or $\boldsymbol{H_0}$, we conclude there is not sufficient evidence to support the alternate hypothesis, or $\boldsymbol{H_A}$

# Hypothesis

- Null Hypothesis ($H_0$)
  - There is no difference in the distribution of agile and non-agile cost, schedule, and performance metrics
- Alternate Hypothesis ($H_A$) :
  - There is a difference in the distribution of agile and non-agile cost, schedule, and performance metrics

# Initial Prediction

| Fail to Reject the Null Hypothesis | Reject the Null Hypothesis | Unsure |
| --- | --- | --- |
| Initial and Final ESLOC Per Hour | Schedule Slip | Software Growth |
| Initial and Final Hours Per Requirement | Hours Growth | Requirements Growth |
| Cost Growth | | |

# Mann-Whitney U Test[3]

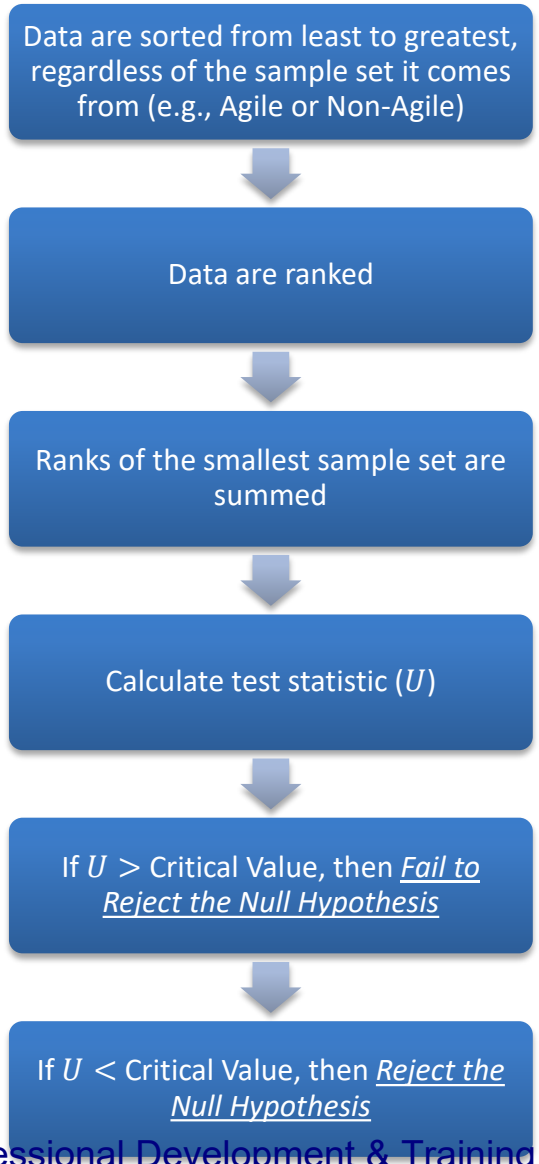
*Compares the distributions of two sample sets (center, shape, spread); does not compare a measure of central tendency (mean, median, mode)*

| ***Nonparametric Analysis Technique*** | ***Data does not need to be normally distributed*** . . . | ***Assumptions*** | ***U – Test Statistic*** | ***Significance Level*** $\alpha = 0.05$ |
|---|---|---|---|---|

- but data can be normally distributed
- Data are randomly selected
- Data are independent
- Ordinal measure scale

3. *7.3 Decision Making in Hypothesis Testing [PNG]. (n.d.). Onlinecourses.science.psu.edu.*

# Mann-Whitney U Test Steps[4]




4. Mann-Whitney U Test - Statistics Solutions. (n.d.). Retrieved from https://www.statisticssolutions.com/mann-whitney-u-test/

# Mann-Whitney U Test Steps[4]

**For Example**

Assume we have a test statistic $U$ where:

$$U = 10$$

and where:

$$Critical\ Value = 15.$$

Since $U$ falls within the critical region, or $U < Critical\ Value$, we reject the null hypothesis.

If $U >$ Critical Value, then *Fail to Reject the Null Hypothesis*

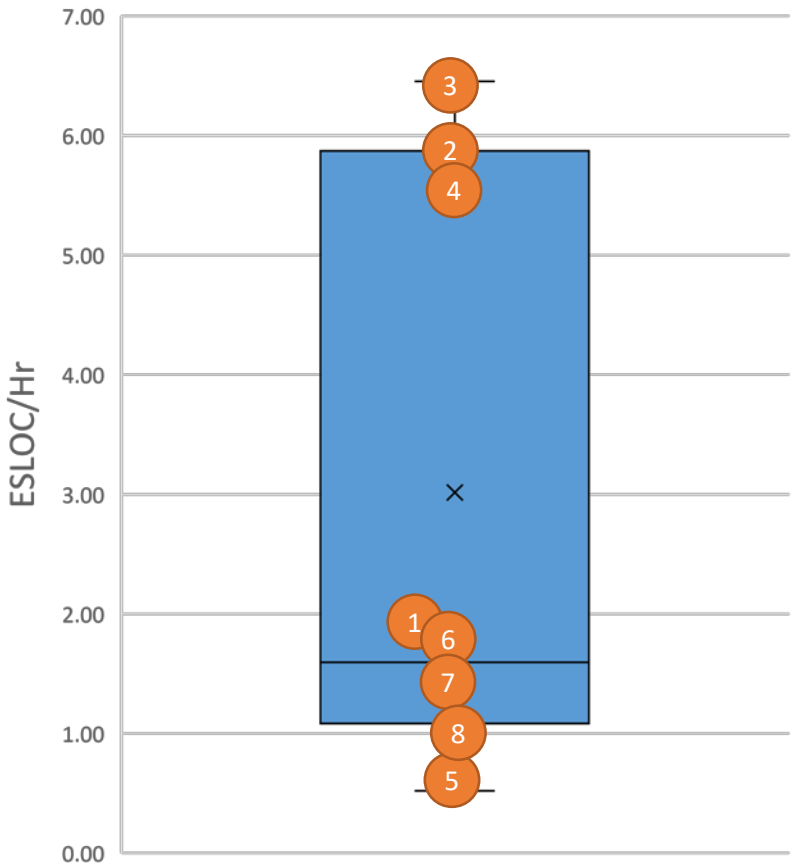
If $U <$ Critical Value, then *Reject the Null Hypothesis*

4. Mann-Whitney U Test - Statistics Solutions. (n.d.). Retrieved from https://www.statisticssolutions.com/mann-whitney-u-test/

# ESLOC per Hour—Initial Report

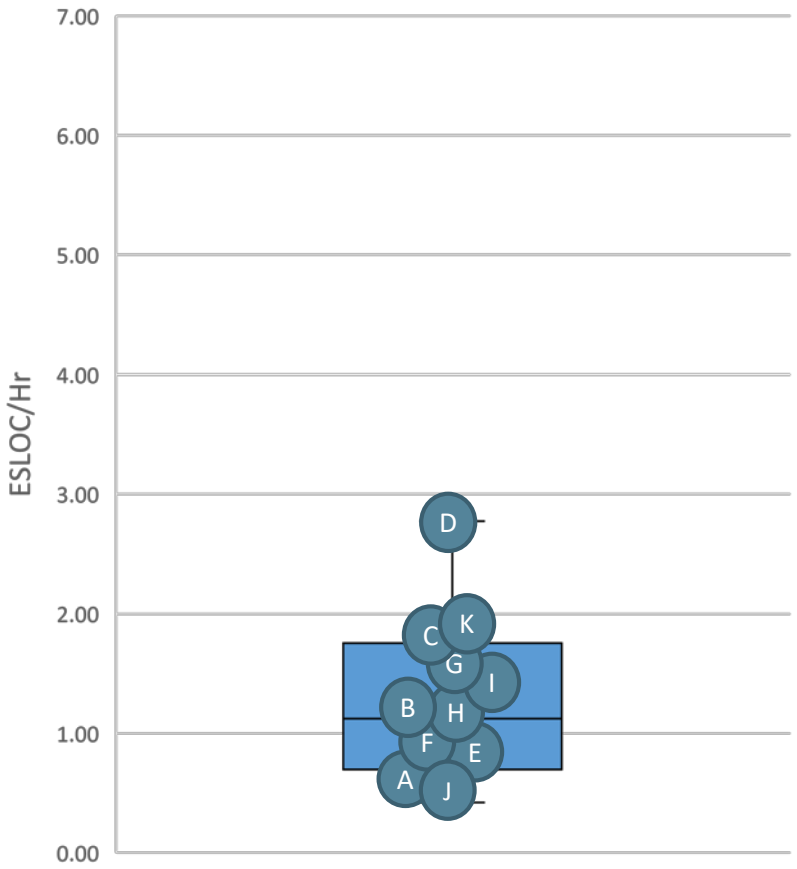
**Agile Projects**

$n_1 = 8$




**Non-Agile Projects**

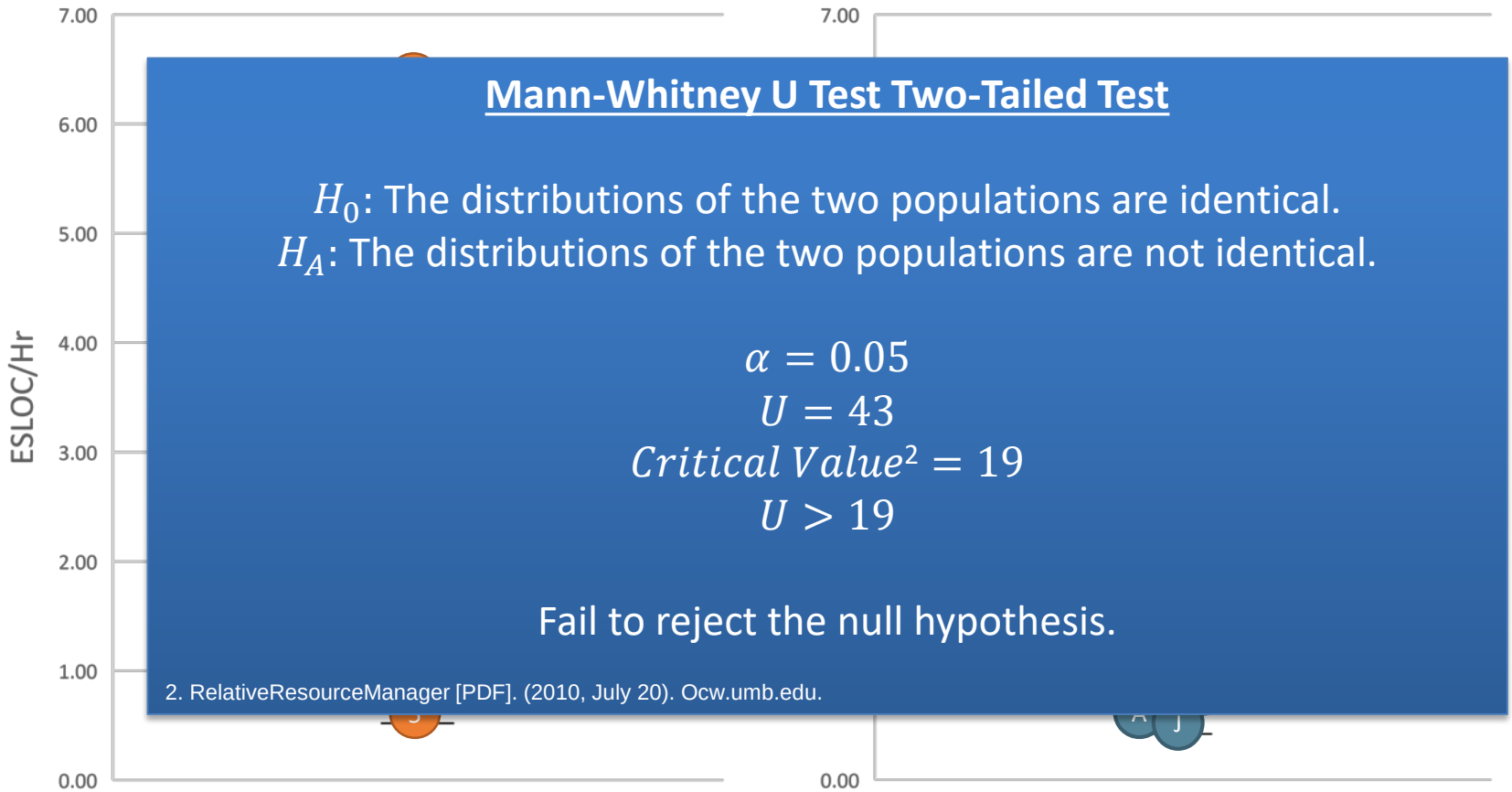
$n_2 = 11$



# ESLOC per Hour—Initial Report

**Agile Projects**
$n_1 = 8$

**Non-Agile Projects**
$n_2 = 11$




**Mann-Whitney U Test Two-Tailed Test**

$H_0$: The distributions of the two populations are identical.
$H_A$: The distributions of the two populations are not identical.

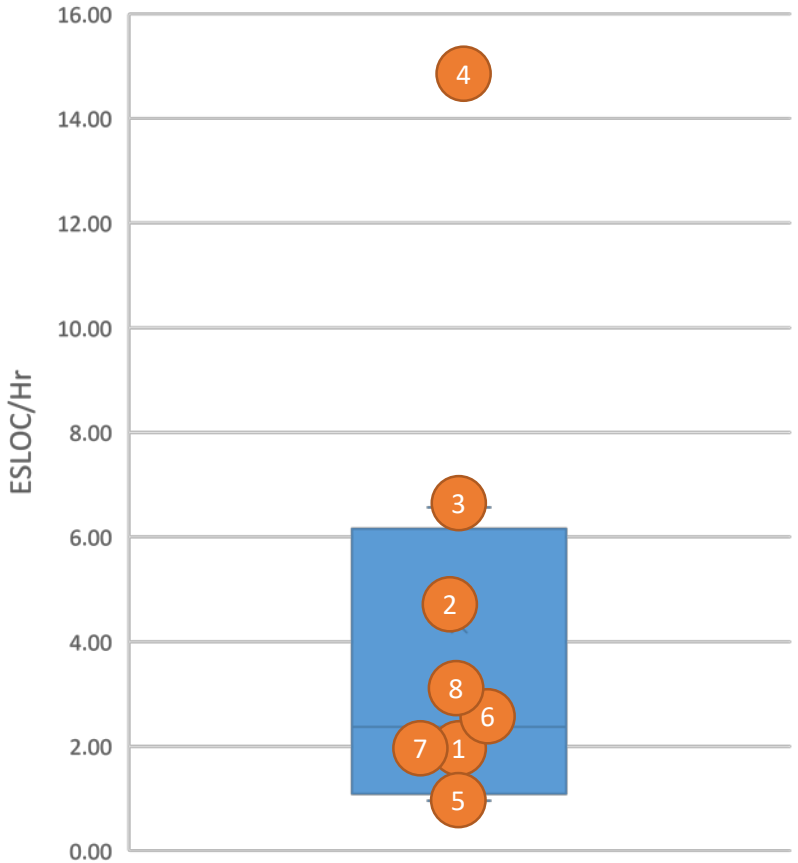
$\alpha = 0.05$
$U = 43$
$Critical\ Value^2 = 19$
$U > 19$

Fail to reject the null hypothesis.

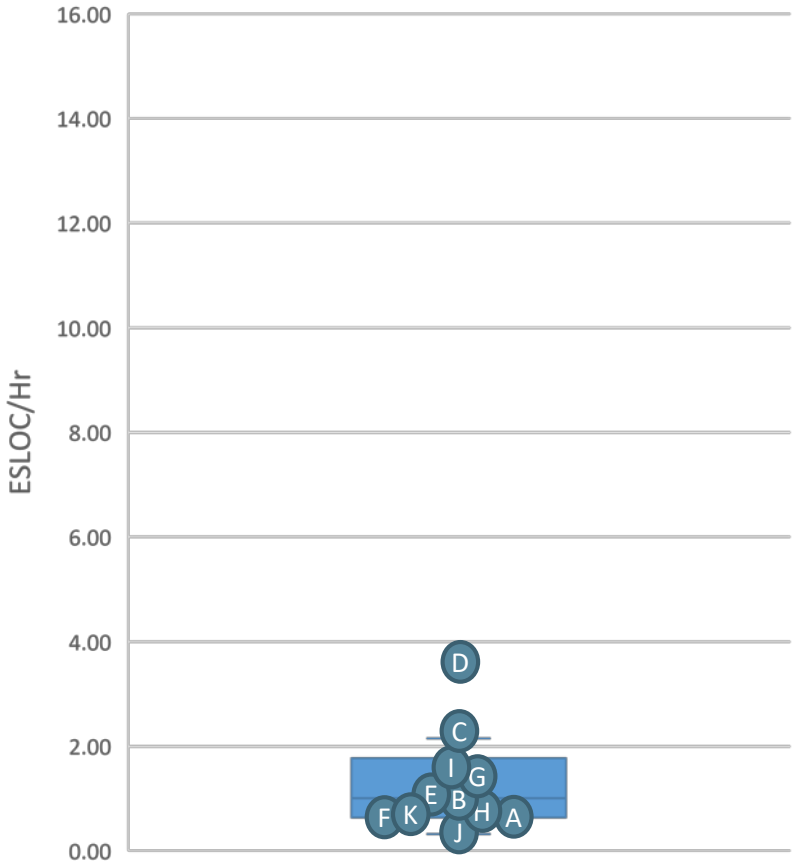
2. RelativeResourceManager [PDF]. (2010, July 20). Ocw.umb.edu.

# ESLOC per Hour—Final Report

**Agile Projects**
$n_1 = 8$



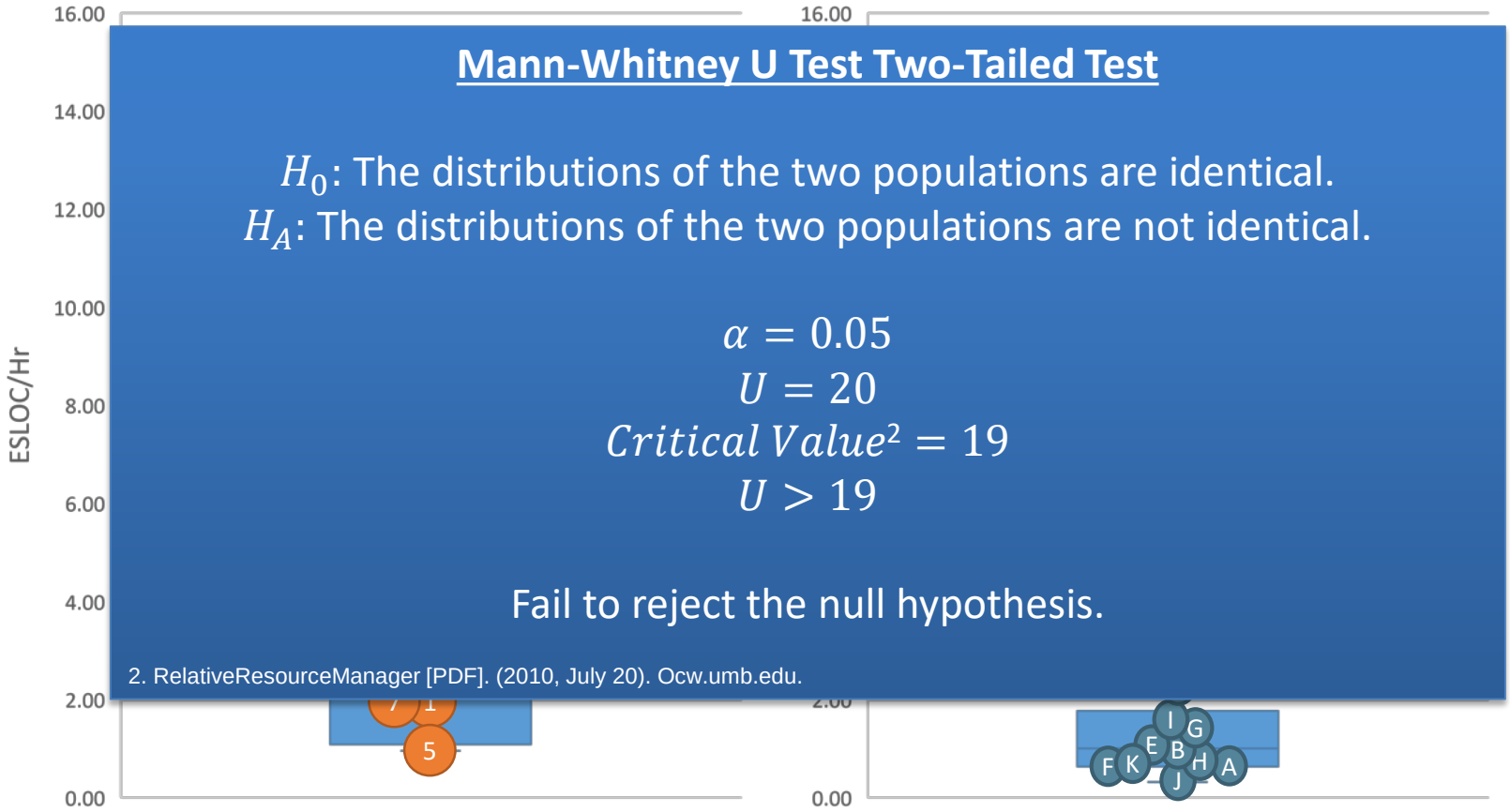

**Non-Agile Projects**
$n_2 = 11$



# ESLOC per Hour—Final Report

**Agile Projects**
$n_1 = 8$

**Non-Agile Projects**
$n_2 = 11$




**Mann-Whitney U Test Two-Tailed Test**

$H_0$: The distributions of the two populations are identical.
$H_A$: The distributions of the two populations are not identical.

$$\alpha = 0.05$$
$$U = 20$$
$$Critical\ Value^2 = 19$$
$$U > 19$$

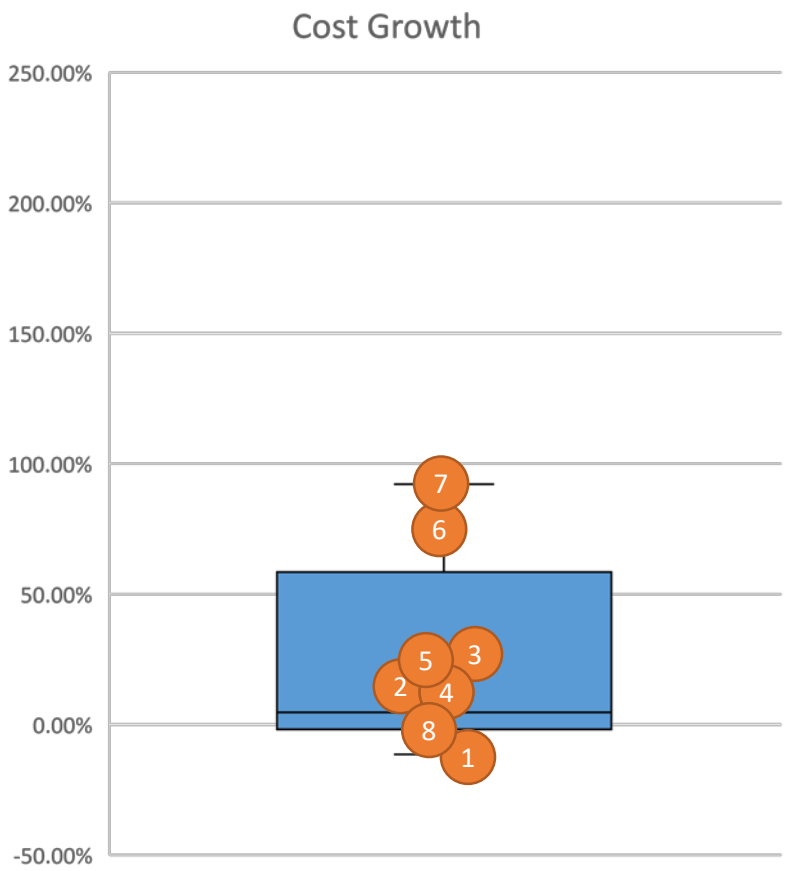
Fail to reject the null hypothesis.

2. RelativeResourceManager [PDF]. (2010, July 20). Ocw.umb.edu.

# Cost Growth

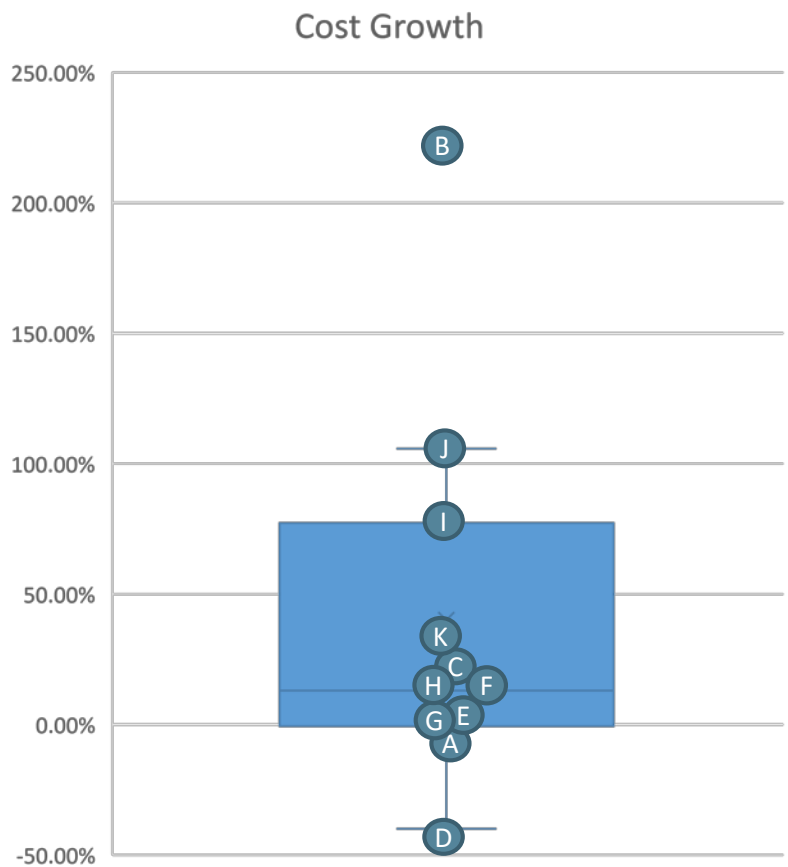
## Agile Projects

$n_1 = 8$




## Non-Agile Projects

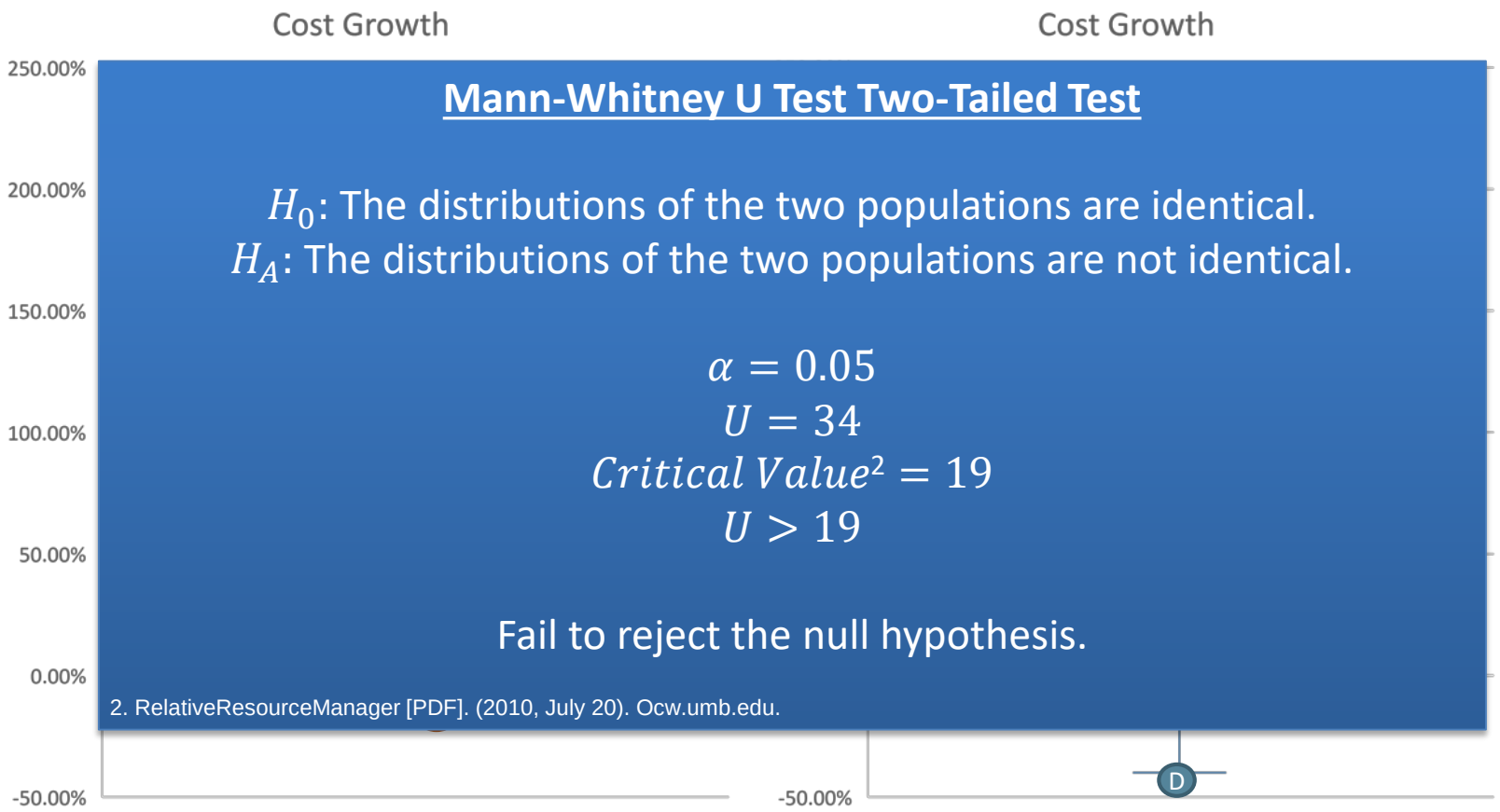
$n_2 = 11$



# Cost Growth

**Agile Projects**
$n_1 = 8$

**Non-Agile Projects**
$n_2 = 11$




**Mann-Whitney U Test Two-Tailed Test**

$H_0$: The distributions of the two populations are identical.
$H_A$: The distributions of the two populations are not identical.

$$\alpha = 0.05$$
$$U = 34$$
$$Critical\ Value^2 = 19$$
$$U > 19$$

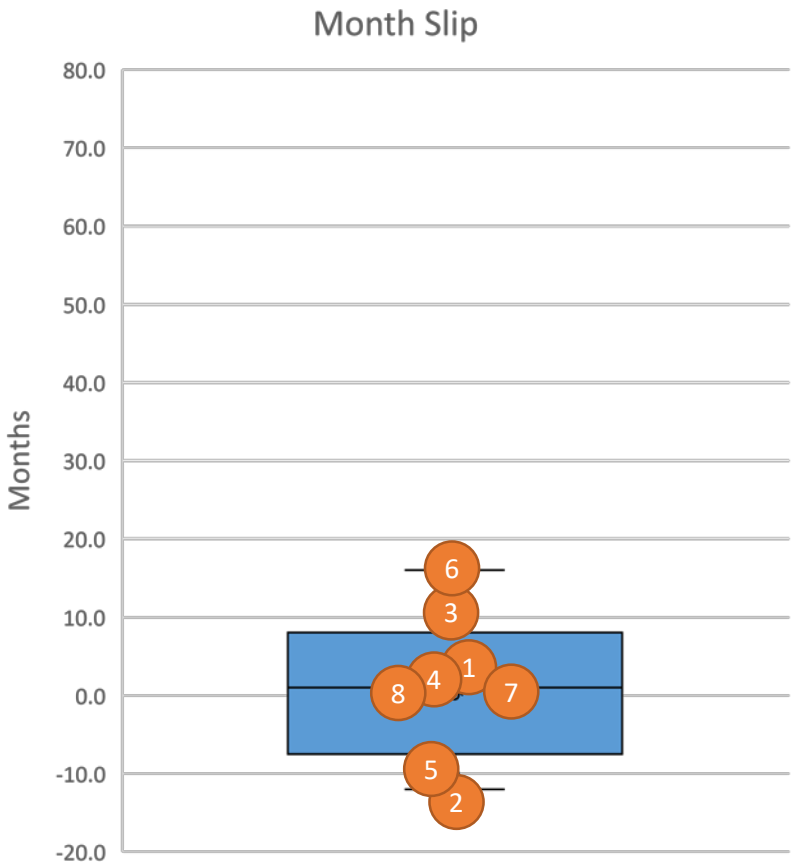
Fail to reject the null hypothesis.

2. RelativeResourceManager [PDF]. (2010, July 20). Ocw.umb.edu.

# Schedule Slip

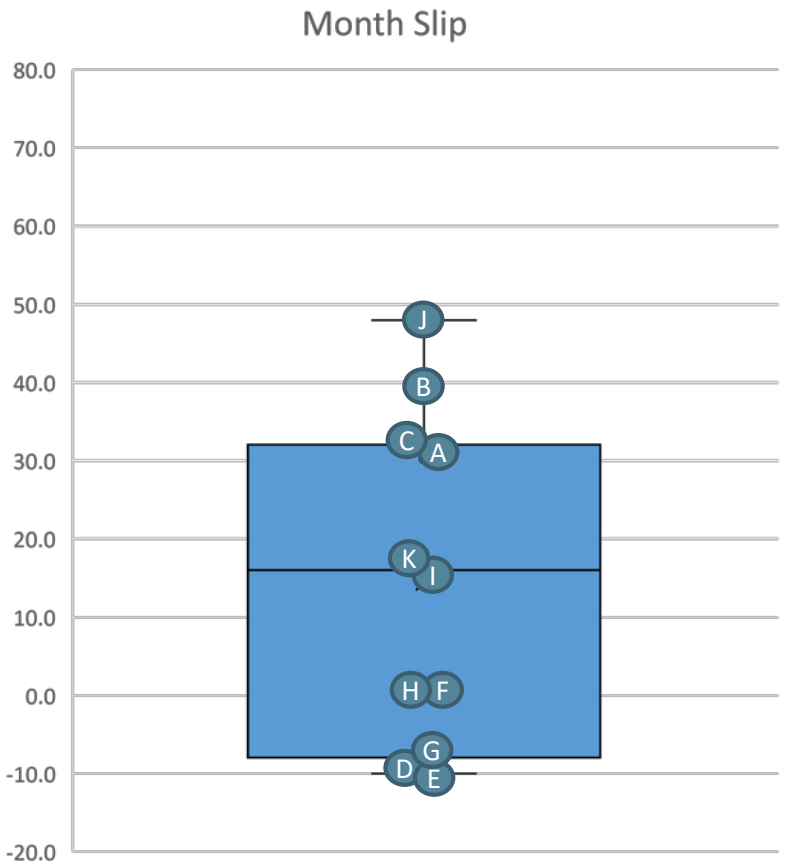
**Agile Projects**

$n_1 = 8$




**Non-Agile Projects**

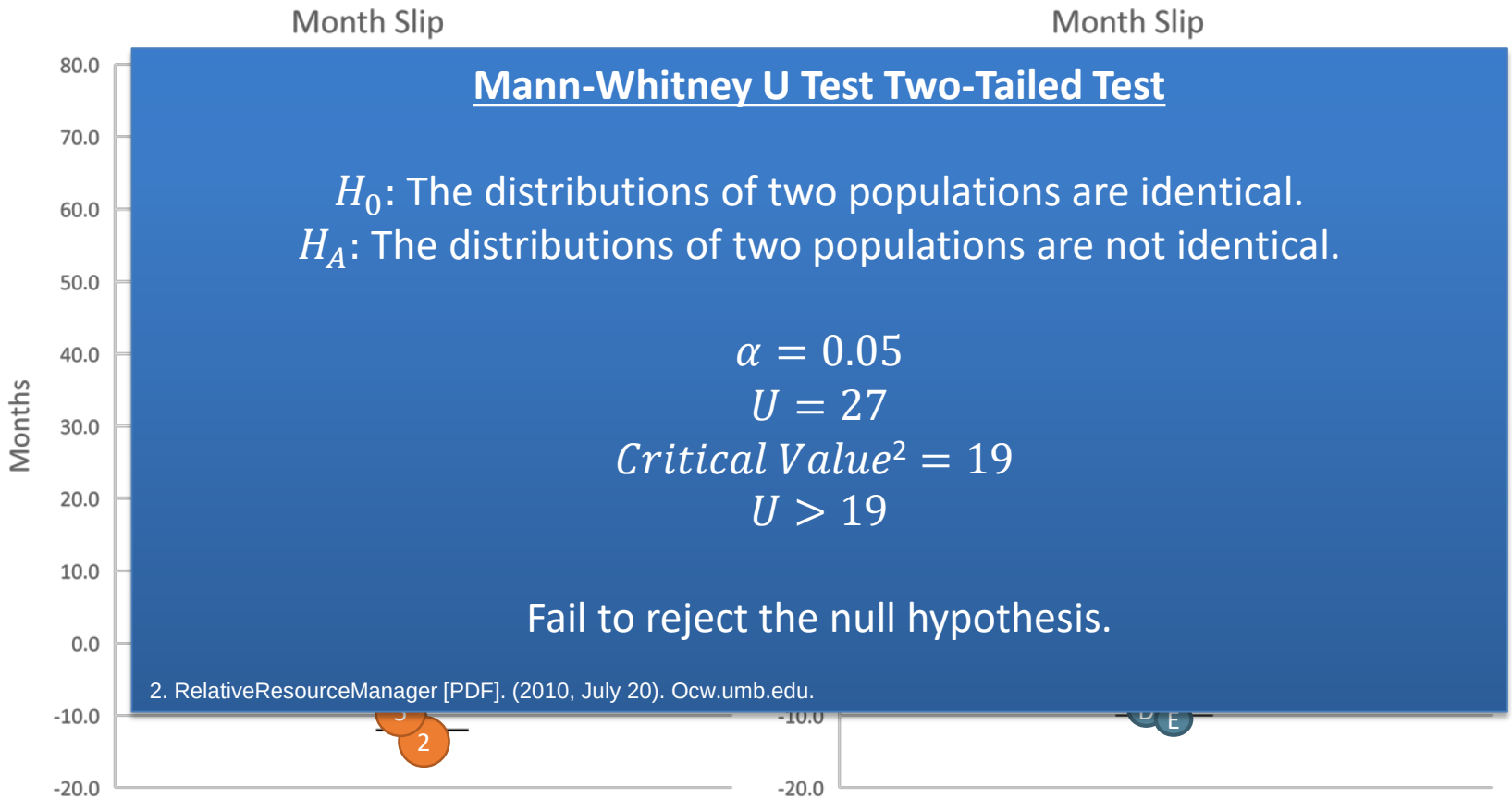
$n_2 = 11$



# Schedule Slip

**Agile Projects**
$n_1 = 8$

**Non-Agile Projects**
$n_2 = 11$




**Mann-Whitney U Test Two-Tailed Test**

$H_0$: The distributions of two populations are identical.
$H_A$: The distributions of two populations are not identical.

$$\alpha = 0.05$$
$$U = 27$$
$$Critical\ Value^2 = 19$$
$$U > 19$$

Fail to reject the null hypothesis.

2. RelativeResourceManager [PDF]. (2010, July 20). Ocw.umb.edu.

# All Results

*Comparison Between Two Samples*

| Metric | *U* Statistic | Critical Value | Result |
|---|---|---|---|
| Initial ESLOC/Hr | 43 | 19 | Fail to reject the null |
| Final ESLOC/Hr | 20 | 19 | Fail to reject the null |
| Initial Hrs/Req | 41 | 19 | Fail to reject the null |
| Final Hrs/Req | 36 | 19 | Fail to reject the null |
| Cost Growth | 34 | 19 | Fail to reject the null |
| Hours Growth | 41 | 19 | Fail to reject the null |
| SW Growth | 28 | 19 | Fail to reject the null |
| Req Growth | 25 | 19 | Fail to reject the null |
| Month Slip | 27 | 19 | Fail to reject the null |

*Comparison Within Each Sample*

| Metric | *U* Statistic | Critical Value | Result |
|---|---|---|---|
| **Agile Initial vs. Final** | | | |
| ESLOC | 23 | 13 | Fail to reject the null |
| ESLOC/Hr | 28 | 13 | Fail to reject the null |
| Hrs/Req | 30 | 13 | Fail to reject the null |
| **Non-Agile Initial vs. Final** | | | |
| ESLOC | 54 | 30 | Fail to reject the null |
| ESLOC/Hr | 57 | 30 | Fail to reject the null |
| Hrs/Req | 50 | 30 | Fail to reject the null |

# Agenda

- Introduction
  - Purpose
  - Agile Project Descriptions
  - Non-Agile Project Descriptions
  - Project Data Summary
- Analysis Details
  - Productivity Initial and Final Reports
  - Cost, Schedule, and Performance Variance
  - Mann-Whitney U Test
- Summary
  - Conclusion
  - Future Research


Presented at the 2019 ICEAA Professional Development & Training Workshop - www.iceaaonline.com

NSWCDD/PN/19/094

# Conclusion

## Fail to Reject the Null Hypothesis

- Initial and Final ESLOC Per Hour
- Initial and Final Hours Per Requirement
- Schedule Slip
- Cost Growth
- Hours Growth
- Software Growth
- Requirements Growth

## Reject the Null Hypothesis


Presented at the 2019 ICEAA Professional Development & Training Workshop - www.iceaaonline.com

NSWCDD/PN/19/094

# So . . . How does this affect me?

**Remember**

- Fail to reject means there is not sufficient evidence that the difference is non-zero

**Thus**

- With the data we have, there is not sufficient evidence to show there is a difference between agile and non-agile programs in the metrics analyzed

**So . . .**

- At this time, there is no reason to believe there is a need for separate methodologies to estimate agile programs

# Way Ahead

- Future Research
  - Collect more data and rerun analysis
    - If the results change, calculate effect size and practical significance
  - Analyze scope of agile and non-agile projects to compare initial and final scope
  - More detailed analysis on the comparability of the two samples in other technical parameters (e.g. product type, software language)
  - Partner with industry to analyze non-government acquisition agile development and compare with government acquisition agile development metrics

# Way Ahead

- Data Concerns
  - As more agile software development acquisition programs are completed, more data can be collected and analyzed
  - Due to limited data points, assumption of independence within datasets as well as between data sets was violated
    - Some data points within agile and between agile and non-agile were either developed by the same contractor for the same program or different contractors for the same program
  - Multiple data points are radar software development programs; prior analysis has shown radar software development is statistically different from other software development efforts[5]

5. Popp, Michael. (2013, 08). How I continued to stop worrying and love the Software Resources Data Report.

# Final Thoughts

**What we are saying**

- Agile software development may not require separate metrics for cost estimating purposes

**What we aren't saying**

- Agile software development is no better than non-agile software development methods


Presented at the 2019 ICEAA Professional Development & Training Workshop - www.iceaaonline.com

NSWCDD/PN/19/094

# Thank You


**Katelyn Barbre**
**Cost Engineering & Analysis Branch**
**(V11)**
**katelyn.barbre@navy.mil**
**540-653-6668**

# *Back ups*

# References


1) OSD CADE. (2018, 05). SRDR Data Compilation. *SRDR Data Compilation as of 20180516*. Retrieved from https://www.osd.cade.mil
2) RelativeResourceManager [PDF]. (2010, July 20). Ocw.umb.edu.
3) 7.3 Decision Making in Hypothesis Testing [PNG]. (n.d.). Onlinecourses.science.psu.edu.
4) Mann-Whitney U Test - Statistics Solutions. (n.d.). Retrieved from https://www.statisticssolutions.com/mann-whitney-u-test/
5) Popp, Michael. (2012, 02). *How I learned to stop worrying and love the Software Resources Data Report*.

# V&V Quality Tag Definitions[1]

- Definition of “Good”:
  - This is a data point that is complete for both hours and SLOC and has correct demographic, reporting event, personnel, and AD information. It also is not a TD or EMD effort (in other words the data point represents the totality of the software effort and does not have the artificial split created by TD/EMD), did not require an allocation of hours associated with support elements like CM, QA, SW Program Management or integration, or did not require combining build or phase information to make the data point complete.
- Definition of “Good – Allocation”
  - This is a data point that meets the criteria of good, but it has allocated hours associated with it to distribute things like QA, CM, SW PM and integration that were reported at the total effort level back to the lowest level CSCIs or work breakdown structure.