

# *Operational Test Design and Analysis using Microsoft Excel*

(FOR SPACE PROGRAMS)

*Dr. Jeffrey Strickland*



# **Test Designs for Space Operational Test & Evaluation using Excel**

By

Dr. Jeffrey Strickland



Test Designs for Space Operational Test & Evaluation using Excel

By Dr. Jeffrey Strickland, Chief Analyst, 1TES, Delta 12

Copyright © 2022, United States Space Force

ISBN 978-1-387-90547-8

This work is licensed under a Creative Commons (CC) license CC-BY-NC-SA. BY mean attribution is required, NC means no commercial use. SA means share alike, which means that any derivative work must adopt the same license type.

Printed in the United States of America



# Acknowledgements

"Where no man has gone before" is a phrase made popular through its use in the title sequence of the original 1966–1969 Star Trek science fiction television series, describing the mission of the starship Enterprise. The complete introductory speech, spoken by William Shatner as Captain James T. Kirk at the beginning of each episode, is:

*Space: the final frontier. These are the voyages of the starship Enterprise. Its five-year mission: to explore strange new worlds. To seek out new life and new civilizations. To boldly go where no man has gone before!*

— James T. Kirk, Captain, Enterprise

I never came very close to space, but I did contribute to the Areas I programs at NASA for a year, and spent a few months build a satellite constellation model for another space agency. I was privileged to teach alongside a future astronaut in the Department of Mathematical Sciences at West Point. I am not enjoying a job as the Chief Analyst for the First Test and Evaluation Squadron of Star Command, United States Space Force. So, to all the people, including James T. Kirk, who have helped me move closer to space, I say thanks.

I owe a great deal of gratitude to my family for their support and endurance, Laurie, Mariah, and Evie. This is my 37<sup>th</sup> book along this life journey.



# Table of Contents

ACKNOWLEDGEMENTS .....	I
TABLE OF CONTENTS.....	III
PREFACE .....	IX
1. SIMPLE COMPARATIVE EXPERIMENTS.....	1
DATA SCALES.....	1
<i>Nominal Data</i> .....	2
<i>Ordinal Data</i> .....	2
<i>Interval Data</i> .....	2
<i>Ratio Data</i> .....	3
BASIC CONCEPTS OF T DISTRIBUTION.....	3
<i>Other properties</i> .....	4
<i>Observations</i> .....	4
<i>Excel Functions</i> .....	5
CONFIDENCE LIMITS FOR THE MEAN .....	6
<i>Example 1.1: Confidence Interval for Heat flow meter</i> .....	8
<i>Potential Questions</i> .....	9
SAMPLE SIZE DETERMINATION .....	9
ESTIMATING SAMPLE SIZE FOR CONFIDENCE INTERVALS ESTIMATES.....	10
<i>Sample Size for One Sample, Continuous Outcome</i> .....	13
<i>Example 1.2: O-Ring Pressure Tolerance</i> .....	13
<i>Sample Size for One Sample, Dichotomous Outcome</i> .....	14
<i>Example 1.3: Space Debris</i> .....	15
<i>Sample Sizes for Two Independent Samples, Continuous Outcome</i> .....	15
<i>Example 1.4: Radiation Degradation</i> .....	16
ESTIMATING SAMPLE SIZE FOR HYPOTHESIS TESTING.....	17
<i>Ensuring That a Test Has High Power</i> .....	22
SAMPLE SIZE FOR MATCHED SAMPLES, CONTINUOUS OUTCOME.....	22
EXAMPLE 1.5: PAIN TREATMENT AND COMPUTER SCREENS .....	23
SAMPLE SIZES FOR TWO INDEPENDENT SAMPLES, DICHOTOMOUS OUTCOMES.....	25
EXAMPLE 1.6: TESTING FOR COVID-19 AMONG GYM USERS .....	26
<i>Sample Size for One Sample, Continuous Outcome</i> .....	27
<i>Example 1.7: Voice Command Translation</i> .....	28
<i>Sample Size for One Sample, Dichotomous Outcome</i> .....	29
<i>Example 1.8: Guardian LDL</i> .....	30
<i>Sample Sizes: Two Independent Samples, Continuous Outcome</i> .....	31

<i>Example 1.9: Water Displacing Formula .....</i>	32
SAMPLE SIZE DETERMINATION USING T-DISTRIBUTION.....	33
<i>Sample Size for t-test based on Confidence Interval.....</i>	35
<i>Example 1.10: Sample Size with Goal Seek.....</i>	35
<i>Required Sample Size for Binomial Testing.....</i>	37
EXAMPLE 1.11: BINOMIAL DISTRIBUTED MANUFACTURING PROCESS.....	37
<i>Determining Power.....</i>	38
<i>Example 1.12: Power Calculations.....</i>	39
TECHNIQUES FOR COMPARING MEANS .....	41
<i>Non-Normal Data .....</i>	42
TESTING: THE TWO-SAMPLE T-TEST .....	42
<i>How to Construct a Probability Plot with Excel .....</i>	45
TWO-SAMPLE T-TEST FOR EQUAL MEANS .....	45
<i>Example 1.13: Two-Sample t-Test, Equal Variances.....</i>	47
<i>Example 1.14: Two-Sample t-Test, Unequal Variances .....</i>	50
<i>Paired t-Test .....</i>	52
<i>Example 1.15: Paired Processor Test Samples.....</i>	53
<i>How to Create a Boxplot with Excel.....</i>	57
<i>Equivalence Testing (TOST).....</i>	65
<i>Example 1.16: Satellite Propagators .....</i>	66
SIGN TEST .....	68
<i>Example 1.17: C4ISR software .....</i>	68
WILCOXON RANK SUM TEST FOR INDEPENDENT SAMPLES .....	71
EXAMPLE 1.18 NEW SOFTWARE VS LEGACY.....	72
<i>Example 1.19: Orbital Elements with Data Analysis Add-in .....</i>	74
<b>2. THE BASIC PRINCIPLES OF DOE .....</b>	<b>81</b>
<i>t-Test Example .....</i>	81
RANDOMIZATION.....	82
REPLICATION.....	82
BLOCKING .....	83
MULTI-FACTOR DESIGNS .....	83
CONFOUNDING .....	84
FACTORS .....	85
<i>Experimental Factors.....</i>	85
<i>Classification Factors .....</i>	85
QUANTITATIVE VS. QUALITATIVE FACTORS.....	85
<i>Quantitative Factors.....</i>	85
<i>Qualitative Factors .....</i>	85
<i>Example 2.1: Factors and Levels Example .....</i>	85

<i>Model</i> .....	86
EXAMPLE 2.2: LEO PROLIFERATION DISCRIMINATION SOFTWARE.....	88
$\Sigma$ -ING.....	92
EXAMPLE 2.3: BY-HAND CALCULATION OF EXAMPLE 2.2.....	93
EXAMPLE 2.4: POWER FOR ANOVA.....	95
EXAMPLE 2.5: RBD SATELLITE CONTROLLERS .....	97
EXAMPLE 2.6: RBD OF EXAMPLE 2.4 USING REGRESSION.....	101
<b>3. ANOVA IN DESIGN OF EXPERIMENTS .....</b>	<b>107</b>
<i>Steps for Planning, Conducting and Analyzing an Experiment</i> .....	107
EXPERIMENTS WITH A SINGLE FACTOR .....	107
<i>One-way ANOVA — Completely Randomized Design (CRD)</i> .....	107
ONE-FACTOR ANOVA .....	108
<i>Definitions</i> .....	108
<i>Model Validation</i> .....	109
<i>Example 3.1: One-Way ANOVA for Gear Data</i> .....	109
<i>Example 3.2 The Gear Table</i> .....	110
TUKEY HONESTLY SIGNIFICANT DIFFERENCE (HSD) TEST.....	116
<i>Example 3.3: Testing a New Software System</i> .....	117
DUNNETT'S TEST .....	124
<i>Example 3.4: Dunnett's Ad Hoc Test</i> .....	124
<i>Example 3.5: Kruskal-Wallis H test</i> .....	130
DUNN'S TEST AFTER KW .....	133
<i>Example 3.6: Dunn's Test for the WD post comparison</i> .....	134
<b>4. MULTI-FACTOR ANALYSIS OF VARIANCE .....</b>	<b>139</b>
EXAMPLE 4.1: WD-40 .....	141
<i>ANOVA: Multi-Factor Without Replication</i> .....	142
<i>Multi-Factor ANOVA Followup Analysis</i> .....	145
<i>Example 4.2:Contrasts for Two Factor ANOVA</i> .....	145
<i>Example 4.3: Multi-Factor ANOVA with replication</i> .....	147
EXAMPLE 4.4: TWO-WAY ANOVA WITH CONTRASTS .....	153
<i>Summary of the ANOVA Analysis</i> .....	156
POST-ANOVA COMPARISON OF MEANS.....	156
<i>Simple Effects</i> .....	156
<i>Example 4.5: Simple Effects: Rows</i> .....	157
<i>Example 4.6: Simple Effects: Columns</i> .....	157
<i>Contrasts</i> .....	159
<i>Example 4.7: WD Contrasts</i> .....	159
EXAMPLE 4.8: ANOVA USING PRESSION .....	164

LATIN SQUARES DESIGN .....	171
<i>Example 4.9: Aircraft Parts</i> .....	171
TWO-FACTOR ART ANOVA .....	175
<i>Two Factor ART ANOVA</i> .....	175
<i>Example 4.10: ART ANOVA Likert Scale Questionnaire</i> .....	176
<b>2k FACTORIAL DESIGN BASIC CONCEPTS</b> .....	187
<i>Basic Concepts</i> .....	187
<i>Example 4.11: Factorial Designs</i> .....	187
<i>Effect Sizes</i> .....	189
<i>Standard Error</i> .....	190
<i>Example 4.12: Factorial Regression Model</i> .....	190
<b>5. BLOCKING SCENARIOS</b> .....	195
<i>Example 5.1: WD-40 Revisited</i> .....	196
<i>Example 5.2: RCBD Follow-up Testing</i> .....	198
<i>Example 5.3: Hardness Testing</i> .....	202
THE ANOVA FOR RANDOMIZED COMPLETE BLOCK DESIGN (RCBD).....	203
<i>ANOVA Display for the RCBD</i> .....	205
OTHER ASPECTS OF THE RCBD.....	210
<b>6. LOGISTIC REGRESSION</b> .....	211
BASICS.....	211
LOGISTIC FUNCTION, ODDS RATIO, AND LOGIT.....	213
<i>Multiple explanatory variables</i> .....	215
EXAMPLE 6.1: ORBIT CLASSIFICATION SW .....	216
<i>Test the Logistic Regression Function</i> .....	222
<i>Validate the Logistic Regression Function</i> .....	222
<b>APPENDIX A: STATISTICAL TABLES</b> .....	227
STUDENT T-DISTRIBUTION .....	227
WILCOXON RANK-SUM TABLE.....	231
STUDENT Q TABLE .....	240
DUNNET'S CRITICAL VALUES .....	252
Z-TABLE (UPPER-TAIL) .....	258
STANDARD NORMAL MARGIN OF ERROR (E) TABLE.....	259
CHI-SQUARE TABLE (UPPER-TAIL) .....	262
F CRITICAL VALUE TABLE, $F_1 - \alpha, df_1, df_2$ .....	266
OFFSET CIRCULAR NORMAL DENSITY FUNCTION .....	280
<b>REFERENCES</b> .....	287





# Preface

To write a single book about operational testing might be absurd, especially since it is not a blossoming discipline and may seem to be confined to the armed services as they acquire new systems to enhance and perform their missions. However, operational testing is not the sole domain of the military or even other government agencies, like NASA. Sometimes operational testing is not called operational testing, like when Ford produces a new line of SUVs, but it's still operational testing.

Now, a lot of people think they know operational testing, but if they think it is about advanced experimental designs (DOE) and multiple analysis of variance (MANOVA), then they might just miss the mark. It's really about data, pure unadulterated data and what the data tells you to do, i.e., how to analyze it—data talks.

*Data really powers everything that we do.*

— Jeff Weiner, LinkedIn

## What is Adaptive Operational Testing?

Actually, it's not anything. There is only operational testing. There's no lean testing or agile testing. Lean and Agile are manufacturing and development ideologies. When we test something, regardless of whether it is a piece of software or the Starship Enterprise, operational testing adapts. It's still all about data—planning for it, collecting it, processing it, analyzing it, and drawing conclusions.

**Data Preprocessing.** We do not want processed data. Data that is processed is already molded into a story, predesigned. We want it raw so that we can see what the unbiased data has to say to us, and we should be prepared for the fact that it may not tell the story we wanted or expected. Care must be taken, when we plan a test, not to bias the data before collecting it. We process the data after the collecting it, not before.

**Data Mining.** Data mining is the semi-automatic or automatic process of taking large quantities of data and extracting data that is pertinent to an organization's data operations. Such mining pulls out interesting patterns such as groups of data records (cluster analysis); identifies items, events or observations which do not conform to an expected pattern (anomaly detection or outlier detection); finds frequent co-occurring associations among a collection of items (association rule mining or market basket analysis); and discovers statistically relevant patterns between data examples where the values are delivered in a sequence (sequential pattern mining). These mining tasks usually involve using database techniques to find patterns, which may be used in further analysis or, for example, in machine learning and predictive analytics. However, the data collection, data preparation, as well as the result interpretation and reporting is not part of the data mining step.

*In God we trust. All others must bring data.*

— W. Edwards Deming, statistician

**Machine Learning.** Machine learning (ML) is not the same thing as artificial intelligence (AI). ML is the science of getting computers to act without being explicitly programmed (the domain of AI). In fact, many researchers think that ML is the best way to make progress towards human-level AI. In data science, ML or the algorithms of ML are used to mine data, explore mined data, and predict future outcome, including how effective a new GPS satellite will be. I will not cover ML here, as it is a more advanced topic, but do be aware of it.

*I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.*

— Alan Turing, Computing machinery and intelligence

**Data Analytics.** Data analytics is more than data analysis. It is a process of examining, cleaning, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and providing decision support. Data analytics encompasses multiple methods and

approaches within the realms of descriptive, predictive, and prescriptive analysis, while being used in different business, science, medical, psychological, and social science domains. The goals of data analytics are discovering useful information, informing conclusions, and supporting decision-making. Once the data is cleaned, data analysts apply a variety of techniques, referred to as exploratory data analysis, to begin understanding the information contained in the data.

Data exploration can result in additional data cleaning or additional data requirements, so these activities may be iterative. In descriptive analytics, unfolding the characteristics of the data with such measures as the mean and variance may help in understanding the data. Predictive analytics is concerned with forecasting future events based on the data that has been mined and explored. Prescriptive analytics is concerned with telling the story of why the data describes the present or predicts the future to help decision makers choose the best courses of action.

We (operational testers) do this.

Descriptive Analytics, which use data aggregation and data mining to provide insight into the past and answer: "What has happened?"

Predictive Analytics, which use statistical models and forecasts techniques to understand the future and answer: "What could happen?"

*The goal is to turn data into information, and information into insight.*

— Carly Fiorina, former CEO, Hewlett-Packard

**Statistics** (Statistics, Applied Mathematics). In my assessment, this is the foundation of data science. It is the science that deals with the collection, classification, analysis, and interpretation of numerical facts or data. Supported directly by use of mathematical theories of probability, statistics imposes order and regularity on collections of dissimilar elements. This is the first time I have included a fair coverage of inferential statistics, regression models, and generalized linear models (GLMs) in a data analytics book. But, people generally do not

know how to read, interpret, and use statistics, which must be governed by the following inquiries:

- How was the data collected?
- Does the evidence come from reliable sources?
- What is the data's background?
- Are all data reported?
- Have the data been interpreted correctly?

*Torture the data, and it will confess to anything.*

– Ronald Coase, winner of the Nobel Prize in Economics

My free workbooks are available at : <https://github.com/stricje1>.

and [1 TES SharePoint](#) if you have a government email account.

## Related Books by the Author

### Related Books by the Author

---

*The R Guide for New Data Scientists*, Copyright © 2022, Lulu, Inc. ISBN 978-1-6780-0244-2

*The Python Guide for New Data Scientists*, Copyright © 2022, Lulu, Inc. ISBN 978-1-4583-2161-9

*Time Series Analysis and Forecasting using Python & R*. Copyright© 2020, Lulu, Inc. ISBN 978-1-716-45113-3

*Data Science Applications using Python and R*. Copyright© 2020, Lulu, Inc. ISBN 978-1-716-89644-6

*Data Science Applications using R*. Copyright© 2020, Lulu, Inc. ISBN 978-0-359-81042-0

*Logistic Regression Inside-Out*. Copyright © 2020 by Jeffrey S. Strickland. Lulu, Inc. ISBN 978-1-365-81915-5

*Predictive Crime Analysis using R*. Copyright© 2018, Jeffrey Strickland. Lulu, Inc. ISBN 978-0-359-43159-5

*Time Series Analysis using Open-Source Tools*. Copyright© 2016, Jeffrey Strickland. Glasstree, Inc. ISBN 978-1-5342-0100-2

*Predictive Analytics using R*. Copyright © 2016 by Jeffrey S. Strickland. Lulu.com. ISBN 978-1-312-84101-7

*Introduction to Crime Analysis and Mapping*. © 2016 by Jeffrey S. Strickland. Lulu.com. ISBN 978-1-312-19311-6

*Data Analytics Using Open-Source Tools*. Copyright © 2015 by Jeffrey Strickland. Lulu Inc. ISBN 978-1-365-21384-7

*Data Science and Analytics for Ordinary People*. Copyright © 2015 by Jeffrey S. Strickland. Lulu.com. ISBN 978-1-329-28062-5

*Operations Research using Open-Source Tools.* Copyright © 2015 by Jeffrey Strickland. Lulu Inc. ISBN 978-1-329-00404-7

*Missile Flight Simulation - Surface-to-Air Missiles, 2<sup>nd</sup> Edition.* Copyright © 2015 by Jeffrey S. Strickland. Lulu.com. ISBN 978-1-329-64495-3

*Predictive Modeling and Analytics.* © 2014 by Jeffrey S. Strickland. Lulu.com. ISBN 978-1-312-37544-4

*Verification and Validation for Modeling and Simulation.* Copyright © 2014 by Jeffrey S. Strickland. Lulu.com. ISBN 978-1-312-74061-7

*Using Math to Defeat the Enemy: Combat Modeling for Simulation.* © 2011 by Jeffrey S. Strickland. Lulu.com. ISBN 978-1-257-83225-5

*Mathematical Modeling of Warfare and Combat Phenomenon.* Copyright © 2011 by Jeffrey S. Strickland. Lulu.com. ISBN 978-1-45839255-8

*Simulation Conceptual Modeling.* Copyright © 2011 by Jeffrey S. Strickland. Lulu.com. ISBN 978-1-105-18162-7.

*Discrete Event Simulation using ExtendSim 8.* Copyright © 2010 by Jeffrey S. Strickland. Lulu.com. ISBN 978-0-557-72821-3

# 1. Simple Comparative Experiments

Simple comparative experiments are often overlooked when performing operational testing. Here, we will look at both hypothesis testing and estimation and from comparative perspectives, and we will look at sample size determination for various scenarios. We start with some preliminaries.

Comparative analysis seeks to see the difference in the average value of the dependent variable between two or more groups. The comparison test (dependent variable) for the two sample groups is the t-test. The test for more than two groups of samples is the F-test (ANOVA) (Global, 2022). Furthermore, comparative analysis requires normal distribution, or the ability to invoke the Central Limit Theorem.

The one-sample hypothesis test for data samples assumed to be approximately normal is appropriate when one knows the standard deviation,  $\sigma$ , of the population distribution and the population is either normally distributed or the sample is sufficiently large that the Central Limit Theorem applies. However,  $\sigma$  is seldom known. Hence, we often use the Student's t-distribution to perform most operations, like sample size determination, confidence interval (CI) construction, and sample analysis.

Before we dive into analysis of test data, let's look at the different types or scales of data.

## Data Scales

In statistics we know the scale of data measurement : nominal data, ordinal data, interval, and ratio. In general, it can be said that the purpose of an observation is to obtain about the condition of an object in various circumstances. Among the various measurements for objects, which are numbers, rank, length, volume, time, weight, and physical-chemical measurements (Glen, 2022).

## **Nominal Data**

First, nominal Scale is the simplest measurement scale. It groups objects into several groups, which have similar features will be in one group. Nominal scale measurement results not to sort but can distinguish. Common examples commonly used are gender variables. Examples of nominal scale applications: user feedback, types of satellites, orbital regimes. Nominal data may be expressed as features form larger datasets. These may arrive as n-grams, where a bigram may be string of two frequently used words.

## **Ordinal Data**

Second, ordinal Data describes the position or rank but do not measure the distance between ranks. Size on an ordinal scale does not give an absolute value to an object, but only a relative sequence. Furthermore, the distance between rank 1 and 2 does not have to be the same as the distance between rank 2 and 3. On an ordinal scale, the rank does not have a unit of measurement. For example: social status (high, low, medium), measurement results that classify people into high, low or medium social status. In this case, we can know the level, but the difference between social statuses (high-low, low-medium, high-medium etc.) is not necessarily the same. Example application: preference level, management position, career path.

## **Interval Data**

Third, interval scale of data gives us the numeric characteristic to objects that lie between over and beyond nominal and ordinal scales. With intervalscale data, the distance between objects carry meaning. Interval scale is the level of this scale above the ordinal and nominal scale. Therefore, an important feature of this scale: we can add, subtract, duplicate, and share without affecting the relative distance of the scores.

Furthermore, this scale does not have an absolute zero. We cannot interpret in full the value of a certain ratio. In interval measurement switches, the ratio between two arbitrary intervals does not depend on the value of zero and the unit of measurement, For example,

measurement of temperature on a Celsius scale. If a water bath is full of 0 degrees C, 50 degrees C, and 100 degrees C, the difference between 0-50 and 50-100 degrees C is the same. We cannot say that water at 100 degrees C is twice as hot as water 50 degrees C. Example application: Guardian performance appraisal (on a scale)

## Ratio Data

Data with ratio scale has all the properties of the interval scale plus one trait to give information about the absolute value of the object. It aims to distinguish, sort, certain distance, and we can compare (the most complete, including all of the scales above). Example: If we want to compare the weight of two people. A weight 40 kg and B 80 kg. We can know that A is twice as heavy as A. Because the value of the numerical variable weight expresses the ratio with zero as its default. Other examples: physical force, value for money, height, etc.

## Basic Concepts of t Distribution

Since the standard deviation of the population is generally not known, we use different statistical approaches when analyzing samples. One approach for addressing this is to use the standard deviation  $s$  of the sample as an approximation for the standard deviation  $\sigma$  for the population. A good approach is to use the so-called (Student's)  $t$ -distribution.

### ***Definition 1-1: Student's t Distribution***

The (Student's)  $t$  distribution with  $k$  degrees of freedom, abbreviated  $T(k)$  has the probability distribution function (pdf)

$$f(x) = \frac{r(\frac{k+1}{2})}{\sqrt{\pi k} \Gamma(\frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}$$

where  $\Gamma(y)$  is the **gamma function**.

Key statistical properties of the  $t$  distribution are:

- Mean = 0 for  $k > 0$

- Median = 0
- Mode = 0
- Range =  $(-\infty, \infty)$
- Variance =  $k / (k - 2)$  for  $k > 2$
- Skewness = 0 for  $k > 3$
- Kurtosis =  $6 / (k - 4)$  for  $k > 4$

The overall shape of the probability density function of the t-distribution resembles the bell shape of a normally distributed random variable with mean 0 and variance 1, except that it is a bit lower and wider. As the number of degrees of freedom grows, the  $t$  distribution approaches the standard normal distribution, and in fact, the approximation is quite close for  $k \geq 30$ .

It may be without saying that the  $t$ -distribution is perhaps the most important probability distribution in the entirety of operational testing, as pertains to nearly all of the statistical tests we apply in this handbook, either directly or under the surface.

## Other properties

**Property 1:** If  $x$  has the normal distribution  $N(\mu, \sigma^2)$ , then for samples of size  $n$  with mean  $\bar{x}$  and standard deviation  $s$ , the random variable

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

has distribution  $T(n-1)$ .

**Property 2:** For samples of sufficiently large size  $n$  with mean  $\bar{x}$  and standard deviation  $s$ , the random variable

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

has distribution  $T(n-1)$ .

## Observations

The test statistic in Properties 1 and 2 are the same as

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

from the *Central Limit Theorem* with the population standard deviation  $\sigma$  replaced by the sample standard deviation  $s$ . What makes this so useful is that usually, the standard deviation of the population is unknown while the standard deviation of the sample is known.

When the sample makes up a substantial portion of a (finite) population (e.g., more than 5%), then the standard error  $s/\sqrt{n}$  in Properties 1 and 2 should be replaced by

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

where  $N$  is the population size.

## Excel Functions

Excel provides the following functions for the  $t$ -distribution:

- $T.DIST(x, \text{deg\_freedom}, \text{cumulative})$  = the probability density function  $f(x)$  for the  $t$ -distribution when cumulative = FALSE. We plot the  $t$ -distribution density function  $f(x)$  for several values of  $n$  and df in **Figure 1.1**.
- $T.DIST(x, df, \text{cumulative})$  = the probability cumulative function  $f(x)$  for the  $t$ -distribution when cumulative = TRUE.
- $T.INV(probability, \text{deg\_freedom},)$  = the value  $x$  such that  $T.DIST(x, df, \text{TRUE}) = p$ , i.e., the inverse of the cumulative distribution.
- $T.DIST.2T$  and  $T.DIST.RT$  are two-tail and right-tail distributions, respectively. Symmetry of the  $t$ -distribution ensures that:

$$T.DIST.T(-x, df, \text{TRUE}) =$$

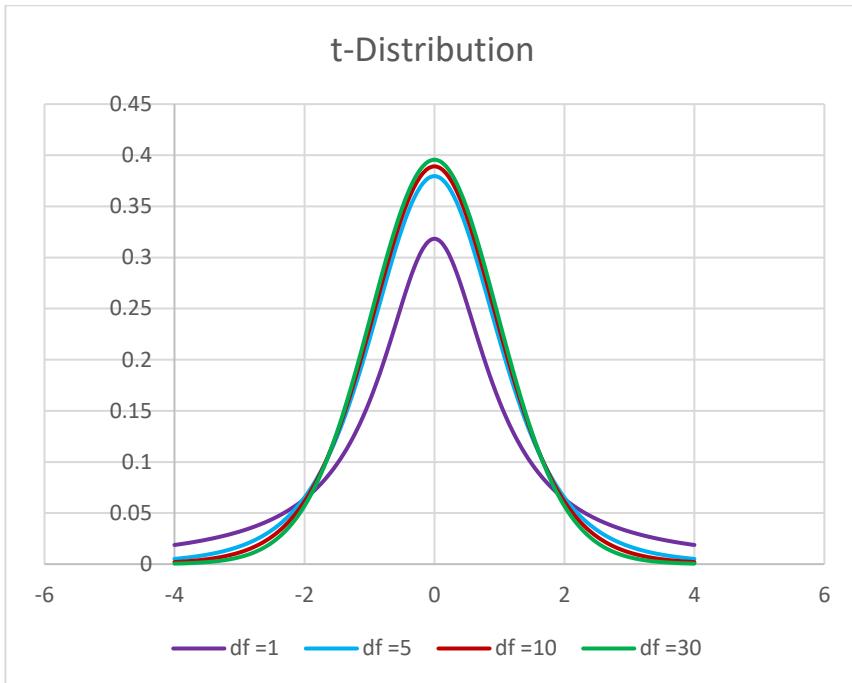
$$1 - T.DIST(x, df, \text{TRUE}) = T.DIST.RT(X, df)$$

*and*

$$T.DIST.RT - x, df =$$

$$1 - T.DIST.RT(x, df) = T.DIST(x, df, \text{TRUE})$$

- $T.INV.2T$  is the two-tail inverse t-distribution. Moreover, since the t-distribution is symmetrical,  $T.INV.2T(2p, df)$  for  $p < .5$  is equivalent to  $T.INV(1 - p, df)$ .



**Figure 1.1.** Student's t-distribution for several values of  $n$  and  $df$

## Confidence Limits for the Mean

Confidence limits (interval estimates) for the mean ( (Nelson & Nelson, 2021)) are an interval estimate for the mean. Interval estimates are often desirable because the estimate of the mean varies from sample to sample. Instead of a single estimate for the mean, a confidence interval generates a lower and upper limit for the mean. The interval estimate gives an indication of how much uncertainty there is in our estimate of the true mean. The narrower the interval, the more precise is our estimate.

Confidence limits are expressed in terms of a confidence coefficient. Although the choice of confidence coefficient is somewhat arbitrary, in

practice 90 %, 95 %, and 99 % intervals are often used, with 95 % being the most commonly used.

As a technical note, a 95 % confidence interval does not mean that there is a 95 % probability that the interval contains the true mean. The interval computed from a given sample either contains the true mean or it does not. Instead, the level of confidence is associated with the method of calculating the interval. The confidence coefficient is simply the proportion of samples of a given size that may be expected to contain the true mean. That is, for a 95 % confidence interval, if many samples are collected and the confidence interval computed, in the long run about 95 % of these intervals would contain the true mean.

***Definition 1-2: Confidence Interval***

Confidence limits are defined as:

$$\bar{Y} \pm \left( t_{1-\frac{\alpha}{2}, N-1} \right) \frac{s}{\sqrt{N}}$$

where  $\hat{Y}$  is the sample mean,  $s$  is the sample standard deviation,  $N$  is the sample size,  $\alpha$  is the desired significance level, and  $t_{1-\frac{\alpha}{2}, N-1}$  is the  $100(1 - \alpha/2)$  percentile of the t distribution with  $N - 1$  degrees of freedom. Note that the confidence coefficient is  $1 - \alpha$ .

From the formula, it is clear that the width of the interval is controlled by two factors:

1. As  $N$  increases, the interval gets narrower from the  $\sqrt{N}$  term. That is, one way to obtain more precise estimates for the mean is to increase the sample size.
2. The larger the sample standard deviation, the larger the confidence interval. This simply means that noisy data, i.e., data with a large standard deviation, are going to generate wider intervals than data with a smaller standard deviation.

### **Definition 1-3: T-Test for a population Mean**

The hypothesis test to test whether the population mean has a specific value,  $\mu_0$ , against the two-sided alternative that it does not have a value  $\mu_0$ , utilized the confidence interval converted to hypothesis-test form. The test is a one-sample t-test, and it is defined as:

$$H_0: \mu = \mu_0$$

$$H_A: \mu \neq \mu_0$$

Test Statistic:

$$T = \frac{\bar{Y} - \mu_0}{\frac{s}{\sqrt{N}}}$$

where  $\bar{Y}$ ,  $N$ , and  $s$  are defined as above.

**Significance Level:**  $\alpha$ . The most commonly used value for  $\alpha$  is 0.05.

**Critical Region:** Reject the null hypothesis that the mean is a specified value,  $\mu_0$ , if

$$T < t_{\frac{\alpha}{2}, N-1}$$

or

$$T > t_{1-\frac{\alpha}{2}, N-1}$$

## **Example 1.1: Confidence Interval for Heat flow meter**

### *Background and Data*

This data set was collected by Bob Zarr of NIST in January, 1990 from a heat flow meter calibration and stability analysis. The response variable is a calibration factor.

The motivation for studying this data set is to illustrate a well-behaved process where the underlying assumptions hold and the process is in statistical control.

We generated a 95 %, two-sided confidence interval for the ZARR13.DAT data set based on the following information.

**Table 1-1. Table of parameters for CI calculation**

N	= 195
MEAN	= 9.261460
STANDARD DEVIATION	= 0.022789
$t_{1-0.025,N-1}$	= 1.9723

$$LOWER\ LIMIT = 9.261460 - 1.9723 * \frac{0.022789}{\sqrt{195}}$$

$$UPPER\ LIMIT = 9.261460 + 1.9723 * \frac{0.022789}{\sqrt{195}}$$

Thus, a 95 % for the mean is (9.258242, 9.264679).

## Potential Questions

Confidence limits for the mean can be used to answer the following questions:

1. What is a reasonable estimate for the mean?
2. How much variability is there in the estimate of the mean?
3. Does a given target value fall within the confidence limits?

## Sample Size Determination

An important aspect of designing an experiment is to know how many observations are needed to make conclusions of sufficient accuracy and with sufficient confidence. We review what we mean by this statement. The sample size needed depends on lots of things; including what type of experiment is being contemplated, how it will be conducted, resources, and desired sensitivity and confidence.

Sensitivity refers to the difference in means that the experimenter wishes to detect, i.e., sensitive enough to detect important differences in the means.

Generally, **increasing** the number of **replications increases** the **sensitivity** and makes it easier to detect small differences in the means. Both power and the margin of error are a function of  $n$  and a function of the error variance. Most of this course is about finding techniques to reduce this unexplained residual error variance, and thereby improving the power of hypothesis tests, and reducing the margin of error in estimation.

## Estimating Sample Size for Confidence Intervals Estimates

Recall that confidence intervals for every parameter take the following general form:

$$\text{Point Estimate} + \text{Margin of Error}$$

The formula for the confidence interval for  $\mu$  is

$$\bar{X} \pm Z \frac{\sigma}{\sqrt{n}}$$

In practice we use the sample standard deviation to estimate the population standard deviation. Note that there is an alternative formula for estimating the mean of a continuous outcome in a single population, and it is used when the sample size is small ( $n < 30$ ). It involves a value from the  $t$  distribution, as opposed to one from the standard normal distribution, to reflect the desired level of confidence. When performing sample size computations, we use the large sample formula shown here. [Note: The resultant sample size might be small, and in the analysis stage, the appropriate confidence interval formula must be used.]

The point estimate for the population mean is the sample mean and the margin of error is

$$Z \frac{\sigma}{\sqrt{n}}$$

In planning studies, we want to determine the sample size needed to ensure that the margin of error is sufficiently small to be informative. For example, suppose we want to estimate the mean latency of a routine data query from a cloud-based Data as a Service (DaaS). We conduct a study and generate a 95% confidence interval as follows  $180 \text{ sec} \pm 40$

sec, or 140 to 220 seconds. The margin of error is so wide that the confidence interval is uninformative. To be informative, an investigator might want the margin of error to be no more than 10 or 12 seconds (meaning that the 95% confidence interval would have a width (lower limit to upper limit) of 20 or 24 seconds). In order to determine the sample size needed, **the investigator must specify the desired margin of error**. It is important to note that this is not a statistical issue, but a clinical or a practical one. For example, suppose we want to estimate the mean latency for a mathematical conversion using a premium Intel processor in operational configured analysis tool suite. Internal processor speed clearly has a much more restricted range than DaaS query latency. Therefore, we would probably want to generate a confidence interval for the processor latency that has a margin of error not exceeding 2 or 3 seconds.

#### ***Definition 1-4: Margin of Error***

The margin of error in the one sample confidence interval for  $\mu$  can be written as follows:

$$E = Z \frac{\sigma}{\sqrt{n}}$$

Our goal is to determine the sample size,  $n$ , that ensures that the margin of error,  $E$ , does not exceed a specified value. We can take the formula above and, with some algebra, solve for  $n$ .

#### ***Definition 1-5: Sample Size***

The sample size,  $n$ , is defined by

$$n = \left( \frac{Z\sigma}{E} \right)^2$$

This result is found by multiplying both sides **Definition 1-4** by the square root of  $n$ . Then cancel out the square root of  $n$  from the numerator and denominator on the right side of the equation (since any number divided by itself is equal to 1). This leaves:

$$\sqrt{n}E = Z\sigma$$

Now divide both sides by  $E$  and cancel out  $E$  from the numerator and denominator on the left side. This leaves:

$$\sqrt{n} = \frac{Z\sigma}{E}$$

Finally, square both sides of the equation to get:

$$n = \left( \frac{Z\sigma}{E} \right)^2$$

This formula generates the sample size,  $n$ , required to ensure that the margin of error,  $E$ , does not exceed a specified value. To solve for  $n$ , we must input  $Z$ ,  $\sigma$ , and  $E$ .

$Z$  is the value from the table of probabilities of the standard normal distribution for the desired confidence level (e.g.,  $Z = 1.96$  for 95% confidence)

$E$  is the margin of error that the investigator specifies as important from a clinical or practical standpoint.

$\sigma$  is the standard deviation of the outcome of interest.

Sometimes it is difficult to estimate  $\sigma$ . When we use the sample size formula above (or one of the other formulas that we will present in the sections that follow), we are **planning** a study to estimate the unknown mean of a particular outcome variable in a population. It is unlikely that we would know the standard deviation of that variable. In sample size computations, investigators often use a value for the standard deviation from a previous study or a study done in a different, but comparable, population. The sample size computation is not an application of statistical inference and therefore it is reasonable to use an appropriate estimate for the standard deviation. The estimate can be derived from a different study that was reported in the literature; some investigators perform a small pilot study to estimate the standard deviation. A pilot study usually involves a small number of participants (e.g.,  $n = 10$ ) who are selected by convenience, as opposed to by random sampling. Data from the participants in the pilot study can be used to compute a sample standard deviation, which serves as a good estimate for  $\sigma$  in the sample size formula. Regardless of how the estimate of the variability of the

outcome is derived, it should always be conservative (i.e., as large as is reasonable), so that the resultant sample size is not too small.

The formula  $n = (Z\sigma/E)^2$  produces the minimum sample size to ensure that the margin of error in a confidence interval will not exceed  $E$ . In planning studies, investigators should also consider attrition or loss to follow-up. The formula above gives the number of participants needed with complete data to ensure that the margin of error in the confidence interval does not exceed  $E$ . We will illustrate how attrition is addressed in planning studies through examples in the following sections.

## Sample Size for One Sample, Continuous Outcome

In studies where the plan is to estimate the mean of a continuous outcome variable in a single population, the formula for determining sample size is given by **Definition 1-5**:

$$n = \left( \frac{Z\sigma}{E} \right)^2$$

where  $Z$  is the value from the standard normal distribution reflecting the confidence level that will be used (e.g.,  $Z = 1.96$  for 95%),  $\sigma$  is the standard deviation of the outcome variable and  $E$  is the desired margin of error. The formula above generates the minimum number of subjects required to ensure that the margin of error in the confidence interval for  $\mu$  does not exceed  $E$ .

### Example 1.2: O-Ring Pressure Tolerance

An investigator wants to estimate the mean pressure tolerance of a combustion chamber O-ring in Molniya/HEO surveillance satellite. How many O-ring tests should be performed in the study? The investigator plans on using a 95% confidence interval (so  $Z = 1.96$ ) and wants a margin of error of 5 psi. The standard deviation of pressure tolerance is unknown, but the investigators conduct a literature search and find that the standard deviation of similar O-ring pressure tolerances in other HEO satellites is 15 psi and 20 psi. To estimate the sample size, we consider

the larger standard deviation in order to obtain the most conservative (largest) sample size.

The calculation is easily done by hand, but we'll use the formula in the Sample Size tab of our Excel workbook as shown in **Figure 1.2**. We see that the resulting sample size required is 61. So, we'll perform the pressure tolerance test in a controlled vacuum environment using 61 O-rings.

	A	B	C	D	E
1	<b>Sample Size for One Sample, Continuous Outcome</b>				
2	Confidence level	95%			
3	Estimated sigma	20		n=	61
4	Margin of error E	5.00			

**Figure 1.2. Sample Size calculator for One Sample, Continuous Outcome**

Name	Cell	Formula
n	E3	=((NORM.S.INV(B2+(1-B2)/2)*B3)/B4)^2

### Sample Size for One Sample, Dichotomous Outcome

In studies where the plan is to estimate the proportion of successes in a dichotomous outcome variable (yes/no) in a single population, the formula for determining sample size is:

$$n = p(1 - p) \left(\frac{Z}{E}\right)^2$$

where **Z** is the value from the standard normal distribution reflecting the confidence level that will be used (e.g.,  $Z = 1.96$  for 95%) and **E** is the desired margin-of-error. **p** is the proportion of successes in the population. Here we are planning a study to generate a 95% confidence interval for the unknown population proportion, **p**. The equation to determine the sample size for determining **p** seems to require knowledge of **p**, but this is obviously this is a circular argument, because if we knew the proportion of successes in the population, then a study would not be necessary! What we really need is an approximate value of **p** or an anticipated value. The range of **p** is 0 to 1, and therefore the range of  $p(1 - p)$  is 0 to 1. The value of **p** that maximizes  $p(1 - p)$  is

$p = 0.5$ . Consequently, if there is no information available to approximate  $p$ , then  $p = 0.5$  can be used to generate the most conservative, or largest, sample size.

### Example 1.3: Space Debris

As space commerce continues to grow, so does the density of space assets in “popular” orbits (low earth orbit, polar orbits, sun synchronous orbits and geostationary orbits). These assets mostly include spent rocket bodies and satellites (many have been inactive for years or decades). Similar to the meticulous measurements that astronomers take to map the location of stars in the sky, space researchers and military organizations measure and track orbital parameters of most RSOs larger than 5 cm in diameter. We want to determine the proportion of active satellites to debris in close proximity (within 1 km). How many LEO satellites do we need to sample to ensure that a 95% confidence interval estimate the proportion of RSOs in close proximity to active satellites.

Because we have no information on the proportion, we use 0.5 to estimate the sample size as shown in Figure 1.3. The outcome dictates that we sample 384 LEO satellites.

	A	B	C	D	E
6	Sample Size for One Sample, Dichotomous Outcome				
7	Confidence level	95%			
8	Proportion	0.5		n=	384
9	Margin of error E	0.05			
10					

Figure 1.3. Sample Size calculation for One Sample, Dichotomous Outcome

Table 1-2. Formula for cell E8.

Name	Cell	Formula
n	E8	=B8*(1-B8)*((NORM.S.INV(B7+(1-B7)/2))/B9)^2

### Sample Sizes for Two Independent Samples, Continuous Outcome

In studies where the plan is to estimate the difference in means between two independent populations, the formula for determining the sample sizes required in each comparison group is given below:

$$n_i = 2 \left( \frac{Z\sigma}{ES} \right)^2$$

where  $n_i$  is the sample size required in each group ( $i = 1,2$ ),  $Z$  is the value from the standard normal distribution reflecting the confidence level that will be used and  $ES$  is the desired margin of error.  $\sigma$  again reflects the standard deviation of the outcome variable.

### **Example 1.4: Radiation Degradation**

An investigator wants to plan an operational test to evaluate the minimization of radiation degradation of a new space-based sensor unit using (a) a new parts selection and (2) new shielding. The plan is to expose each mitigating factor to radiation in a vacuum., for 6 weeks. Based on prior experience with similar trials, the investigator expects that 10% of all components will be lost to Degradation or damage to surfaces and electronic components. A 95% confidence interval will be estimated to quantify the difference in mean radiation levels between new parts and new shielding. The investigator would like the margin of error to be no more than 3 units. How many units from each group should be selected for the evaluation? Figure 1.4 shows that we need 55 samples from each group or 121 total to account for attrition.

	A	B	C	D	E	F
11	<b>Sample Sizes for Two Independent Samples, Continuous Outcome</b>					
12	Confidence level	95%		n1=	55	
13	Estimated sigma	8		n2=	55	
14	Margin of error E	3.00		N=	121	
15	Attrition Rate	0.1				

*Figure 1.4. Sample Sizes for Two Independent Samples, Continuous Outcome*

**Table 1-3. Formulas for column E, rows 12, 13, and 14.**

Name	Cell	Formula
n1	E12	=2*((NORM.S.INV(B12+(1-B12)/2)*B13)/B14)^2
n2	E13	=2*((NORM.S.INV(B12+(1-B12)/2)*B13)/B14)^2
N	E14	=2*E12/(1-B15)

## Estimating Sample Size for Hypothesis Testing

There are common elements to each test. For example, in each test of a hypothesis, there are two errors that can be committed. The first is called a Type I error and refers to the situation where we incorrectly reject  $H_0$  when in fact it is true. In the first step of any test of a hypothesis, we select a level of significance,  $\alpha$ , and  $\alpha = P(\text{Type I error}) = P(\text{Reject } H_0 | H_0 \text{ is true})$ . Because we purposely select a small value for  $\alpha$ , we control the probability of committing a Type I error. The second type of error is called a Type II error and it is defined as the probability we do not reject  $H_0$  when it is false.

The probability of a Type II error is denoted  $\beta$ , and  $\beta = P(\text{Type II error}) = P(\text{Do not Reject } H_0 | H_0 \text{ is false})$ . In hypothesis testing, we usually focus on power, which is defined as the probability that we reject  $H_0$  when it is false, i.e.,  $\text{power} = 1 - \beta = P(\text{Reject } H_0 | H_0 \text{ is false})$ . Power is the probability that a test correctly rejects a false null hypothesis. A good test is one with low probability of committing a Type I error (i.e., small  $\alpha$ ) and high power (i.e., small  $\beta$ , high power).

Here we present formulas to determine the sample size required to ensure that a test has high power. The sample size computations depend on the level of significance,  $\alpha$ , the desired power of the test (equivalent to  $1 - \beta$ ), the variability of the outcome, and the effect size. The effect size is the difference in the parameter of interest that represents a clinically meaningful difference. Similar to the margin of error in confidence interval applications, the effect size is determined based on clinical or practical criteria and not statistical criteria.

The concept of statistical power can be difficult to grasp. Before presenting the formulas to determine the sample sizes required to

ensure high power in a test, we will first discuss power from a conceptual point of view.

Suppose we want to test the following hypotheses at  $\alpha = 0.05$ :  $H_0: \mu = 90$  versus  $H_1: \mu \neq 90$ . To test the hypotheses, suppose we select a sample of size  $n = 100$ . For this example, assume that the standard deviation of the outcome is  $\sigma = 20$ . We compute the sample mean and then must decide whether the sample mean provides evidence to support the alternative hypothesis or not. This is done by computing a test statistic and comparing the test statistic to an appropriate critical value. If the null hypothesis is true ( $\mu = 90$ ), then we are likely to select a sample whose mean is close in value to 90. However, it is also possible to select a sample whose mean is much larger or much smaller than 90. Recall from the Central Limit Theorem, that for large  $n$  (here  $n = 100$  is sufficiently large), the distribution of the sample means is approximately normal with a mean of

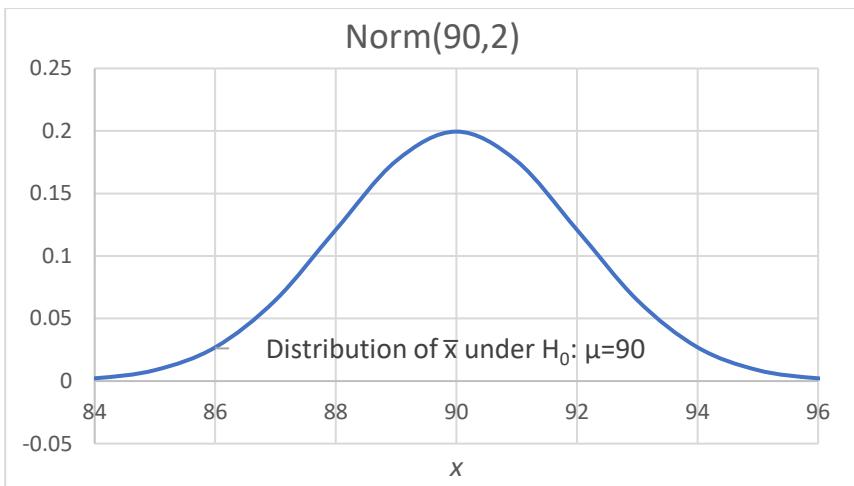
$$\mu_{\bar{X}} = \mu = 90$$

and

$$\text{Standard Deviation} = \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{100}} = 2.0$$

If the null hypothesis is true, it is possible to observe any sample mean shown in **Figure 1.5**; all are possible under  $H_0: \mu = 90$ .

When we set up the decision rule for our test of hypothesis, we determine critical values based on  $\alpha = 0.05$  and a two-sided test. When we run tests of hypotheses, we usually standardize the data (e.g., convert to  $Z$  or  $t$ ) and the critical values are appropriate values from the probability distribution used in the test.

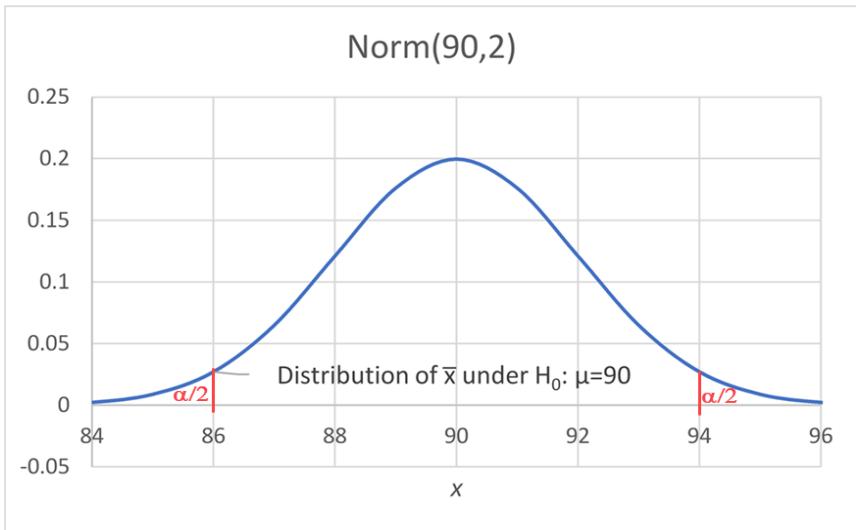


**Figure 1.5.** Distribution of  $\bar{x}$  under  $H_0: \mu=90$

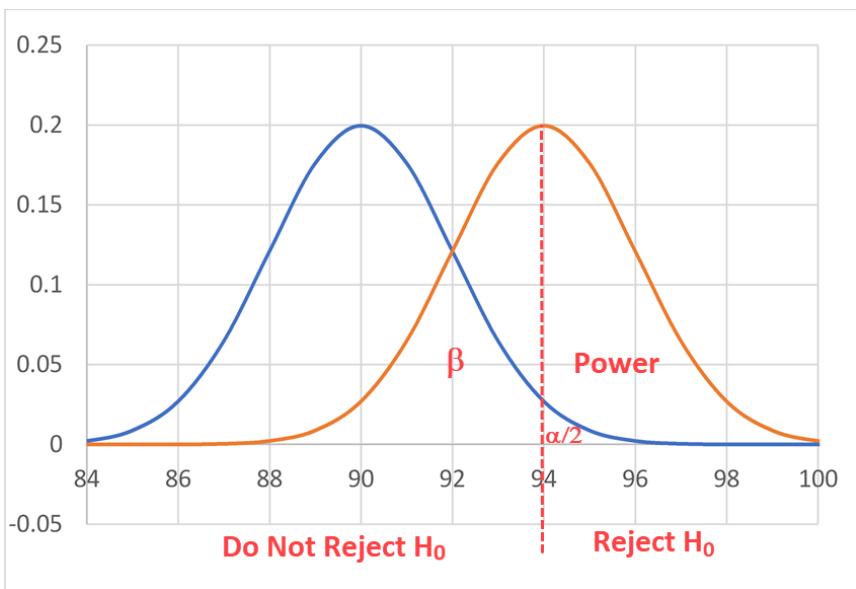
To facilitate interpretation, we will continue this discussion with  $\bar{x}$  as opposed to  $Z$ . The critical values for a two-sided test with  $\alpha = 0.05$  are 86.06 and 93.92 (these values correspond to -1.96 and 1.96, respectively, on the  $Z$  scale), so the decision rule is as follows: Reject  $H_0$  if  $\bar{x} < 86.06$  or if  $\bar{x} > 93.92$ . The rejection region is shown in **Figure 1.6**.

The areas in the two tails of the curve represent the probability of a Type I Error,  $\alpha = 0.05$ . Now, suppose that the alternative hypothesis,  $H_1$ , is true (i.e.,  $\mu \neq 90$ ) and that the true mean is actually 94. **Figure 1.7** shows the distributions of the sample mean under the null and alternative hypotheses. The values of the sample mean are shown along the horizontal axis.

As an aside, recall that we discussed some question that could be answered with confidence intervals. It is important to note that hypothesis testing builds off of confidence intervals (CIs). We also saw that Type I and Type II errors are key concepts. Moreover, that  $\alpha$  is the probability of making a Type I error and  $(1-\beta)$  is the probability of making a Type II error. However, when we speak of something like 95% confidence, this is not exactly a probability; it is a percentage. For example, when 95% confidence means that if we drew 100 samples from the population, 95 percent of the means would be captured by the CI.



**Figure 1.6.** Rejection Region for Test  $H_0: \mu = 90$  versus  $H_1: \mu \neq 90$  at  $\alpha = 0.05$



**Figure 1.7.** Distribution of  $\bar{x}$  under  $H_0: \mu = 90$  and under  $H_1: \mu = 98$ .

If the true mean is 94, then the alternative hypothesis is true. In our test, we selected  $\alpha = 0.05$  and reject  $H_0$  if the observed sample mean exceeds 93.92 (focusing on the upper tail of the rejection region for

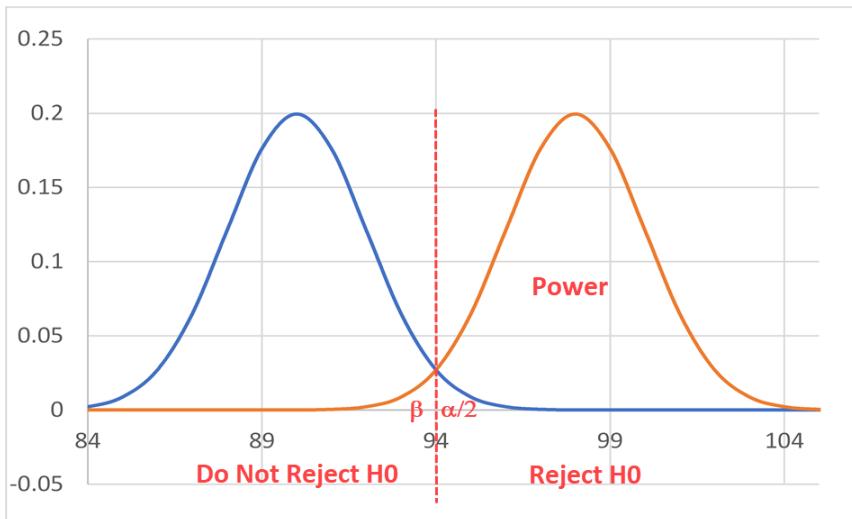
now). The critical value (93.92) is indicated by the vertical line. The probability of a Type II error is denoted  $\beta$ , and  $\beta = P(\text{Do not Reject } H_0 | H_0 \text{ is false})$ , i.e., the probability of not rejecting the null hypothesis if the null hypothesis were true.  $\beta$  is shown in Figure 1.7 above as the area under the rightmost curve ( $H_1$ ) to the left of the vertical line (where we do not reject  $H_0$ ). Power is defined as  $1 - \beta = P(\text{Reject } H_0 | H_0 \text{ is false})$  and is shown in the figure as the area under the rightmost curve ( $H_1$ ) to the right of the vertical line (where we reject  $H_0$ ).

Note that  $\beta$  and power are related to  $\alpha$ , the variability of the outcome and the effect size. From the **Figure 1.7** we can see what happens to  $\beta$  and power if we increase  $\alpha$ . Suppose, for example, we increase  $\alpha$  to  $\alpha = 0.10$ . The upper critical value would be 92.56 instead of 93.92. The vertical line would shift to the left, increasing  $\alpha$ , decreasing  $\beta$  and increasing power. While a better test is one with higher power, it is not advisable to increase  $\alpha$  as a means to increase power. Nonetheless, there is a direct relationship between  $\alpha$  and power (as  $\alpha$  increases, so does power).

$\beta$  and power are also related to the variability of the outcome and to the effect size. The effect size is the difference in the parameter of interest (e.g.,  $\mu$ ) that represents a clinically meaningful difference. The figure above graphically displays  $\alpha$ ,  $\beta$ , and power when the difference in the mean under the null as compared to the alternative hypothesis is 4 units (i.e., 90 versus 94). Figure 1.8 shows the same components for the situation where the mean under the alternative hypothesis is 98.

Notice that there is much higher power when there is a larger difference between the mean under  $H_0$  as compared to  $H_1$  (i.e., 90 versus 98). A statistical test is much more likely to reject the null hypothesis in favor of the alternative if the true mean is 98 than if the true mean is 94. Notice also in this case that there is little overlap in the distributions under the null and alternative hypotheses. If a sample mean of 97 or higher is observed it is very unlikely that it came from a distribution whose mean is 90. In the previous figure for  $H_0: \mu = 90$  and  $H_1: \mu = 94$ , if we observed a sample mean of 93, for example, it would not be as

clear as to whether it came from a distribution whose mean is 90 or one whose mean is 94.



**Figure 1.8. Distribution of under  $H_0: \mu = 90$  and under  $H_1: \mu = 94$**

### Ensuring That a Test Has High Power

In designing studies most people consider power of 80% or 90% (just as we generally use 95% as the confidence level for confidence interval estimates). The inputs for the sample size formulas include the desired power, the level of significance and the effect size. The effect size is selected to represent a clinically meaningful or practically important difference in the parameter of interest, as we will illustrate.

The formulas we present below produce the minimum sample size to ensure that the test of hypothesis will have a specified probability of rejecting the null hypothesis when it is false (i.e., a specified power). In planning studies, investigators again must account for attrition or loss to follow-up. The formulas shown below produce the number of participants needed with complete data, and we will illustrate how attrition is addressed in planning studies.

### Sample Size for Matched Samples, Continuous Outcome

In studies where the plan is to perform a test of hypothesis on the mean difference in a continuous outcome variable based on matched data, the hypotheses of interest are:

$$H_0: \mu_d = 0$$

Versus

$$H_1: \mu_d \neq 0$$

where  $\mu_d$  is the mean difference in the population. The formula for determining the sample size to ensure that the test has a specified power is given below:

$$n = \left( \frac{Z_{1-\alpha/2} + Z_{1-\beta}}{ES} \right)^2$$

where  $\alpha$  is the selected level of significance and  $Z_{1-\alpha/2}$  is the value from the standard normal distribution holding  $1 - \alpha/2$  below it,  $1 - \beta$  is the selected power and  $Z_{1-\beta}$  is the value from the standard normal distribution holding  $1 - \beta$  below it and  $ES$  is the effect size, defined as follows:

$$ES = \frac{\mu_d}{\sigma_d}$$

where  $\mu_d$  is the mean difference expected under the alternative hypothesis  $H_1$ , and  $\sigma_d$  is the standard deviation of the **difference** in the outcome (e.g., the difference based on measurements over time or the difference between matched pairs).

### Example 1.5: Pain Treatment and Computer Screens

We've been asked to evaluate the efficacy of an acupuncture treatment for reducing pain in patients with chronic migraine headaches due to increase hours in computer screen exposure, resulting in training on a new satellite control system. The plan is to enroll Guardians who suffer from migraine headaches. Each will be asked to rate the severity of the pain they experience with their next migraine before any treatment is administered. Pain will be recorded on a scale of 1-100 with higher scores indicative of more severe pain. Each Guardian will then undergo the acupuncture treatment. On their next migraine (post-treatment),

each Guardian will again be asked to rate the severity of the pain. The difference in pain will be computed for each Guardian. A two-sided test of hypothesis will be conducted, at  $\alpha = 0.05$ , to assess whether there is a statistically significant difference in pain scores before and after treatment. How many Guardians should be involved in the study to ensure that the test has 80% power to detect a difference of 10 units on the pain scale? Assume that the standard deviation in the difference scores is approximately 20 units.

First compute the effect size:

$$ES = \frac{\mu_d}{\sigma_d} = \frac{10}{20} = 0.50$$

Then substitute the effect size and the appropriate Z-values for the selected  $\alpha$  and power to compute the sample size, as shown in **Figure 1.9**.

	A	B	C	D	E
18	Sample Size for Matched Samples, Continuous Outcome				
19	Confidence level	95%		n=	32
20	sigma diff scores	80%			
21	Effect Size (ES)	0.5			

*Figure 1.9. Sample size estimate for a matched sample with continuous outcome.*

A sample of size  $n = 32^1$  Guardians with migraine will ensure that a two-sided test with  $\alpha = 0.05$  has 80% power to detect a mean difference of 10 points in pain before and after treatment, assuming that all 32 Guardians complete the treatment. **Table 1-4** provides the sample size formula.

*Table 1-4. Formula for sample size (cell E19) in Figure 1.9.*

Name	Cell	Formula
n	E19	=ROUNDUP(((ABS(NORM.S.INV((1-B19)/2)) +ABS(NORM.S.INV(1-B20))))/B21)^2,0)

---

<sup>1</sup> The calculated value of n is 31.4 but we've used a ROUNDUP (to 0 decimal places) function in Excel

We often do not know the parameters required to calculate ES. The effect size, ES, is more of a subjective evaluation than a statistical measure. There are generally three categories of effect size but the values assigned are not consistent. As a general rule of thumb, there are as follows:

- Small = 10%
- Medium = 30%
- Large = 50 %

These are often expressed as defects that are acceptable, like 4 out of 30.

### **Sample Sizes for Two Independent Samples, Dichotomous Outcomes**

In studies where the plan is to perform a test of hypothesis comparing the proportions of successes in two independent populations, the hypotheses of interest are:

$$H_0: p_1 = p_2 \text{ versus } H_1: p_1 \neq p_2$$

where  $p_1$  and  $p_2$  are the proportions in the two comparison populations. The formula for determining the sample sizes to ensure that the test has a specified power is given below:

$$n_i = 2 \left( \frac{Z_{1-\alpha/2} + Z_{1-\beta}}{EZ} \right)^2$$

where  $n_i$  is the sample size required in each group ( $i = 1,2$ ),  $\alpha$  is the selected level of significance and  $Z_{1-\alpha/2}$  is the value from the standard normal distribution holding  $1 - \alpha/2$  below it, and  $1 - \beta$  is the selected power and  $Z_{1-\beta}$  is the value from the standard normal distribution holding  $1 - \beta$  below it.  $ES$  is the effect size, defined as follows:

$$ES = \frac{|p_1 - p_2|}{\sqrt{p(1-p)}}$$

where  $|p_1 - p_2|$  is the absolute value of the difference in proportions between the two groups expected under the alternative hypothesis,  $H_1$ .

and  $p$  is the overall proportion, based on pooling the data from the two comparison groups ( $p$  can be computed by taking the mean of the proportions in the two comparison groups, assuming that the groups will be of approximately equal size).

### **Example 1.6: Testing for COVID-19 among Gym Users**

The Centers for Disease Control (CDC) hypothesizes that there is a higher incidence of COVID-19 among persons who use their fitness facilities more regularly than their counterparts who do not. SpOC has asked for a test to be conducted in the spring at the Peterson-Schriever Garrison to determine if the problem exists as noted by the CDC. Each Guardian using the Garrison gymnasiums will be asked if they used the facility regularly over the past 6 months and whether or not they had been diagnosed as having COVID-19. A test of hypothesis will be conducted to compare the proportion of Guardians who used the gymnasium facilities regularly and contracted COVID-19 with the proportion of Guardians who did not use the facilities and contracted COVID-19. During a previous year, approximately 35% of the Guardian experience COVID-19. The command believes that a 30% increase (margin of error) in the disease among those who used the gymnasium facilities regularly would be clinically meaningful. We need to determine how many Guardians should be enrolled in the test to ensure that the power of the test is 80% to detect this difference in the proportions. A two-sided test will be used with a 5% level of significance.

We first enter the Confidence level of 95% and the proportions of Guardians in each group who are expected to develop COVID-19,  $p_1 = 0.35$  and  $p_2 = 0.35$ . Next, we enter the power we want to achieve. Finally, we enter the 30% margin of error as shown in **Figure 1.10**.

Samples of size  $n_1 = 324$  and  $n_2 = 324$  will ensure that the test of hypothesis will have 80% power to detect a 30% difference in the proportions of Guardians who develop flu between those who do and do not use the athletic facilities regularly.

A	B	C	D	E	F
<b>23 Sample Sizes for Two Independent Samples, Dichotomous Outcome</b>					
24 Confidence level	95%		n=	324	
25 Proportion p1	35%		n1		
26 Proportion p2	35%		n2		
27 Power	80%	(Helper p)	0.35		
28 Margin of Error	30%				

**Figure 1.10.** Calculation of the sample size for the COVID-19 test.

**Table 1-5** provides the formulas used in for producing **Figure 1.10**.

**Table 1-5. Formulas for the sample size and helper<sup>2</sup> calculations in Figure 1.10.**

Name	Cell	Formula
n	E24	=2*((ABS(NORM.S.INV((1-B24)/2))+ABS(NORM.S.INV(1-B27)))/(ABS((B25*(1+B28))-B26)/SQRT(E27*(1-E27))))^2
(Helper p)	0.35	=(B25+B26)/2

## Sample Size for One Sample, Continuous Outcome

In studies where the plan is to perform a test of hypothesis comparing the mean of a continuous outcome variable in a single population to a known mean, the hypotheses of interest are:

$H_0: \mu = \mu_0$  and  $H_1: \mu \neq \mu_0$  where  $\mu_0$  is the known mean (e.g., a historical control). The formula for determining sample size to ensure that the test has a specified power is given below:

$$n = \left( \frac{Z_{1-\alpha/2} + Z_{1-\beta}}{ES} \right)^2$$

where  $\alpha$  is the selected level of significance and  $Z_{1-\alpha/2}$  is the value from the standard normal distribution holding  $1 - \alpha/2$  below it. For example, if  $\alpha = 0.05$ , then  $1 - \alpha/2 = 0.975$  and  $Z = 1.960$ .  $1 - \beta$  is the selected power, then  $Z_{1-\beta}$  is the value from the standard normal distribution

---

<sup>2</sup> Helper functions aid in calculation for complex formulas. In this case, we used the helper, p, for the total proportions under the radical in the ES calculation.

holding  $1 - \beta$ . Sample size estimates for hypothesis testing are often based on achieving 80% or 90% power. The  $Z_{1-\beta}$  values for these popular scenarios are given below:

- For 80% power  $Z_{0.80} = 0.84$
- For 90% power  $Z_{0.90} = 1.282$

$ES$  is the effect size, defined as follows:

$$\text{Effect Size} = EX = \frac{|\mu_1 - \mu_0|}{\sigma}$$

where  $\mu_0$  is the mean under  $H_0$ ,  $\mu_1$  is the mean under  $H_1$  and  $\sigma$  is the standard deviation of the outcome of interest. The numerator of the effect size, the absolute value of the difference in means  $|\mu_1 - \mu_0|$ , represents what is considered a clinically meaningful or practically important difference in means. Similar to the issue we faced when planning studies to estimate confidence intervals, it can sometimes be difficult to estimate the standard deviation. In sample size computations, investigators often use a value for the standard deviation from a previous study or a study performed in a different but comparable population. Regardless of how the estimate of the variability of the outcome is derived, it should always be conservative (i.e., as large as is reasonable), so that the resultant sample size will not be too small.

### **Example 1.7: Voice Command Translation**

A new software system for satellite maneuver control takes voice commands from trained operators and translates them into coded instructions received by the satellites to execute the maneuvers. An operational evaluation is established to test the hypothesis that the translation software. A cross-sectional study is planned to assess the mean error rate for persons whose native language is English. The mean error rate is about 5 out of 100 words with a standard deviation of 1.6 If the mean error rate for native English speakers is 20 out of 100 words, this would be important operationally. How many 100-word instructions should be sampled in the evaluation to ensure that the power of the test is 80% to detect this difference? A two-sided test will be used with a 5%

level of significance. **Figure 1.11** shows that a sample of 489 messages are required for this test.

	A	B	C	D	E
28	<b>Sample Size for One Sample, Continuous Outcome</b>				
29	Confidence level	95%		n=	489
30	Power Z_(1-B)	80%			
31	Effect Size	0.126667			
32	sigma	1.5			
33	$\mu_0$	0.01			
34	$\mu_1$	0.2			

**Figure 1.11.** Sample Size calculation for One Sample, Continuous Outcome, hypothesis test based

The formulas used to produce **Figure 1.11** are provided in **Table 1-6**.

**Table 1-6. Formulas for cells E29 and B31.**

Name	Cell	Formula
n	E29	=((NORM.S.INV(B29+(1-B29)/2)+NORM.S.INV(B30))/B31)^2
ES	B31	=ABS(B34-B33)/B32

### Sample Size for One Sample, Dichotomous Outcome

In studies where the plan is to perform a test of hypothesis comparing the proportion of successes in a dichotomous outcome variable in a single population to a known proportion, the hypotheses of interest are:

$$H_0: p = p_0$$

versus

$$H_1: p \neq p_0$$

where  $p_0$  is the known proportion (e.g., a historical control). The formula for determining the sample size to ensure that the test has a specified power is given below:

$$n = \left( \frac{Z_{1-\alpha/2} + Z_{1-\beta}}{ES} \right)^2$$

where  $\alpha$  is the selected level of significance and  $Z_{1-\alpha/2}$  is the value from the standard normal distribution holding  $1 - \alpha/2$  below it.  $1 - \beta$  is the selected power and  $Z_{1-\beta}$  is the value from the standard normal distribution holding  $1 - \beta$  below it, and  $ES$  is the effect size, defined as follows:

$$ES = \frac{p_1 - p_0}{\sqrt{p_0(1 - p_0)}}$$

where  $p_0$  is the proportion under  $H_0$  and  $p_1$  is the proportion under  $H_1$ . The numerator of the effect size, the absolute value of the difference in proportions  $|p_1 - p_0|$ , again represents what is considered a clinically meaningful or practically important difference in proportions.

### Example 1.8: Guardian LDL

A recent report from the Tricare Heart Study indicated that 26% of Guardians free of cardiovascular disease had elevated LDL cholesterol levels, defined as  $LDL > 159$  mg/dL. An investigator hypothesizes that a higher proportion of Guardians with a history of cardiovascular disease will have elevated LDL cholesterol. How many Guardians should be studied to ensure that the power of the test is 90% to detect a 5% difference in the proportion with elevated LDL cholesterol? A two-sided test will be used with a 5% level of significance. The inputs and outputs for this test is shown in **Figure 1.12**.

	A	B	C	D	E	F
38	<b>Sample Size for One Sample, Dichotomous Outcome</b>					
39	Confidence level	95%		n=	809	
40	Power Z_(1-B)	90%				
41	Effect Size					
42	Proportion p0	0.26				
43	Proportion p1	0.31 (this is 0.26 + the 5% detection difference)				

**Figure 1.12. Sample Size calculation for One Sample, Dichotomous Outcome, hypothesis test based**

So, 809 Guardians (USSF-wide) are required for this study for 95% confidence and 90% power to detect a 5% difference. **Table 1-7** provides the formulas used for our calculations in **Figure 1.12**.

**Table 1-7. Formulas for cells E37 and B39.**

Name	Cell	Formula
n	E37	=((NORM.S.INV(B37+(1-B37)/2) + NORM.S.INV(B38)) / ABS((B41-B40)/SQRT(B40*(1-B40))))^2
ES	B39	=(B41-B40)/SQRT(B40*(1-B40))

## Sample Sizes: Two Independent Samples, Continuous Outcome

In studies where the plan is to perform a test of hypothesis comparing the means of a continuous outcome variable in two independent populations, the hypotheses of interest are:

$$H_0: \mu_1 = \mu_2$$

Versus

$$H_1: \mu_1 \neq \mu_2$$

where  $\mu_1$  and  $\mu_2$  are the means in the two comparison populations. The formula for determining the sample sizes to ensure that the test has a specified power is:

$$n_i = 2 \left( \frac{Z_{1-\alpha/2} + Z_{1-\beta}}{ES} \right)^2$$

where  $n_i$  is the sample size required in each group ( $i = 1,2$ ),  $\alpha$  is the selected level of significance and  $Z_{1-\alpha/2}$  is the value from the standard normal distribution holding  $1 - \alpha/2$  below it, and  $1 - \beta$  is the selected power and  $Z_{1-\beta}$  is the value from the standard normal distribution holding  $1 - \beta$  below it.  $ES$  is the effect size, defined as:

$$ES = \frac{|\mu_1 - \mu_2|}{\sigma}$$

where  $|\mu_1 - \mu_2|$  is the absolute value of the difference in means between the two groups expected under the alternative hypothesis,  $H_1$ .  $\sigma$  is the standard deviation of the outcome of interest. Recall from the module on Hypothesis Testing that, when we performed tests of hypothesis comparing the means of two independent groups, we used  $S_p$ , the pooled estimate of the common standard deviation, as a measure of variability in the outcome.  $S_p$  is computed as follows:

$$S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}}$$

If data are available on variability of the outcome in each comparison group, then  $S_p$  can be computed and used to generate the sample sizes. However, it is more often the case that data on the variability of the outcome are available from only one group, usually the untreated (e.g., placebo control) or unexposed group. When planning a clinical trial to investigate a new drug or procedure, data are often available from other trials that may have involved a placebo or an active control group (i.e., a standard medication or treatment given for the condition under study). The standard deviation of the outcome variable measured in patients assigned to the placebo, control or unexposed group can be used to plan a future trial, as illustrated.

Also, the formula shown above generates sample size estimates for samples of equal size. If a study is planned where different numbers of patients will be assigned or different numbers of patients will comprise the comparison groups, then alternative formulas can be used.

### **Example 1.9: Water Displacing Formula**

Operational testers are planning an operational trial to evaluate the efficacy of a new formula designed to displace water from silo-based missiles. The plan is to randomly assign missiles to receive either the new water displacing formula or a control formula. Effectiveness will be measured in silos 12 weeks on the assigned treatment by cubic centimeters of water displacement. Based on prior experience with similar trials, the investigator expects that 5% of all silos to experience problems that cannot be controlled will be lost for recording consistent data and will drop out of the study. If the new formula shows a 10-unit reduction in water condensation, this would represent an operationally meaningful reduction. How many missiles should be sampled in the test to ensure that the power of the test is 80% to detect this difference? A two-sided test will be used with a 5% level of significance.

In order to compute the effect size, an estimate of the variability in water reduction is needed. Analysis of data from an earlier study showed that the standard deviation of water displacement was 19.0 cc. This value can be used to plan the test. The calculations are carried out as shown in **Figure 1.13**. The output shows that sample of 60 missiles are need to be randomly assigned to receive either the new formula or control.

	A	B	C	D	E	F
43	<b>Sample Sizes for Two Independent Samples, Continuous Outcome</b>					
44	Confidence level	95%		n=	57	
45	Power Z_(1-B)	80%		N=	60	
46	Effect Size	0.526316				
47	Sigma	19				
48	$\mu_1$	10				
49	$\mu_2$	n/a				
50	Attrition	5%				

*Figure 1.13. Sample Size calculation two Independent Samples, Continuous Outcome*

**Table 1-8** provides the formulas used to calculate the results in **Figure 1.13**.

*Table 1-8. Formulas for cells E44, E45, and B46.*

Name	Cell	Formula
n	E44	=2*((NORM.S.INV(B44+(1-B44)/2)+NORM.S.INV(B45)) /B46)^2
N	E45	=E44/(1-B50)
ES	B46	=B48/B47

### Sample Size Determination using t-Distribution

The estimation approach to determining sample size addresses the question: "How accurate do you want your estimate to be?" In this case, we are estimating the difference in means. This approach requires us to specify how large a difference we are interested in detecting, say  $B$  for the Bound on the margin of error, and then to specify how certain we want to be that we can detect a difference that large. Recall that when we assume equal sample sizes of  $n$ , a confidence interval for is given by:

$$\{ \bar{Y}_1 - \bar{Y}_2 \pm t(1 - \alpha; df) \cdot s \cdot \sqrt{2/n} \},$$

where  $n$  is the sample size for each group, and  $df = n + n - 2 = 2(n - 1)$  and  $s$  is the pooled standard deviation. Therefore, we first specify  $B$  and then solve this equation:

$$B = t \left(1 - \frac{\alpha}{2}; df\right) \cdot s \cdot \sqrt{\frac{2}{n}}$$

for  $n$ . Therefore,

$$n = \left[ t \left(1 - \frac{\alpha}{2}; df\right) \cdot s \cdot \frac{\sqrt{2}}{B} \right]^2 = \left[ \frac{t^2 \left(1 - \frac{\alpha}{2}; df\right) \cdot s^2 \cdot 2}{B^2} \right]$$

Since in practice, we don't know what  $s$  will be, prior to collecting the data, we will need a guesstimate to substitute into this equation. To do this by hand and we use  $z$  rather than  $t$ , since we don't know the  $df$  (because we don't know the sample size  $n$ ). The computer will iteratively update the  $df$ , as it computes the sample size, giving a slightly larger sample size when  $n$  is small.

So, we need to have an estimate of  $\sigma^2$ , a desired margin of error bound  $B$ , that we want to detect, and a confidence level  $1 - \alpha$ . With this, we can determine the sample size in this comparative type of experiment. We may or may not have direct control over  $\sigma^2$ , but by using different experimental designs we do have some control over this and we will address this later in this course. In most cases, an estimate of  $\sigma^2$  is needed in order to determine the sample size.

One special extension of this method is when we have a binomial situation. In this case where we are estimating proportions rather than some quantitative mean level, we know that the worst-case variance,  $p(1 - p)$ , is where  $p$  (the true proportion) is equal to 0.5 and then we would have an approximate sample size formula that is simpler, namely  $n = 2/B^2$  for  $\alpha = 0.05$ .

## Sample Size for t-test based on Confidence Interval

In Sample Size Requirements for t-tests, we show how to determine the minimum sample size for a one-sample *t-test* based on achieving a statistical power objective. There is another way to determine the required sample size, namely by achieving a sufficiently narrow confidence interval.

### Example 1.10: Sample Size with Goal Seek

Determine the sample size necessary to achieve a 95% confidence interval for the population mean that is no wider than .5 when the standard deviation of the data is 1.0.

In general,  $\text{CONFIDENCE.T}(\alpha, s, n) = k$  such that  $(\bar{x} - k, \bar{x} + k)$  is the confidence interval of the population mean. Thus, we seek the smallest value of  $n$  such that  $\text{CONFIDENCE.T}(0.05, 1, n) \leq .5/2$ . We can use repeated guessing to find this value of  $n$ , as shown in **Figure 1.14**.

	A	B	C	D	E	F	G	H	I	J
1	Sample size based on confidence interval									
2										
3	stdev	1								
4	alpha	0.05								
5	tails	2								
6										
7	n	20	40	60	62	63	64	65	70	80
8	k	0.46801	0.31982	0.25833	0.25395	0.25185	0.24979	0.24779	0.23844	0.22254

**Figure 1.14. Finding the sample size**

We see from the figure that a sample size of 64 is required. Table 1-9 provides the formula used to k in Figure 1.14.

**Table 1-9. Formulas for cells B8 to J8.**

Name	Cell(s)	Formula(s)
k	B8:J8	=CONFIDENCE.T(\$B\$4,\$B\$3,B7)

Here, we select cell B8 and copy it to the neighboring cells.

We can avoid the guessing by using Goal Seek. This is done by selecting Goal Seek from the Data ribbon and filling in the dialog box as shown in **Figure 1.15**.

	A	B	C	D	E	F	G	H	I
1									
2									
3	stdev	1							
4	alpha	0.05							
5	tails	2							
6									
7	n	20							
8	k	0.46801							
9									

Goal Seek

Set cell: \$B\$8  
To value: .20  
By changing cell: \$B\$7

OK Cancel

		64	65	70
		0.24979	0.24779	0.23844

**Figure 1.15. Finding the sample size using Goal Seek**

After clicking on the OK button, the values in cells B7 and B8 change as shown in **Figure 1.16**.

	A	B	C	D	E	F	G	H	I
1									
2									
3	stdev	1							
4	alpha	0.05							
5	tails	2							
6									
7	n	98.831							
8	k	0.20049							
9									

Goal Seek Status

Goal Seeking with Cell B8 found a solution.

Target value: 0.2  
Current value: 0.20049

Step Pause

OK Cancel

		65	70
		0.24779	0.23844

**Figure 1.16. Sample size results**

Rounding up, we conclude that the required sample size is 65.

Note that for a sample drawn from a finite population of size  $np$ , especially when the sample makes up a substantial portion of the population, the required sample size  $n$  needs to be adjusted to  $n'$  as follows:

$$n' = \frac{n}{1 + (n - 1)/np}$$

## Required Sample Size for Binomial Testing

We now show how to determine the sample size required to achieve a specified power objective.

### Example 1.11: Binomial Distributed Manufacturing Process

A company has made a major improvement in its manufacturing process. They want to test whether this improvement will result in 80% of the components passing their quality assurance requirements instead of 35%. What sample size do they need to achieve 90% power based on a one-tailed test with a significance level of  $\alpha = .01$ ?

We capture the situation on the left side of Figure 1 (where  $n = 0$  is an initial guess of the sample size) and use Excel's Goal Seek capability, which is accessed as described in Goal Seeking and Solver. We fill in the dialog box as shown on the right side of **Figure 1.17**.

#### *Goal Seek input*

A	B	C	D	E	F	G
1	Sample size requirements, one-tailed test					
2						
3	p0	0.35				
4	p1	0.8				
5	alpha	0.01				
6	x-crit	7 =BINOM.INV(B7,B3,1-B5)				
7	n	10				
8	1-β	0.6778 =1-BINOM.DIST(B6,B7,B4,TRUE)				

Goal Seek      ?      X

Set cell: \$B\$8  
To value: .9  
By changing cell: \$B\$7

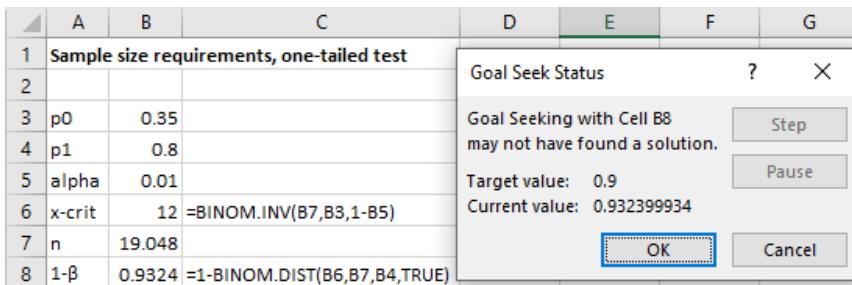
OK Cancel

**Figure 1.17.** Goal Seek set-up to find required sample size

We now press the OK button on the dialog box. The results are shown in **Figure 1.18**.

Generally, we would round up the value in cell range B7:B17, but as you can see 16.03 is pretty close to 16. Thus, we should try putting 16 in cell B7 to see if the value in cell B8 is still at least 0.90. It turns out that it is. So, a sample size of 16 is required.

## Goal Seek results



**Figure 1.18.** Goal Seek finds required sample size

Considering both of these scenarios, although there is no difference between the minimums and the maximums, the quantities  $\sum \tau_i^2$  are very different.

Of the two scenarios, the second is the least favorable configuration (LFC). It is the configuration of means for which you get the least power. The first scenario would be much more favorable. But generally, you do not know which situation you are in. The usual approach is not to guess exactly what all the values will be but simply to specify, which is the maximum difference between the true means.

## Determining Power

We begin this part by defining the power of a hypothesis test (see **Table 1-10**). This also provides another way of determining the sample size. The power is the probability of achieving the desired outcome. What is the desired outcome of a hypothesis test? Usually rejecting the null hypothesis. Therefore, power is the probability of rejecting the null hypothesis when in fact the alternative hypothesis is true.

**Table 1-10.** Hypotheses decisions shown with associated errors.

Decision	$H_0$	$H_A$
Reject Null Hypothesis	Type I Error – $\alpha$	OK
Accept Null Hypothesis	OK	Type II Error - $\beta$

Note:

$$P(\text{Reject } H_0 | H_0 \text{ is true}) = \alpha : P(\text{Type I Error})$$

$$P(\text{Accept } H_0 | H_A \text{ is true}) = \beta : P(\text{Type II Error})$$

Therefore, the power of the test is  $P(\text{Reject } H_0 | H_a \text{ is true}) = 1 - \beta$ .

Before any experiment is conducted you typically want to know how many observations you will need to run. If you are performing a study to test a hypothesis, for instance in the blood pressure example where we are measuring the efficacy of the blood pressure medication, if the drug is effective there should be a difference in the blood pressure before and after the medication. Therefore, we want to reject our null hypothesis, and thus we want the power (i.e., the probability of rejecting  $H_0$  when it is false) to be as high as possible.

We will describe an approach to determine the power, based on a set of operating characteristic curves traditionally used in determining power for the  $t$ -test. Power depends on the level of the test,  $\alpha$ , the actual true difference in means, and  $n$  (the sample size). Figure 2.13 (2.12 in 7th ed) in the text gives the operating characteristic curves where  $\beta$  is calculated for  $n *= 2n - 1$  for an  $\alpha = 0.05$  level test. When you design a study, you usually plan for equal sample size, since this gives the highest power in your results. We will look at special cases where you might deviate from this but generally, this is the case.

To use the Figure in the text, we need to first calculate the difference in means measured in numbers of standard deviation, i.e.,  $|\mu_1 - \mu_2|/\sigma$ . You can think of this as a signal to noise ratio, i.e., how large or strong is the signal,  $|\mu_1 - \mu_2|$ , in relation to the variation in the measurements,  $\sigma$ . We are not using the symbols in the text, because the 2 editions define  $d$  and  $\delta$  differently. Different software packages or operating characteristic curves may require either  $|\mu_1 - \mu_2|/\sigma$  or  $|\mu_1 - \mu_2|/2\sigma$  to compute sample sizes or estimate power, so you need to be careful in reading the documentation.

### Example 1.12: Power Calculations

Let's consider an example in the two-sample situation. We will let  $\alpha = 0.05$ ,  $|\mu_1 - \mu_2| = 8$  (the difference between the two means), and the

sigma (assumed true standard deviation) would equal 12, and finally, let the number of observations in each group  $n = 5$ .

In this case,  $|\mu_1 - \mu_2|/\sigma = 8/12 = 0.66$ , and  $n^* = 2n - 1 = 9$ .

If you look at the Figure you get approximately a  $\beta$  of about 0.9. Therefore, power - or the chance of rejecting the null hypothesis prior to doing the experiment is  $1 - \beta$  or  $1 - 0.9 = 0.1$  about ten percent of the time. With such low power we should not even do the experiment!

If we were willing to do a study that would only detect a true difference of, let's say,  $|\mu_1 - \mu_2| = 18$  then and  $n^*$  would still equal 9, then figure 2-12 the Figure shows that looks to be about 0.5 and the power or chance of detecting a difference of 18 is also 5. This is still not very satisfactory since we only have a 50/50 chance of detecting a true difference of 18 even if it exists.

Finally, we calculate the power to detect this difference of 18 if we were to use  $n = 10$  observations per group, which gives us  $n^* = 19$ . For this case  $\beta = 0.1$  and thus  $power = 1 - \beta = 0.9$  or 90%, which is quite satisfactory.

Another way to improve power is to use a more efficient procedure - for example, if we have paired observations we could use a paired *t*-test. For instance, if we used the paired *t*-test, then we would expect to have a much smaller sigma – perhaps somewhere around 2 rather than 12. So, our signal to noise ratio would be larger because the noise component is smaller. We do pay a small price in doing this because our *t*-test would now have degrees of freedom  $n - 1$ , instead of  $2n - 2$ .

The take-home message here is: If you can reduce variance or noise, then you can achieve an incredible savings in the number of observations you have to collect. Therefore, the benefit of a good design is to get a lot more power for the same cost or much-decreased cost for the same power.

You can use these dialog boxes to plug in the values that you have assumed and have Excel calculate the sample size for a specified power or the power that would result, for a given sample size.

## Techniques for Comparing Means

There are many cases in statistics where you'll want to compare means for two populations or samples. Which technique you use depends on what type of data you have and how that data is grouped together.

The four major ways of comparing means from data that is assumed to be normally distributed are:

1. **Independent Samples T-Test.** Use the independent samples t-test when you want to compare means for two data sets that are independent from each other. There are actually two variants of this:
  - a. Assuming equal variances
  - b. Assuming unequal variances

However, we can really use the test with unequal variances for both cases.

2. **One sample T-Test.** Choose this when you want to compare means between one data set and a specified constant (like the mean from a hypothetical normal distribution). [Click here](#) for a step-by-step article.
3. **Paired Samples T-Test.** Use this test if you have one group tested at two different times. In other words, you have two measurements on the same item, person, or thing. The groups are “paired” because their intrinsic connections between them (i.e., they are not independent). This comparison of means is often used for groups of patients before treatment and after treatment, or for trainees tested before remediation and after remediation. [Click here](#) for a step-by-step article.
4. **One way Analysis of Variance (ANOVA).** Although not really a test for comparison of means, ANOVA is the main option when you have more than two levels of independent variable. For example, if your independent variable was “brand of coffee” your levels might be ~~Starbucks (it's not really coffee)~~, Seattle’s Best, Peets and Trader Joe’s. Use this test when you have a group of individuals randomly split into smaller groups and completing different tasks (like drinking different coffee, or using different software for maintaining the Space Catalog).

## Non-Normal Data

If you have non-normal data (or if you don't know what distribution your data comes from), you can't use any of the above tests for comparison of means. You must use a non-parametric test (non-parametric basically means that you don't know the distribution's parameters):

1. For independent samples, use the **Mann-Whitney U test**. This test is essentially the same as the t-test for independent samples.
2. For paired groups, use **Wilcoxon Signed Rank test**. This test compares medians, not means.

## Testing: The two-sample t-test

For the two-sample  $t$ -test, both samples are assumed to come from Normal populations with (possibly different) means  $\mu_i$  and variances  $\sigma^2$ . When the variances are not equal, we will generally try to overcome this by transforming the data. Using a metric where the variation is equal, or we can use complex ANOVA models, which also assume equal variances. (There is a version of the two-sample t-test which can handle different variances, but unfortunately, this does not extend to more complex ANOVA models.) We want to test the hypothesis that the means  $\mu_i$  are equal.

We estimate the mean and the sample variance using formulas:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \text{ and } s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

We divide by  $n - 1$  so we can get an unbiased estimate of  $\sigma^2$ . These are the summary statistics for the two-sample problem. We basically need to know the sample size,  $n$ , the sample mean, and the sample standard deviation (or the variance) for each sample, and these will be sufficient for performing statistical inference. However, it is dangerous to not look at the data and only look at the summary statistics because these summary statistics do not tell you anything about the shape or distribution of the data or about potential outliers, both things you'd want to know about to determine if the assumptions are satisfied.

The two-sample t-test is basically looking at the difference between the sample means relative to the standard deviation of the difference of the sample means. Engineers would express this as a signal-to-noise ratio (SNR) for the difference between the two groups.

If the underlying distributions are normal then the z-statistic is the difference between the sample means divided by the true population variance of the sample means. Of course, if we do not know the true variances—we have to estimate them. Therefore, we use the *t*-distribution and substitute sample quantities for population quantities, which is something we do frequently in statistics. This ratio is an approximate z-statistic—Gosset published the exact distribution under the pseudonym "Student" and the test is often called the "Student *t*" test. If we can assume that the variances are equal, an assumption we will make whenever possible, then we can pool or combine the two sample variances to get the pooled standard deviation shown below.

Our pooled statistic is the pooled standard deviation  $s_p$  times the square root of the sum of the inverses of the two sample sizes. The *t*-statistic is a signal-to-noise ratio, a measure of how far apart the means are for determining if they are really different.

So, we ask does the data provide evidence that the true means differ, or does  $\mu_1 = \mu_2$ ?

However, we need to know if the variances are equal, and we do that with an F-test. Here, we test:

$$H_0: s_1^2 = s_2^2$$

We will now calculate the F-test statistic:

$$F = \frac{s_1^2}{s_2^2}$$

If we see that the F-statistic is greater than the F critical value then we conclude that variances are not equal. Consequently, we have to use a *t* statistic that uses pooled standard deviations,  $S_p$ , for testing

$$H_0: \mu_1 = \mu_2$$

**Definition 1-6: Two-Sample t-Test for Sample Means with Equal Variances**

The 2-sample t-test statistic for equal variances is given by:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

This is always a relative question. Are they different relative to the variation within the groups? Perhaps, they look a bit different. Our  $t$ -statistic turns out to be 8.3383. If you know the  $t$ -distribution, you should then know that this is a borderline value and therefore requires that we examine carefully whether these two samples are really far apart.

We compare the sample  $t$  to the distribution with the appropriate  $df$ . We typically will calculate just the  $p$ -value which is the probability of finding the value at least as extreme as the one in our sample. This is under the assumption of the null hypothesis that our means are equal. The  $p$ -value in our example is essentially zero, 1.17598E-15, as shown in the Excel output below.

$H_0$ : There is not difference in the average city vs highway mpg

If the test result is significant, then the alternative to reject the null hypothesis.

$H_A$ : There is a significant difference in city vs highway mpg

So, using the  $p$ -value, we would reject the null hypothesis and conclude that the means are not equal.

Confidence intervals involve finding an interval, in this case, the interval is about the difference in means. We want to find upper and lower limits that include the true difference in the means with a specified level of confidence, typically we will use 95%.

In the cases where we have a two-sided hypothesis test which rejects the null hypothesis, then the confidence interval will not contain 0. In our example above we can see in the Minitab output that the 95% confidence interval does not include the value 0, the hypothesized value

for the difference, when the null hypothesis assumes the two means are equal.

## How to Construct a Probability Plot with Excel

**Step 1.** Open a data set in Excel, with values starting in cell A2 (cell A1 should contain a heading).

**Step 2.** Starting in cell B2, calculate a Z-value for each data element in Column A, using:

```
=NORM.S.INV((RANK(A2, $A$2:$A$16, 1)-0.5)/COUNT(A:A))
```

**Step 3:** Create a scatterplot comprised of the values in columns A and B, using **Insert| Charts | Scatter | Scatter**.

## Two-Sample t-Test for Equal Means

The two-sample t-test (Snedecor and Cochran, 1989) is used to determine if two population means are equal. A common application is to test if a new process or treatment is superior to a current process or treatment. There are several variations on this test.

The data may either be paired or not paired. By paired, we mean that there is a one-to-one correspondence between the values in the two samples. That is, if  $X_1, X_2, \dots, X_n$  and  $Y_1, Y_2, \dots, Y_n$  are the two samples, then  $X_i$  corresponds to  $Y_i$ . For paired samples, the difference  $X_i - Y_i$  is usually calculated. For unpaired samples, the sample sizes for the two samples may or may not be equal. The formulas for paired data are somewhat simpler than the formulas for unpaired data.

The variances of the two samples may be assumed to be equal or unequal. Equal variances yield somewhat simpler formulas, although with computers this is no longer a significant issue.

In some applications, you may want to adopt a new process or treatment only if it exceeds the current treatment by some threshold. In this case, we can state the null hypothesis in the form that the difference between the two populations means is equal to some constant  $\mu_1 - \mu_2 = d_0$  where the constant is the desired threshold.

**Definition 1-7: Generalized Two-Sample t-Test**

The two-sample t-test for the sample means with unequal variances is defined as:

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

**Test Statistic:**

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

where  $N_1$  and  $N_2$  are the sample sizes,  $\bar{Y}_1$  and  $\bar{Y}_2$  are the sample means, and  $s_1^2$  and  $s_2^2$  are the sample variances.

If equal variances are assumed, then the formula reduces to:

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{sp \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

where

$$sp^2 = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}$$

**Significance Level:**  $\alpha$ .

**Critical Region:** Reject the null hypothesis that the two means are equal if

$$|T| > t_{1-\frac{\alpha}{2}, v}$$

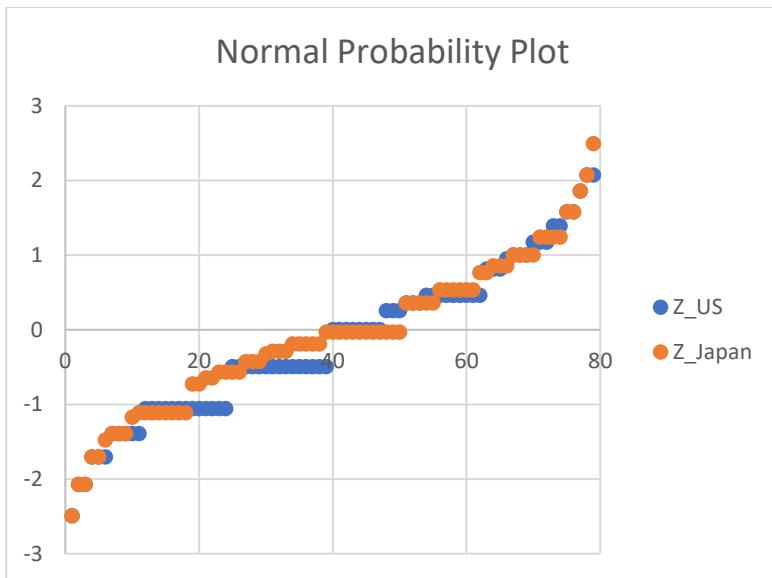
where  $t_{1-\frac{\alpha}{2}, v}$  is the critical value of the  $t$  distribution with  $v$  degrees of freedom. The degrees of freedom are given by:

$$v = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\left(\left(\frac{s_1^2}{N_1}\right)^2 / (N_1 - 1)\right) + \left(\left(\frac{s_2^2}{N_2}\right)^2 / (N_2 - 1)\right)}$$

If equal variances are assumed, then  $v = N_1 + N_2 - 2$ .

### Example 1.13: Two-Sample t-Test, Equal Variances

A diagnostic to check whether we can perform this test is a normal probability plot. The normal probability plots in **Figure 1.19** look reasonable, so we can say that the distributions are approximately normal.



*Figure 1.19. Normal probability plot for Mpg\_US and Mpg\_Japan.*

Confidence intervals involve finding an interval, in this case, the interval is about the difference in means. We want to find upper and lower limits that include the true difference in the means with a specified level of confidence, typically we will use 95%.

The data set AUTO83B.DAT contains miles per gallon for U.S. cars (sample 1) and for Japanese cars (sample 2). The data was used as a Bivariate Analysis of 1983 EPA Auto MPG (for autos 1970 to 1982).

Note that the -999 for the Japan data indicates the data should be deleted, and this is not a paired comparison.

The summary statistics for each sample are shown in **Table 1-11**.

**Table 1-11. Summary statistics for the Mpg\_US and Mph\_Japan data.**

	Mpg_US	Mpg_Japan
Mean	20.14458	30.48101
Variance	41.14837	37.30412
Observations	249	79

We are testing the hypothesis that the population means are equal for the two samples. We do not assume that the variances for the two samples are equal. Instead, we perform a test to check for equal variance. This is done using a F-Test in **Table 1-11**. The null hypothesis is:

$$H_0: S_1^2 = S_2^2$$

The result shows the F-statistic is less than the critical values (see **Table 1-12**), so we fail to reject the null hypothesis.

**Table 1-12. F-Test Two-Sample for Variances**

	Mpg_US	Mpg_Japan
Mean	20.14	30.48
n	249	79
Variance	41.1483677	37.30412204
df	248	78
F	1.10305150	
P(F<=f) one-tail	0.30968188	Fail to reject Ho
F Critical one-tail	1.37347056	Fail to reject Ho

The formulas used for calculating the test for equal variances are provided in **Table 1-13**.

**Table 1-13. Formulas for the F-Test Two-Sample for Variances**

	Mpg_US	Mpg_Japan
Mean	=AVERAGE(A5:A253)	=AVERAGE(B5:B83)
n	=COUNT(A5:A253)	=COUNT(B5:B83)
Variance	=VAR.S(A5:A253)	=VAR.S(B5:B83)
df	=E11-1	=F11-1
F	=E12/F12	
P(F<=f) one-tail	=F.TEST(A5:A253,B5:B83)/2	=IF(E15<0.05,"Reject Ho", "Fail to reject Ho")

<b>F Critical one-tail</b>	=F.INV.RT(0.05,E13,F13)
----------------------------	-------------------------

Since we failed to reject the hypothesis that the variances are not different, we proceed with a t-test for the equality of means given equal variances (see **Table 1-14**). The null and alternative hypotheses are:

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

**Table 1-14. t-Test: Two-Sample Assuming Equal Variances**

	Mpg_US	Mpg_Japan
Mean	20.14	30.48
n	249	79
Variance	41.14836766	37.304122
df	326	
Hypothesized Mean Difference	0	
t Stat	12.94627327	
P(T<=t) one-tail	2.63647E-30	Reject Ho
t Critical one-tail	1.649541157	Reject Ho
P(T<=t) two-tail	5.27294E-30	Reject Ho
t Critical two-tail	1.967267522	Reject Ho

The absolute value of the test statistic for our example, 12.9463, is greater than the critical value of 1.97743, so we reject the null hypothesis and conclude that the population means are different at the 0.05 significance level.

In general, there are three possible alternative hypotheses and rejection regions for the one-sample t-test, as shown in **Table 1-15**:

**Table 1-15. Table of Alternative hypotheses with associated rejection regions**

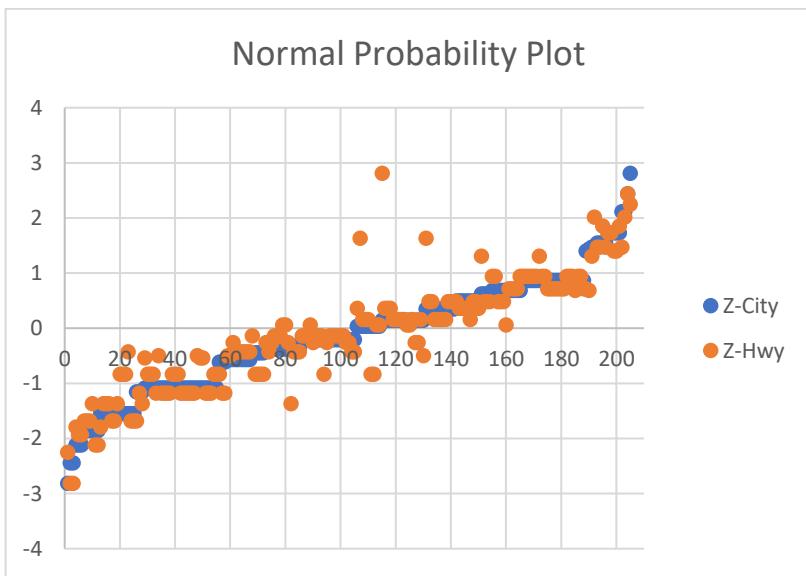
Alternative Hypothesis	Rejection Region
$H_a: \mu_1 \neq \mu_2$	$ T  > t_{1-\frac{\alpha}{2},v}$
$H_a: \mu_1 > \mu_2$	$T > t_{1-\alpha,v}$
$H_a: \mu_1 < \mu_2$	$T < t_{\alpha,v}$

For our two-tailed t-test, the critical value is  $t_{1-\frac{\alpha}{2}, \nu} = 1.97743$ , where  $\alpha = 0.05$  and  $\nu = 328$ . If we were to perform an upper, one-tailed test, the critical value would be  $t_{1-\alpha, \nu} = 1.65605$ . The rejection regions for three possible alternative hypotheses using our example data are shown below.

### Example 1.14: Two-Sample t-Test, Unequal Variances

The data set `Automobile_data` contains city and highway miles per gallon for 205 automobile makes and models. We want to test to see if there is a statistically significant difference between the city and highway miles per gallon.. The data was used as a Bivariate Analysis of 1983 EPA Auto MPG (for autos 1970 to 1982).

Again, we check whether we can perform this test is a normal probability plot. The normal probability plots in **Figure 1.20** look reasonable, so we can say that the distributions are approximately normal.



**Figure 1.20.** Normal probability plot for City and Highway MPG.

The summary statistics for each sample are shown in **Table 1-16**.

**Table 1-16. Summary statistics for the Mpg\_US and Mph\_Japan data.**

	Mpg_US	Mpg_Japan
Mean	25.22	30.75
Variance	42.80	47.42
Observations	205.00	205.00

We are testing the hypothesis that the population means are equal for the two samples. We do not assume that the variances for the two samples are equal. Instead, we perform a test to check for equal variance. This is done using a F-Test in **Table 1-17**. The null hypothesis is:

$$H_0: S_1^2 = S_2^2$$

The result shows the F-statistic is greater than the critical values (see **Table 1-17**), so we reject the null hypothesis.

**Table 1-17. F-Test Two-Sample for Variances**

	City	Highway
Mean	24.80	30.77
Variance	35.57	47.85
Observations	205.00	205.00
df	204.00	204.00
F	1.34551244	Reject Ho
P(F<=f) one-tail	0.017302908	Reject Ho
F Critical one-tail	0.793847997	

Since we reject the hypothesis that the variances are not different, we proceed with a t-test with pooled variance (unequal variances) for the equality of means (see **Table 1-18**). The null and alternative hypotheses are:

$$H_0: \mu_1 = \mu_2$$

And

$$H_A: \mu_1 \neq \mu_2$$

The pooled variance is:

$$sp^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 + n_2 - 2)}$$

$$sp^2 = 41.7099$$

**Table 1-18. T-test for equal city and highway mpg means with unequal variances.**

	City	Highway
Mean	24.80	30.77
n	205	205
Variance	35.56566236	47.85404113
Pooled Var	41.70985175	
df	204	204
Hypothesized Mean Difference	0	
t Stat	9.34464894	
P(T<=t) one-tail	2.55572E-19	Reject Ho
t Critical one-tail	1.648596901	Reject Ho
P(T<=t) two-tail	5.11144E-19	Reject Ho
t Critical two-tail	1.971660889	Reject Ho

The absolute value of the test statistic for our example, 9.3446, is greater than the critical value of 1.97166, so we reject the null hypothesis and conclude that the population means are different at the 0.05 significance level.

## Paired t-Test

In Paired T-Test, they compare the means of two groups of observations. The observations must be randomly assigned to each of the two groups so that the difference in response seen is due to the treatment and not because of any other factors. If two samples are given, then the observation of one sample can be paired with the observation of the other sample. This test can be used in making observations on the same sample before and after an event. Now, let us discuss what is paired t-test, its formula, table and the procedure to perform the paired t-test in detail.

The paired t-test gives a hypothesis examination of the difference between population means for a set of random samples whose variations are almost normally distributed. Subjects are often tested in a before-after situation or with subjects as alike as possible. The paired t-test is a test that the differences between the two observations are zero.

Let's assume two paired sets, such as  $X_i$  and  $Y_i$  for  $i = 1, 2, \dots, n$  such that their paired-difference are independent which are identically and normally distributed (iid). Then the paired t-test concludes whether they notably vary from each other.

**Definition. Paired T-test** is a test based on the differences between the values of a single pair, that is one deducted from the other.

$$t - Stat = \frac{\sum d}{\sqrt{\frac{n \sum d^2 - (\sum d)^2}{n - 1}}}$$

where  $d$  is the difference between pairs, and  $n$  is the number of pairs.

### Example 1.15: Paired Processor Test Samples

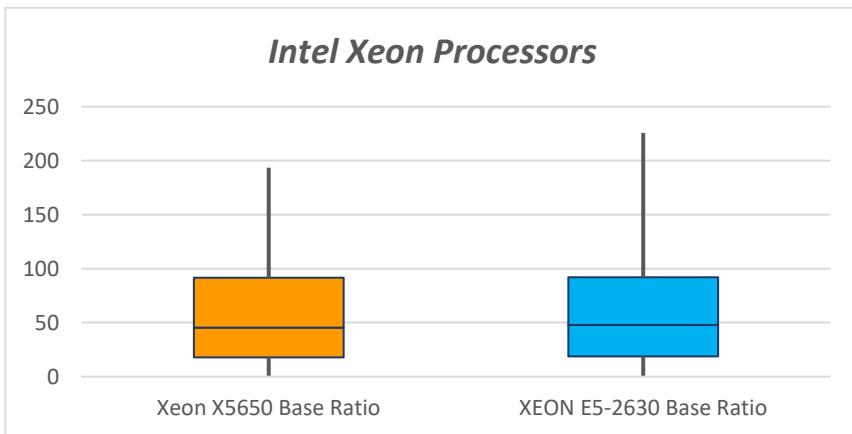
In this example, we are interested in replacing a chip in an onboard processor for a PNT satellite currently in production, if the chip significantly improves the processor performance. The manufacturer of Intel Xeon chips claims a significant performance difference between Intel Xeon X5650, 3.06 GHz and Intel Xeon E5-2630 v3, 2.80 GHz, which is the chip currently used. We have paired data for the Base Run Time for paired Benchmarks and Base Reference Time, partly shown in **Figure 1.21**.

To start, we construct boxplots of the data to get a graphical representation, shown in **Figure 1.22**.

The paired  $t$ -test is our first example of a blocking design. In this context, the benchmark is used as a *block*, and the results from the paired  $t$ -test are identical to what we will find when we analyze this as a Randomize Complete Block Design.

	A	B	C	D
3	Benchmark	Base Ref Time	Xeon X5650 Base Ratio	XEON E5-2630 Base Ratio
4	410.bwaves	13590	172.718497	284.096836
5	410.bwaves	13590	148.874665	281.741415
6	410.bwaves	13590	153.225873	282.915843
7	416.gamess	19580	21.815368	24.498048
8	416.gamess	19580	21.740361	24.497887
9	416.gamess	19580	21.79301	24.639893
10	433.milc	9180	49.574538	44.437179
11	433.milc	9180	49.594733	44.649153
12	433.milc	9180	49.657526	44.620707
13	434.zeusmp	9100	83.282103	118.742068
14	434.zeusmp	9100	107.504325	118.743665
15	434.zeusmp	9100	105.992229	118.744259
16	435.gromacs	7140	20.180133	26.31455
17	435.gromacs	7140	20.191496	26.451011
18	435.gromacs	7140	20.100497	26.529716
19	436.cactusADM	11950	265.914138	299.175809
20	436.cactusADM	11950	244.140156	296.278004
21	436.cactusADM	11950	226.560174	300.747649

*Figure 1.21. Partial data view of the benchmarked base ratio data for two intel chips, the Xeon X5650 (new) and ES-2630 (current).*



*Figure 1.22. Boxplots for Intel processors*

We will perform a paired t-test under the null hypothesis:

$H_0$ : The base ratio of run times for two Intel processors are the same

The alternative hypothesis is:

$H_A$ : The base ratio of run times for two Intel processors are different

To implement the Paired t-Test, given by our definition

$$t - Stat = \frac{\sum d}{\sqrt{\frac{n \sum d^2 - (\sum d)^2}{n - 1}}}$$

We need to calculate the difference,  $d$ , between pairs and the squared differences,  $d^2$ , between pairs. We show this in **Figure 1.23**.

	F	G	H	I
3	$diff(X5650, E5-2630)$	$diff(X5650, E5-2630)^2$	<i>Paired-T Test Differences</i>	<i>X5650 vs. E5-2630</i>
4	-111.378339	12405.1344	$\sum(dif)$	1145.136511
5	-132.86675	17653.57326	$\sum(dif^2)$	87992.92313
6	-129.68997	16819.48832	$\sum(dif)^2$	1311337.629
7	-2.68268	7.196771982		
8	-2.757526	7.603949641		
9	-2.846883	8.104742816		
10	5.137359	26.39245749		
11	4.94558	24.45876154		
12	5.036819	25.36954564		
13	-35.459965	1257.409118		
14	-11.23934	126.3227636		
15	-12.75203	162.6142691		

**Figure 1.23.** Differences and squared differences for the X5650 vs. ES-2630 chip data.

Note that our test statistic does not have any reference to variance, but depends only on  $d$  and  $n$ . Also, note that  $\Sigma$  (upper-case sigma) is the notation for the summation operator, or the Excel function =SUM().

Also, note the distinction in the definition between  $\sum d^2$  and  $\sum(d)^2$ , in the former, the  $ds$  are squared and then added, and in the latter, the  $ds$  are added and then squared. **Table 1-19** provides the results of the Paired t-Test, calculated with built-in Excel function only (no add ins).

Also, we'll perform the test at the 0.05 level of significance. The formulas used for these calculations are given in **Table 1-20**.

**Table 1-19. t-Test: Paired Two Sample for Means**

	Xeon X5650 Base Ratio	XEON E5-2630 Base Ratio
Mean	74.46324445	96.91690153
Variance	5190.566858	9531.995747
Observations	51	51
Pooled Variance (not used)	7361.281303	102
Hypothesized Mean Difference	0	
df	50	
t Stat	4.543404538	
P(T<=t) one-tail	1.76056E-05	Reject Ho
t Critical one-tail	1.675905025	Reject Ho
P(T<=t) two-tail	3.52111E-05	Reject Ho
t Critical two-tail	2.008559112	Reject Ho

Since the Test Statistic is greater than the Critical Value, we can reject the null hypothesis. Alternatively, we can compare the p-value with the level of significance of the test ( $\alpha=0.05$ ) and since it is less than the 0.05, we can reject the null hypothesis. It is tempting to look at a boxplot and say that the means do not look very different, but since we are talking about micro-processor speeds, minor differences can be major. Hence, we should always check for statistical significance. Examples of drawing incorrect conclusions from charts appears in media articles more frequent than we might expect.

The workbook also shows two other Paired t-Test using Excel's Data Analysis add-in. These tests compare Xeon X5650, 3.06 GHz vs. Intel Xeon E5-2620, 2.40 GHz, and Intel Xeon E5-2630, 2.80 vs. Intel Xeon E5-2620, 2.40. We do not show those results here.

**Table 1-20. Formulas used to calculate the Paired t-Test for computer chips**

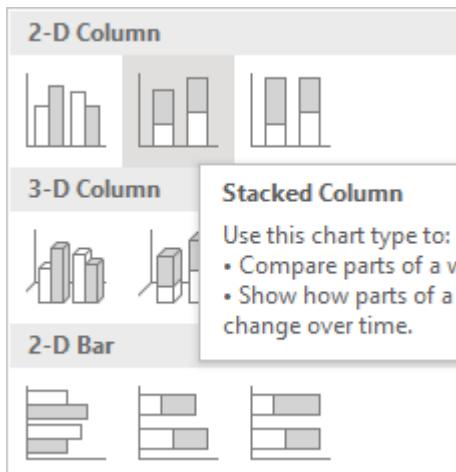
	<i>Xeon X5650 Base Ratio</i>	<i>XEON E5-2630 Base Ratio</i>
<b>Mean</b>	=AVERAGE(C\$4:C\$54)	=AVERAGE(D\$4:D\$54)
<b>Variance</b>	=VAR(C\$4:C\$54)	=VAR(D\$4:D\$54)
<b>Observations</b>	=COUNT(C\$4:C\$54)	=COUNT(D\$4:D\$54)
<b>Pooled Variance (not used)</b>	=((L9-1)*L8 + (M9-1)*M8) / (L9+M9-2)	=L9+M9
<b>Hypothesized Mean Difference</b>	0	
<b>df</b>	=L9-1	
<b>t Stat</b>	=I4/SQRT((L9*I5-I6)/(L9-1))	
<b>P(T&lt;=t) one-tail</b>	=T.TEST(C4:C54,D4:D54 ,1,1)	=IF(L14<0.05,"Reject Ho","Fail to reject Ho")
<b>t Critical one-tail</b>	=-T.INV(0.05,L12)	=IF(L13>L15,"Reject Ho","Fail to reject Ho")
<b>P(T&lt;=t) two-tail</b>	=T.TEST(C4:C54,D4:D54 ,2,1)	=IF(L16<0.05,"Reject Ho","Fail to reject Ho")
<b>t Critical two-tail</b>	=T.INV.2T(0.05,L12)	=IF(L13>L17,"Reject Ho","Fail to reject Ho")

## How to Create a Boxplot with Excel

**Step 1:** Calculate the following fields:

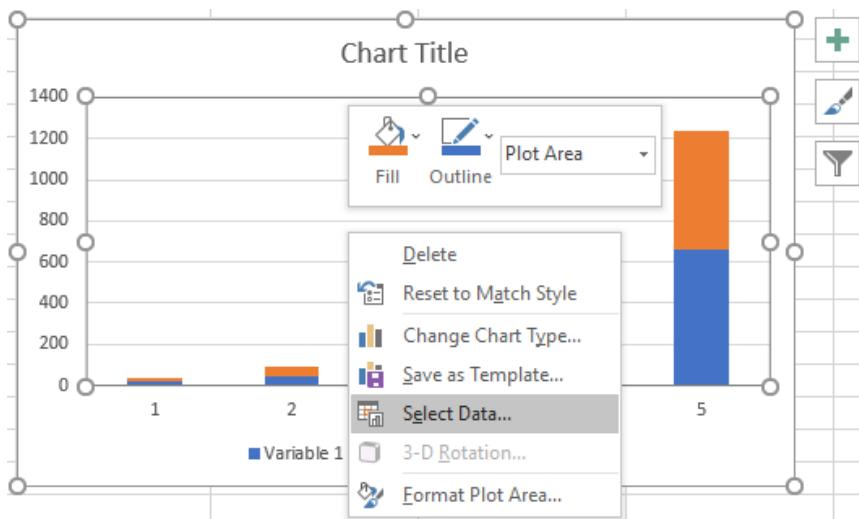
Value	Variable 1 Name	Variable 2 Name
Minimum value	=MIN(C2:C52)	=MIN(D2:D52)
First quartile	=QUARTILE.INC(C2:C52, 1)	=QUARTILE.INC(D2:D52, 1)
Median value	=QUARTILE.INC(C2:C52, 2)	=QUARTILE.INC(D2, 2)
Third quartile	=QUARTILE.INC(C2:C52, 3)	=QUARTILE.INC(D2:D52, 3)
Maximum value	=MAX(C2:C52)	=MAX(D2:D52)

**Step 2:** Referring to **Figure 1.24**, create a stacked column chart comprised of Variable 1's and Variables 2's Minimum value, First quartile, Median value, Third quartile, and Maximum value.



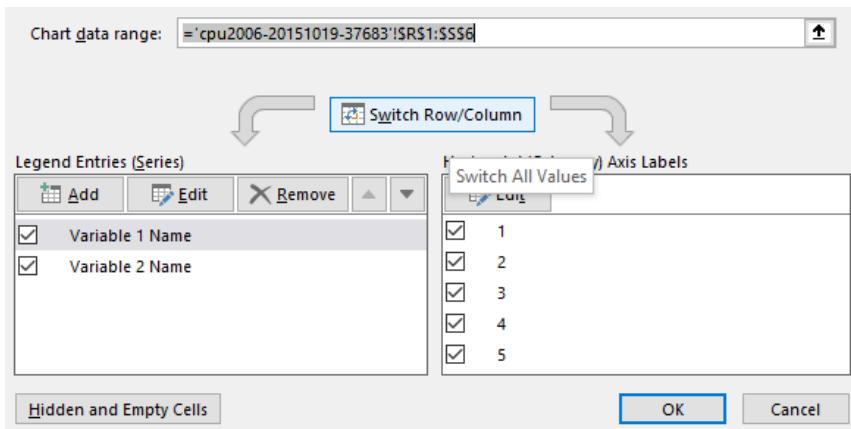
**Figure 1.24.** Column plot menu item with Stacked Column highlighted

**Step 3:** Right-click on the chart area and choose **Select Data...** from the dropdown menu, as shown in **Figure 1.25**.



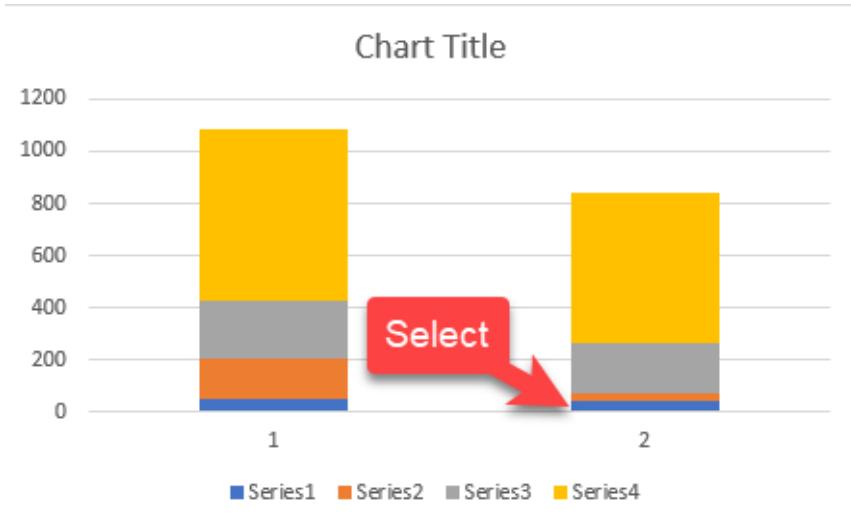
**Figure 1.25.** 2-D stacked bar plot of the intel micro processors

**Step 4.** In the **Select Data Source** dialog, shown in **Figure 1.26**, click on **Switch Row/Column** and select **OK**.



**Figure 1.26.** Select Data dialog window

**Step 5:** Referring to **Figure 1.27**, click on the bottom most block of the bar chart on the right (the one on the left will be selected also). Note: You may have to enlarge the chart to select the piece.



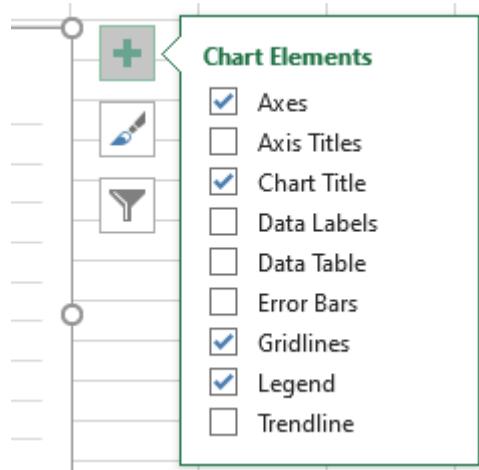
**Figure 1.27.** 2-D stacked bar plot of the intel microprocessors with switched rows and columns

**Step 6:** On the menu ribbon, shown in **Figure 1.28**, select (1) **Chart Tools | Chart Design**, and on the far left, select (2) **Chart Layouts | Add Chart Element**, as shown in Figure 1.28.



**Figure 1.28. Chart design tool bar and Add Chart Element menu**

Note: you can also select the “+” that appears when the chart is selected (also see Figure 1.25) as shown in **Figure 1.29**.



**Figure 1.29. Alternative Chart Elements addition**

**Step 7:** From the Add Chart Element choose **Select Error Bars | Standard Deviation**, shown in **Figure 1.30**.

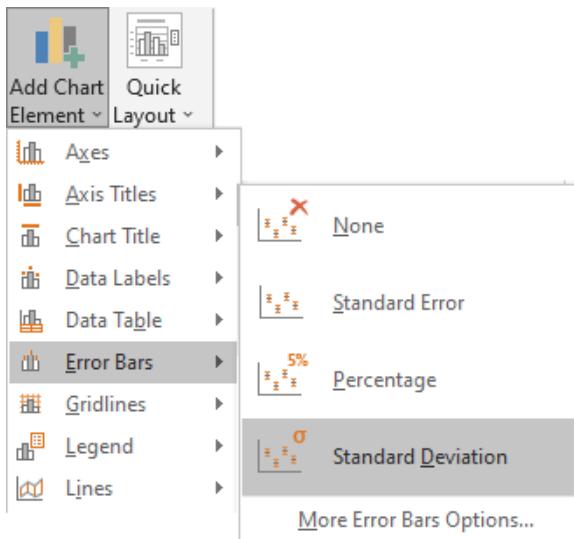


Figure 1.30. Add Chart drop-down menu with Error Bars and Standard deviation selected

**Step 8:** Select the bottom block again, as shown in **Figure 1.31**, right-click, and select **No Fill**.

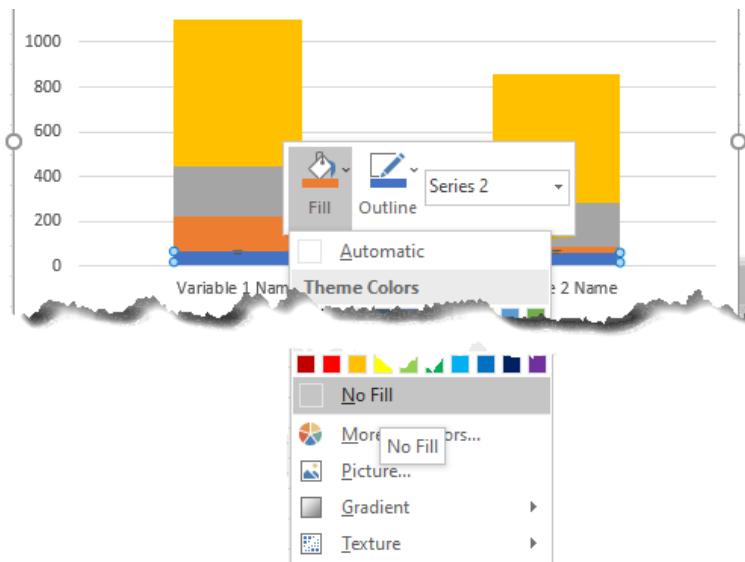
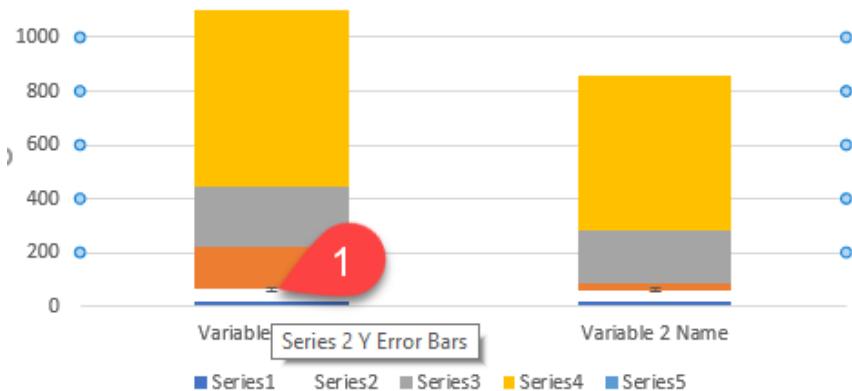


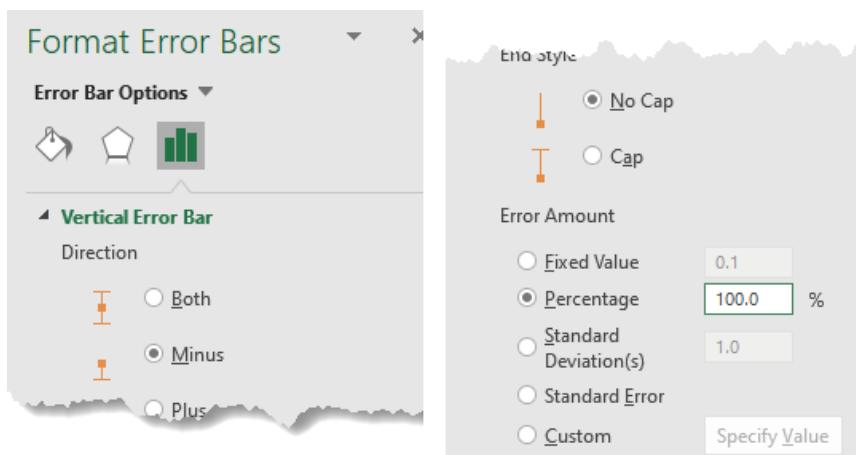
Figure 1.31. Intel micro-processor bar plot with bottom boxes selected and no Fill selected

**Step 9:** Select the **Series 2 Y Error Bars** at (1) as shown in **Figure 1.32**, it may be very small, and then right-click to select Format Error Bars



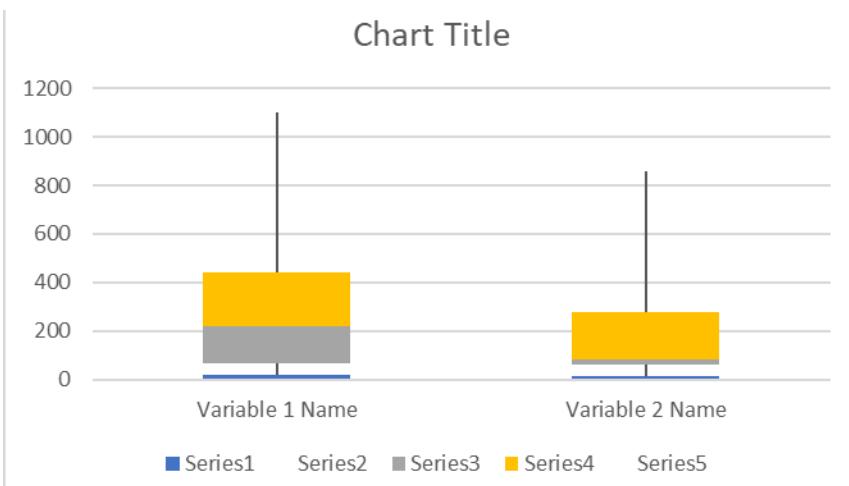
*Figure 1.32. Intel micro-processor bar plot with bottom boxes showing standard deviation error bars*

**Step 10.** From the **Format Error Bars** on the right of the main Excel window, choose **Vertical Error Bar**, and then select **Minus**, **No Cap**, **Percentage**, and enter **100.0**, as seen in **Figure 1.33**.



*Figure 1.33. Error bar formatting menu*

**Step 11.** Next, repeat Steps 5 through 10 to the top bar of the bar chart until you get the chart shown in **Figure 1.34**.



*Figure 1.34. Intel micro-processor bar plot with lower and upper error bars*

**Step 12.** Click on one of the middle boxes of the bar chart and under **Format Data Series | Series Options | Border |Solid Line**, as shown in **Figure 1.35**. Repeat for the remaining box.



*Figure 1.35. Format window for the setting a solid border for the center boxes*

**Step 13.** With a middle box still selected, under **Format Data Series | Series Options | Border |Solid Fill**, as shown in **Figure 1.36**.

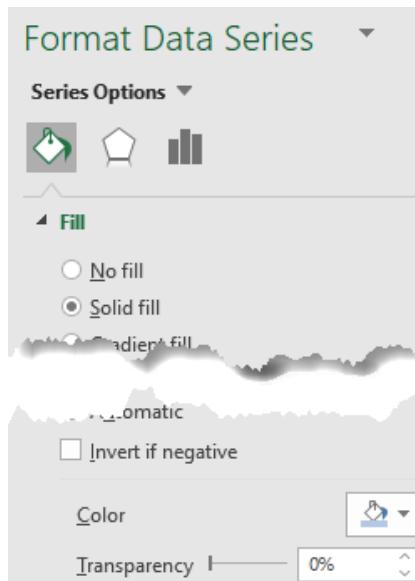


Figure 1.36. Format window for the setting a solid fill for the center boxes

**Step 14.** Select and delete the legend (1) and then at a Chart Title (2), as shown in Figure 1.37.

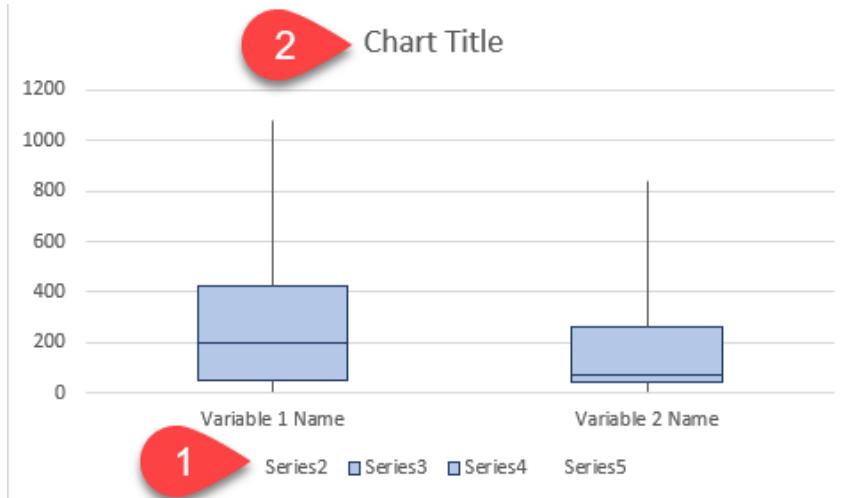


Figure 1.37. Nearly completed box plot.

## Equivalence Testing (TOST)

The objective of a two-sample equivalence test is to determine whether the means of two populations are equivalent based on two independent samples from these populations; here “equivalent” means that the two means differ by a small pre-defined amount. This margin of equivalence is determined by knowledge of the domain under study and represents the tolerance that is acceptable.

A two-sided t-test (TOST) is used to make this determination. Essentially, TOST reverses the roles of the null and alternative hypotheses in a two-sided t-test.

If  $\theta$  represents the margin of equivalence, then we test the hypotheses:

$$H_0: \mu_2 - \mu_1 \leq -\theta \text{ or } \mu_2 - \mu_1 \geq \theta$$

$$H_A: -\theta < \mu_2 - \mu_1 < \theta$$

This is done by conducting two one-sided  $t$ -tests, each of which is based on a null hypothesis that is one of the parts of the above null hypothesis. If the null hypothesis of both tests is rejected then the difference falls within the equivalence interval and we can claim that the two population means are equivalent. The larger p-value of the two  $t$ -tests is used as the p-value of the TOST.

Another way of looking at this method is to conduct a two-sided t-test and determine the  $1 - 2\alpha$  confidence interval  $I$ . If the confidence interval lies completely within the interval  $(-\theta, \theta)$  then we accept that the two means are equivalent.

We can also use two different limits for the margin of equivalence, an upper-value  $\theta_U$  and a lower-value  $\theta_L$ . In this case,  $\theta_U$  replaces  $\theta$  in the TOST method described above and  $\theta_L$  replaces  $-\theta$ .

The TOST approach can also be used for a one-sample test (and similarly for a two dependent sample test). In this case, we test the equivalence between the mean of a single population and some hypothetical mean  $\mu_0$ .

$$H_0: \mu - \mu_0 \leq \theta_L \text{ or } \mu - \mu_0 \geq \theta_U$$

$$H_A: \theta_L < \mu - \mu_0 < \theta_U$$

This time we conduct two one-sample t-tests. Alternatively, if the  $1 - 2\alpha$  confidence interval for the two-tailed one-sample t-test lies completely within the interval  $(\theta_L, \theta_U)$  then we accept that the population mean and hypothetical mean are equivalent.

### **Example 1.16: Satellite Propagators**

A contractor that markets a premium propagator used to determine the position of satellite at any instance of time, with given acceleration and initial velocity. The propagator is embedded in maintenance agreement update that is still in effect. Using the random sample of 12 readings from the new propagator and 12 readings from the legacy propagator as shown in **Table 1-21**. Determine whether the sources are equivalent based on  $\alpha = 0.1$ .

*Table 1-21. Satellite propagator scores*

#### **Equivalence Testing (TOST)**

Original	New
2311	2298
2274	2260
2262	2250
2297	2242
2291	2302
2319	2297
2263	2283
2329	2286
2289	2270
2287	2213
2290	2305
2301	2290

The readings from the propagators are scored based on parameters from the 2-line ELSET and compared requiring the scores to be within  $\pm 25$

points. Moreover, by convention, lower scores are better. Even though the new propagator passed developmental testing, an operational test is required to determine if the new propagator works in an operationally configured system, and whether it performed better than the original one, i.e., a one tailed t-test is required.

**Figure 1.38** shows the analysis for a two independent sample t-test. Since the 90% confidence interval (0.7044, 35.4623) and (0.6278, 35.5389) is not completely contained in the interval (-25, 25), we conclude that the two propagators are not equivalent.

	D	E	F	G	H	I	J	K	L
3	T Test: Two Independent Samples								
4									
5	SUMMARY			Hyp Mean	0				
6	Groups	Count	Mean	Variance	Cohen d				
7	Original	12	2292.75	423.8409					
8	New	12	2274.67	805.3333					
9	Pooled			614.5871	0.7294				
10									
11	TTEST: Equal Variances			Alpha	0.1				
12		std err	t-stat	df	p-value	t-crit	lower	upper	sig
13	One Tail	10.1208	1.7867	22	0.0439	1.3212			yes
14	Two Tail	10.1208	1.7867	22	0.0878	1.7171	0.7044	35.4623	yes
15									
16	TTEST: Unequal Variances			Alpha	0.1				
17		std err	t-stat	df	p-value	t-crit	lower	upper	sig
18	One Tail	10.1208	1.7867	20.0670	0.0446	1.3253			yes
19	Two Tail	10.1208	1.7867	20.0670	0.0891	1.7247	0.6278	35.5389	yes

*Figure 1.38. Analysis for a two independent sample t-test*

Note that if we test the two one-sided null hypotheses directly, we would obtain p-values of 0.0439 and 0.0446. Since both are greater than  $\alpha = 0.1$ , we again conclude that the scores are not equivalent (with  $p - value = .036$ ).

Note too that a two-sided t-test would yield a  $p - value = 0.0878$  and 0.0891, so we can also conclude that there is a significant difference between the two suppliers.

## Sign Test

The **sign test** is a basic non-parametric test that can be applied when the conditions for the single sample t-test are not met. The test itself is very simple: perform a binomial test (or use the normal distribution approximation when the sample is sufficiently large) on the signs of the data elements as described in the following example.

### Example 1.17: C4ISR software

A C4ISR software system claims that the data latency for routine messaging is less than the threshold value for a legacy system that it is replacing. Given data from a test of this software in an operational configuration, determine whether the new system performs at least as good as the legacy system in terms of routine message latency over the same period of time. Using a threshold 25 seconds for the median latency, a test team collected 15 sample messages over an 8-hour period to evaluate the new system's performance as shown in Figure 1.39.

If we view the QQ Plot in **Figure 1.39**, we see that the data is not normally distributed, which makes the use of the nonparametric sign test appropriate. This also apparent in the boxplot in **Figure 1.40**. The null hypothesis here is:

$$H_0: \text{the population median} \geq 25 \text{ seconds}$$

To perform the test, we count the number of data elements whose value is larger than 25 and the number of data elements whose value is smaller than 25, dropping any data elements with a value of 25 from the sample. This is accomplished by putting a **+1** in column C if the corresponding data element in column B is  $< 25$ , a **-1** if the data element is  $> 25$  and **0** if the data element is  $= 25$ .

The number  $N+$  of **data elements  $< 25$  (cell B21)** is calculated by the formula **=COUNTIF(C4:C18,1)**.

Similarly, the number  $N-$  of **data elements  $> 25$  (cell B22)** is calculated by the formula **=COUNTIF(C4:C18,-1)**.

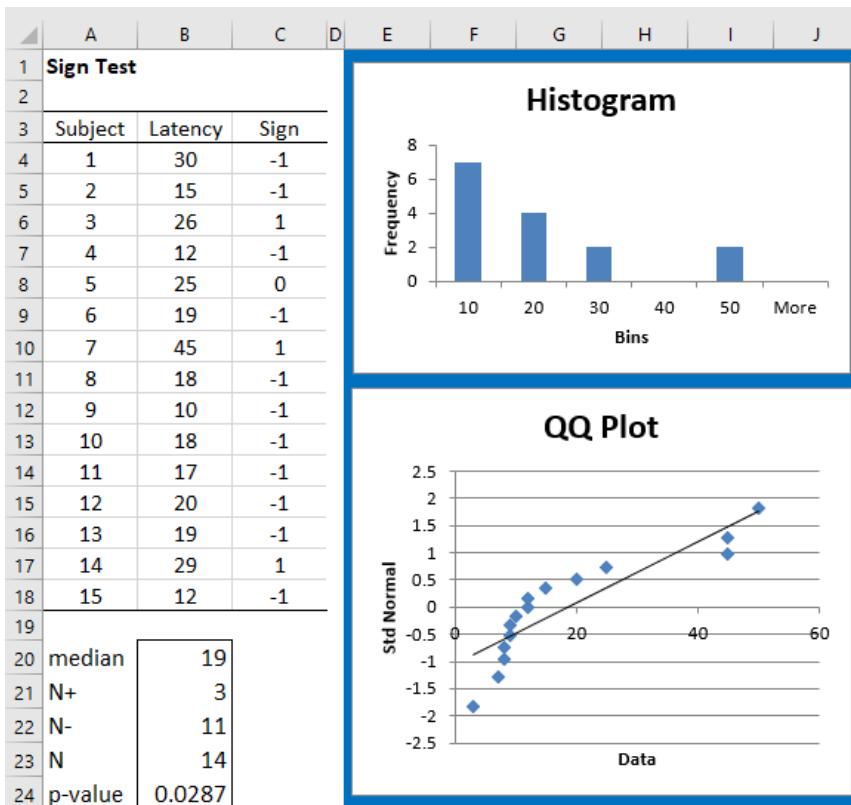


Figure 1.39. Excel workbook with histogram and normal probability plot

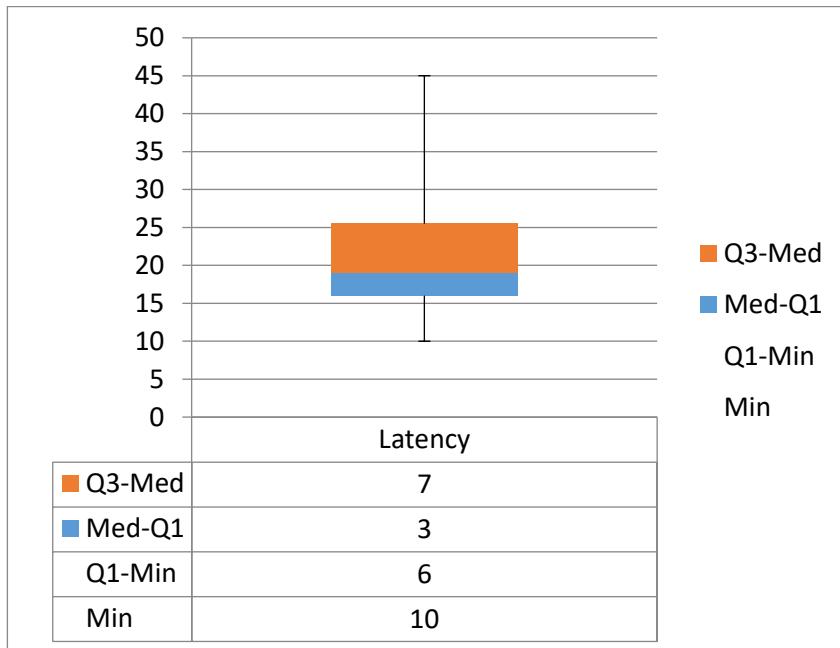
The revised sample size  $N$  (cell B23) is given by the formula =B21+B22 (where the elements with a value of 25 are dropped).

If the null hypothesis is true then the probability that a data element is  $< 25$  is 0.5, and so we need to determine the probability that 3 out of 14 data elements are greater than the median given that the probability on any trial is 0.5, i.e.

$$p\text{-value} = \text{BINOMDIST}(3, 14, 0.5, \text{TRUE}) = 0.0287 < 0.05 = \alpha$$

Since the p-value  $< \alpha$ , (one-tailed test) we can reject the null hypothesis, and conclude with 95% confidence that the median latency using the new C4ISR software is less than the threshold median latency value of 25 seconds.

Note that we have used a one-tail test. If we had used a two-tail test instead then we would double the p-value calculated above. Also, note that in performing a two-tail test we would perform the test using the smaller of  $N_+$  and  $N_-$ , which for this example is  $N_+ = 3$  (since  $N_- = 11$  is larger).



**Figure 1.40. Boxplot of processing latency**

Thus, in terms of latency of routine reports, the new C4ISR software is effective.

**Table 1-22** and **Figure 1.41** provide the formulas used to create the Excel worksheet.

**Table 1-22. Formulas for creating the box plot in Excel**

Min	=MIN(B4:B18)
Q1-Min	=QUARTILE(B4:B18,1)-S4
Med-Q1	=MEDIAN(B4:B18)-QUARTILE(B4:B18,1)
Q3-Med	=QUARTILE(B4:B18,3)-MEDIAN(B4:B18)
Max-Q3	=MAX(B4:B18)-QUARTILE(B4:B18,3)

	A	B	C
1	Sign Test		
3	Subject	Latency	Sign
4	1	30	=IF(B4>25,-1,IF(B4<25,1,0))
5	=A4+1	15	=IF(B5>25,1,IF(B5<25,-1,0))
6	=A5+1	26	=IF(B6>25,1,IF(B6<25,-1,0))
7	=A6+1	12	=IF(B7>25,1,IF(B7<25,-1,0))
8	=A7+1	25	=IF(B8>25,1,IF(B8<25,-1,0))
9	=A8+1	19	=IF(B9>25,1,IF(B9<25,-1,0))
10	=A9+1	45	=IF(B10>25,1,IF(B10<25,-1,0))
11	=A10+1	18	=IF(B11>25,1,IF(B11<25,-1,0))
12	=A11+1	10	=IF(B12>25,1,IF(B12<25,-1,0))
13	=A12+1	18	=IF(B13>25,1,IF(B13<25,-1,0))
14	=A13+1	17	=IF(B14>25,1,IF(B14<25,-1,0))
15	=A14+1	20	=IF(B15>25,1,IF(B15<25,-1,0))
16	=A15+1	19	=IF(B16>25,1,IF(B16<25,-1,0))
17	=A16+1	29	=IF(B17>25,1,IF(B17<25,-1,0))
18	=A17+1	12	=IF(B18>25,1,IF(B18<25,-1,0))
19			
20	median	=MEDIAN(B4:B18)	
21	N+	=COUNTIF(C4:C18,1)	
22	N-	=COUNTIF(C4:C18,-1)	
23	N	=B21+B22	
24	p-value	=BINOMDIST(B21,B23,0.5,TRUE)	

Figure 1.41. Sign Test formulas in Excel

## Wilcoxon Rank Sum Test for Independent Samples

When the requirements for the t-test for two independent samples are not satisfied, the **Wilcoxon Rank-Sum** non-parametric test can often be used provided the two independent samples are drawn from populations with an ordinal distribution. For this test we use the following null hypothesis:

$$H_0: \text{the observations come from the same population}$$

From a practical point of view, this implies:

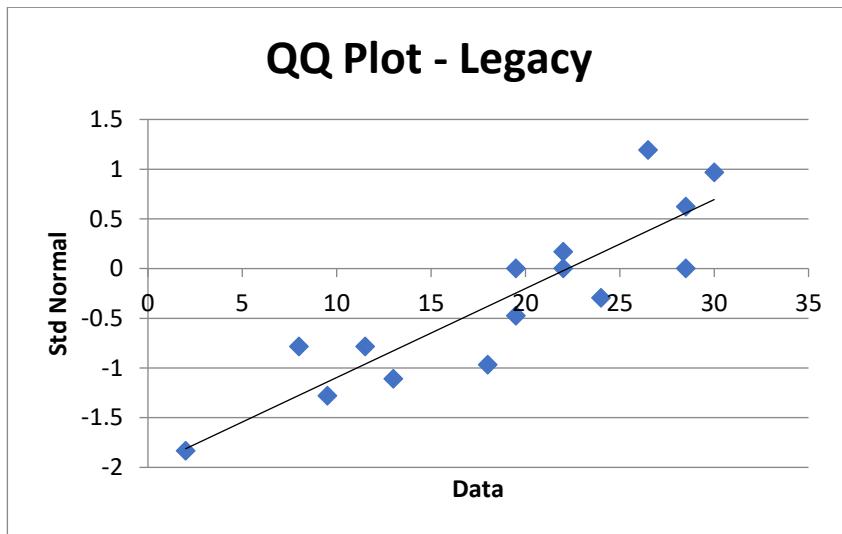
$H_0$ : if one observation is made at random from each population (call them  $x_0$  and  $y_0$ ), then the probability that  $x_0 > y_0$  is the same as the probability that  $x_0 < y_0$ .

If, in addition, the populations for each sample have the same shape, then this would mean that the two populations have the same median. You would normally judge whether the populations have the same shape based on looking at the shapes of the two samples (e.g., by using histograms or box plots).

We illustrate the technique with the following examples.

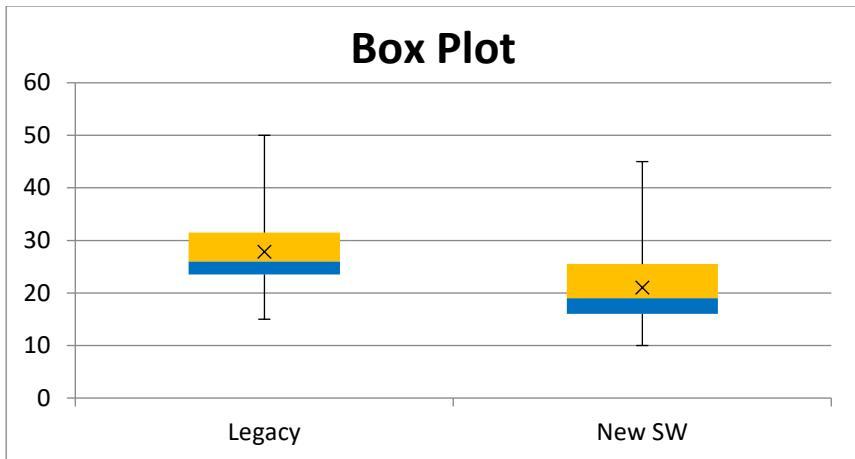
### Example 1.18 New Software vs Legacy

When we look at the QQ Plot for the Legacy group in **Figure 1.42**, we see that it is not very normal.



**Figure 1.42. Normal probability plot for processor latency**

More concerning is that the Box Plot for the group that we sampled the New SW latencies is not very symmetric (see **Figure 1.43**). We, therefore, decide to use the Wilcoxon Sign-Rank test instead of the t-test.



*Figure 1.43. Box plots for the two processors' latencies*

The results of the Wilcoxon Rank-Sum test are displayed in **Figure 1.44**.

	A	B	C	D	E	F	G	H	I	
1	<b>Wilcoxon Rank-Sum Test</b>									
2										
3	Legacy	New SW		Legacy	New SW		Legacy	New SW		
4	31	30		6.5	8		Count	15	15	
5	15	15		26.5	26.5		Rank Sum	182.5	282.5	
6	31	26		6.5	11.5		a	0.05		
7	24	12		15	28.5		tails	2		
8	34	25		4	13		W	182.5		
9	24	19		15	19.5		W-crit	184		
10	50	45		1	2		sig	no		
11	24	18		15	22		Results			
12	16	10		25	30		Reject Ho			
13	23	18		17	22					
14	26	17		11.5	24					
15	32	20		5	18					
16	29	19		9.5	19.5					
17	41	29		3	9.5					
18	18	12		22	28.5					

*Figure 1.44. Excel workbook of the Wilcoxon Rank-Sum Test*

We begin by calculating the ranks of the combined 30 raw scores using the standard RANK.AVG function in Excel 2010 or later. The contents of

cell D4 is the rank of the first participant in the Control group, namely **RANK.AVG(A6,\$A\$4:\$B\$18)** (all the formulas for this test are provided in **Table 1-23**). We copy and paste this formula down and to the right in the range D4:E18 (see **Figure 1.44**).

We then calculate the sum of the ranks for each group to arrive at the rank sums  $R_1 = 182.5$  and  $R_2 = 282.5$ . Since the sample sizes are equal, the value of the test statistic  $W =$  the smaller of  $R_1$  and  $R_2$ , which for this example means that  $W = 182.5$  (cell H9).

We next compare  $W$  with the critical value  $W_{crit}$ , which can be found in the Wilcoxon Rank-Sum Table. Since the sample sizes are both 15, we look up the critical value in the table for  $\alpha = .05$  (two-tail) where  $n_1 = n_2 = 15$ , and find that  $W_{crit} = 184$ . This represents the smallest value we could expect to obtain for  $W$  if the null hypothesis were true. Since  $W = 182.5 < 184 = W_{crit}$ , we can reject the null hypothesis, and so conclude there is a statistically significant difference between the effectiveness of the legacy and new SW, and that the new SW is more effective in terms of routine message latency.

**Table 1-23. Excel formulas for the Wilcoxon Rank-Sum Test**

		Legacy	New SW
<b>Count</b>	H4	=COUNT(D4:D18)	=COUNT(E4:E18)
<b>Rank Sum</b>	H5	=SUM(D4:D18)	=SUM(E4:E18)
<b>a</b>	H7	0.05	
<b>tails</b>	H8	2	
<b>W</b>	H9	=MIN(H5:I5)	
<b>W-crit</b>	H10	=VLOOKUP(H4,M54:V76,3)	
<b>sig</b>	H11	no	
<b>Results</b>	H14	=IF(H9>H10,"Cannot reject Ho", "Reject Ho")	

### Example 1.19: Orbital Elements with Data Analysis Add-in

It is well-known that locating resident space objects (RSOs) is highly sensitive to errors in eccentricity (pertaining to the shape of an orbit. In this example we want to compare the orbital data produced by aa truth source versus the orbital data from new orbital data production for DoD.

There are quite a few field in a two-line element set, so we chose the eccentricity to begin. The data collect was on the communications satellite, Globalstar.

Globalstar.csv contains the data we are interested in. The data from our truth source is indicated by the variable name followed by a 1, i.e., Eccentricity1. The same data from the new program is Eccentricity2. A portion of the data is shown in **Figure 1.45**.

	E	F	G	H
1	ECCENTRICITY1	ECCENTRICITY2	INCLINATION1	INCLINATION2
2	0.0002316	0.000145073	52.0054	50.7279
3	0.0002815	0.000225364	51.9985	50.6965
4	0.000165	0.000082677	51.9976	50.4838
5	0.000732	0.000053252	51.9595	50.8735
6	0.0000869	0.000063616	51.9826	50.5870
7	0.0003507	0.000324812	51.9916	50.8015
8	0.0013023	0.001301195	51.9786	50.8955
9	0.0002452	0.000133564	51.9946	50.6979
10	0.0010550	0.00006527	52.01	50.5725
78	0.000231	0.00019225	51.9856	50.0232
79	0.0001337	0.000075330	51.9867	50.7980
80	0.000073	0.000026261	52.006	50.1667
81	0.0000798	0.000000626	51.9824	50.6178
82	0.0000833	0.000050182	51.9975	50.3300
83	0.000084	0.000057936	52.0098	50.8671
84	0.0000462	0.000005517	51.9862	50.6398
85	0.0000492	0.000032584	51.9787	50.6621

*Figure 1.45. A portion of the orbital data from two sources.*

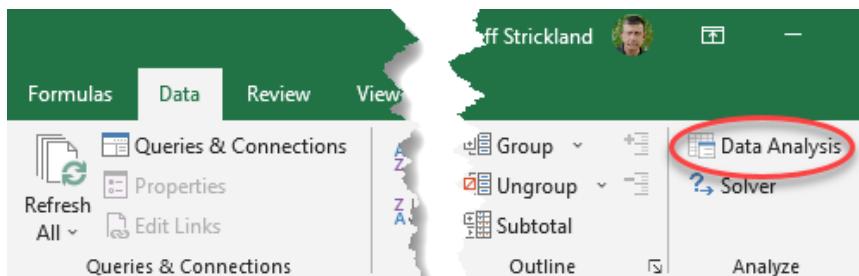
We are interested in whether there is a significant difference in the average values for eccentricity between the two data sources. Assuming that the sample variances may differ due to algorithm differences, we perform a t-test on sample means with unequal variances (so we have to use a pooled variance).

To perform our analysis, we'll use Excel's Data Analysis add-in. To set up the problem for comparing the eccentricity means, we hypothesize the following:

$$H_0: \text{the sample means are equal}, \mu_1 = \mu_2$$

$$H_1: \text{the sample means are significantly different}, \mu_1 \neq \mu_2$$

To run the analysis, we open the Data Analysis add-in from the Data tab as shown in **Figure 1.46**



**Figure 1.46.** Excel's Data tab with the Data Analysis add-in denoted by the red ellipse.

Opening the add-in yields the dialog box opens as shown in Figure 1.47. First, we want to test our assumption of unequal variances using an F-test. An F-test (Snedecor & Cochran, 1989) is used to test if the variances of two populations are equal. This test can be a two-tailed test or a one-tailed test. The two-tailed version tests against the alternative that the variances are not equal. Our hypotheses for this test are:

$$H_0: \text{the sample means are equal}, \sigma_1 = \sigma_2$$

$$H_1: \text{the sample means are significantly different}, \sigma_1 \neq \sigma_2$$

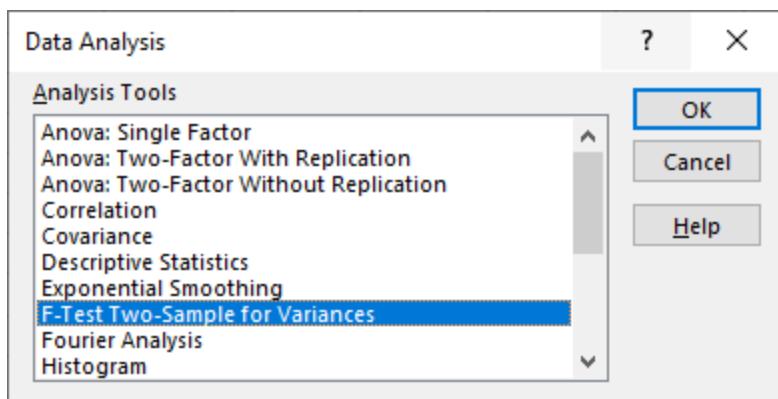


Figure 1.47. The Data Analysis add-in Analysis Tool selection window, with F-Test Two-Sample for Variance selected.

Clicking OK yields the window shown in Figure 1.48. So, we have selected the two columns holding our eccentricity values, namely cells \$E\$1:\$E\$85 and \$F\$1:\$F\$85 and indicated that the columns have headers (Labels). We have also selected an alpha level of 0.05, and have elected to show the results in a new tab, “Variance Comparison”. When we complete the input, we click the OK button.

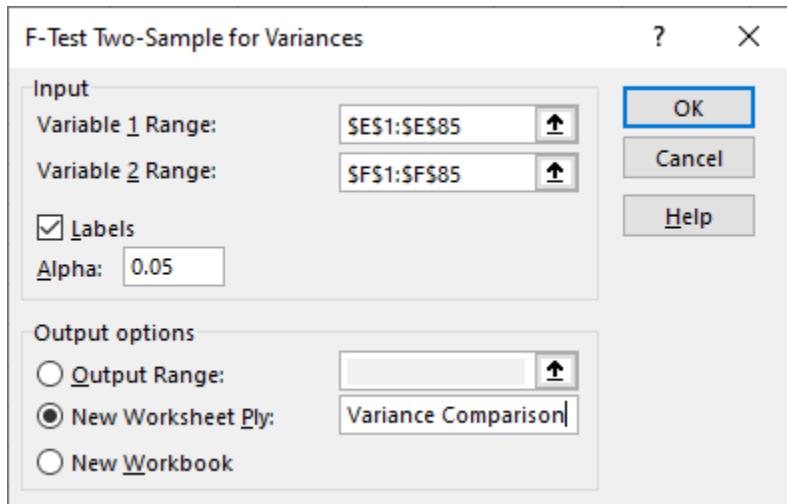


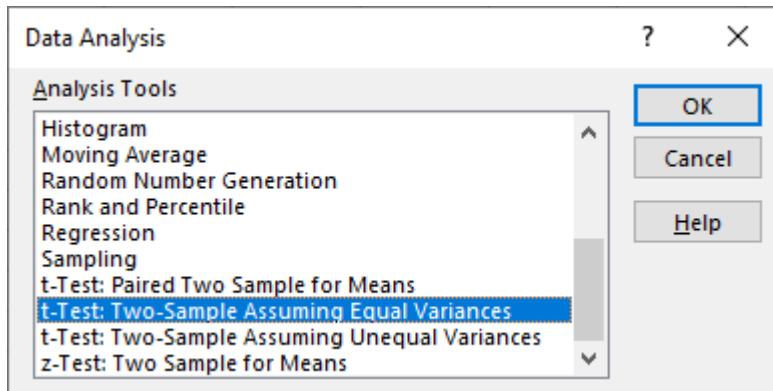
Figure 1.48. Setup for the F-Test for two variances.

The results are shown in **Figure 1.49**. Since the test statistic is less than the critical value, or the p-value (0.0742) is greater than the level of significance (0.05), we fail to reject the null hypothesis and conclude that the variances for the eccentricity data are not significantly different.

	A	B	C
1	F-Test Two-Sample for Variances		
2			
3		ECCENTRICITY1	ECCENTRICITY2
4	Mean	0.00042523	0.000249323
5	Variance	6.37441E-07	4.63417E-07
6	Observations	84	84
7	df	83	83
8	F	1.375523462	
9	P(F<=f) one-tail	0.074235713	
10	F Critical one-tail	1.437878961	

**Figure 1.49. F-Test results for comparing eccentricity variances.**

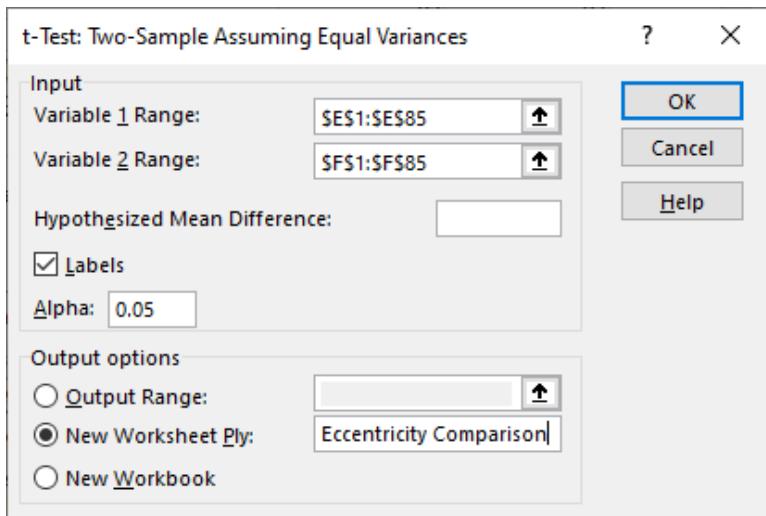
Next, we run the Data Analysis add-in again and choose the t-Test for two samples having equal variances, and click OK. This leads to the next dialog shown in **Figure 1.50**.



**Figure 1.50. The Data Analysis add-in Analysis Tool selection window.**

Referring to **Figure 1.51**, we have selected the two columns holding our eccentricity values, namely cells \$E\$1:\$E\$85 and \$F\$1:\$F\$85 and indicated that the columns have headers (Labels). We have also selected an alpha level of 0.05, and have elected to show the results in a new tab,

“Eccentricity Comparison”. When we complete the input, we click the OK button. The results are shown in **Figure 1.52**.



*Figure 1.51. F-Test results for comparing eccentricity variance.*

	A	B	C
1	t-Test: Two-Sample Assuming Unequal Variances		
2			
3		ECCENTRICITY1	ECCENTRICITY2
4	Mean	0.00042523	0.000249323
5	Variance	6.37441E-07	4.63417E-07
6	Observations	84	84
7	Hypothesized Mean Difference	0	
8	df	162	
9	t Stat	1.536582356	
10	P(T<=t) one-tail	0.063173217	
11	t Critical one-tail	1.654313957	
12	P(T<=t) two-tail	0.126346434	
13	t Critical two-tail	1.974715786	

*Figure 1.52. t-Test results for comparing eccentricity means.*

Since the test statistic is less than the critical value, or the p-value (0.1263) is greater than the level of significance (0.5), we fail to reject the null hypothesis and conclude that the means are not significantly different.

Based on our test results, we conclude that there is not enough evidence to convince us that the mean eccentricities between the new program and the truth source are different. So, we are left with ensuring the same is true for the remaining variables in the two-line element set.

Note that we cannot say that the two means are equal. Recall that if the p-value is greater than the level of significance, we can only fail to reject  $H_0$ , that the means are equal; we do not accept that the means are equal.

## 2. The Basic Principles of DOE

### t-Test Example

In Example 1.1 (p.8), we performed a two-sided, one-sample t-test using the calibration factor data in the ZARR13.DAT data set to test the null hypothesis that the sample mean is equal to 9.0, the manufacturer's specification.

$$H_0: \mu = 9.0$$

$$H_a: \mu \neq 9.0$$

*Table 2-1. t-test output for the Calibration factor*

Calibration Factor

$$H_0: \mu_1 = \mu_2$$

Mean	9.261461
Std Dev	0.022789
N	195
SEM	0.001632
df	194
Hypo Mean	9.0

t-Test

$$\text{t-Statistic} \quad 160.2147$$

$$\alpha \quad 0.05$$

$$t_{1-\frac{\alpha}{2}, \nu} \quad 1.972268$$

$$\text{p-value} \quad 4.4\text{E}-208$$

*Reject  $H_0$  if  $|T| > 1.9723$*

Result      Reject  $H_0$

We reject the null hypotheses for our two-tailed t-test because the absolute value of the test statistic is greater than the critical value. If we were to perform an upper, one-tailed test, the critical value would be  $t_{1-\alpha, \nu} = 1.6527$ , and we would still reject the null hypothesis.

The confidence interval provides an alternative to the hypothesis test. If the confidence interval contains 5, then  $H_0$  cannot be rejected. In our example, the confidence interval (9.258242, 9.264679) does not contain 5, indicating that the population mean does not equal 5 at the 0.05 level of significance.

In general, there are three possible alternative hypotheses and rejection regions for the one-sample t-test:

**Table 2-2. Table of Alternative hypotheses with associated rejection regions**

Alternative Hypothesis	Rejection Region
$H_a: \mu \neq \mu_0$	$ T  > t_{1-\frac{\alpha}{2},v}$
$H_a: \mu > \mu_0$	$T > t_{1-\alpha,v}$
$H_a: \mu < \mu_0$	$T < t_{\alpha,v}$

The rejection regions for three possible alternative hypotheses using our example data are shown in the following graphs.

Now, when we think about designing an experiment to perform this type of analysis, yet where we control such things as treatments, factors, levels of the factors, and actual control elements, we speak of experimental design or **design of experiments** (DOE). And, we'll next discuss the basic principles of DOE.

## Randomization

This is an essential component of any experiment that is going to have validity. If you are doing a comparative experiment where you have two treatments, a treatment and a control, for instance, you need to include in your experimental process the assignment of those treatments by some random process. An experiment includes experimental units. You need to have a deliberate process to eliminate potential biases from the conclusions, and random assignment is a critical step.

## Replication

Replication is some in sense the heart of all of statistics. To make this point... Remember what the standard error of the mean is? It is the square root of the estimate of the variance of the sample mean, i.e.,  $\sqrt{s^2/n}$ . The width of the confidence interval is determined by this

statistic. Our estimates of the mean become less variable as the sample size increases.

Replication is the basic issue behind every method we will use in order to get a handle on how precise our estimates are at the end. We always want to estimate or control the uncertainty in our results. We achieve this estimate through replication. Another way we can achieve short confidence intervals is by reducing the error variance itself. However, when that isn't possible, we can reduce the error in our estimate of the mean by increasing  $n$ . Another way is to reduce the size or the length of the confidence interval is to reduce the error variance — which brings us to blocking.

## **Blocking**

Blocking is a technique to include other factors in our experiment which contribute to undesirable variation. Much of the focus in this class will be to creatively use various blocking techniques to control sources of variation that will reduce error variance. For example, in human studies, the gender of the subjects is often an important factor. Age is another factor affecting the response. Age and gender are often considered nuisance factors which contribute to variability and make it difficult to assess systematic effects of a treatment. By using these as blocking factors, you can avoid biases that might occur due to differences between the allocation of subjects to the treatments, and as a way of accounting for some noise in the experiment. We want the unknown error variance at the end of the experiment to be as small as possible. Our goal is usually to find out something about a treatment factor (or a factor of primary interest), but in addition to this, we want to include any blocking factors that will explain variation.

## **Multi-factor Designs**

We will spend at least half of this course talking about multi-factor experimental designs:  $2^k$  designs,  $3^k$  designs, response surface designs, etc. The point to all of these multi-factor designs is contrary to the scientific method where everything is held constant except one factor which is varied. The one factor at a time method is a very inefficient way of making scientific advances. It is much better to design an experiment

that simultaneously includes combinations of multiple factors that may affect the outcome. Then you learn not only about the primary factors of interest but also about these other factors. These may be blocking factors which deal with nuisance parameters or they may just help you understand the interactions or the relationships between the factors that influence the response.

## Confounding

Confounding is something that is usually considered bad! Here is an example. Let's say we are doing a medical study with drugs *A* and *B*. We put 10 subjects on drug *A* and 10 on drug *B*. If we categorize our subjects by gender, how should we allocate our drugs to our subjects? Let's make it easy and say that there are 10 male and 10 female subjects. A balanced way of doing this study would be to put five males on drug *A* and five males on drug *B*, five females on drug *A* and five females on drug *B*. This is a perfectly balanced experiment such that if there is a difference between male and female at least it will equally influence the results from drug *A* and the results from drug *B*.

An alternative scenario might occur if patients were randomly assigned treatments as they came in the door. At the end of the study, they might realize that drug *A* had only been given to the male subjects and drug *B* was only given to the female subjects. We would call this design totally confounded. This refers to the fact that if you analyze the difference between the average response of the subjects on *A* and the average response of the subjects on *B*, this is exactly the same as the average response on males and the average response on females. You would not have any reliable conclusion from this study at all. The difference between the two drugs *A* and *B*, might just as well be due to the gender of the subjects since the two factors are totally confounded.

Confounding is something we typically want to avoid but when we are building complex experiments, we sometimes can use confounding to our advantage. We will confound things we are not interested in order to have more efficient experiments for the things we are interested in. This will come up in multiple-factor experiments later on. We may be interested in main effects but not interactions so we will confound the interactions in this way in order to reduce the sample size, and thus the

cost of the experiment, but still have good information on the main effects.

## **Factors**

We usually talk about “treatment” factors, which are the factors of primary interest to you. In addition to treatment factors, there are nuisance factors which are not your primary focus, but you have to deal with them. Sometimes these are called blocking factors, mainly because we will try to block on these factors to prevent them from influencing the results. There are other ways that we can categorize factors.

### **Experimental Factors**

These are factors that you can specify, set the levels, and then assign at random as the treatment to the experimental units. Examples would be temperature, level of an additive fertilizer amount per acre, etc.

### **Classification Factors**

These can't be changed or assigned, these come as labels on the experimental units. The age and sex of the participants are classification factors which can't be changed or randomly assigned. But you can select individuals from these groups randomly.

### **Quantitative vs. Qualitative Factors**

#### **Quantitative Factors**

You can assign any specified level of a quantitative factor. Examples: percent or pH level of a chemical.

#### **Qualitative Factors**

These factors have categories which are different types. Examples might be species of a plant or animal, a brand in the marketing field, gender--these are not ordered or continuous but are arranged perhaps in sets.

### **Example 2.1: Factors and Levels Example**

Questions we should ask about factors in a particular application include

1. What factors have (significant) effect on the response (yield)
2. What effect (small/large) (pos/neg)?
3. Do factors influence each other (interaction)?
4. Find a model that predicts the response for a given combination of factors?

So, suppose we are evaluating strength of materials. Possible factors and levels include:

Factor 1 = Lab (8 levels: 1 to 8)  
Factor 2 = Bar id within lab (1 to 30)  
Factor 3 = Set xxx (4 levels: 1 to 4)  
Factor 4 = Table Speed (2 levels: slow and fast)  
Factor 5 = Down Feed Rate (2 levels: slow & fast)  
Factor 6 = Wheel Grit (2 levels: fine & medium)

## Model

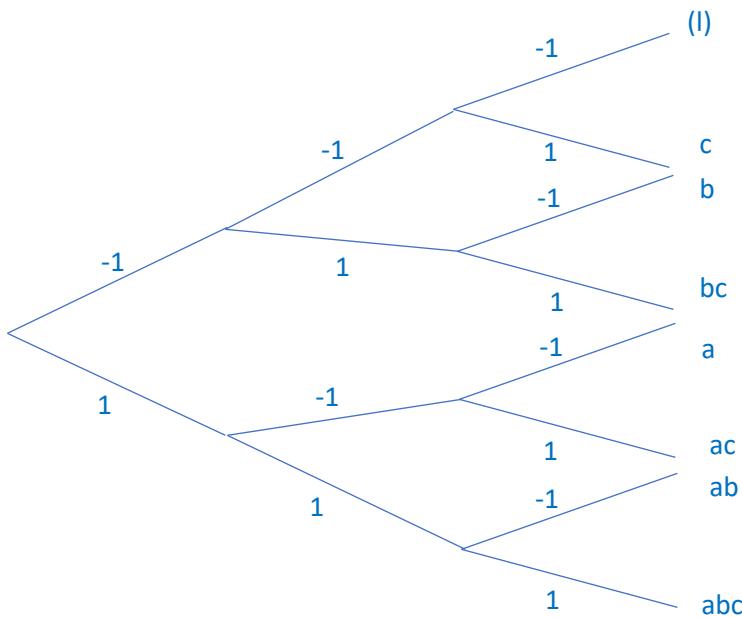
Suppose we have three factors, then our model could include each factor and each interaction, resulting in the intercept and seven terms.

$$y = \beta_1 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{123} X_1 X_2 X_3$$

8 parameters:  $B_1, B_2, \dots, B_{123}$

	Factor A	Factor B	Factor C
Level 1	-1	-1	-1
Level 2	1	1	1

The factor and level combinations can be shown in a tree charts as shown below. Along the edges are the levels and at the end of each branch is a representation of the combinations, using the alphabet. The first branch of the tree are represented by "a". If the level at that branch is a negative 1, then the letter is not assigned, so the top branches, with -1'a results in no value at the end, represented by (L). At the bottom branches, the values are all +1 and the result at the end is "abc". And this pattern continues.



*Figure 2.1. Branch diagram of the factorial design process*

This taxonomy can also be shown in a table.

*Table 2-3. Table of the factorial design process*

Name	Factors			Interactions				Response	
	A	B	C	AB	AC	BC	ABC	Yield 1	Yield 2
(L)	-1	-1	-1	1	1	1	-1	$Y_1$	$Y_9$
1	1	-1	-1	-1	-1	1	1	$Y_2$	$Y_{10}$
b	-1	1	-1	-1	1	-1	1	$Y_3$	$Y_{11}$
ab	1	1	-1	1	-1	-1	-1	$Y_4$	$Y_{12}$
c	-1	-1	1	1	-1	-1	1	$Y_5$	$Y_{13}$
ac	1	-1	1	-1	1	-1	-1	$Y_6$	$Y_{14}$
bc	-1	1	1	-1	-1	1	-1	$Y_7$	$Y_{15}$
abc	1	1	1	1	1	1	1	$Y_8$	$Y_{16}$

## Analysis of Variance

Now, what we just described with Example 1.1 is an analysis that we'll present more formally from here on, but by example. This strategy for

controlling and measuring factors and levels and such, leads us to a method of statistical analysis described an analysis of variance or ANOVA.

We'll build this idea of ANOVA with Example 2.2, where we will perform the ANOVA using the Excel add-in we've used thus far. Then we'll interject some more formal definitions, and repeat Example 2.2 "by-hand" as Example 2.3. We'll use Excel to perform the "by-hand" operations.

### **Example 2.2: LEO Proliferation Discrimination Software**

Suppose we will test the effectiveness of two new software apps that will perform LEO proliferation. The test consists of 10 independent samples of LEO proliferation objects detected by each System Under Test (SUT) and Control (known) Set. Determine whether there is a perceived significant difference between the three systems.

	A	B	C
5	DATA		
6	Control	SUT 1	SUT 2
7	13	12	11
8	17	15	19
9	19	8	15
10	11	16	14
11	20	10	18
12	15	14	16
13	18	10	18
14	9	4	11
15	12	6	14
16	16	11	11

*Figure 2.2. Test data in Excel format*

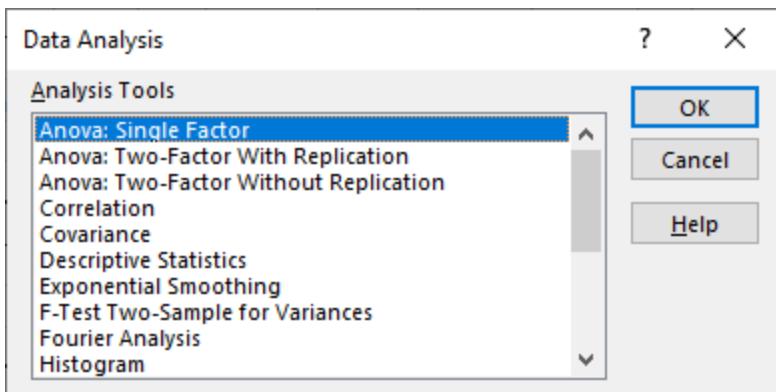


Figure 2.3. Tata Analysis tool selection highlighting Nova: Single Factor.

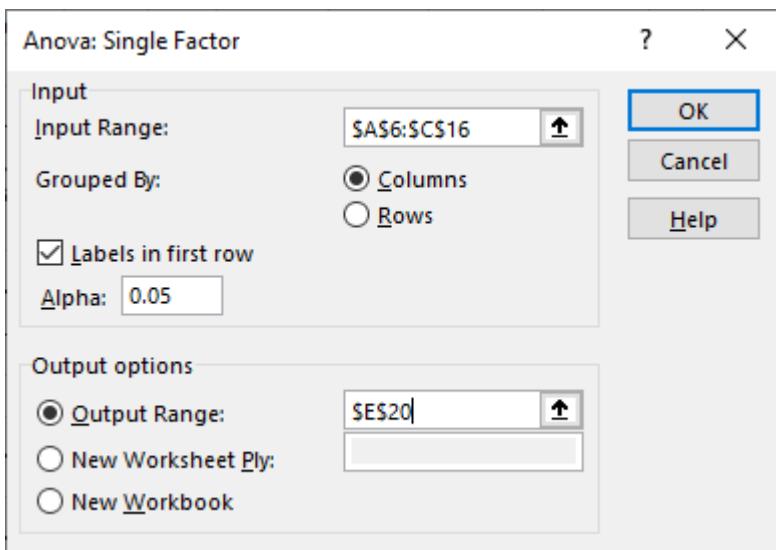


Figure 2.4. Single Factor ANOVA dialog window setup.

	E	F	G	H	I
22	SUMMARY				
23	Groups	Count	Sum	Average	Variance
24	SUT 1	10	150	15	13.3333
25	SUT 2	10	111	11.1	18.7667
26	SUT 3	10	135	13.5	14.0556

Figure 2.5. Single Factor summary statistics (using Data Analysis Add-in)

	E	F	G	H	I	J	K
28	ANOVA						
29	Source of Variation	SS	df	MS	F	P-value	F crit
30	Between Groups	77.4	2	38.7	2.5154	0.0996	3.3541
31	Within Groups	415.4	27	15.3852			
32							
33	Total	492.8	29				

Figure 2.6. ANOVA: Single Factor results (using Data Analysis Add-in)

That's pretty straight-forward using the Data Analysis add-in, but let's step back and look at what's really going on underneath the hood. Here's a definition to start with.

**Definition:** For one-way ANOVA, suppose we have  $k$  samples, which we will call groups (or treatments); these are the columns in our analysis (corresponding to the 3 flavors in the above example). We will use the index  $j$  for these. Each group consists of a sample of size  $n_j$ . The sample elements are the rows in the analysis. We will use the index  $i$  for these.

Suppose the  $j$ th group sample is  $\{x_{1j}, \dots, x_{nj}\}$  and so the total sample consists of all the elements  $\{x_{ij}: 1 \leq i \leq n_j, 1 \leq j \leq k\}$ .

We will use the abbreviation  $\bar{x}_j$  for the mean of the  $j$ th group sample (called the group mean) and  $\bar{x}$  for the mean of the total sample (called the total or grand mean).

Let the sum of squares for the  $j$ th group be

$$SS_j = \sum_i (x_{ij} - \bar{x}_j)^2$$

We now define the following terms:

$$SS_T = \sum_j \sum_i (x_{ij} - \bar{x})^2$$

$$SS_W = \sum_j SS_j = \sum_j \sum_i (x_{ij} - \bar{x}_j)^2$$

$$SS_B = \sum_j n_j (\bar{x}_j - \bar{x})^2$$

$SS_T$  is the sum of squares for the total sample, i.e., the sum of the squared deviations from the grand mean.  $SS_W$  is the sum of squares within the groups, i.e., the sum of the squared means across all groups.  $SS_B$  is the sum of the squares between group sample means, i.e., the weighted sum of the squared deviations of the group means from the grand mean,

where

$$n = \sum_{j=1}^k n_j$$

we also define the following degrees of freedom

$$df_T = n - 1, \quad df_B = k - 1, \quad df_{W \sum_{j=1}^k (n_j - 1)} = n - k$$

Finally, we define the mean square as  $MS = SS/df$

and so

$$MS_T = \frac{SS_T}{df_T}, \quad MS_B = \frac{SS_B}{df_B}, \quad MS_W = \frac{SS_W}{df_W},$$

We now summarize with **Table 2-4**:

**Table 2-4. Summary table for one-way ANOVA.**

	$df$	$SS$	$MS$
Total ( $T$ )	$n - 1$	$SS_T = \sum_j \sum_i (x_{ij} - \bar{x})^2$	$\frac{SS_T}{df_T}$
Between ( $B$ )	$k - 1$	$SS_B = \sum_j n_j (\bar{x}_j - \bar{x})^2$	$\frac{SS_B}{df_B}$
Within ( $W$ )	$n - k$	$SS_W = \sum_j \sum_i (x_{ij} - \bar{x}_j)^2$	$\frac{SS_W}{df_W}$

## $\Sigma$ -ing

In my years of teaching, writing, and doing math and stats, there is one thing that seems to get in the way, this symbol:  $\Sigma$ . It's just a harmless uppercase Greek sigma, or "S" for us, and it does not bite. Yet even in grad school, it often stumped me. But, let us not allow it to prevent us from testing. We call it a summation sign.

That object we labeled the sum of squares is so important, all of DOE and ANOVA, and Regression, and ... depend upon it. So, let's take a moment and get control over it.

Let's start with something easy

$$n = \sum_{j=1}^{k=3} n_j = n_1 + n_2 + n_3$$

This says take the n's with the j index for  $j = 1, 2$ , and  $3$ , and sum them. So, if we have three samples and each is of size 5 then, with  $n_1 = 5$ ,  $n_2 = 5$ , and  $n_3 = 5$ , we have  $n = 5 + 5 + 5 = 15$ .

But, wait! What if there are two Summation signs together?

$$SS_T = \sum_j \sum_i (x_{ij} - \bar{x})^2$$

Let's dissect this,  $\sum_i (x_{ij} - \bar{x})^2$ . Sum of squares is short for Sum of Squared Deviations and the squared deviations are just the difference between the observed value and the sample mean value. So, let's say we have three observed values 3.2, 4.1, and 3.5, and ignore the second index,  $j$ , for now. So, the mean is 3.6. Then the squared differences are  $(3.2 - 3.6)^2$ ,  $(4.1 - 3.6)^2$ , and  $(3.5 - 3.6)^2$ .

Now, we can't really ignore the  $j$  when we need the sum of squares total or the sum of squares within, so in the table we show  $i=3$  and  $j=2$

**Table 2-5. Calculating the summation of squared deviations.**

	j=1	j=2
i=1	3.2=x_1	3.3=x_1
i=2	4.1=x_2	3.7=x_2

i=3	3.5=x_3	3.1=x_3
$\sum_i$	$\sum_i(x_i - \bar{x})^2 =$ $(3.2 - 3.6)^2 + (4.2 - 3.6)^2 +$ $(3.5 - 3.6)^2$ =0.42	$\sum_i(x_i - \bar{x})^2 =$ $(3.3 - 3.4)^2 + (3.7 - 3.4)^2 +$ $(3.2 - 3.4)^2$ =0.14

Now, let's try the  $j$  index. That's just  $\sum_j = 0.42 + 0.14 = 0.56$  So, in this case,  $SS_T = \sum_j \sum_i (x_{ij} - \bar{x})^2 = 0.56$ . Hopefully, the SIGMA is demystified.

### Example 2.3: By-Hand Calculation of Example 2.2

Let's look at the same example problem, with the data

	A	B	C
5	DATA		
6	Control	SUT 1	SUT 2
7	13	12	11
8	17	15	19
9	19	8	15
10	11	16	14
11	20	10	18
12	15	14	16
13	18	10	18
14	9	4	11
15	12	6	14
16	16	11	11

Figure 2.7. The test data set in Excel format.

Here, we need some summary values to use in calculating the ANOVA table:

1. n, which we get by the COUNT of the columns
2. mean, which we get using the AVERAGE of the columns
3. sum-of-squares, which we get using DEVSQ. DEVSQ returns the sum of squares of deviations of data points from their sample means.

Technically, we only need the sum-of-squares and the counts. When Excel is calculating the sum-of-squares with DEVSQ, the calculations for the means are “built-in.”

	E	F	G	H	J
7	DESCRIPTION				
8	Groups	Count	Sum	Mean	SS
9	=A6	=COUNT(A7:A16)	=SUM(A7:A16)	=AVERAGE(A7:A16)	=DEVSQ(A7:A16)
10	Control	10	150	15	120
11	SUT 1	10	106	10.6	134.4
12	SUT 2	10	147	14.7	84.1

*Figure 2.8. The summary statistics we made need for calculating the ANOVA table.*

We rearrange and abbreviated **Table 2-4** as **Table 2-6**, as this aligns with the ANOVA table in Excel.

*Table 2-6. A rehash of Table 2-4 n the Excel format.*

	SS	df	MS
Between	$SS_B = SS_T - SS_W$	$k - 1$	$\frac{SS_B}{df_B}$
Within	$SS_W = SS_{Control} + SS_{SUT1} + SS_{SUT2}$	$n - k$	$\frac{SS_W}{df_W}$
Total	$SS_T = DEVSQ(all\ observations)$	$n - 1$	$\frac{SS_T}{df_T}$

In the manner below and using Figure 2.9, we can get the results of Figure 2.10:

1. Referring to Table 2-6.**Table 2-6**, we can use DEVSQ to get the total sum-of-squares,  $SS_T$ , of all the data using, DEVSQ(A7:C16).
2. To get  $SS_W$  we use SUM(J10:J12) from **Figure 2.8**.
3. We get  $SS_B$  by subtracting  $SS_W$  fro  $SS_T$ .
4. We get  $n$  by the total count, either COUNT(A7:C16) or SUM(F10:F12) from **Figure 2.8**.
5. We get  $k$  by the COUNT(F10:F112), yields  $k = 3$ .
6. Then we can get the  $dfs$  using  $n$  and  $k$ .

7. Referring to Table 2-6.**Table 2-6**, we divide the SSs by the  $dfs$  to get the  $MSs$ .

	E	F	G	H
14	ANOVA			Alpha
15	Sources	SS	df	MS
16	Between Groups	=F18-F17	=COUNTA(E10:E12)-1	=F16/G16
17	Within Groups	=SUM(J10:J12)	=G18-G16	n - k
18	Total	=DEVSQ(A7:C16)	=COUNT(A7:C16)-1	=F18/G18

Figure 2.9. Illustration of the ANOVA table calculations

	E	F	G	H
14	ANOVA			
15	Sources	SS	df	MS
16	Between Groups	120.8667	2	60.4333
17	Within Groups	338.5	27	12.537
18	Total	459.3667	29	15.8402

Figure 2.10. The finished ANOVA table.

#### Example 2.4: Power for ANOVA

We can use our ANOVA results to calculate the power ( $1-\beta$ ) of the test. The results are shown in **Figure 2.11**. Note that the calculation of  $\lambda$  (cell A30, **Figure 2.11**) is used to calculate  $\beta$ , the Type II error, and  $\beta$  is used to calculate the power ( $1-\beta$ ) in cell A34. However, the appropriate distribution needed is noncentral F distribution, so the power here is an approximation using F.DIST.RT, the right tail of the F distribution. If we used the 2-tail F value, the power (86.5%) is an over estimate. Our calculation resulting in 84.5% is a conservative estimate. This meets the default power value we desire (80%).

	A	B
21	<b>Power</b>	
22	dfB	2
23	dfE	27
24	SSB	120.8667
25	MSE	15.3852
26	n	30
27	k	3
28	f	0.51173
29	RMSSE	0.626739
30	$\lambda$	2
31	$\alpha$	0.05
32	F-crit	3.354131
33	$\beta$	0.154892
34	1- $\beta$	0.845108

*Figure 2.11. Test Power of the ANOVA .*

The formulas used in Excel are provided in **Table 2-7**.

*Table 2-7. Formulas used to calculate Power in Figure 2.11.*

Power Formulas		
dfB	B37	=G30
dfE	B38	=G31
SSB	B39	=F16
MSE	B40	=H31
n	B41	=B22+B23+1
k	B42	=B22+1
f	B43	=SQRT(B24/(B25*B26))
RMSSE	B44	=B28*SQRT(B27/(B27-1))
$\lambda$	B45	=F30/H30
$\alpha$	B46	=I14
F-crit	B47	=FINV(B31,B22,B23)
$\beta$	B48	=F.DIST.RT(B30,B22,B23)
1- $\beta$	B49	=1-B33

## Example 2.5: RBD Satellite Controllers

LEO Satellites are being deployed as constellations to better support users. A disadvantage of this is that it makes the satellites in greater risk of conjunction. This makes the type of controllers used on these satellites a critical design decision.

Referring to

Figure 2.12 we consider four alternative controllers rated from 0 to 100, the higher rating being better. The satellites labeled  $X_1$  and  $X_2$  is a new generation controller with considerable cost increases over the three legacy controllers,  $X_3$  and  $X_4$ .

- $X_1$ . An altitude-weighted controller that took altitude commands.
- $X_2$ . An altitude-weighted controller that took along-track commands.
- $X_3$ . An along-track weighted controller that took altitude commands.
- $X_4$ . An along-track weighted controller that took along-track commands.

The factors that are rated for these controllers include:

- tracking accuracy
- tracking speed
- response time
- overshooting
- along-track errors
- fuel consumption
- slot exiting

The  $X_1$  controller has performed better than all others in trials; however, its high cost must be considered in program budgets if its advantages do not clearly out-weigh other controllers.

Using a randomized block design, blocking on  $X_1$ , we will test the hypothesis that the  $X_1$  controller advantages do not out-weigh the others. That is,

$$H_0: \text{Controller } X_1 \text{ is not as effective as the alternatives}$$

The model for our Randomized Block Design is

$$y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$$

	A	B	C	D	E
1		X1	X2	X3	X4
2	Tracking accuracy	90	70	40	60
3	Tracking speed	95	60	30	50
4	Response time	85	85	40	60
5	Overshooting	80	70	40	80
6	along-track errors	85	70	35	60
7	fuel consumption	80	90	30	90
8	slot exiting	60	90	40	50

Figure 2.12. Table representing the randomized block design for the LEO satellite controllers

Figure 2.13 shows the summary statistics for the model

	G	H	I	J	K
1	Anova: Two-Factor Without Replication				
2					
3	SUMMARY	Count	Sum	Average	Variance
4	Tracking accuracy	4	260	65.000	433.333
5	Tracking speed	4	235	58.750	739.583
6	Response time	4	270	67.500	475.000
7	Overshooting	4	270	67.500	358.333
8	along-track errors	4	250	62.500	441.667
9	fuel consumption	4	290	72.500	825.000
10	slot exiting	4	240	60.000	466.667
11					
12	X1	7	575	82.143	123.810
13	X2	7	535	76.429	139.286
14	X3	7	255	36.429	139.286
15	X4	7	450	64.286	228.571

Figure 2.13. ANOVA table for Two-Factor Without Replication

**Figure 2.14** shows the ANOVA results revealing significant differences in the treatments (columns), but not among the rows.

	G	H	I	J	K	L	M
18	ANOVA						
19	Source of Variation	SS	df	MS	F	P-value	F crit
20	Rows	555.3571	6	92.55952	0.6584	0.6833	2.6613
21	Columns	8688.393	3	2896.131	20.6020	0.0000	3.1599
22	Error	2530.357	18	140.5754			
23	Total	11774.11	27				

*Figure 2.14. ANOVA results table*

The columns correspond to the treatments. We are really only interested in the columns factor, and see that there is a significant difference between the controllers (*p*-value = 0.0000).

**Table 2-8** provides the formulas used in the column and row summary statistics. **Table 2-9** provides the formulas used for the ANOVA results.

*Table 2-8. Formulas for column and row summary statistics.*

Name	Cell(s)	Formula
<b>Tracking accuracy Count</b>	H4	=COUNT(B2:E2)
<b>Tracking speed Count</b>	H5	=COUNT(B3:E3)
<b>Response time Count</b>	H6	=COUNT(B4:E4)
<b>Overshooting Count</b>	H7	=COUNT(B5:E5)
<b>Along-track errors Count</b>	H8	=COUNT(B6:E6)
<b>Fuel consumption Count</b>	H9	=COUNT(B7:E7)
<b>Slot exiting Count</b>	H10	=COUNT(B8:E8)
<b>Tracking accuracy Sum</b>	I4	=SUM(B2:E2)
<b>Tracking speed Sum</b>	I5	=SUM(B3:E3)
<b>Response time Sum</b>	I6	=SUM(B4:E4)
<b>Overshooting Sum</b>	I7	=SUM(B5:E5)
<b>Along-track errors Sum</b>	I8	=SUM(B6:E6)
<b>Fuel consumption Sum</b>	I9	=SUM(B7:E7)
<b>Slot exiting Sum</b>	I10	=SUM(B8:E8)
<b>Tracking accuracy Average</b>	J4	=AVERAGE(B2:E2)
<b>Tracking speed Average</b>	J5	=AVERAGE(B3:E3)

Name	Cell(s)	Formula
<b>Response time Average</b>	J6	=AVERAGE(B4:E4)
<b>Overshooting Average</b>	J7	=AVERAGE(B5:E5)
<b>Along-track errors Average</b>	J8	=AVERAGE(B6:E6)
<b>Fuel consumption Average</b>	J9	=AVERAGE(B7:E7)
<b>Slot exiting Average</b>	J10	=AVERAGE(B8:E8)
<b>Tracking accuracy Variance</b>	K4	=VAR(B2:E2)
<b>Tracking speed Variance</b>	K5	=VAR(B3:E3)
<b>Response time Variance</b>	K6	=VAR(B4:E4)
<b>Overshooting Variance</b>	K7	=VAR(B5:E5)
<b>Along-track errors Variance</b>	K8	=VAR(B6:E6)
<b>Fuel consumption Variance</b>	K9	=VAR(B7:E7)
<b>Slot exiting Variance</b>	K10	=VAR(B8:E8)
<b>X1 COUNT</b>	H12	=COUNT(B2:B8)
<b>X2 COUNT</b>	H13	=COUNT(C2:C8)
<b>X3 COUNT</b>	H14	=COUNT(D2:D8)
<b>X4 COUNT</b>	H15	=COUNT(E2:E8)
<b>X1 SUM</b>	I12	=SUM(B2:B8)
<b>X2 SUM</b>	I13	=SUM(C2:C8)
<b>X3 SUM</b>	I14	=SUM(D2:D8)
<b>X4 SUM</b>	I15	=SUM(E2:E8)
<b>X1 AVERAGE</b>	J12	=AVERAGE(B2:B8)
<b>X2 AVERAGE</b>	J13	=AVERAGE(C2:C8)
<b>X3 AVERAGE</b>	J14	=AVERAGE(D2:D8)
<b>X4 AVERAGE</b>	J15	=AVERAGE(E2:E8)
<b>X1 VAR</b>	K12	=VAR(B2:B8)
<b>X2 VAR</b>	K13	=VAR(C2:C8)
<b>X3 VAR</b>	K14	=VAR(C2:C8)
<b>X4 VAR</b>	K15	=VAR(E2:E8)

Table 2-9. Formulas for ANOVA results.

Name	Cell(s)	Formula
<b>Rows SS</b>	H20	=DEVSQ(I4:I10)/H10
<b>Columns SS</b>	H21	=DEVSQ(I12:I15)/H15
<b>Error SS</b>	H22	=H23-H21-H20

Total SS	H23	=DEVSQ(B2:E8)
Rows df	I20	=COUNT(B2:B8)-1
Columns df	I21	=COUNT(B2:E2)-1
Error df	I22	=I23-I21-I20
Total df	I23	=COUNT(B2:E8)-1
Rows MS	J20	=H20/I20
Columns MS	J21	=H21/I21
Error MS	J22	=H22/I22
Rows F	K20	=J20/J22
Columns F	K21	=J21/J22
Rows p-value	L20	=FDIST(K20,I20,\$I\$23)
Columns p-value	L21	=FDIST(K21,I21,\$I\$23)
Rows F-crit	M20	=FINV(0.05,I20,\$I\$22)
Columns F-crit	M21	=FINV(0.05,I21,\$I\$22)

### Example 2.6: RBD of Example 2.4 using Regression

Recall that we discussed the principle of DOE at the beginning of this chapter, randomization and blocking. In this example, we'll introduce the **Randomized Block Design** (RBD). To build the regression model for the controller data, we need to reformat the data in Error! Reference source not found. for use with Excel's Data Analysis add-in. The nuisance factor in this case is **orbital slot exiting**.

A	B	C	D	E
1	X1	X2	X3	X4
2 Tracking accuracy	90	70	40	60
3 Tracking speed	95	60	30	50
4 Response time	85	85	40	60
5 Overshooting	80	70	40	80
6 along-track errors	85	70	35	60
7 fuel consumption	80	90	30	90
8 slot exiting	60	90	40	50

Figure 2.15. Transposed controller data.

The model contains four treatment group variables  $X_1 = D_1$ ,  $X_2 = D_2$ , and  $X_3 = D_3$ , while  $X_4$  represents our nuisance variable (factor).

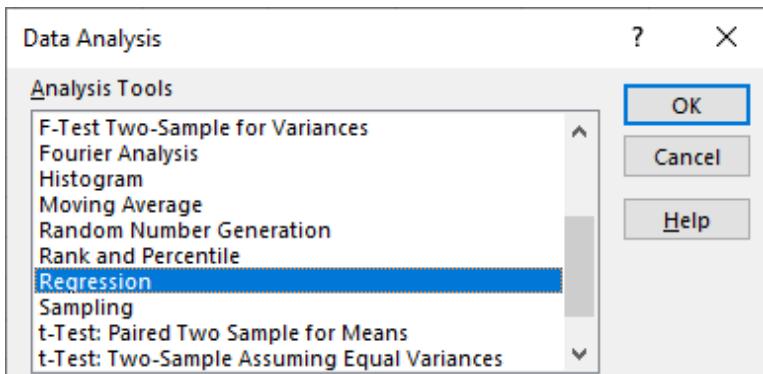
$$y = b_0 + b_1 D_1 + b_2 D_2 + b_3 D_3 + b_4 T_1 + b_5 T_2 + b_6 T_3 + b_7 T_4 \\ + b_8 T_5 + b_9 T_6 + b_{10} T_7$$

Next, we put the data into the Excel format (see **Figure 2.16**) required for performing regression analysis using the Excel Data Analysis add-in.

	I	J	K	L	M	N	O	P	Q	R
2	D1	D2	D3	T1	T2	T3	T4	T5	T6	Yield
3	0	0	0	1	0	0	0	0	0	90
4	0	0	0	0	1	0	0	0	0	95
5	0	0	0	0	0	1	0	0	0	85
6	0	0	0	0	0	0	1	0	0	80
7	0	0	0	0	0	0	0	1	0	85
8	0	0	0	0	0	0	0	0	1	80
9	0	0	0	0	0	0	0	0	0	60
10	1	0	0	1	0	0	0	0	0	70
11	1	0	0	0	1	0	0	0	0	60
12	1	0	0	0	0	1	0	0	0	85
13	1	0	0	0	0	0	1	0	0	70
14	1	0	0	0	0	0	0	1	0	70
15	1	0	0	0	0	0	0	0	1	90
16	1	0	0	0	0	0	0	0	0	90
17	0	1	0	1	0	0	0	0	0	40
18	0	1	0	0	1	0	0	0	0	30
19	0	1	0	0	0	1	0	0	0	40
20	0	1	0	0	0	0	1	0	0	40
21	0	1	0	0	0	0	0	1	0	35
22	0	1	0	0	0	0	0	0	1	30
23	0	1	0	0	0	0	0	0	0	40
24	0	0	1	1	0	0	0	0	0	60
25	0	0	1	0	1	0	0	0	0	50
26	0	0	1	0	0	1	0	0	0	60
27	0	0	1	0	0	0	1	0	0	80
28	0	0	1	0	0	0	0	1	0	60
29	0	0	1	0	0	0	0	0	1	90
30	0	0	1	0	0	0	0	0	0	50

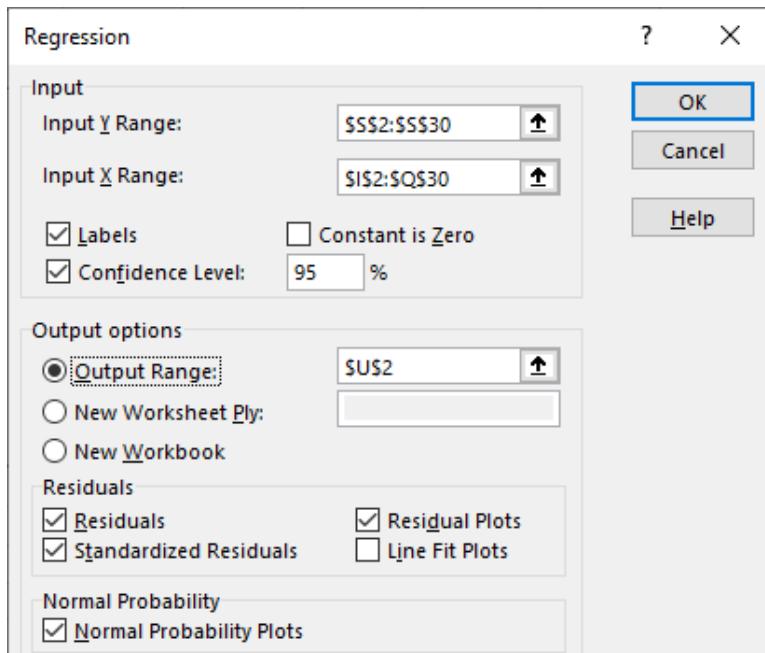
**Figure 2.16.** Controller data in Excel format for regression analysis.

For this analysis we used Excel's Data Analysis add-in and its Regression function, as shown in **Figure 2.17**. When we click okay, the dialog window shown in Error! Reference source not found. will appear.



**Figure 2.17.** Excel's Data Analysis add-in with the regression function selected.

Error! Reference source not found. shows how we input our data and modeling options. We populate the Input Y Range with \$R\$2:\$R\$30. For the Input X Range, we enter \$I\$2:\$Q\$30.



**Figure 2.18.** Regression settings for our analysis.

We included the headers in the data input ranges, so we check the Labels box and accept the default value for Confidence Level. We also set the output to appear in our current sheet starting in cell \$R\$3.

For output, we also want the Residuals, Standardized Residuals, and Normal Probability Plots; we'll make our own residuals plot.

When we finish the setup, we click the OK button and the regression runs. The results are shown in **Figure 2.19**, **Figure 2.20** and **Figure 2.21**.

T	U
2	SUMMARY OUTPUT
3	
4	<i>Regression Statistics</i>
5	Multiple R 0.88605382
6	R Square 0.78509138
7	Adjusted R Square 0.67763707
8	Standard Error 11.8564496
9	Observations 28

**Figure 2.19. Summary output**

In **Figure 2.19**, we highlight the Multiple R coefficient, which shows the regression model is a pretty good fit. **Figure 2.20** shows the ANOVA output. Note that the between and within sums of square is aggregated in the regression ANOVA. This significance result is the same.

T	U	V	W	X	Y
11	ANOVA				
12		df	SS	MS	F Significance F
13	Regression	9	9243.7500	1027.0833	7.3063 0.0002
14	Residual	18	2530.3571	140.5754	
15	Total	27	11774.1071		

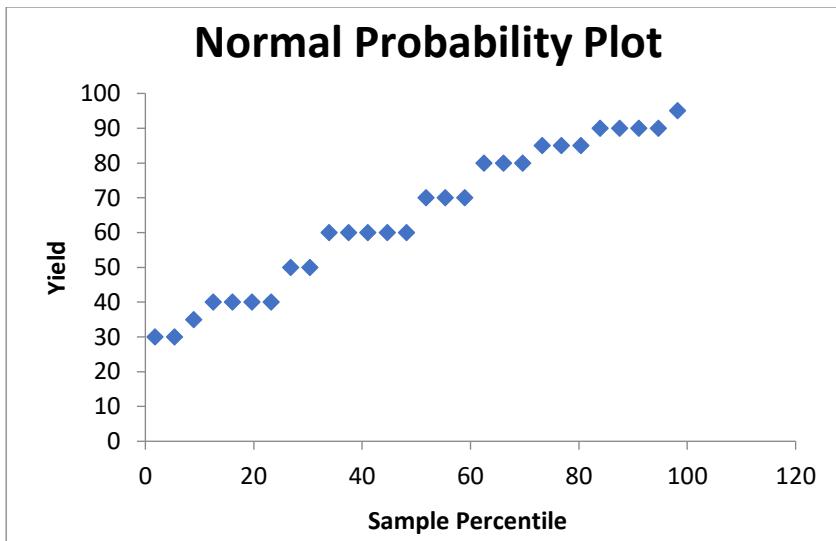
**Figure 2.20. ANOVA output**

**Figure 2.21** shows the regression output. The significant independent variables are the intercept,  $D_2$ , and  $D_3$ , at the 0.05-level.

	T	U	V	W	X	Y	Z	AA	AB
17		Coefficients	Std Err	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
18	Inter	77.3214286	7.0856	10.9125	0.00000	62.4352	92.2077	62.4352	92.2077
19	D1	-5.7142857	6.3375	-0.9017	0.37915	-19.0290	7.6004	-19.0290	7.6004
20	D2	-45.7142857	6.3375	-7.2133	0.00000	-59.0290	-32.3996	-59.0290	-32.3996
21	D3	-17.8571429	6.3375	-2.8177	0.01140	-31.1718	-4.5425	-31.1718	-4.5425
22	T1	5	8.3838	0.5964	0.55834	-12.6137	22.6137	-12.6137	22.6137
23	T2	-1.25	8.3838	-0.1491	0.88313	-18.8637	16.3637	-18.8637	16.3637
24	T3	7.5	8.3838	0.8946	0.38282	-10.1137	25.1137	-10.1137	25.1137
25	T4	7.5	8.3838	0.8946	0.38282	-10.1137	25.1137	-10.1137	25.1137
26	T5	2.5	8.3838	0.2982	0.76897	-15.1137	20.1137	-15.1137	20.1137
27	T6	12.5	8.3838	1.4910	0.15328	-5.1137	30.1137	-5.1137	30.1137

**Figure 2.21.** Regression output

**Figure 2.22** shows the normal probability plot and does not indicate any issues with the model.



**Figure 2.22.** Normal probability plot.

**Figure 2.23** shows the standard residuals plot that also does not indicate any issues with the regression model.

## Standard Residuals

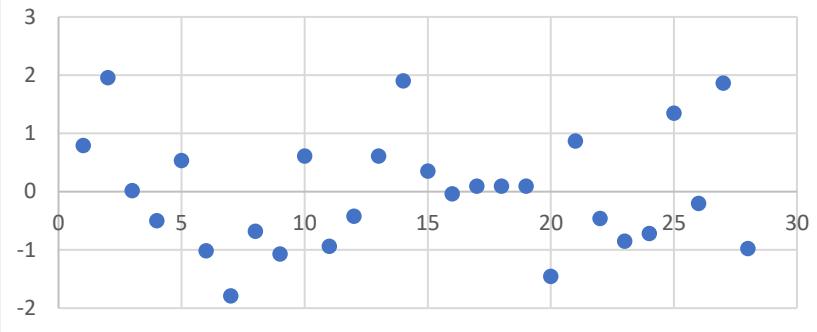


Figure 2.23. Standardized residuals plot.

### **3. ANOVA in Design of Experiments**

In this chapter we will review the application of ANOVA if DOE, starting with the steps for planning, conducting and analyzing an experiment, followed by applications of the simple and more complex part of DOE.

#### **Steps for Planning, Conducting and Analyzing an Experiment**

The practical steps needed for planning and conducting an experiment include: recognizing the goal of the experiment, choice of factors, choice of response, choice of the design, analysis and then drawing conclusions. This pretty much covers the steps involved in the scientific method.

1. Recognition and statement of the problem
2. Choice of factors, levels, and ranges
3. Selection of the response variable(s)
4. Choice of design
5. Conducting the experiment
6. Statistical analysis
7. Drawing conclusions, and making recommendations

What this course will deal with primarily is the choice of the design. This focus includes all the related issues about how we handle these factors in conducting our experiments.

#### **Experiments with a Single Factor**

##### **One-way ANOVA — Completely Randomized Design (CRD)**

A one-way ANOVA is a type of statistical test that compares the variance in the group means within a sample whilst considering only one independent variable or factor. It is a hypothesis-based test, meaning that it aims to evaluate multiple mutually exclusive theories about our data. Before we can generate a hypothesis, we need to have a question about our data that we want an answer to. A one-way ANOVA compares three or more than three categorical groups to establish whether there is a difference between them. Within each group there should be three or more observations, and the means of the samples are compared.

In a one-way ANOVA there are two possible hypotheses.

- The null hypothesis ( $H_0$ ) is that there is no difference between the groups and equality between means.
- The alternative hypothesis ( $H_1$ ) is that there is a difference between the means and groups.

What are the assumptions and limitations of a one-way ANOVA?

- Normality – that each sample is taken from a normally distributed population
- Sample independence – that each sample has been drawn independently of the other samples
- Variance equality – that the variance of data in the different groups should be the same

## One-Factor ANOVA

### Definitions

The model for the analysis of variance can be stated in two mathematically equivalent ways. We explain the model for a two-way ANOVA (the concepts are the same for additional factors). In the following discussion, each combination of factors and levels is called a cell. In the following, the subscript  $i$  refers to the level of factor 1,  $j$  refers to the level of factor 2, and the subscript  $k$  refers to the  $k^{th}$  observation within the  $(i, j)^{th}$  cell. For example,  $Y_{235}$  refers to the fifth observation in the second level of factor 1 and the third level of factor 2.

**Definition:** The first ANOVA model is defined by

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

This model decomposes the response,  $Y_{ijk}$ , into a mean for each cell,  $\mu_{ij}$ , and an error term,  $\varepsilon_{ijk}$ .

The analysis of variance provides estimates for each cell mean. These cell means are the predicted values of the model and the differences between the response variable and the estimated cell means are the

residuals. That is, the predicted values are given by,  $\hat{Y}_{ijk} = \hat{\mu}_{ij}$ , while the residuals are given by,  $R_{ijk} = Y_{ijk} - \hat{\mu}_{ij}$ .

**Definition:** The second ANOVA model is

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

This model decomposes the response into an overall (grand) mean,  $\mu$ , factor effects ( $\hat{\alpha}_i$  and  $\hat{\beta}_j$  represent the effects of the  $i^{th}$  level of the first factor and the  $j^{th}$  level of the second factor, respectively), and an error term,  $\varepsilon_{ijk}$ .

The ANOVA provides estimates of the grand mean and the factor effects. The predicted values are  $\hat{Y}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j$  and the residuals of the model are,  $R_{ijk} = Y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j$ .

The distinction between these models is that the second model divides the cell mean into an overall mean and factor effects. This second model makes the factor effect more explicit, so we will emphasize this approach.

## Model Validation

Note that the ANOVA model assumes that the error term,  $E_{ijk}$ , should follow the assumptions for a univariate measurement process. That is, after performing an analysis of variance, the model should be validated by analyzing the residuals.

### Example 3.1: One-Way ANOVA for Gear Data

A one-way analysis of variance was generated for the GEAR.DAT data set. The data set contains 10 measurements of gear diameter for ten different batches for a total of 100 measurements.

Engineering development and manufacturing for the aviation and space industries depend upon precision parts that function under extreme conditions. Satellites require gear parts for to meet the precision standards of ISO 9001 + AS9100. Whether it is standard catalog components such as couplings, bearings, and hardware to custom gears,

shafting, and intricate gear assemblies, attention is paid to every detail ensuring it will meet expectations for form, fit and function. The mechanical components and assemblies include:

- Spur Gears
- Helical Gears
- Worm Gears
- Planetary Gearheads
- Worm Drive Gearboxes

The database, GEAR.DAT, is comprised of data on the precision of gear diameters. The following is a summary of the data.

NUMBER OF OBSERVATIONS = 100

NUMBER OF VARIABLES = 2

ORDER OF VARIABLES ON A LINE IMAGE—

1. DIAMETER
2. BATCH

### **Example 3.2 The Gear Table**

The Gear data is in the GEAR Tab and is in column two-format. We inserted a column for Lookup Value (1 – 100) and used the excel function “VLOOKUP” used in Cells F3 to O12. Cells F1 to O1 contain a factor that increases by 10 and Cells E3 to E12 contain a factor that increases by 1. So, Cell G5 contains the value at the intersection of G1 (=10) and E5 (=3), or the  $10+3=13^{\text{th}}$  value from Column B. The VLOOKUP function does this by designating a lookup value in Column E starting in Row 3 (position 1), looks up that value in the first column of the designated array, \$A\$2:\$C\$101, and returns the value in the  $2^{\text{nd}}$  column of the array. When the formula is copied to the right to Cell G3, the lookup value changes to Row position 1 and column position 10, or the  $1+10=11^{\text{th}}$  value in Column B = 0.998. When the formula in Cell F3 is copied down to F4, the lookup value changes to row position 2 column and position 0, =r the  $2+0=2^{\text{nd}}$  value in Column B = 0.996. So, we copy the formula in cell F3 down to F12 and the over to O12. We need the table to be in this Excel format shown in **Figure 3.1** to run a Single Factor ANOVA using Excel’s Data Analysis add-in tool, shown in **Figure 3.2**.

	E	F	G	H	I	J	K	L	M	N	O
1		0	10	20	30	40	50	60	70	80	90
2	BATCH	1	2	3	4	5	6	7	8	9	10
3	1	1.006	0.998	0.991	1.005	0.998	1.009	0.99	0.998	1.002	0.991
4	2	0.996	1.006	0.987	1.002	0.998	1.013	1.004	1	0.998	0.995
5	3	0.998	1	0.997	0.994	0.982	1.009	0.996	1.006	0.996	0.984
6	4	1	1.002	0.999	1	0.99	0.997	1.001	1	0.995	0.994
7	5	0.992	0.997	0.995	0.995	1.002	0.988	0.998	1.002	0.996	0.997
8	6	0.993	0.998	0.994	0.994	0.984	1.002	1	0.996	1.004	0.997
9	7	1.002	0.996	1	0.998	0.996	0.995	1.018	0.998	1.004	0.991
10	8	0.999	1	0.999	0.996	0.993	0.998	1.01	0.996	0.998	0.998
11	9	0.994	1.006	0.996	1.002	0.98	0.981	0.996	1.002	0.999	1.004
12	10	1	0.988	0.996	0.996	0.996	0.996	1.002	1.006	0.991	0.997

Figure 3.1. Gear data in Excel format

Table 3-1. Formula used to construct the table shown in Figure 3.1

Name	Cell	Formula
BATCH 1	F3 to O12	=VLOOKUP(\$E3+F\$1,\$A\$2:\$C\$101,2)

We want to investigate the precision of the gears using batch (10 levels) as a single factor and diameter as the yield or response. We can do this using a single-factor ANOVA as seen in **Figure 3.3** and a null hypothesis:

$H_0$ : All individual batch means are equal.

$H_a$ : At least one batch mean is not equal to the others.

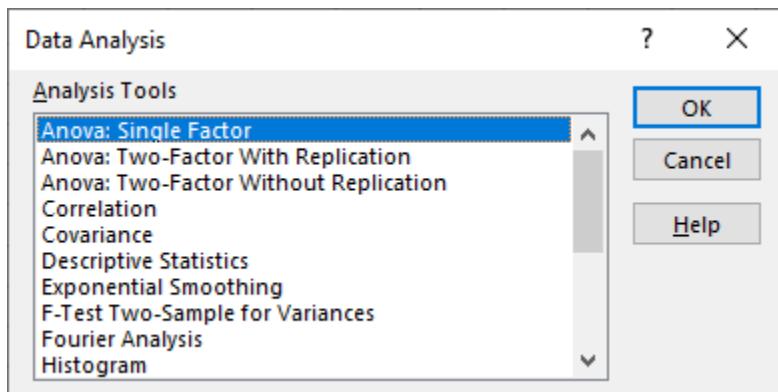


Figure 3.2. Data Analysis tool dialog window

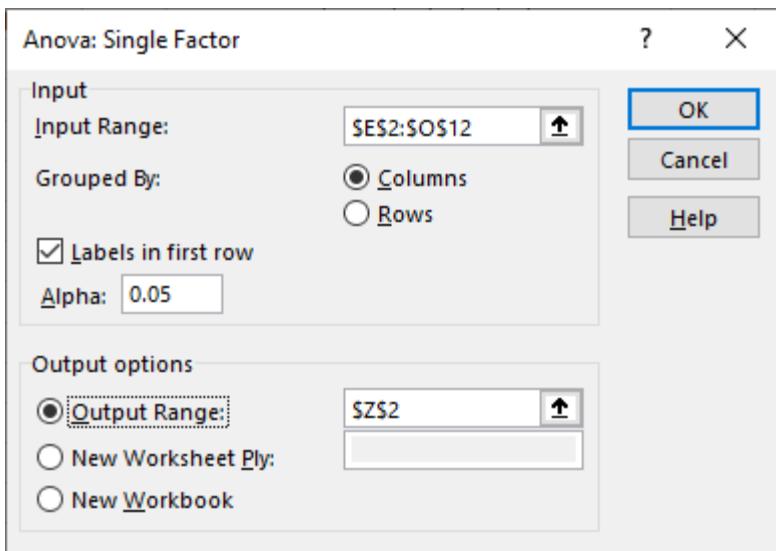


Figure 3.3. Single factor ANOVA window with Input Range entered, \$E\$2:\$O\$12, Grouped By Columns, Label in first row checked, Alpha 0.05 entered, and Output Range \$Z\$2 entered.

When the OK button is selected, the ANOVA is executed and produces the ANOVA Summary table shown as **Table 3-2** and the ANOVA results table shown as **Table 3-3**.

Table 3-2. Single factor ANOVA Data Summary Output

SUMMARY				
Groups	Count	Sum	Average	Variance
1	10	9.98	0.998	1.89E-05
2	10	9.991	0.9991	2.72E-05
3	10	9.954	0.9954	1.58E-05
4	10	9.982	0.9982	1.48E-05
5	10	9.919	0.9919	5.74E-05
6	10	9.988	0.9988	9.77E-05
7	10	10.015	1.0015	6.21E-05
8	10	10.004	1.0004	1.32E-05
9	10	9.983	0.9983	1.71E-05
10	10	9.948	0.9948	2.84E-05

**Table 3-3. Sige Factor ANOVA Results Table**

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.000729	9	8.1E-05	2.29691	0.022661	1.98560
Within Groups	0.003174	90	3.53E-05			
Total	0.003903	99			Reject Ho	

The F statistic is greater than the test critical value, so we can reject the null hypothesis and conclude that the diameters between batches is not consistent.

The ANOVA decomposes the variance into the following component sum of squares:

- **Total sum of squares.** The degrees of freedom for this entry is the number of observations minus one.
- **Sum of squares** for each of the factors. The degrees of freedom for these entries are the number of levels for the factor minus one. The mean square is the sum of squares divided by the number of degrees of freedom.
- **Residual sum of squares.** The degrees of freedom is the total degrees of freedom minus the sum of the factor degrees of freedom. The mean square is the sum of squares divided by the number of degrees of freedom.

The analysis of variance summarizes how much of the variance in the data (total sum of squares) is accounted for by the factor effects (factor sum of squares) and how much is due to random error (residual sum of squares). Ideally, we would like most of the variance to be explained by the factor effects. The ANOVA table provides a formal F test for the factor effects. To test the overall batch effect in our example we use the following hypotheses.

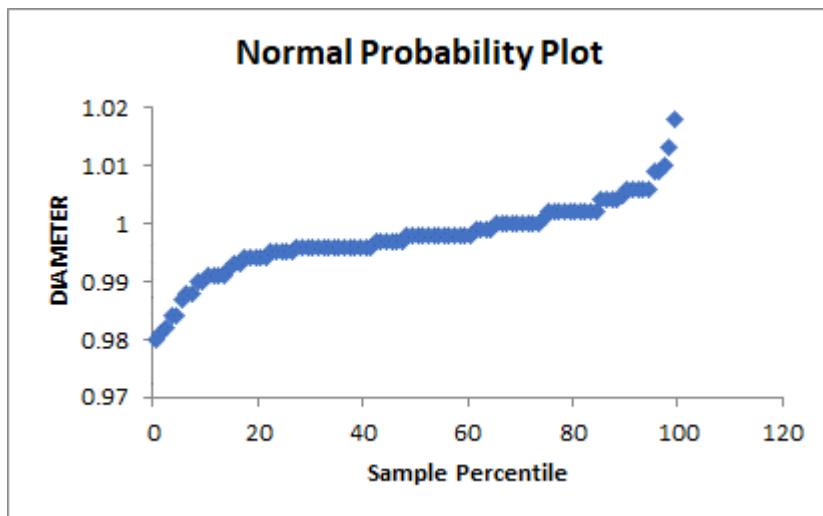
The F statistic is the mean square for the factor divided by the residual mean square. This statistic follows an F distribution with  $(k - 1)$  and  $(n - k)$  degrees of freedom where k is the number of levels for the given factor. Here, we see that the size of the "direction" effect dominates the

size of the other effects. For our example, the critical F value (upper tail) for  $\alpha = 0.05$ ,  $(k - 1) = 1$ , and  $(-k) = 475$  is 3.86111. Thus, "table speed" and "batch" are significant at the 5% level while "down feed rate" and "wheel grit size" are not significant at the 5 % level.

In addition to the quantitative ANOVA output, it is recommended that any analysis of variance be complemented with model validation. At a minimum, this should include

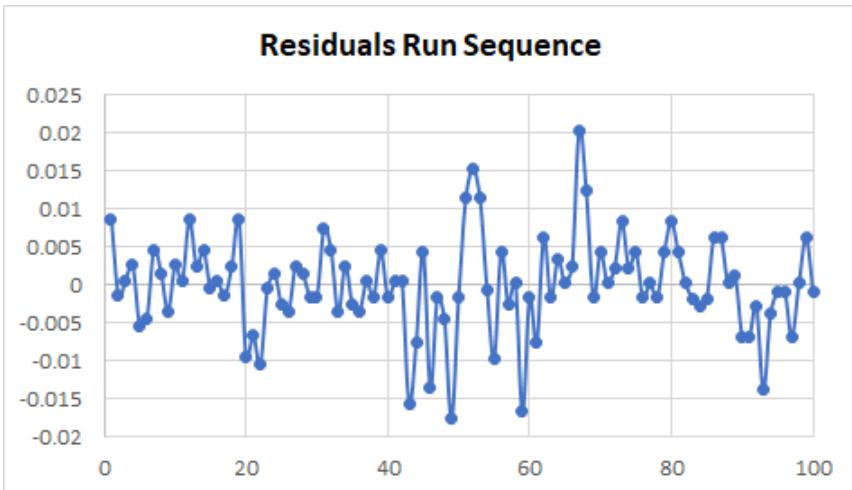
1. A normal probability plot of the residuals (**Figure 3.4**).
2. A run sequence plot of the residuals (**Figure 3.5**).
3. A scatter plot of the predicted values against the residuals (**Figure 3.6**).

We should check if the data are normal—they should be approximately normal—they should certainly have constant variance among the groups.



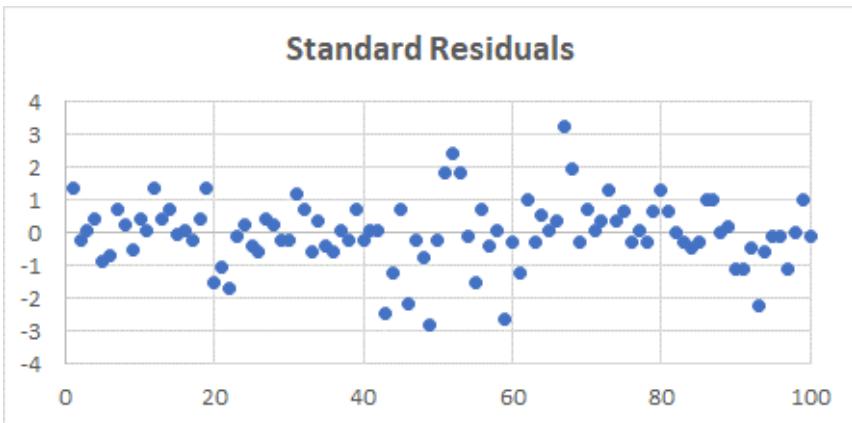
*Figure 3.4. Normal probability plot*

Independence is hard to check but plotting the residuals in the order in which the operations are done can sometimes detect if there is lack of independence. If the run-sequence plot does not show a pattern then we can assume the residuals independent.



**Figure 3.5.** Run-sequence chart of the standardized residuals

Let's examine the residuals, which are just the observations minus the predicted values, in this case, treatment means. Hence,  $e_{ij} = y_{ij} - \bar{y}_i$ .



**Figure 3.6.** Standardized residual plot

These plots don't look exactly normal but at least they don't seem to have any wild outliers. The normal scores plot looks reasonable. The residuals show significant outliers (those beyond  $\pm 3$  are not shown). This looks a little suspect as the positive residuals do not show the same. These are the kinds of clues that you look for... if you are conducting this experiment you would want to find out what was causing the irregularity.

## Tukey Honestly Significant Difference (HSD) Test

The idea behind the Tukey HSD (Honestly Significant Difference) test is to focus on the largest value of the difference between two group means. The relevant statistic is:

$$q = \frac{\bar{x}_{\max} - \bar{x}_{\min}}{\sqrt{MS_w/n}}$$

where  $n$  = the size of each of the group samples. The statistic  $q$  has a distribution called the studentized range  $q$ . The critical values for this distribution are presented in the Studentized Range  $q$  Table based on the values of  $\alpha, k$  (the number of groups) and  $df_W$ . If  $q > q_{crit}$  then the two means are significantly different.

This test is equivalent to

$$\bar{x}_{\max} - \bar{x}_{\min} > q_{crit} \sqrt{MS_w/n}$$

Picking the largest pairwise difference in means allows us to control the experiment-wise error rate for all possible pairwise contrasts; in fact, Tukey's HSD keeps experiment-wise  $\alpha = 0.05$  for the largest pairwise contrast, and is conservative for all other comparisons.

Note that the statistic  $q$  is related to the usual  $t$  statistic by  $q = \sqrt{2} t$ . Thus we can use the following  $t$  statistic

$$t = \frac{\bar{x}_{\max} - \bar{x}_{\min}}{\sqrt{2 \cdot MS_w/n}}$$

The critical value for  $t$  is now given by  $t_{crit} = q_{crit}/\sqrt{2}$ . If  $t > t_{crit}$  then we reject the null hypothesis that  $H_0: \mu_{\max} = \mu_{\min}$  similarly for other pairs.

As described above, to control type I error, we can't simply use the usual critical value for the distribution, but instead, use a critical value based on the largest difference of the means.

From these observations we can calculate confidence intervals in the usual way:

$$(\bar{x}_i - \bar{x}_j) \pm t_{crit} \sqrt{2 \cdot MS_w/n}$$

or equivalently

$$(\bar{x}_i - \bar{x}_j) \pm q_{crit} \sqrt{MS_w/n}$$

### Example 3.3: Testing a New Software System

Here, we'll implement a comparison test for the operational efficiency of a new software system vs a truth source (control system) to replace the current legacy system. Historical data for the legacy system's efficiency is also used for comparison. This test is accomplish using a one-way ANOVA design, with planned follow-on means comparisons using contrast testing and Tukey's HSD tests. The data in **Table 3-4** is already in Excel for use with The Analysis Tool.

*Table 3-4. Data for the comparison testing in Excel format*

DATA		
System A	System B	Control
3	2	3
5	4	5
6	3	3
4	5	4
5	3	5
6	4	3
4	2	2
3	3	2
7	6	4
4	2	4
5	4	2
2	1	3

For this test, we are investigating the null hypothesis:

$H_0$ : *There is no significant difference in the efficiencies of the systems*

The ANOVA summary of descriptive statistics is shown in **Table 3-5**. The formulas for this table appear in **Table 3-13**.

The ANOVA analysis in **Table 3-6** shows significance between the Systems, including the Control. The test statistic,  $F = 3.3396$  is greater than the critical value,  $F_{crit} = 3.2849$ . Therefore, we have enough evidence to reject the null hypothesis. However, the ANOVA does not tell us how the factors are significant, so we will perform follow-on tests using Tukey's HSD Test. The formulas for this table appear in **Table 3-14**.

**Table 3-5. ANOVA Summary table for the software efficiency experiment**

SUMMARY					alpha: 0.05
Groups	Count	Sum	Average	Variance	SS
System A Data	12	54	4.5000	2.0909	23.0000
System B Data	12	39	3.2500	2.0227	22.2500
Control	12	40	3.3333	1.1515	12.6667

**Table 3-6. ANOVA results table for the software efficiency experiment**

**ANOVA**

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	11.7222	2	5.8611	3.3396	4.778E-02	3.2849
Within Groups	57.9167	33	1.7551			

We've implemented Tukey's HSD Test in Excel. **Table 3-7** contains the parameters needed for looking up Q values in a lookup table. The table of Q values is located at Appendix. The Q critical function values are shown in **Table 3-8**.

The table also shows the Q critical difference values that we'll use in an alternative for of the test. The formulas for these tables appear in **Table 3-17** and **Table 3-18**.

**Table 3-7. Q value parameter**

$$Q = (\text{Mean}_{\text{Largest}} - \text{Mean}_{\text{Smallest}})/SE$$

$$SE = \text{SQRT}(\text{MS}_{\text{within}}/n)$$

Q Test Table Lookup	
$\alpha$	0.05
df	33
k	3

**Table 3-8. Q critical values**

	Q Critical function		Q critical difference	
Level ( $\alpha = 0.05$ )	q-crit	3.470	diff-crit	1.3271
Level ( $\alpha = 0.01$ )	q-crit	4.422	diff-crit	1.6912

The first Tukey test is for comparison of means between System A and System B. The null hypothesis for this test if:

*H<sub>0</sub>: The means of System A and System B are not different.*

The test results in **Table 3-9** show that there is a significant difference between the efficiencies of System A and System B, so we can conclude that System A is more efficient than System B. The formulas used for this table appear in **Table 3-15**.

**Table 3-9. Tukey's HSD Test Results for comparison of System A-System B**

Tukey's HSD	$\mu_{WD} - \mu_{WC}$							
mean	std error	q-stat	q-crit	t-crit adj	sig	lower	upper	
3.2686	0.38243	8.54676	3.470	2.4538	yes		2.3301	4.2070

The second Tukey test is for comparison of means between System A and the control. The null hypothesis for this test if:

*H<sub>0</sub>: The means of System A and System B are not different.*

The test results in **Table 3-10** show that there is a significant difference between the efficiencies of System A and Control, so we can conclude that System A is more efficient than Control. The formulas used for this table appear in **Table 3-16**.

**Table 3-10. Tukey's HSD Test Results for comparison of System A-Control**

Tukey's HSD	$\mu_{WD} - \mu_{MC}$							
mean	std error	q-stat	q-crit	t-crit adj	sig	lower	upper	
3.0507	0.38243	7.97698	3.470	2.4538	yes		2.1122	3.9891

The next Tukey's HSD Test repeats the same test employs the Q critical difference values, but yields the same results as shone in **Table 3-11**. The formulas for this table appear in **Table 3-19**.

**Table 3-11. Tukey's HSD at  $\alpha = 0.05$**

	<i>mean</i>	<i>diff-crit</i>	<i>sig</i>
<i>System A-System B</i>	3.2686	1.3271	Yes
<i>System A-Control</i>	3.0507	1.3271	Yes
<i>System B-Control</i>	0.2179	1.3271	No

The next Tukey's HSD Test repeats the same test employs the Q critical difference value at the 0.01 level of significance, but yields the same results as shown in **Table 3-12**. This allows us to reject the null hypotheses with a 99% level of confidence. The formulas for this table appear in **Table 3-20**.

**Table 3-12. Tukey's HSD at  $\alpha = 0.01$**

	<i>mean</i>	<i>diff-crit</i>	<i>sig</i>
<i>System A-System B</i>	3.2686	1.6912	Yes
<i>System A-Control</i>	3.0507	1.6912	Yes
<i>System B-Control</i>	0.2179	1.6912	No

**Table 3-13. Formulas used for the One-Way ANOVA Summary Statistics**

<b>Factors</b>	<b>Cells</b>	<b>Formulas</b>
<i>Count<sub>SysA</sub></i>	G10	=COUNT(A7:A100)
<i>Count<sub>SysB</sub></i>	G11	=COUNT(B7:B100)
<i>Count<sub>Control</sub></i>	G12	=COUNT(C7:C100)
<i>Sum<sub>SysA</sub></i>	H10	=SUM(A7:A100)
<i>Sum<sub>SysB</sub></i>	H11	=SUM(B7:B100)
<i>Sum<sub>Control</sub></i>	H12	=SUM(C7:C100)
<i>Mean<sub>SysA</sub></i>	I10	=AVERAGE(A7:A100)
<i>Mean<sub>SysB</sub></i>	I11	=AVERAGE(B7:B100)
<i>Mean<sub>Control</sub></i>	I12	=AVERAGE(C7:C100)
<i>Var<sub>SysA</sub></i>	J10	=VAR(A7:A100)
<i>Var<sub>SysB</sub></i>	J11	=VAR(B7:B100)
<i>Var<sub>Control</sub></i>	J12	=VAR(C7:C100)
<i>DevSq<sub>SysA</sub></i>	K10	=DEVSQ(A\$7:A\$18)
<i>DevSq<sub>SysB</sub></i>	K11	=DEVSQ(B\$7:B\$18)
<i>DevSq<sub>Control</sub></i>	K12	=DEVSQ(C\$7:C\$18)

**Table 3-14.** Formulas used for the One-Way ANOVA results

Factors	Cells	Formulas
$SS_{Between}$	G16	=G19-G17
$SS_{Within}$	G17	=SUM(K10:K12)
$df_{Between}$	H16	=COUNTA(A6:C6)-1
$df_{Within}$	H17	=H19-H16
$MS_{Between}$	I16	=G16/H16
$MS_{Within}$	I17	=G17/H17
$F_{stat}$	J16	=I16/I17
$p - value$	K16	=FDIST(J16,H16,H17)
$F_{crit}$	L16	=FINV(K8,H16,H17)
$SS_{Total}$	G19	=DEVSQ(A7:C18)
$df_{Total}$	H19	=COUNTA(A7:C18)-1

**Table 3-15.** Tukey's HSD Test results for System A vs System B

System A - System B		
<b>Mean</b>	N9	=ABS(I10-I11)/O9
<b>Std Error</b>	O9	=SQRT(\$I\$17/\$G\$10)
$q_{stat}$	P9	=N9/O9
$q_{crit}$	Q9	=P17
$t_{crit,adj}$	R9	=Q9/SQRT(2)
<b>Significance</b>	S9	=IF(ABS(P9)>R9,"yes","no")
<b>Lower CI</b>	T9	=N9-O9*R9
<b>Upper CI</b>	U9	=N9+O9*R9

**Table 3-16.** Tukey's HSD Test results for System A vs Control

System A - Control		
<b>Mean</b>	N13	=ABS(I10-I12)/O13
<b>Std Error</b>	O13	=SQRT(\$I\$17/\$G\$10)
$q_{stat}$	P13	=N13/O13
$q_{crit}$	Q13	=Q9
$t_{crit,adj}$	R13	=Q13/SQRT(2)
<b>Significance</b>	S13	=IF(ABS(P13)>=R13,"yes","no")
<b>Lower CI</b>	T13	=N13-O13*R13
<b>Upper CI</b>	U13	=N13+O13*R13

**Table 3-17.** Q critical function value lookup formulas

Q Critical function		
$\alpha = 0.05 q_{crit}$	Cell P17	=VLOOKUP(L23,'Stud. Q Table 2'!\$A\$65:\$W\$185,L24,TRUE)
$\alpha = 0.01 q_{crit}$	Cell P18	=VLOOKUP(L23,'Stud. Q Table'!\$A\$105:\$T\$225,L24,TRUE)

**Table 3-18.** Q critical difference value lookup formulas

Q critical difference		
$\alpha = 0.05 diff_{crit}$	Cell U17	=P17*SQRT(\$I\$17/\$G\$10)
$\alpha = 0.01 diff_{crit}$	Cell U18	=P18*SQRT(\$I\$17/\$G\$10)

**Table 3-19.** Alternative Tukey's HSD Test at  $\alpha = 0.05$

Tukey's HSD at $\alpha = 0.05$		
$\mu_{SysA} - \mu_{SysB}$	Cell O22	=ABS(I9-I10)/O9
$\alpha = 0.05 diff_{crit}$	Cell P22	=\$U\$17
<b>Significance</b>	Cell Q22	=IF(O22>P22,"Yes","No")
$\mu_{SysA} - \mu_{Control}$	Cell O23	=ABS(I9-I11)/O13
$\alpha = 0.05 diff_{crit}$	Cell P23	=\$U\$17
<b>Significance</b>	Cell Q23	=IF(O23>P23,"Yes","No")
$\mu_{SysB} - \mu_{Control}$	Cell O24	=ABS(I10-I11)/O13
$\alpha = 0.05 diff_{crit}$	Cell P24	=\$U\$17
<b>Significance</b>	Cell Q24	=IF(O24>P24,"Yes","No")

**Table 3-20.** Alternative Tukey's HSD Test at  $\alpha = 0.01$

Tukey's HSD at $\alpha = 0.01$		
$\mu_{SysA} - \mu_{SysB}$	Cell T22	=ABS(I9-I10)/O9
$\alpha = 0.05 diff_{crit}$	Cell U22	=\$U\$18
<b>Significance</b>	Cell V22	=IF(T22>U22,"Yes","No")
$\mu_{SysA} - \mu_{Control}$	Cell T23	=ABS(I9-I11)/O13
$\alpha = 0.05 diff_{crit}$	Cell U23	=\$U\$18
<b>Significance</b>	Cell V24	=IF(T23>U23,"Yes","No")
$\mu_{SysB} - \mu_{Control}$	Cell T24	=ABS(I10-I11)/O13
$\alpha = 0.05 diff_{crit}$	Cell U25	=\$U\$18
<b>Significance</b>	Cell V25	=IF(T24>U24,"Yes","No")

**Figure 3.7** shows the Excel spreadsheet for the one-way ANOVA and **Figure 3.8** shows the Excel spreadsheet for the Tukey's HSD Test.

	F	G	H	I	J	K	L
7	Anova: Single Factor						
8	SUMMARY				alpha: 0.05		
9	Groups	Count	Sum	Average	Variance	SS	
10	System A	12	54	4.5000	2.0909	23.0000	
11	System B	12	39	3.2500	2.0227	22.2500	
12	Control	12	40	3.3333	1.1515	12.6667	
13							
14	ANOVA						
15	Source of Variation	SS	df	MS	F	P-value	F crit
16	Between Groups	11.722	2	5.8611	3.3396	4.778E-02	3.2849
17	Within Groups	57.917	33	1.7551			
18							
19	Total	69.639	35			sig: TRUE	

**Figure 3.7. Excel spreadsheet for the one-way ANOVA**

	N	O	P	Q	R	S	T	U	V
5	Unplanned Comparisons				$q_{stat} = \frac{\bar{x}_{max} - \bar{x}_{min}}{\sqrt{MS_w/n}}$				
6	System A-System B								
7	Tukey's HSD			$\mu_{WD} - \mu_{WC}$					
8	mean	std error	q-stat	q-crit	t-crit adj	sig	lower	upper	
9	3.2686	0.38243	8.5468	3.47	2.4538	yes	2.3301	4.2070	
10	System A-Control								
11	Tukey's HSD			$\mu_{WD} - \mu_{MC}$					
12	mean	std error	q-stat	q-crit	t-crit adj	sig	lower	upper	
13	3.0507	0.38243	7.9770	3.47	2.4538	yes	2.1122	3.9891	
14									
15		Q Critical function				Q critical difference			
16	Level ( $\alpha = 0.05$ )	q-crit	3.470			diff-crit	1.327		
17	Level ( $\alpha = 0.01$ )	q-crit	4.422			diff-crit	1.691		
18									
19		Tukey's HSD at $\alpha = 0.05$				Tukey's HSD at $\alpha = 0.01$			
20		mean	diff-crit	sig		mean	diff-crit	sig	
21	System A-System B	3.2686	1.3271	Yes	System A-System B	3.2686	1.6912	Yes	
22	System A-Control	3.0507	1.3271	Yes	System A-Control	3.0507	1.6912	Yes	
23	System B-Control	0.2179	1.3271	No	System B-Control	0.2179	1.6912	No	

**Figure 3.8. Excel spreadsheet for the Tukey's HSD Test**

## Dunnett's Test

Dunnett's test is used when we want to compare one group (usually the control treatment) with the other groups, by interval estimation or hypothesis testing, all active treatments with a control when sampling from a distribution where the normality assumption is reasonable.

**Definition.** In Dunnett's test, we check whether

$$|\bar{x}_j - \bar{x}_0| > t_d \sqrt{\frac{2MS_w}{n}}$$

where  $n$  = size of the group samples,  $\bar{x}_0$  = mean of the control group,  $\bar{x}_j$  is the mean of any other group and  $t_d$  is the (two-tailed) Dunnett's critical value given in the Dunnett's Table.  $MS_w$  is the mean squares of within group of the ANOVA output.

Unlike Tukey's HSD, which considered all kinds of pairwise comparisons, Dunnett's test only compares one group with the others, addressing a special case of multiple comparisons problem — pairwise comparisons of multiple treatment groups with a single control group. Because its comparisons are more focused, it is considered a stronger test. The

Dunnet's critical value table contains the values  $t_d(k, df_W, \alpha)$  where  $k$  = the number of groups (treatments) including the control (see Appendix A, Dunnet's Critical Values fro  $\alpha = 0.01$  and  $\alpha = 0.05$ ;  $\alpha = 0.10$  is included in the Excel workbook).

### Example 3.4: Dunnett's Ad Hoc Test

We perform the test on the data from the previous example. The data is in cells A5:C12, as shown in **Figure 3.9**. This example is in the Excel workbook I tab *Dunnett's Test*. So, we test he null hypothesis:

$$H_0: \mu_{System1} = \mu_{System2} = \mu_{Control}$$

Assuming that Control (C5:C12) is the control group, we perform one-way ANOVA for comparison/validation of Between group significance (using SS between and MS between), resulting in the output in the top of **Figure 3.10**.

	A	B	C
3	DATA		
4	System 1	System 2	Control
5	3	2	3
6	5	4	5
7	6	3	3
8	4	5	4
9	5	3	5
10	6	4	3
11	4	2	2
12	3	3	2
13	7	6	4
14	4	2	4
15	5	4	2
16	3	1	3

Figure 3.9. Data for the experiment

Then, we compare the means of this method with the means using Dunnett's test. Since we do have significance,  $t_d > d\text{-crit}$  we reject the null hypothesis. The table of Dunnett's,  $t_d$  values are in Appendix A.

	H	I	J	K	L	M	N
3	ANOVA						
4	Source of Variation	SS	df	MS	F	P-value	F crit
5	Between Groups	13.3889	2	6.6944	4.1037	0.0205	3.2849
6	Within Groups	53.8333	33	1.6313			
7	Total	67.2222	35				
8							
9	DUNNETT'S TEST			alpha	0.05		
10	group	mean	size	ss	df	$t_d$	$d\text{-crit}$
11	System 1	4.5833	12	18.917			
12	System 2	3.25	12	22.25			
13	Control	3.3333	12	12.6667			
14							
15				36	53.8333	33	2.50
							1.3036

Figure 3.10. ANOVA output and Dunnett's Test for comparing means between groups

Finally, we perform t-tests to determine which systems and Control are significantly different. The null hypotheses are:

$$H_0: \mu_{System1} = \mu_{Control}$$

and

$$H_0: \mu_{System2} = \mu_{Control}$$

The result of the t-tests in **Figure 3.11** show that the means of System 1 and the Control are significantly different (i.e., we reject the null hypothesis), and the means of System 2 and Control are not significantly different (i.e., we fail to reject the null hypothesis).

T-TEST									
group	mean	std err	d-stat	lower	upper	p-value	mean-crit	Cohen d	Sig
System 1	1.25	0.52143	2.39727	-1.25	3.75	0.0432	0.6797	0.9787	Yes
System 2	0.0833	0.52143	0.15982	-2.4167	2.5833	0.3614	0.6797	0.0652	No

**Figure 3.11.** Follow-on t-test for comparing pairs of group means

The Excel formulas for each test follow in **Table 3-21** through **Table 3-23**

**Table 3-21.** Excel formulas for ANOVA table

Name	Cell	Formula
$SS_{Between}$	I5	=I7-I6
$SS_{Within}$	I6	=SUM(K11:K13)
$SS_{Total}$	I7	=DEVSQ(A5:C16)
$df_{Between}$	J5	=COUNTA(A4:C4)-1
$df_{Within}$	J6	=J7-J5
$df_{Total}$	J7	=COUNT(A5:C16)-1
$MS_{Between}$	K5	=I5/J5
$MS_{Within}$	K6	=I6/J6
$F_{Stat}$	Lf	=K5/K6
$p - Value$	M5	=F.DIST(L5,J5,J6,FALSE)
$F_{Crit}$	N5	=F.INV.RT(L9,J5,J6)

**Table 3-22. Excel formulas for Dunnett's test**

Name	Cell	Formula
$\mu_{System1}$	I11	=AVERAGE(A5:A16)
$\mu_{System2}$	I12	=AVERAGE(B5:B16)
$\mu_{Control}$	I13	=AVERAGE(C5:C16)
$n_{System1}$	J11	=COUNT(A5:A16)
$n_{System2}$	J12	=COUNT(B5:B16)
$n_{Control}$	J13	=COUNT(C5:C16)
$ss_{System1}$	K11	=DEVSQ(A5:A16)
$ss_{System2}$	K12	=DEVSQ(B5:B16)
$ss_{Control}$	K13	=DEVSQ(C5:C16)
$n_{Total}$	J15	=SUM(J11:J14)
$ss_{Total}$	K15	=SUM(K11:K14)
$df_{Total}$	L15	=J15-COUNT(J11:J14)
$t_d$	M15	=VLOOKUP(L15,'Dunnett Table'!A56:U101,COUNT(Sheet2!J11:J13))
$d\text{-}Crit$	N15	=ABS(M15*SQRT((2*\$K\$6)/COUNT(\$A\$5:\$A\$16 ))))

**Table 3-23. Excel formulas for t-test**

Name	Cell	Formula
<i>mean</i>	I19	=ABS(I\$13-I11)
<i>Std err</i>	J19	=SQRT(K\$15/L\$15*(1/J\$11+1/J13))
<i>d-stat</i>	K19	=ABS(I19/J19)
<i>lower</i>	L19	=I19-\$M\$15
<i>upper</i>	M19	=I19+\$M\$15
<i>p-value</i>	N19	=T.DIST(K19,3,FALSE)
<i>mean-crit</i>	O19	=J19*N\$15
<i>Cohen d</i>	OP19	=ABS(I19)*SQRT(L\$15/K\$15)
<i>Significance</i>	Q19	=IF(N19<L9,"Yes","No")

The **Kruskal-Wallis H test** is a non-parametric test that is used in place of a one-way ANOVA. Essentially it is an extension of the Wilcoxon Rank-Sum test to more than two independent samples.

Although, as explained in Assumptions for ANOVA, one-way ANOVA is usually quite robust, there are many situations where the assumptions are sufficiently violated and so the Kruskal-Wallis test becomes quite useful: in particular, when:

- Group samples strongly deviate from normal; this is especially relevant when sample sizes are small and unequal and data are not symmetric.
- Group variances are quite different because of the presence of outliers

If the assumptions of ANOVA are satisfied, then the Kruskal-Wallis test is less powerful than ANOVA, and so you should use ANOVA. This is also the case when a transformation can be used to meet the ANOVA assumptions. When the homogeneity assumption fails, Welch's ANOVA is often preferred over the Kruskal-Wallis test.

Some characteristics of Kruskal-Wallis test are:

- The assumptions are similar to those for the Mann-Whitney test: independent group samples, data in each group is randomly selected and data is at least ordinal
- No assumptions are made about the type of underlying distribution, although see below
- Each group sample has at least 5 elements.
- No population parameters are estimated, and so there are no confidence intervals.

The Kruskal-Wallis test is actually testing the null hypothesis that the populations from which the group samples are selected are equal in the sense that none of the group populations is **dominant** over any of the others. A group is dominant over the others if when one element is drawn at random from each of the group populations, it is more likely that the largest element is in that group.

$$H_0: \text{the group populations have equal dominance}$$

For instance, when one element is drawn at random from each group population, the largest (or smallest, or second smallest, etc.) element is equally likely to come from any one of the group populations

*H<sub>1</sub>: At least one of the groups is dominant over the others*

When the group samples have the same shape (and so presumably this is reflective of the corresponding population distributions), then the null hypothesis can be viewed as a statement about the group medians.

*H<sub>0</sub>: the group population medians are equal*

*H<sub>1</sub>: the group population medians are not equal*

An indication that the population distributions have the same shape (except that possibly there is a shift to the right or left among them) is that the box plots are similar, except that the box and whiskers among them may be at different heights. Another indication is that the group histograms or QQ plots look similar (although not necessarily indicating normality).

We define the test statistic as follow.

**Definition.** The test statistic,  $H$ , is defined as

$$H = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(n+1)$$

where  $k$  = the number of groups,  $n_j$  is the size of the  $j$ th group,  $R_j$  is the rank sum for the  $j$ th group and  $n$  is the total sample size, i.e.

$$n = \sum_{j=1}^k n_j$$

Then

$$H \sim \chi^2(k-1)$$

provided  $n_j \geq 5$  based on the following null hypothesis:

$H_0$ : The distribution of scores is equal across all groups

An alternative expression for  $H$  is given by

$$H = \frac{12}{n^3 - n} SS'_B$$

where  $SS'_B$  is the sum of squares between groups using the ranks instead of raw data. This is based on the fact that  $\frac{12(k-1)}{n(n+1)}$  is the expected value (i.e., mean) of the distribution of  $SS'_B$ .

### Example 3.5: Kruskal-Wallis H test

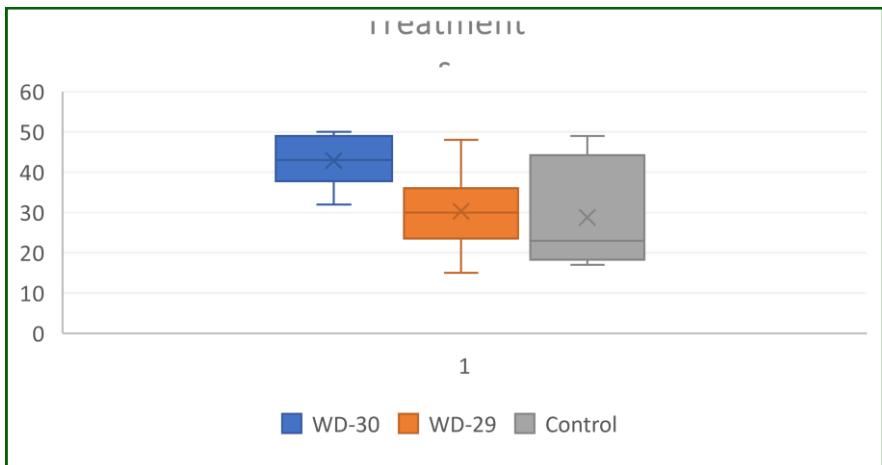
We return to our WD-40 problem, but this time there are only one group of silos. Additionally, when the test was performed, one sample from the previous formula, WD-29, was invalid and the results cannot be used. From the control group, two samples could not be used. The data from the test are shown in **Figure 3.12**.

	A	B	C
3	DATA		
4	WD-30	WD-29	Control
5	156	133	145
6	177	156	160
7	147	112	176
8	146	100	120
9	181	168	155
10	144	130	144
11	183	132	156
12	147	115	143
13	149	140	
14	161		

**Figure 3.12.** Data from the WD experiment

We now face a situation where the samples may not come from population with normal distributions, so we use a boxplot to develop some insight. The boxplot in **Figure 3.13**, shows that the WD-30 sample

is probably not from a normal distribution and the control sample is not, likewise. The WD-29 sample may be from a normal distribution, but we cannot conclude that it is based on the boxplot.



**Figure 3.13.** Boxplot of the three treatments in the WD experiment.

There are several tests we could perform for normality, but these are difficult to implement in Excel. Due to their complexity, we just show the results here in **Figure 3.14**.

	E	F	G	H
3	Shapiro-Wilk Normality Test			
4		WD-30	WD-29	Control
5	W-stat	0.82183036	0.027601	0.041238
6	W-crit	0.829	0.818	0.803
7	alpha	0.05	0.05	0.05
8	normal	No	No	No

**Figure 3.14.** Results of the Shapiro-Wilk Normality test.

Now, we have a situation where we might use the Kruskal-Wallis H test. **Figure 3.15** shows the results. The formulas used for **Figure 3.15**, are provided in **Table 3-24**. The outcome for the test shows that we can reject the null hypothesis, that is, the difference in means is statistically significant. To determine which of the formulas' means are significantly

different, we perform follow-on tests like we did with single-factor ANOVA.

	A	B	C	D	E
16	<b>Kruskal-Wallis test</b>				
17	mean	159.1	131.78	149.875	440.753
18	var	242.1	456.6944444	263.8392857	962.634
19	median	152.5	132	150	434.5
20	Rank Sums R	187.5	74.0	116.5	378
21	Group Size n	10	9	8	27
22	R <sup>2</sup> /n	3515.625	608.444	1696.53125	5820.6
23	H				8.39049
24	df				2
25	p				0.01507
26	$\alpha$				0.05
27	sig				yes

Figure 3.15. Output of the non-parametric Kruskal-Wallis test.

Table 3-24. Excel formulas used for the Kruskal-Wallis test shown in Figure 3.15

Cell Name	Cell(s)	Formula
mean	B17 to D17	=AVERAGE(A5:A14)
var	B18 to D18	=VAR(A5:A14)
median	B19 to D19	=MEDIAN(A5:A14)
Rank Sums R	B20 to D20	=B40
Group Size n	B21 to D21	=COUNT(A5:A14)
R <sup>2</sup> /n	B22 to D22	=B20^2/B21
sum of means	E17	=SUM(B17:D17)
sum of vars	E18	=SUM(B18:D18)
sum of medians	E19	=SUM(B19:D19)
sum of rank sums	E20	=SUM(B20:D20)
sum of group sizes	E21	=SUM(B21:D21)
sum of R <sup>2</sup> /n	E22	=SUM(B22:D22)

<b>H</b>	E23	=12*E22/(E21*(E21+1))-3*(E21+1)
<b>df</b>	E24	=COUNTA(A7:C7)-1
<b>p</b>	E25	=CHIDIST(E23,E24)
<b>α</b>	E26	0.05
<b>sig</b>	E27	=IF(E25<E26,"yes","no")

## Dunn's Test after KW

**Dunn's test** relies on the normal distribution and it includes a **ties correction**. This test is in the Excel workbook at tab *Dunn's Test*.

The term **tie** is used in connection with rank order statistics. Tied observations are observations having the same value, which prohibits the assignment of unique rank numbers. As a way out tied observations are assigned to the average of their hypothetical ranks (Amerise & Tarsitano, 2015). This will impact the calculation of the standard errors, so we use the one defined below.

Dunn's test uses the statistic

$$z = \frac{|\bar{R}_i - \bar{R}_j|}{\text{std. err.}}$$

where

$$\bar{R}_j = \frac{\bar{R}_j}{n_j}$$

and the standard error is

$$\text{std. err.} = \sqrt{\frac{n(n+1)}{12} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

Here  $n$  = the total sample size and  $n_i$  and  $n_j$  are the sizes of the groups being compared. If we are using the test statistic and critical values for determining significant, then we use

$$\bar{R}_{\max} - \bar{R}_{\min} > z_{crit} \cdot \text{std. err.}$$

where

$$z_{crit} = \text{NORM.S.INV} \left( \frac{\alpha^*}{2} \right).$$

Otherwise, we use the p-value.

Here,  $\alpha^*$  is the significance level using some form of familywise error correction. For example, if you want to account for all possible pairwise comparisons using a Bonferroni-like correction, then you would set

$$\alpha^* = \frac{\alpha}{\frac{k(k-1)}{2}}$$

where  $k$  = the number of groups. Note that if you know in advance that only certain pairwise tests will be interesting, then you can reduce the correction factor. For example, if  $\alpha = .05$  and  $k = 4$ , then  $\alpha^* = 0.05/6 = 0.0833$ . If you are certain (before looking at the data) that you will only perform comparisons of groups A vs. B and A vs. C, then you can use  $\alpha^* = 0.05/2 = 0.025$  instead.

### Example 3.6: Dunn's Test for the WD post comparison

Now, we'll use Dunn's test to see which of the WD sample means, from the Kruskal-Wallis H test, are significantly different from each other and the control. The data is shown again in **Figure 3.16**.

	A	B	C
3	DATA		
4	WD-30	WD-29	Control
5	156	133	145
6	177	156	160
7	147	112	176
8	146	100	120
9	181	168	155
10	144	130	144
11	183	132	156
12	147	115	143
13	149	140	
14	161		

*Figure 3.16. Test data from the WD formula experiment.*

Before we can run the test, we need to get the summary statistics, namely the Counts and Variance (left side of **Figure 3.17**). We will also

need the Sum of Ranks statistic (right side of **Figure 3.17**) The formulas for these results are provided in **Table 3-25**.

Before we move on with the test, let's take a closer look at the ranks in **Figure 3.17**.

```
=IF(ISNUMBER(A5)=TRUE,RANK.AVG(A5,$A$5:$C$14,1),)
```

The function ISNUMBER() is used to check that the referenced cell contains a number. If the cell does contain a number, we use RANK.AVG function that returns the rank of a number in a list of numbers, that is its size relative to the other number in the list. If there is a more than one value with the same rank, then the average rank is used. The last parameter of the RANK.AVG function is 1, telling the function to rank in ascending order; we would use a 0 for descending order.

	F	G	H	I	J	K	L
3	COUNT	WD-30		10		Sum of Ranks	
4		WD-20		9	WD-30	WD-29	Control
5		Control		8	18.0	7.0	12.0
6					25.0	18.0	20.0
7	MEAN	WD-30	161.1		22.0	2.0	24.0
8		WD-20	131.7778		13.0	1.0	4.0
9		Control	149.875		26.0	23.0	16.0
10					10.5	5.0	10.5
11	VARIANCE	WD-30	228.3222		27.0	6.0	18.0
12		WD-20	456.6944		14.0	3.0	9.0
13		Control	263.8393		15.0	8.0	0.0
14					21.0	0.0	0.0
15					191.5	73.0	113.5

*Figure 3.17. Summary statistics (left) and Sum of Ranks (right).*

In **Figure 3.18**, we show the results of Dunn's test and provide the formulas we used in **Table 3-26**. Remember that in cells Q11, Q12, and Q13, we used the standard error we defined earlier.

	N	O	P	Q	R	S	T	U
3	DUNN's TEST			alpha	0.05	0.0167		
4	group	R-sum	size	R-mean	z-crit			
5	WD-30	191.5	10	19.15				
6	WD-29	73.0	9	8.11111				
7	Control	113.5	8	14.1875				
8			27		1.96			
9	D TEST							
10	group 1	group 2	R-mean	std err	z-stat	R-crit	p-value	sig?
11	WD-30	WD-29	11.0389	3.64692	3.0269	7.1478	0.00247	yes
12	WD-30	Control	4.9625	3.76497	1.3181	7.3792	0.18748	no
13	WD-29	Control	6.07639	3.85681	1.5755	7.5592	0.11514	no

*Figure 3.18. Results of the Dunn's Test performed ad hoc.*

The results tell us that the formula WD-30 has a mean value that is statistically different from the other treatment and control.

*Table 3-25. Dunn's Test inputs: summary statistics and rank sums.*

Cells	Item	Formula
<b>Summary Statistics</b>		
H3 to H5	COUNT	=COUNT(A5:A14)
H7 to H9	MEAN	=AVERAGE(A5:A14)
H11 to H13	VARIANCE	=VAR(A5:A14)
<b>Sum of Ranks</b>		
Cells	Item	Formula
J5	Rank Sum WD-30	=IF(ISNUMBER(A5)=TRUE,RANK.AVG(A5,\$A\$5:\$C\$14,1),)
K5	Rank Sum WD-29	=IF(ISNUMBER(B5)=TRUE,RANK.AVG(B5,\$A\$5:\$C\$14,1),)
L5	Rank Sum Control	=IF(ISNUMBER(C5)=TRUE,RANK.AVG(C5,\$A\$5:\$C\$14,1),)
J15	Sum of Ranks	=SUM(J5:J14)
K15	Sum of Ranks	=SUM(K5:K14)
L15	Sum of Ranks	=SUM(L5:L14)

**Table 3-26. Dunn's Test**

Cells	Item	Formula
S3	$\alpha^*$	=R3/(COUNT(P5:P7)*(COUNT(P5:P7)-1)/2)
O5,O6,O7	R1	=J15, J16, J17
P5 to P7	n1	=COUNT(A5:A14)
Q5 to Q7	R1 mean	=O5/P5
P8	n	=SUM(P5:P7)
R8	z-crit	=NORM.S.INV(1-R3/2)
P11 to P13	diff means	=ABS(Q5-Q6)
Q11 to Q13	std err	=SQRT((P8*(P8+1)/12)*(1/P5+1/P6))
R11 to R13	z-stat	=P11/Q11
S11 to S13	R-crit	=\$R\$8*Q11
T11 to T13	p-value	=2*(1-NORM.S.DIST(R11,TRUE))

There are other KW post comparison tests we could perform to compare the treatment, including:

- Nemenyi test
- Pairwise Mann-Whitney Tests
- Steel's test
- Schaich-Hamerle Test
- Conover test
- $\chi^2$  test

Of these tests, the Pairwise Mann-Whitney Tests is the most appropriate to use, but it cannot be easily implemented in Excel (Conroy, 2012). However, we provide the in the workbook in the tab marker WK-Post.



## 4. Multi-factor Analysis of Variance

The analysis of variance (ANOVA) (Neter, Wasserman, and Kutner, 1990) is used to detect significant factors in a multi-factor model. In the multi-factor model, there is a response (dependent) variable and one or more factor (independent) variables. This is a common model in designed experiments where the experimenter sets the values for each of the factor variables and then measures the response variable.

Each factor can take on a certain number of values. These are referred to as the levels of a factor. The number of levels can vary between factors. For designed experiments, the number of levels for a given factor tends to be small. Each factor and level combination is a cell. Balanced designs are those in which the cells have an equal number of observations and unbalanced designs are those in which the number of observations varies among cells. It is customary to use balanced designs in designed experiments.

### *Assumptions and limitations of a two-way ANOVA*

Our dependent variable – here, “material strength”, should be continuous – that is, measured on a scale which can be subdivided using increments (i.e., grams, milligrams)

- Our independent variables – here, “table speed” and “batch”, should be in categorical, independent groups.
- Sample independence – that each sample has been drawn independently of the other samples
- Variance Equality – That the variance of data in the different groups should be the same
- Normality – That each sample is taken from a normally distributed population

### *Hypotheses of a two-way ANOVA?*

Because the two-way ANOVA consider the effect of two categorical factors, and the effect of the categorical factors on each other, there are three pairs of null or alternative hypotheses for the two-way ANOVA.

The first categorical variable hypotheses:

*H<sub>0</sub>: The means of variable X<sub>1</sub> groups are equal*

*H<sub>1</sub>: The mean of at least one X<sub>1</sub> group is different*

The second categorical variable hypotheses:

*H<sub>0</sub>: The means of variable X<sub>2</sub> groups are equal*

*H<sub>1</sub>: The means of variable X<sub>2</sub> groups are different*

The interaction hypotheses:

*H<sub>0</sub>: There is no interaction between the batch and table speed*

*H<sub>1</sub>: There is interaction between the batch and table speed*

#### *Interactions in two-way ANOVA*

These last two hypotheses, of there being (or not being) interactions in a two-way ANOVA, refer to how the two variables in the study affect each other.

This is most easily explained by going back to our material strength. If the researchers found that material strength in batch 2 significantly decreased at a slow table speed, but material strength of batch 5 remained steady or slightly increased with a slow table speed, subsequent statistical analysis may conclude that there was an interaction between the two independent variables of batch and table speed.

These effects are not to be ignored. If we put the interactions to one side, with the results mentioned above, an incomplete analysis might conclude that material strength in general increases after early batches, which would ignore a reality that the increase was driven by small changes to table speed over time, assuming that batches are run sequentially over time. (slow can represent a continuous range of speeds). Another example could be the efficacy of a candidate drug for a disease; you can see how proper modeling of interaction effects would be important.

A three-way ANOVA is not much different than the two-way, except that the results are a little more difficult to interpret. However, this is where we can employ simple effects and contrast for a better understanding.

Four-way ANOVA and above are rarely used because the results of the test are complex and very difficult to interpret.

### *Summary: One-way and Two-way ANOVA Differences*

The key differences between one-way and two-way ANOVA are summarized clearly below.

1. A one-way ANOVA is primarily designed to enable the equality testing between three or more means. A two-way ANOVA is designed to assess the interrelationship of two independent variables on a dependent variable.
2. A one-way ANOVA only involves one factor or independent variable, whereas there are two independent variables in a two-way ANOVA.
3. In a one-way ANOVA, the one factor or independent variable analyzed has three or more categorical groups. A two-way ANOVA instead compares multiple groups of two factors.
4. One-way ANOVA need to satisfy only two principles of design of experiments, i.e. replication and randomization. As opposed to two-way ANOVA, which meets all three principles of design of experiments which are replication, randomization and local control.

### **Example 4.1: WD-40**

In 1953, a fledgling company called Rocket Chemical Company and its staff of three set out to create a line of rust-prevention solvents and degreasers for use in the aerospace industry. Working in a small lab in San Diego, California, it took them 40 attempts to get the water displacing formula worked out. But they must have been really good, because the original secret formula for WD-40® Multi-Use Product - which stands for Water Displacement perfected on the 40th try—is still in use today.

Convair, an aerospace contractor, first used WD-40 Multi-Use Product to protect the outer skin of the Atlas Missile from rust and corrosion. The product actually worked so well that several employees snuck some cans out of the plant to use at home.

The early experiments with the WD formula, beginning with WD-1, did not go well. The water displacement factors were low and inconsistent.

Using hypothetical data, we want to determine if the mean readings from four different silos using three formulations are not significantly different. More specifically, we want to see if there is a difference between formulations. The data are shown in **Figure 4.1**.

	A	B	C	D	E
3		Silo 1	Silo 2	Silo 3	Silo 4
4	WD-20	123	138	110	151
5	WD-21	145	165	140	167
6	WD-22	156	176	185	175

**Figure 4.1.** Data for the WD formula experiment

So, we run a multifactor ANOVA without replacement. Here, our focus is on the analysis of an experiment using ANOVA, rather than the experimental design itself. The point being that before we design and experiment, we consider how we are going to analyze the data it produces.

## ANOVA: Multi-Factor Without Replication

An experiment with more than one factor (simple or one-way) includes experiments with two, three, etc., with multiple levels of each factor, which is more complex than the simple ANOVA we have seen so far. However, the mathematics behind the scene are really not much more complex, as we will see, for example with Sums of Squares. The results are shown in **Figure 4.2** and **Figure 4.3**.

	G	H	I	J	K	L
3	SUMMARY	Count	Sum	Average	Variance	SS
4	WD-20	4	522	130.5	317.6667	953
5	WD-21	4	617	154.25	188.9167	566.75
6	WD-22	4	692	173	148.6667	446
7	Silo 1	3	424	141.3333	282.3333	564.6667
8	Silo 2	3	479	159.6667	382.3333	764.6667
9	Silo 3	3	435	145	1425	2850
10	Silo 4	3	493	164.3333	149.3333	298.6667

**Figure 4.2.** ANOVA descriptive statistics for the WD formula experiment

	G	H	I	J	K	L	M
11	ANOVA					alpha = 0.05	
12	Source of Variation	SS	df	MS	F	P-value	F crit
13	Rows	3629.167	2	1814.583	12.8264	0.0013	5.1433
14	Columns	1116.917	3	372.3056	2.6317	0.1023	4.7571
15	Error	848.8333	6	141.4722			
16	Total	5594.917	11				

**Figure 4.3.** ANOVA results for WD formula experiment

There are two null hypotheses: one for the rows and the other for the columns. Let's look first at the rows:

*$H_0$ : there is no significant difference in yield between the means of the blends*

Since the p-value for the rows =  $.0068 < .05 = \alpha$  (or  $F = 12.83 > 5.14 = F\text{-crit}$ ) we reject the null hypothesis, and so at the 95% level of confidence we conclude there is significant difference in the yields produced by the three blends. The null hypothesis for the columns is:

*$H_0$ : there is no significant difference in yield between the means for the silos*

Since the p-value for the columns =  $.1446 > .05 = \alpha$  (or  $F = 2.63 < 4.76 = F\text{-crit}$ ) we can't reject the null hypothesis, and so at 95% level of confidence we conclude there is no significant difference in the yields for the four crops studied.

**Observation:** Although the analysis in Figure 2 was produced automatically by Excel's data analysis tool, the same result can be produced using Excel formulas, just as we were able to do in Basic Concepts of ANOVA for one-way ANOVA. The most interesting cells are the ones corresponding to the four sum squares. We show how to calculate the values for each of those cells in **Table 4-1** and **Table 4-2**.

**Table 4-1.** Key formulas ANOVA Summary

SUMMARY	Count	Sum	Average	Variance
<b>WD-20</b>	=COUNT(B4:E4)	=SUM(B4:E4)	=AVERAGE(B4:E4)	=VAR(B4:E4)
<b>WD-21</b>	=COUNT(B5:E5)	=SUM(B5:E5)	=AVERAGE(B5:E5)	=VAR(B5:E5)

<b>Wd-22</b>	=COUNT(B6:E6)	=SUM(B6:E6)	=AVERAGE(B6:E6)	=VAR(B6:E6)
<b>Silo 1</b>	=COUNT(B4:B6)	=SUM(B4:B6)	=AVERAGE(B4:B6)	=VAR(B4:B6)
<b>Silo 2</b>	=COUNT(C4:C6)	=SUM(C4:C6)	=AVERAGE(C4:C6)	=VAR(C4:C6)
<b>Silo 3</b>	=COUNT(D4:D6)	=SUM(D4:D6)	=AVERAGE(D4:D6)	=VAR(D4:D6)
<b>Silo 4</b>	=COUNT(E4:E6)	=SUM(E4:E6)	=AVERAGE(E4:E6)	=VAR(E4:E6)

**Table 4-2.** Key formulas for ANOVA output

ANOVA					
Source of Variation	SS	df	MS	F	P-value
<b>Rows</b>	=DEVSQ(J6:J8)*H6	=H10-1	=H18/I18	=J18/J20	=FDIST(K18,I18,I20)
<b>Columns</b>	=DEVSQ(J10:J13)*H10	=H6-1	=H19/I19	=J19/J20	=FDIST(K19,I19,I20)
<b>Error</b>	=H22-H18-H19	=I18*I19	=H20/I20		
<hr/>					
<b>Total</b>	=DEVSQ(B4:E6)	=H6*H10-1			

The formulas for calculating  $SS_{Row}$  and  $SS_{Col}$  in Definition 2 involve taking squared deviations of the group means. For Example,  $SS_{Row}$  can be calculated via the formula  $=DEVSQ(I6:I8)/H6$ . Alternatively, we can take squared deviations from the sums of each group, as is done in **Table 4-3**.

**Table 4-3.** Error term formulas for two-way ANOVA

Sum of Squares	Degrees of Freedom	Mean Squares
$SS_T = \sum_j \sum_i (x_{ij} - \bar{x}^2)^2$	$df_T = n - 1$	$MS_T = \frac{SS_T}{df_T}$
$SS_A = c \sum_i (x_i - \bar{x})^2$	$df_A = r - 1$	$MS_A = \frac{SS_A}{df_a}$
$SS_B = r \sum_j (x_j - \bar{x})^2$	$df_B = c - 1$	$MS_B = \frac{SS_B}{df_B}$
$SS_E = \sum_j \sum_i (x_{ij} - \bar{x}_i - \bar{x}_j - \bar{x}^2)^2$	$df_E = (r - 1)(c - 1)$	$MS_E = \frac{SS_E}{df_E}$

## Multi-Factor ANOVA Followup Analysis

The tests following multi-factor ANOVA without replication are similar to those following two-factor ANOVA with replication, except that clearly, no tests for the interaction of factors are necessary.

We will show how to perform Contrasts and Tukey's HSD tests for the data in WD-40 Example of Multi-Factor ANOVA without Replications. That analysis showed that there was a significant difference in the Rows factor (the difference WD formulations), but not in the Columns factor, and so it is appropriate to pinpoint where the differences lie in the Rows factor. For completeness, we also show how to perform follow-up testing on the Columns factor, even though this isn't really necessary.

### Example 4.2:Contrasts for Two Factor ANOVA

First, we determine whether there is a significant difference between WD-22 and the mean of WD-20 and WD-21. To conduct the analysis, enter the formulas in **Table 4-4** on the same worksheet as the original WD-40 data.

*Table 4-4. Excel formulas for two factor ANOVA contrasts*

Cells	Entity	Formula
H3:K7 <sup>3</sup>	transpose	=TRANSPOSE(A3:E6)
Q5 <sup>4</sup>	contrast	=SUMPRODUCT(\$N\$4:\$P\$4,N5:P5)
Q9, M15	mean	=AVERAGE(Q5:Q8)
Q10	std deviation	=STDEV(Q5:Q8)
Q11, N15	Std err	=STDEV(Q5:Q8)
O15	t-stat	=M15/N15
P15	df	=COUNTA(Q5:Q8)-1
Q15	p-value	=T.DIST.2T(ABS(O15),P15)
R15	significant	=IF(Q15<Q13,"Yes","No")
S15	Cohen d	=ABS(Q9)/Q10

<sup>3</sup> To enter the transpose, highlight the range A3:E6, place the cursor in the input box, enter the formula, and press Ctrl + Shift + Enter)

<sup>4</sup> After entering the formula in cell Q5, copy it down through cell Q8

Cells	Entity	Formula
T15	effect r	=SQRT(M15^2/(M15^2+P15))

Now manually fill in the contrast coefficients shown in range N4:P4 (-0.5, -0.5, and 1) and the other values in the figure will change to reflect these contrast coefficients. We see (cell R15) that there is no significant difference between WD-22 and the mean of WD-20 and WD-21.

The output should look like

**Table 4-5 and Table 4-6:**

Transpose Matrix

	WD-20	WD-21	WD-22
Silo 1	123	145	156
Silo 2	138	165	176
Silo 3	110	140	185
Silo 4	151	167	175

*Table 4-5. Contrast: Multi-Factor ANOVA w/o Replacement*

Contrast	WD-20	WD-21	WD-22	Diff
	-0.5	-0.5	1	
1	123	145	156	22
2	138	165	176	24.5
3	110	140	185	60
4	151	167	175	16
	mean		30.625	
			19.9055	
	std dev		1	
			9.95275	
	std err		3	

*Table 4-6. t-test results table for the WD formulas experiment*

T TEST		Alpha		0.05			
mean	std err	t-stat	df	p-value	sig	effect d	effect r
30.63	9.95275	3.07704	3	0.054262	No	1.53852	0.99840

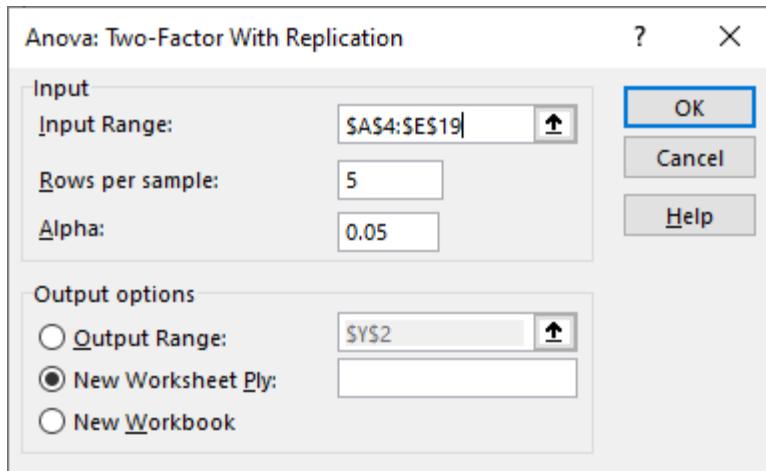
### Example 4.3: Multi-Factor ANOVA with replication

Let's repeat the WD-40 experiment, but this time we'll gather five WD formula readings from each silo and WD formulation, as seen in **Table 4-7**. The data is shown in Excel format so as to ease the process for the built-in analysis tool.

*Table 4-7. WD formulation and Missile Silo performance data.*

Formula	Missile Silos			
	Silo 1	Silo 2	Silo 3	Silo 4
WD-20	123	128	166	151
	156	150	178	125
	112	174	187	117
	100	116	153	155
	168	109	195	158
WD-21	135	175	140	167
	130	132	145	183
	176	120	159	142
	120	187	131	167
	155	184	126	168
WD-22	156	186	185	175
	180	138	206	173
	147	178	188	154
	146	176	165	191
	193	190	188	169

We set up this analysis with Excel's analysis tool add-in as shown in **Figure 4.4**. The range includes the WD formula and silo headings (labels). The rows per sample correspond the observations made for each WD formulation. We accept the default 0.05 ( $\alpha$  value). Where we direct the output is optional, but its best to put on a new worksheet to prevent overwriting existing data.



**Figure 4.4.** Data Analysis tool dialog window with ANOVA settings

The summary statistics and ANOVA results are shown in **Table 4-8** and **Table 4-9**.

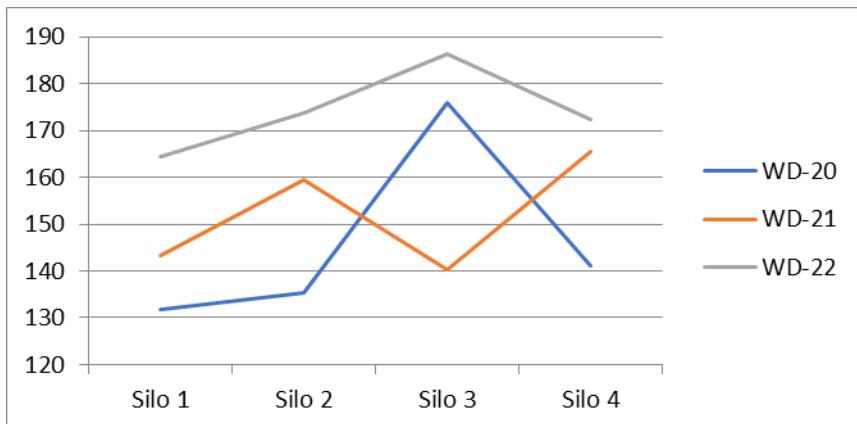
**Table 4-8.** ANOVA summary statistics Multi-Factor with Replication

SUMMARY	Silo 1	Silo 2	Silo 3	Silo 4	Total
<i>WD-20</i>					
Count	5	5	5	5	20
Sum	659	677	879	706	2921
Average	131.8	135.4	175.8	141.2	146.05
Variance	844.2	707.8	278.7	354.2	782.3658
<i>WD-21</i>					
Count	5	5	5	5	20
Sum	716	798	701	827	3042
Average	143.2	159.6	140.2	165.4	152.1
Variance	498.7	978.3	165.7	217.3	511.0421
<i>WD-22</i>					
Count	5	5	5	5	20
Sum	822	868	932	862	3484
Average	164.4	173.6	186.4	172.4	174.2
Variance	443.3	428.8	212.3	175.8	330.6947
<i>Total</i>					
Count	15	15	15	15	60
Sum	2197	2343	2512	2395	9447
Average	146.4667	156.2	167.4667	159.6667	157.45
Variance	705.8381	871.0286	605.981	404.9524	671.8788

**Table 4-9. ANOVA results table for WD formulation experiment**

Source of Variation	SS	df	MS	F	P-value	F crit
Formula (Rows)	8782.9	2	4391.45	9.93335	0.000245	3.19073
Silos (Columns)	3411.65	3	1137.217	2.57236	0.064944	2.79806
Interaction	6225.9	6	1037.65	2.34714	0.045555	2.29460
Within	21220.4	48	442.0917			
Total	39640.85	59				

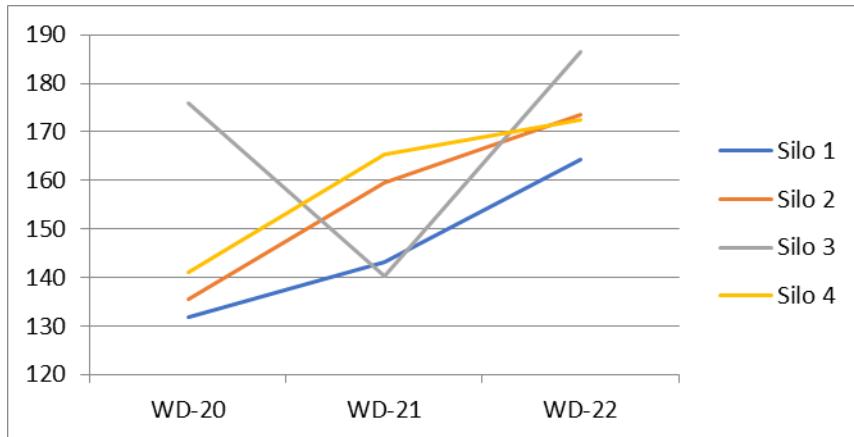
The output that the rows (WD formulas) and interactions are significant, based on the F statistics and critical values. However, the columns (missile silos) are not significant. We will learn that not rejecting a null hypothesis, as with  $H_0$  for the columns, does not mean we accept it. Not rejecting merely means we do not have enough evidence to reject it. We can also get an idea of what interactions are occurring by plotting them. The plot in **Figure 4.5** shows the row (WD formulas) interactions.



**Figure 4.5. Interaction plot for WD formulation. The plot shows an interaction between WD-20 and WD-21.**

The Second plot in **Figure 4.6** shows the interaction of missile silos. Here, we are not saying that the physical silos interact with one another. Rather, the conditions of the silos vary, and that at one time or another they are the same (i.e., same humidity and same physical design). That is, the coefficient of the regressor term for missile silos. Just because we do not reject the null hypothesis does not mean we accept it. It just

means that we do not have enough statistical evidence to reject it. Hence, there may be significant simple effects and significant contrasts. In our next example we demonstrate this.



*Figure 4.6. Interaction plot for the missile silos.*

The Excel formulas used in the analysis are shown in **Table 4-10** through **Table 4-13**.

*Table 4-10. Part 1 of the summary statistics table formulas*

SUMMARY	Silo 1	Silo 2	Silo 3
<b>WD-20</b>			
Count	=COUNT(B5:B9)	=COUNT(C5:C9)	=COUNT(D5:D9)
Sum	=SUM(B5:B9)	=SUM(C5:C9)	=SUM(D5:D9)
Average	=AVERAGE(B5:B9)	=AVERAGE(C5:C9)	=AVERAGE(D5:D9)
Variance	=VAR(B5:B9)	=VAR(C5:C9)	=VAR(D5:D9)

*Table 4-11. Part 2 of the summary statistics table formulas*

Silo 4	Total
=COUNT(E5:E9)	=SUM(H5:K5)
=SUM(E5:E9)	=SUM(B5:E9)
=AVERAGE(E5:E9)	=AVERAGE(B5:E9)
=VAR(E5:E9)	=VAR(B5:E9)

**Table 4-12.** Part 1 of the ANOVA output table formulas

Source of Variation	SS	df
<b>Formula (Rows)</b>	=DEVSQ(H38:H40)/L5	=COUNTA(G38:G40)-1
<b>Silos (Columns)</b>	=DEVSQ(H43:H46)/H23	=COUNTA(G43:G46)-1
<b>Interaction</b>	=H35-H30-H31-H33	=I30*I31
<b>Within</b>	=SUM(K39:N41)	=COUNT(B5:E19)- COUNTA(G38:G40)*COUNTA(G43:G46)
<b>Total</b>	=DEVSQ(B5:E19)	=COUNT(B5:E19)-1

**Table 4-13.** Part 2 of the ANOVA output table formulas

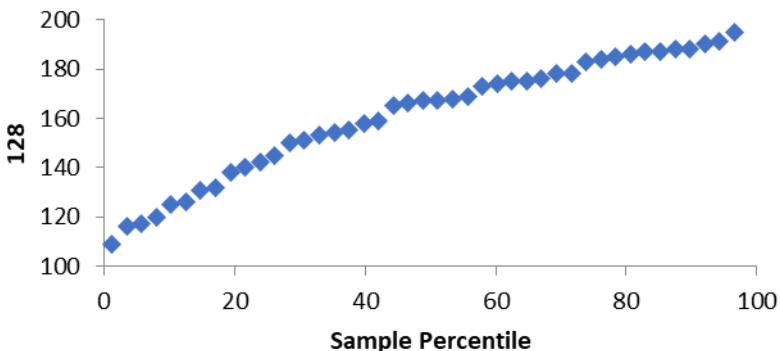
MS	F	P-value	F crit
=H30/I30	=J30/J\$33	=FDIST(K30,I30,I\$33)	=FINV(0.05,I30,I\$33)
=H31/I31	=J31/J\$33	=FDIST(K31,I31,I\$33)	=FINV(0.05,I31,I\$33)
=H32/I32	=J32/J\$33	=FDIST(K32,I32,I\$33)	=FINV(0.05,I32,I\$33)
=H33/I33			

Whenever we perform an ANOVA, it is important to explore the residuals, through a standardized residual scatterplot (**Figure 4.7**), a normal probability plot (**Figure 4.8**), and standard residuals run sequence chart (**Figure 4.9**). None of the plots show anything to be concerned with.



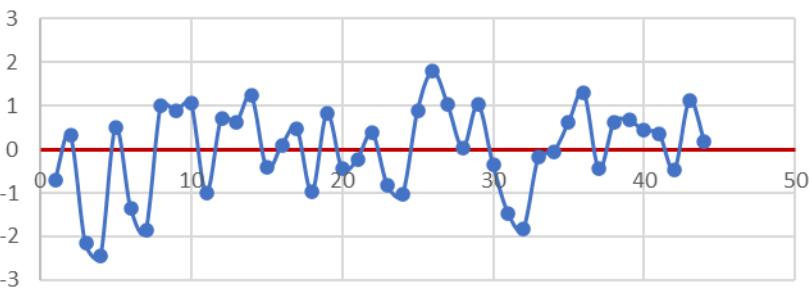
**Figure 4.7.** Standard residual plot from the associated regressions model.

## Normal Probability Plot



**Figure 4.8.** Normal probability plot for the WD formulation experiment

## Standard Residuals Run Sequence



**Figure 4.9.** Sequence-run chart of the WD formulation residuals

Although we will not perform follow-on means analysis for this example, we'll demonstrate how to calculate the means (**Table 4-14**) and their squared deviations (**Table 4-15**) or variance. The next example we introduce will include these kinds of analyses, including simple effects analysis and contrast analysis.

**Table 4-14.** Squared Deviations of the Silos' Means

SSW	Silo 1	Silo 2	Silo 3	Silo 4
WD-20	3376.8	2831.2	1114.8	1416.8
WD-21	1994.8	3913.2	662.8	869.2
WD-22	1773.2	1715.2	849.2	703.2

**Table 4-15. Means of the Silos' Readings**

Means	Silo 1	Silo 2	Silo 3	Silo 4
WD-20	131.8	135.4	175.8	141.2
WD-21	143.2	159.6	140.2	165.4
WD-22	164.4	173.6	186.4	172.4

**Table 4-16. Formulas for Table 4-14. Squared Deviations of the Silos' Means**

SSW	Silo 1	Silo 2	Silo 3	Silo 4
WD-20	=DEVSQ(B5:B9)	=DEVSQ(C5:C9)	=DEVSQ(D5:D9)	=DEVSQ(E5:E9)
WD-21	=DEVSQ(B10:B14)	=DEVSQ(C10:C14)	=DEVSQ(D10:D14)	=DEVSQ(E10:E14)
WD-22	=DEVSQ(B15:B19)	=DEVSQ(C15:C19)	=DEVSQ(D15:D19)	=DEVSQ(E15:E19)

### Example 4.4: Two-way ANOVA with Contrasts

Here we use that same data, minus one silo to demonstrate running an ANOVA with replacement analysis, followed by a more detailed investigation of means. That is, where we have significant main effects, we further investigate the means of the main effects using simple effect and contrast analysis.

**Table 4-17. Partial WD formulation and Missile Silo performance data.**

Formula	Silos		
	Silo 1	Silo 2	Silo 3
WD-20	123	128	166
	156	150	178
	112	174	187
	100	116	153
	168	109	195
WD-21	135	175	140
	130	132	145
	176	120	159
	120	187	131
	155	184	126
WD-22	156	186	185
	180	138	206
	147	178	188
	146	176	165
	193	190	188

**Figure 4.10** shows the descriptive statistics for the combinations of WD formulas and missile silos. Figure 4.11 shows the two-way ANOVA output.

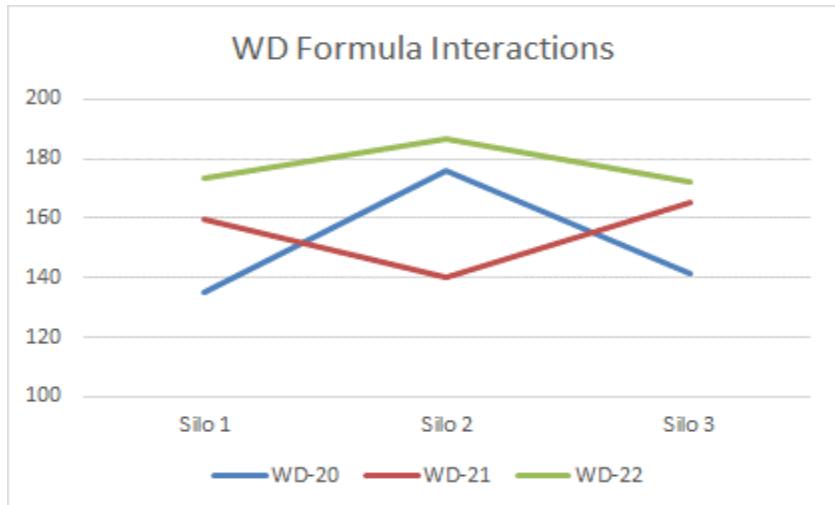
	F	G	H	I	J
2	Descriptive Statistics				
3	COUNT	balanced			
4		Silo 1	Silo 2	Silo 3	
5	WD-20	5	5	5	15
6	WD-21	5	5	5	15
7	WD-22	5	5	5	15
8		15	15	15	45
9					
10	MEAN				
11		Silo 1	Silo 2	Silo 3	
12	WD-20	131.8	135.4	175.8	147.6667
13	WD-21	143.2	159.6	140.2	147.6667
14	WD-22	164.4	173.6	186.4	174.8
15		146.4667	156.2	167.4667	156.7111
16					
17	VARIANCE				
18		Silo 1	Silo 2	Silo 3	
19	WD-20	844.2	707.8	278.7	949.381
20	WD-21	498.7	978.3	165.7	547.2381
21	WD-22	443.3	428.8	212.3	397.0286
22		705.8381	871.0286	605.981	769.8465

**Figure 4.10.** Summary descriptive statistics tables

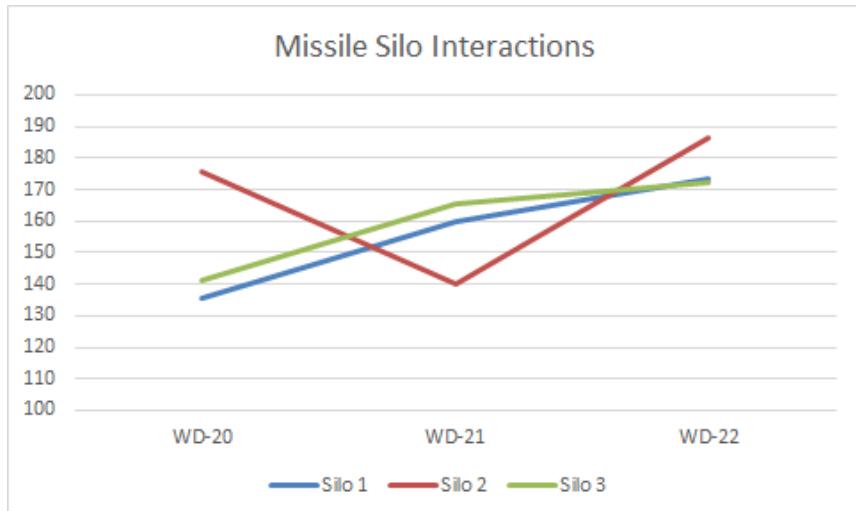
	L	M	N	O	P	Q	R
3	ANOVA				Alpha	0.05	
4		SS	df	MS	F	p-value	sig
5	Rows	7362.178	2	3681.089	7.268814	0.00223	yes
6	Columns	3313.378	2	1656.689	3.271359	0.049498	yes
7	Inter	4966.489	4	1241.622	2.451753	0.063495	no
8	Within	18231.2	36	506.4222			
9	Total	33873.24	44	769.8465			

**Figure 4.11.** Two Factor ANOVA table

The interactions of the WD formulas are shown in **Figure 4.12** and the interactions of the missile silos are shown in **Figure 4.13**. Interactions plot of the missile silos.



*Figure 4.12. Interaction plot of the WD formulas*



*Figure 4.13. Interactions plot of the missile silos*

## **Summary of the ANOVA Analysis**

- In a factorial design, the main effect of an independent variable is its overall effect averaged across all other independent variables. There is one main effect for each independent variable.
- There is an interaction between two independent variables when the effect of one depends on the level of the other. Some of the most interesting research questions and results in psychology are specifically about interactions.
- A simple effects analysis provides a means for researchers to break down interactions by examining the effect of each independent variable at each level of the other independent variable.

## **Post-ANOVA Comparison of Means**

So, we found the means are significantly different for the main effects. Now what? Next, we examine simple effects and contrasts.

### **Simple Effects**

The finding that we have significant main effects for the rows (WD formulas) and the interactions, does not inform as to why they are significant. The interactions, when plotted give some insight but we cannot really draw any conclusions from them. This is where we can investigate simple effects.

Simple effects (sometimes called simple main effects) are differences among particular cell means within the design. More precisely, a simple effect is the effect of one independent variable within one level of a second independent variable. Use a Test of Simple Effects. This will produce a table comparing all pairs of levels of one factor, for each level of all the other factors.

It is not necessary to know whether the simple effects differ from zero in order to understand an interaction because the question of whether simple effects differ from zero has nothing to do with interaction except that if they are both zero there is no interaction. It is not uncommon to see research articles in which the authors report that they analyzed simple effects in order to explain the interaction. However, this is not a valid approach since an interaction does not depend on the analysis of the simple effects.

However, there is a reason to test simple effects following a significant interaction. Since an interaction indicates that simple effects differ, it means that the main effects are not general.

### Example 4.5: Simple Effects: Rows

First, we perform the simple effects of the WD formulas in **Figure 4.14**.

	M	N	O	P	Q	R	S
12	Simple Effects: Rows						
13							
14	ANOVA				Alpha	0.05	
15		SS	df	MS	F	p-value	sig
16	WD-20	5968.533	2	2984.267	5.892843	0.006110604	yes
17	WD-21	1090.533	2	545.2667	1.076704	0.35143506	no
18	WD-22	1220.8	2	610.4	1.205318	0.311399493	no
19	Error	18231.2	36	506.4222			

*Figure 4.14. ANOVA results table for the simple effects of rows*

The simple effects of WD-20 is significant and in opposite direction the interacting formula WD-21, which is not significant as a simple effect. More telling is that WD-22 is not part of the interaction. This does show that the formulation WD-22 is not part of the interaction. A non-significant simple effect does not mean that the simple effect is zero: the null hypothesis should not be accepted just because it is not rejected.

### Example 4.6: Simple Effects: Columns

The simple effects of the missile silos (columns) all interact and are significant as shown in **Figure 4.15**.

	M	N	O	P	Q	R	S
21	Simple Effects: Columns						
22							
23	ANOVA				Alpha	0.05	
24		SS	df	MS	F	p-value	sig
25	Silo 1	2736.933	2	1368.467	2.702225	0.080648955	no
26	Silo 2	3734.8	2	1867.4	3.687437	0.034926653	yes
27	Silo 3	5856.933	2	2928.467	5.782658	0.006640763	yes
28	Error	18231.2	36	506.4222			

*Figure 4.15. ANOVA results table for the simple effects of columns*

The conclusion should be that the simple effects differ, and that at least one of them is not zero. However, no conclusion should be drawn about which simple effect(s) is/are not zero.

Another error that can be made by mistakenly accepting the null hypothesis is to conclude that two simple effects are different because one is significant and the other is not.

The proper conclusion is that the experiment supports the researcher's hypothesis that the difference in silos is significant, but not strongly enough to allow a confident conclusion.

The Excel formulas used for this analysis appear in **Table 4-18** and **Table 4-19**

*Table 4-18. Excel formulas for Simple Effect Rows*

Name	Cell	Formula
$SS_{WD-20}$	N16-N18	=SUMPRODUCT(H6:J6,(H13:J13-K13)^2)
$SS_{Error}$	N19	=SUMPRODUCT(H20:J22,(H6:J8-1))
$df_{WD-20}$	O16-O18	=COUNT(H6:J6)-1
$df_{Error}$	O19	=SUM(H6:J8)-COUNT(H6:J6)*COUNT(H6:H8)
$MS_{(WD-20)}$	P16-P19	=N16/O16
$F_{WD-20}$	Q16-Q18	=P16/P\$19
$p\text{-value}$	R16-R18	=FDIST(Q16,O16,O\$19)
<i>Significant</i>	S16-S18	=IF(R16<R\$14,"yes","no")

*Table 4-19. Excel formulas for Simple Effects Columns*

Name	Cell	Formula
$SS_{Silo\ 1}$	N25-N27	=SUMPRODUCT(H6:H8,(H13:H15-H\$16)^2)
$SS_{Error}$	N28	=SUMPRODUCT(H20:J22,(H6:J8-1))
$df_{Silo\ 1}$	O25-O27	=COUNT(H6:H8)-1
$df_{Error}$	O28	=SUM(H6:J8)-COUNT(H6:J6)*COUNT(H6:H8)
$MS_{Silo\ 1}$	P25-P28	=N25/O25
$F_{Silo\ 1}$	Q25-P27	=P25/P\$28
$p\text{-value}$	R25-R27	=FDIST(Q25,O25,O\$28)
<i>Significant</i>	S25-S27	=IF(R25<R\$23,"yes","no")

## Contrasts

A contrast is a way of testing more general hypotheses about population means. The F-tests used to determine if the main effects and interactions terms are statistically significantly different from zero are based on accumulating information from contrasts.

### Example 4.7: WD Contrasts

For two-way (or higher) ANOVA without interaction, main effects contrasts are constructed separately for each factor, where the population means represent setting a specific level for one factor and ignoring (averaging over) all levels of the other factor. The first contrast is for rows WD-20, with *mean* = 150.80, vs WD-22, with *mean* = 177.467, and is significant with a t-statistic that is greater than the critical value, that is  $t\text{-stat} = 3.69 > t\text{-crit} = 2.03$  (see **Figure 4.16** and **Figure 4.17**).

	AA	AB	AC	AD
2	CONTRAST: ROWS			
3	<i>Groups</i>	<i>c</i>	<i>mean</i>	<i>n</i>
4	WD-20	-1	147.6667	15
5	WD-21		147.6667	15
6	WD-22	1	174.8	15
7		0	27.13333	7.5

**Figure 4.16.** Rows contrast table for WD-20 and WD-22

	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
8	T TEST					Alpha	0.05			
9	<i>std err</i>	<i>t-stat</i>	<i>df</i>	<i>p-value</i>	<i>t-crit</i>	<i>lower</i>	<i>upper</i>	<i>sig</i>	<i>Cohen d</i>	<i>effect r</i>
10	8.2172	3.3020	36	0.00217	2.0281	10.4680	43.7987	yes	1.2057	0.4821

**Figure 4.17.** t-test for the contrast between WD-20 and WD-22

The second contrast is for columns (missile silos) Silo 1 and Silo 3, and is not significant. Here the t-statistic = 0.48 and the critical value of 2.03. This implies that there is no difference between the silos, which would mean that the differences in performance of the WD formulas is due to the formulations (see **Figure 4.18** and **Figure 4.19**).

	AA	AB	AC	AD
12	CONTRAST: COLUMNS			
13	Groups	c	mean	n
14	Silo 1	-1	146.4667	15
15	Silo 2		156.2	15
16	Silo 3	1	167.4667	15
17		0	21	7.5

Figure 4.18. Rows contrast table for Silo 1 and Silo 3

	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
18	T TEST					Alpha	0.05			
19	std err	t-stat	df	p-value	t-crit	lower	upper	sig	Cohen d	effect r
20	8.2172	2.5556	36	0.015	2.028	4.3347	37.665	yes	0.9332	0.3919

Figure 4.19. t-test for the contrast between Silo 1 and Silo 3

The next contrast is for WD-formulas with silos. Here we look at WD-20 in Silo 3 vs. WD-22 in Silo 3. This test is also significant with a p-value of 0.0173. Note that t-statistics is greater than t-critical, as well (see Figure 4.20 and Figure 4.21).

	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
47	CONTRAST					CONTRAST: INTERACTIONS				
48		Silo 1	Silo 2	Silo 3		row	column	c	mean	n
49	WD-20			-1		WD-20	Silo 3	-1	175.8	5
50	WD-21					WD-22	Silo 3	1	186.4	5
51	WD-22			1				0	10.6	2.5

Figure 4.20. Rows contrast table for WD-20 in Silo 3 and WD-22 in Silo 3

	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO
52	T TEST							Alpha	0.05	
53	std err	t-stat	df	p-value	t-crit	lower	upper	sig	Cohen d	effect r
54	14.233	0.745	36	0.46125	2.0281	-18.265	39.465	no	0.47103	0.12318

Figure 4.21. t-test for the contrast between WD-20 in Silo 3 and WD-22 in Silo 3

For the next contrast we examine WD-20 in Silo 1 and silo 2 vs WD-22 in Silo 1 and Silo 2. Here the contrast is significant leading us to believe that there is a significant difference between WD-20 in the designated silos, and WD-22 in the same silos. Also note that when we perform contrasts, the rows and columns sum to one (see Figure 4.22 and Figure 4.23).

	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	
35	CONTRAST					CONTRAST: INTERACTIONS					
36	Silo 1	Silo 2	Silo 3			row	column	c	mean	n	
37	WD-20	-0.5	-0.5			WD-20	Silo 1	-0.5	131.8	5	
38	WD-21					WD-20	Silo 2	-0.5	135.4	5	
39	WD-22	0.5	0.5			WD-22	Silo 1	0.5	164.4	5	
40						WD-22	Silo 2	0.5	173.6	5	
41								0	35.4	5	

Figure 4.22. Rows contrast table for WD-20 in silos 1 and 2 and WD-22 in silos 1 and 2

	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	
42	TTEST					Alpha					0.05
43	std err	t-stat	df	p-value	t-crit	lower	upper	sig	Cohen d	effect r	
44	10.064	3.5175	36	0.0012	2.0281	14.9892	55.811	yes	1.5731	0.5057	

Figure 4.23. t-test for the contrast between WD-20 in silos 1 and 2 and WD-22 in silos 1 and 2

The next contrast is WD-22 in Silo 1 and WD-22 in Silo 3. Here the contrast is not significant, with a p-value of 0.926 (see **Figure 4.24** and **Figure 4.25**).

	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	
57	CONTRAST					CONTRAST: INTERACTIONS					
58	Silo 1	Silo 2	Silo 3			row	column	c	mean	n	
59	WD-20					WD-22	Silo 1	-1	164.4	5	
60	WD-21					WD-22	Silo 3	1	186.4	5	
61	WD-22	-1		1				0	22	2.5	

Figure 4.24. Rows contrast table for WD-22 in Silo 1 and Silo 3

	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	
62	TTEST					Alpha					0.05
63	std err	t-stat	df	p-value	t-crit	lower	upper	sig	Cohen d	effect r	
64	14.233	1.5457	36	0.1309	2.0281	-6.8652	50.865	no	0.9776	0.2495	

Figure 4.25. t-test for the contrast between WD-22 in Silo 1 and Silo 3

Taken together, the contrasts show that formula for WD-22 performs differently than WD-20 in the same silos, and that WD-22's performance is the same regardless of which silo a missile is in.

The Excel formulas for the calculations used in this analysis are presented in **Table 4-20** through **Table 4-22**.

**Table 4-20. CONTRAST: ROWS**

Name	Cell	Formulas
<i>c</i>	AB4	-1
<i>mean</i>	AC4	=K6
<i>n</i>	AD4	=K14
<b><i>mean<sub>Total</sub></i></b>	AC7	=SUMPRODUCT(AB4:AB6,AC4:AC6)
<b><i>n<sub>Total</sub></i></b>	AD7	=1/SUMPRODUCT(AB4:AB6^2,1/AD4:AD6)
<b>T-TEST</b>		
<i>std err</i>	AA10	=SQRT(SUMPRODUCT(H20:J22*(H6:J8-1))/(AD7*AC10))
<i>t-stat</i>	AB10	=AC7/AA10
<i>df</i>	AC10	=SUM(H6:J8)-COUNT(H6:J8)
<i>p-value</i>	AD10	=TDIST(ABS(AB10),AC10,2)
<i>t-crit</i>	AE10	=TINV(AG8,AC10)
<i>lower</i>	AF10	=AC7-AA10*AE10
<i>upper</i>	AG10	=AC7+AA10*AE10
<i>sig</i>	AH10	=IF(AD10<AG8,"yes","no")
<i>Cohen d</i>	AI10	=ABS(AC7)/(AA10*SQRT(AD7))
<i>effect r</i>	AJ10	=SQRT(AB10^2/(AB10^2+AC10))

**Table 4-21. CONTRAST: COLUMNS**

Name	Cell	Formula
<i>c</i>	AB14	-1
<i>mean</i>	AC14	{=TRANSPOSE(H16:J16)}
<i>n</i>	AD1 4	{=TRANSPOSE(H9:J9)}
<b><i>mean<sub>Total</sub></i></b>	AC17	=SUMPRODUCT(AB14:AB16,AC14:AC16)
<b><i>n<sub>Total</sub></i></b>	AD1 7	=1/SUMPRODUCT(AB14:AB16^2,1/AD14:AD16)
<b>T-TEST</b>		
<i>std err</i>	AA2 0	=SQRT(SUMPRODUCT(H20:J22*(H6:J8-1))/(AJ29*AH32))
<i>t-stat</i>	AB20	=AI29/AF32
<i>df</i>	AC20	=SUM(H6:J8)-COUNT(H6:J8)

<i>p-value</i>	AD2	=TDIST(ABS(AG32),AH32,2)
	0	
<i>t-crit</i>	AE20	=TINV(AN30,AH32)
<i>lower</i>	AF20	=AI29-AF32*AJ32
<i>upper</i>	AG2	=AI29+AF32*AJ32
	0	
<i>sig</i>	AH2	=IF(AI32<AN30,"yes","no")
	0	
<i>Cohen d</i>	AI20	=ABS(AI29)/(AF32*SQRT(AJ29))
<i>effect r</i>	AJ20	=SQRT(AG32^2/(AG32^2+AH32))

**Table 4-22. CONTRAST: INTERACTIONS**

Name	Cell	Formula
<i>row</i>	AF37	{=IFERROR(INDEX(\$AA\$37:\$AA\$39,SMALL(IF(\$AB\$37:\$AD\$39<>"",ROW(\$AB\$37:\$AD\$39)-MIN(ROW(\$AB\$37:\$AD\$39))+1,""),ROW(A1))),""})}
<i>column</i>	AG37	{=IFERROR(INDEX(\$AB\$36:\$AD\$36,SMALL(IF(\$AB\$37:\$AD\$39<>"",IF(ROW(\$AB\$37:\$AD\$39)=MATCH(AF37,\$AA\$37:\$AA\$39,0)+ROW(\$AB\$37)-1,COLUMN(\$AB\$37:\$AD\$39)-COLUMN(\$AB\$37)+1,""),""),COUNTIF(\$AF\$37:AF37,AF37))),""})}
<i>c</i>	AH37	=IFERROR(INDEX(\$AB\$37:\$AD\$39,MATCH(AF37,\$G\$13:\$G\$15),MATCH(AG37,\$H\$12:\$J\$12,0)),0)
<i>mean</i>	AI37	=IFERROR(INDEX(\$H\$13:\$J\$15,MATCH(AF37,\$G\$13:\$G\$15),MATCH(AG37,\$H\$12:\$J\$12,0)), "")
<i>n</i>	AJ37	=IFERROR(INDEX(\$H\$6:\$J\$8,MATCH(AF37,\$G\$6:\$G\$8),MATCH(AG37,\$H\$5:\$J\$5,0)),1)
<i>mean<sub>Total</sub></i>	AI41	=SUMPRODUCT(AH37:AH40, AI37:AI40)
<i>n<sub>Total</sub></i>	AJ41	=1/SUMPRODUCT(AH37:AH40^2,1/AJ37:AJ40)
<b>T-TEST</b>		
<i>std err</i>	AA20	=SQRT(SUMPRODUCT(H20:J22*(H6:J8-1))/(AJ41*AH44))
<i>t-stat</i>	AB20	=AI41/AF44
<i>df</i>	AC20	= SUM(H6:J8)-COUNT(H6:J8)
<i>p-value</i>	AD20	=TDIST(ABS(AG44),AH44,2)
<i>t-crit</i>	AE20	=TINV(AN42,AH44)
<i>lower</i>	AF20	=AI41-AF44*AJ44
<i>upper</i>	AG20	=AI41+AF44*AJ44

<i>sig</i>	AH20	=IF(AI44<AN42,"yes","no")
<i>Cohen d</i>	AI20	=ABS(AI41)/(AF44*SQRT(AJ41))
<i>effect r</i>	AJ20	=SQRT(AG44^2/(AG44^2+AH44))

The other interaction contrast calculations are similar.

### Example 4.8: ANOVA Using Pression

We'll now perform Two Factor ANOVA On the WD data using multiple regression. Our objective is to determine whether there is a significant difference between the three WD formulations. Recall the data as shown in **Figure 4.26**.

	S	T	U	V
2		Missile Location		
3	Climate	Silo 1	Silo 2	Silo 3
4	WD-20	128	166	151
5		150	178	125
6		174	187	117
7		116	153	155
8		109	195	158
9		175	140	167
10		132	145	183
11		120	159	142
12		187	131	167
13		184	126	168
14	WD-21	186	185	175
15		138	206	173
16		178	188	154
17		176	165	191
18		190	188	169

*Figure 4.26. Original WD formula data*

First, we define the following two dummy variables and map the original data into the model on the right side of **Figure 4.27**.

- $t1 = 1$  if WD-20;  $= -1$  if WD-21;  $= 0$  otherwise
- $t2 = 1$  if Silo1;  $-1$  if Silo2;  $= 0$  otherwise
- $t3 = 1$  if Silo2;  $-1$  if Silo3;  $= 0$  otherwise

	X	Y	Z	AA	AB	AC
	t1	t2	t3	t1t2	t1t3	Y
3						
4	1	1	0	1	0	128
5	1	1	0	1	0	150
6	1	1	0	1	0	174
7	1	1	0	1	0	116
8	1	1	0	1	0	109
9	-1	1	0	-1	0	175
10	-1	1	0	-1	0	132
11	-1	1	0	-1	0	120
12	-1	1	0	-1	0	187
13	-1	1	0	-1	0	184
14	0	1	0	0	0	186
15	0	1	0	0	0	138
16	0	1	0	0	0	178
17	0	1	0	0	0	176
18	0	1	0	0	0	190
19	1	-1	0	-1	0	166
20	1	-1	0	-1	0	178
21	1	-1	0	-1	0	187
22	1	-1	0	-1	0	153
23	1	-1	0	-1	0	195
24	-1	-1	0	1	0	140
25	-1	-1	0	1	0	145
26	-1	-1	0	1	0	159
27	-1	-1	0	1	0	131
28	-1	-1	0	1	0	126
29	0	-1	0	0	0	185
30	0	-1	0	0	0	206
31	0	-1	0	0	0	188
32	0	-1	0	0	0	165
33	0	-1	0	0	0	188
34	1	0	-1	0	-1	151
35	1	0	-1	0	-1	125
36	1	0	-1	0	-1	117
37	1	0	-1	0	-1	155
38	1	0	-1	0	-1	158
39	-1	0	-1	0	1	167
40	-1	0	-1	0	1	183
41	-1	0	-1	0	1	142
42	-1	0	-1	0	1	167
43	-1	0	-1	0	1	168
44	0	0	-1	0	0	175
45	0	0	-1	0	0	173
46	0	0	-1	0	0	154
47	0	0	-1	0	0	191
48	0	0	-1	0	0	169

Figure 4.27. Dummy variable matrix.

Note that in general, if the original data has  $k$  values the model will require  $k - 1$  dummy variables.

Note that this time we model the interaction of  $t_1$  with  $t_2$  and  $t_3$ , the regression model that we use is of form

$$y = b_0 + b_1 t_2 + b_2 t_3 + b_3 t_1 t_2 + b_4 t_1 t_3 + b_5 t_2 t_3$$

We now build a table of the means for each of the 9 groups (i.e., cells), as described in **Figure 4.28**.

	X	Y	Z	AA	AB	AC	
51	<b>t1</b>	<b>t2</b>	<b>t3</b>	<b>t1*t2</b>	<b>t1*t3</b>	<b>Y</b>	
52	1	1	0	1	0	135.4	
53	-1	1	0	-1	0	159.6	
54	0	1	0	0	0	173.6	
55	1	-1	0	-1	0	175.8	
56	-1	-1	0	1	0	140.2	
57	0	-1	0	0	0	186.4	
58	1	0	-1	0	-1	141.2	
59	-1	0	-1	0	1	165.4	
60	0	0	-1	0	0	172.4	

**Figure 4.28 Regression equation matrix.**

This table can be constructed by calculating the means of each of the above 9 groups from the original data or by applying the AVERAGEIFS function to the transformed data.

We note that the mean in the case for WD-20 Silo1 (i.e., where  $t_1 = t_2 = t_3 = 0$ ) is given by

$$135.4 = \mu_{X,WD20} = E[y] = b_0 + b_1 1 + b_2 1 + b_3 0 + b_4 1 + b_5$$

$$159.6 = \mu_{X,WD20} = E[y] = b_0 - b_1 1 + b_2 1 + b_3 0 - b_4 1 + b_5$$

$$173.6 = \mu_{X,WD20} = E[y] = b_0 + b_1 0 + b_2 1 + b_3 0 + b_4 0 + b_5$$

$$175.8 = \mu_{X,WD20} = E[y] = b_0 + b_1 1 - b_2 1 + b_3 0 - b_4 1 + b_5$$

$$140.2 = \mu_{X,WD20} = E[y] = b_0 - b_1 1 - b_2 1 + b_3 0 + b_4 1 + b_5$$

$$186.4 = \mu_{X,WD20} = E[y] = b_0 + b_1 0 - b_2 1 + b_3 0 + b_4 0 + b_5$$

$$141.2 = \mu_{X,WD20} = E[y] = b_0 + b_1 1 + b_2 0 - b_3 1 + b_4 - b_5 1$$

$$165.4 = \mu_{X,WD20} = E[y] = b_0 - b_1 1 + b_2 0 - b_3 1 + b_4 + b_5 1$$

$$172.4 = \mu_{X,WD20} = E[y] = b_0 + b_1 0 + b_2 0 - b_3 1 + b_4 0 + b_5$$

### Equation Summary

$$E[y] = 135.4 = b_0 + b_2 + b_4$$

$$E[y] = 159.6 = b_0 - b_1 + b_2 - b_4$$

$$E[y] = 173.6 = b_0 + b_2$$

$$E[y] = 175.8 = b_0 + b_1 - b_2 - b_4$$

$$E[y] = 140.2 = b_0 - b_1 - b_2 - b_4$$

$$E[y] = 186.4 = b_0 - b_2$$

$$E[y] = 141.2 = b_0 + b_1 - b_3 - b_5$$

$$E[y] = 165.4 = b_0 - b_1 - b_3 + b_5$$

$$E[y] = 172.4 = b_0 - b_3$$

Solving the simultaneous equations, we get the following values for the coefficients:

$$b_0 = 161.833, b_1 = 2.85, b_2 = -5.633, b_3 = 2.167, b_4 = -14.95, \\ b_5 = 14.95$$

We get the same results when we run the Regression data analysis tool (see **Table 4-23**, **Table 4-24**, **Table 4-25**, and **Table 4-26**).

**Table 4-23. Summary output or the regression model**

Regression Statistics	
Multiple R	0.50938183
R Square	0.25946985
Adjusted R <sup>2</sup>	0.16453008
Standard Error	22.78712337
Observations	45

**Table 4-24. ANOVA table generated by the regression**

**ANOVA**

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significant F</i>
Regression	5	7095.577778	1419.1156	2.73299	0.03281520
Residual	39	20250.86667	519.2530		
Total	44	27346.44444			

**Table 4-25. Coefficient information for the regression model (Part 1)**

	<i>Coefficients</i>	<i>Stand Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	161.833333	4.1603405	38.899060	0.000000
t1	2.85	5.0953557	0.559333	0.579133
t2	-5.63333333	4.1603405	-1.354056	0.183513
t3	2.16666667	7.2059211	0.300679	0.765257
t1*t2	-14.95	5.0953557	-2.934044	0.005580
t1*t3	14.95	8.8254149	1.693971	0.098244

**Table 4-26. Coefficient information for the regression model (Part 1)**

	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	153.418250	170.248416	153.418250	170.248416
t1	-7.456330	13.156330	-7.456330	13.156330
t2	-14.048416	2.781750	-14.048416	2.781750
t3	-12.408685	16.742018	-12.408685	16.742018
t1*t2	-25.256330	-4.643670	-25.256330	-4.643670
t1*t3	-2.901087	32.801087	-2.901087	32.801087

**Figure 4.29** and **Figure 4.30** show the normal probability plot and standardized residual plot, respectively. Neither provides any concern with respect to the normality assumption, i.e., the normal probability plot appears in general as a straight diagonal line. Moreover, the residuals seem to be randomly scattered—a regular pattern would be cause for concern.

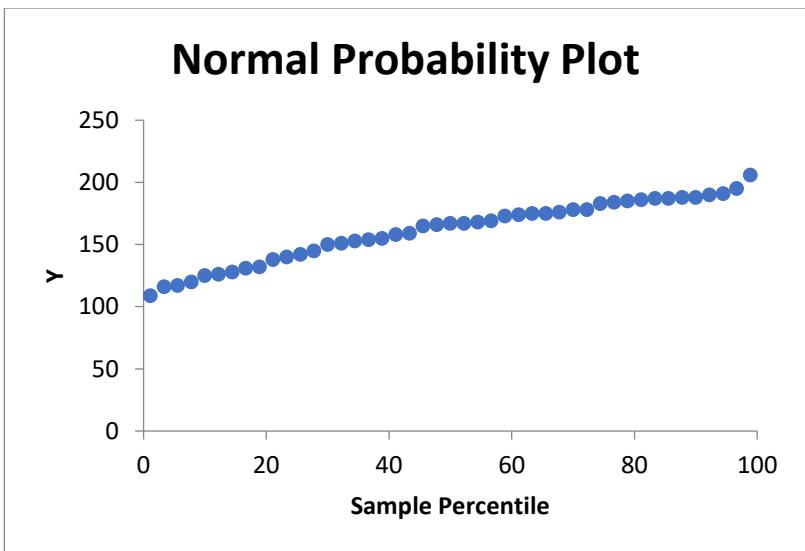


Figure 4.29. Normal probability plot.

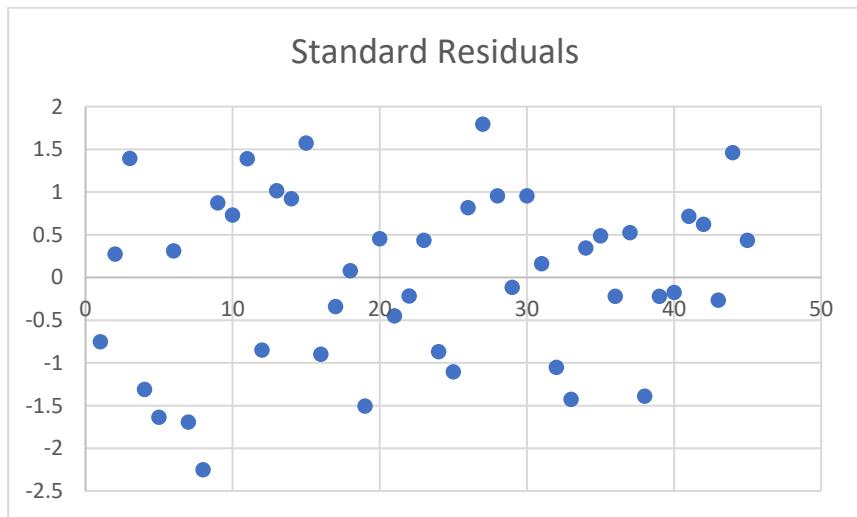


Figure 4.30. Residual plot using standardized residuals.

Table 4-27 provides a partial set of formulas used to generate Figure 4.27.

**Table 4-27. Formulas used for generating Figure 4.27**

t1	t2	t3
=IF(AND(S4="WD-20",T4=[@Y]),1,IF(AND(S4 ="WD-21",T9=AC9),-1,0))	=IF(AND(T4=[@Y],T\$3="Silo"),1,IF(AND(T4=[@Y],T\$3="Silo 2"),-1,0))	=IF(AND(V4=[@Y],V\$3="Silo"),1,IF(AND(V4=[@Y],V\$3="Silo 3"),-1,0))
=IF(AND(S9="WD-20",T9=[@Y]),1,IF(AND(S9 ="WD-21",T14=AC14),-1,0))	=IF(AND(T9=[@Y],T\$3="Silo"),1,IF(AND(T9=[@Y],T\$3="Silo 2"),-1,0))	=IF(AND(V9=[@Y],V\$3="Silo"),1,IF(AND(V9=[@Y],V\$3="Silo 3"),-1,0))
=IF(AND(S14="WD-20",T14=[@Y]),1,IF(AND(S14 ="WD-21",T19=AC19),-1,0))	=IF(AND(T14=[@Y],T\$3="Silo"),1,IF(AND(T14=[@Y],T\$3="Silo 2"),-1,0))	=IF(AND(V14=[@Y],V\$3="Silo"),1,IF(AND(V14=[@Y],V\$3="Silo 3"),-1,0))
=IF(AND(S4="WD-20",U4=[@Y]),1,IF(AND(S4 ="WD-21",U4=AC19),-1,0))	=IF(AND(U4=[@Y],U\$3="Silo"),1,IF(AND(U4=[@Y],U\$3="Silo 2"),-1,0))	=IF(AND(V4=[@Y],V\$3="Silo"),1,IF(AND(V4=[@Y],V\$3="Silo 3"),-1,0))
=IF(AND(S9="WD-20",U9=[@Y]),1,IF(AND(S9 ="WD-21",U9=AC24),-1,0))	=IF(AND(U9=[@Y],U\$3="Silo"),1,IF(AND(U9=[@Y],U\$3="Silo 2"),-1,0))	=IF(AND(V9=[@Y],V\$3="Silo"),1,IF(AND(V9=[@Y],V\$3="Silo 3"),-1,0))
=IF(AND(S14="WD-20",U14=[@Y]),1,IF(AND(S14 ="WD-21",U14=AC29),-1,0))	=IF(AND(U14=[@Y],U\$3="Silo"),1,IF(AND(U14=[@Y],U\$3="Silo 2"),-1,0))	=IF(AND(V14=[@Y],V\$3="Silo"),1,IF(AND(V14=[@Y],V\$3="Silo 3"),-1,0))
=IF(AND(S4="WD-20",V4=[@Y]),1,IF(AND(S4 ="WD-21",V4=AC34),-1,0))	=IF(AND(U4=[@Y],U\$3="Silo"),1,IF(AND(U4=[@Y],U\$3="Silo 2"),-1,0))	=IF(AND(V4=[@Y],V\$3="Silo"),1,IF(AND(V4=[@Y],V\$3="Silo 3"),-1,0))
=IF(AND(S9="WD-20",V9=[@Y]),1,IF(AND(S9 ="WD-21",V9=AC39),-1,0))	=IF(AND(U9=[@Y],U\$3="Silo"),1,IF(AND(U9=[@Y],U\$3="Silo 2"),-1,0))	=IF(AND(V9=[@Y],V\$3="Silo"),1,IF(AND(V9=[@Y],V\$3="Silo 3"),-1,0))
=IF(AND(S14="WD-20",V14=[@Y]),1,IF(AND(S14 ="WD-21",V14=AC44),-1,0))	=IF(AND(U14=[@Y],U\$3="Silo"),1,IF(AND(U14=[@Y],U\$3="Silo 2"),-1,0))	=IF(AND(V14=[@Y],V\$3="Silo"),1,IF(AND(V14=[@Y],V\$3="Silo 3"),-1,0))

## Latin Squares Design

A Latin Square design has two nuisance factors (Rows and Cols) and one treatment factor, each of which has the same number of levels, denoted  $r$ . There are no replications and no interactions. If we denote the possible treatment effects by Latin letters, then all the rows and columns are permutations of these letters (with no repeated rows and no repeated columns). For  $r = 4$  and  $r = 5$ , possible configurations are shown in **Figure 4.31**.

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

B	A	C	E	D
E	C	D	B	A
A	D	E	C	B
C	B	A	D	E
D	E	B	A	C

**Figure 4.31. Latin Square designs in Latin letters for  $r = 4$  and  $r = 5$**

Note that there are many possible  $4 \times 4$  or larger configurations, although many of these are equivalent in the sense that one can be obtained from another by interchanging one or more rows and/or columns. In fact, there are 4 non-equivalent  $4 \times 4$  configurations and 56 non-equivalent  $5 \times 5$  configurations. It turns out that all the  $3 \times 3$  configurations are equivalent.

### Example 4.9: Aircraft Parts

A defense contractor wants to determine whether there is a significant difference between four different methods of manufacturing an aircraft component, based on the number of millimeters of the part from the standard measurement. Four operators and four machines are assigned to the study. A Latin Squares design is used to account for operators and machines nuisance factors.

The representation of a Latin Squares design is shown in **Table 4-28** where A, B, C and D are the four manufacturing methods and the rows

correspond to the operators and the columns correspond to the machines.

**Table 4-28. Data in Latin Squares design format**

	1	2	3	4
1	A=4	B=2	C=5	D=7
2	B=1	C=2	D=6	A=5
3	C=8	D=9	A=6	B=3
4	D=11	A=3	B=7	C=8

We will use the following equivalent representations in **Figure 4.32**.

	A	B	C	D	E
1	Latin Squares - Excel Format				
2		1	2	3	4
3	1	A	B	C	D
4	2	B	C	D	A
5	3	C	D	A	B
6	4	D	A	B	C
7	1	4	2	5	7
8	2	1	6	6	5
9	3	8	9	6	3
10	4	11	3	7	8

**Figure 4.32. Data in Excel and Standard Formats**

The linear model of the Latin Squares design takes the form:

$$x_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \varepsilon_{ijk}$$

As usual,  $\sum \alpha_i = \sum \beta_j = \sum \tau_k = 0$  and  $\varepsilon_{ijk} \sim N(0, \sigma)$ .

An Excel implementation of the design is shown in

**Figure 4.33.** The SNOVA results are shown in **Figure 4.34.**

	G	H	I	J	K
2		1	2	3	4
3	1	4	2	5	7
4	2	1	6	6	5
5	3	8	9	6	3
6	4	11	3	7	8
7		6	5	6	5.75
8					
9	A	B	C	D	
10		4.5	3.25	6.75	8.25

Figure 4.33. Latin squares design in Excel Format

N	O	P	Q	R	S	T
ANOVA						
				Alpha	0.05	
	SS	df	MS	F	p-value	eta-sq
Treat	60.1875	3	20.0625	5.766467	0.03354	0.742483
Between	23.6875	3	7.8958	2.269461	0.18068	0.531557
Within	2.6875	3	0.8958	0.257485	0.85360	0.114058
Error	20.8750	6	3.4792			
Total	107.4375	15	7.1625			

Figure 4.34. Latin Squares design in Excel format

Error! Reference source not found. contains the means of each of the factor levels and **Figure 4.34** contains the ANOVA analysis. The degrees of freedom for all three factors is 3 (cells P4, P5, P6), equal to the number to  $r - 1$ ,  $df_{Treat} = r^2 - 1 = 15$ , while  $df_{Error} = (r - 1)(r - 2) = 6$ .

We see from **Figure 4.34** that there is a significant difference between the four methods ( $p - value = 0.03345 < .04 = \alpha$ ). There is no significant difference between the operators or between the machines, and so blocking on these factors may not have been necessary in this case.

Excel formulas for all of the calculations are shown in **Table 4-29**.

**Table 4-29.** Excel formulas for the Latin Squares analysis

Cell	Factor	Formula
<i>Treatment</i>	O4	=DEVSQ(H11:K11)*(P5+1)
<i>Between</i>	O5	=DEVSQ(L4:L7)*(P6+1)
<i>Within</i>	O6	=DEVSQ(H8:K8)*(P7+1)
<i>Error</i>	O7	=O9-SUM(O5:O7)
<i>Total</i>	O8	=DEVSQ(H4:K7)
<i>Mean<sub>Row1</sub></i>	L3	=AVERAGE(H3:K3)
<i>Mean<sub>Row2</sub></i>	L4	=AVERAGE(H4:K4)
<i>Mean<sub>Row3</sub></i>	L5	=AVERAGE(H5:K5)
<i>Mean<sub>Row4</sub></i>	L6	=AVERAGE(H6:K6)
<i>Mean<sub>Total</sub></i>	L7	=AVERAGE(H3:H6)
<i>Mean<sub>Col1</sub></i>	H7	=AVERAGE(H3:H6)
<i>Mean<sub>Col2</sub></i>	I7	=AVERAGE(I3:I6)
<i>Mean<sub>Col3</sub></i>	J7	=AVERAGE(J3:J6)
<i>Mean<sub>Col4</sub></i>	K7	=AVERAGE(K3:K6)
<i>Mean<sub>TreatA</sub></i>	H10	=AVERAGEIF(B3:E6,H9,B7:E10)
<i>Mean<sub>TreatB</sub></i>	I10	=AVERAGEIF(B3:E6,I9,B7:E10)
<i>Mean<sub>TreatC</sub></i>	J10	=AVERAGEIF(B3:E6,J9,B7:E10)
<i>Mean<sub>TreatD</sub></i>	K10	=AVERAGEIF(B3:E6,K9,B7:E10)
<i>SS<sub>Treat</sub></i>	O4	=DEVSQ(H10:K10)*(P5+1)
<i>SS<sub>Between</sub></i>	O5	=DEVSQ(L3:L6)*(P6+1)
<i>SS<sub>Within</sub></i>	O6	=DEVSQ(H7:K7)*(P6+1)
<i>SS<sub>Error</sub></i>	O7	=O8-SUM(O4:O6)
<i>SS<sub>Total</sub></i>	O8	=DEVSQ(H3:K6)
<i>df<sub>Treat</sub></i>	P4	=COUNT(B7:B10)-1
<i>df<sub>Between</sub></i>	P5	=COUNT(C7:C10)-1
<i>df<sub>Within</sub></i>	P6	=COUNT(D7:D10)-1
<i>df<sub>Error</sub></i>	P7	=(COUNT(E7:E10)-1)* (COUNT(E7:E10)-2)
<i>df<sub>Total</sub></i>	P8	=SUM(P4:P7)
<i>MS<sub>Treat</sub></i>	Q4	=O4/P4
<i>MS<sub>Between</sub></i>	Q1	=O5/P5
<i>MS<sub>Within</sub></i>	Q2	=O6/P6
<i>MS<sub>Error</sub></i>	Q3	=O7/P7
<i>MS<sub>Total</sub></i>	Q4	=O8/P8
<i>F<sub>Treat</sub></i>	R4	=Q4/\$Q\$7
<i>F<sub>Between</sub></i>	R5	=Q5/\$Q\$7
<i>F<sub>Within</sub></i>	R6	=Q6/\$Q\$7
<i>p<sub>Treat</sub></i>	S4	=FDIST(R4,P4,\$P\$7)
<i>p<sub>Between</sub></i>	S5	=FDIST(R5,P5,\$P\$7)
<i>P<sub>Within</sub></i>	S6	=FDIST(R6,P6,\$P\$7)
<i>p eta - sqTreat</i>	T4	=O4/(\$O\$7+O4)

Cell	Factor	Formula
$p\ eta - sq_{Between}$	T5	=O5/(\$O\$7+O5)
$p\ eta - sq_{Within}$	T6	=O6/(\$O\$7+O6)

## Two-Factor ART ANOVA

Aligned Rank Transform (ART) ANOVA is a non-parametric approach to factorial ANOVA that enables you to analyze the interaction as well as the main effects. As usual, ranked data is used, but first, the data for each effect (main or interaction) must be aligned before ranks are calculated.

This approach is useful when the data is not normally distributed. It can be used when the homogeneity of variances assumption is violated, although there is a risk of an inflated alpha value (alpha is up to about 0.07 when set to 0.05 for the interaction effect and up to 0.09 for the main effects).

## Two Factor ART ANOVA

When we have two factors (row and column), we assume that the data can be represented as  $x_{ijk}$  where  $i$  represents the rows,  $j$  represents the columns and  $k$  represents replications. Let  $\bar{x}_{ij} =$  the average of  $\{x_{ijk}: k\}$ ,  $\bar{x}_i =$  the average of  $\{x_{ijk}: j, k\}$ ,  $\bar{x}_j =$  the average of  $\{x_{ijk}: i, k\}$  and  $\bar{x} =$  the average of  $\{x_{ijk}: i, j, k\}$ .

For the rows factor, we adjust each data element as follows:

$$x'_{ijk} = x_{ijk} - \bar{x}_{ij} + \bar{x}_i - \bar{x}$$

For the columns factor, we adjust each data element as follows:

$$x'_{ijk} = x_{ijk} - \bar{x}_{ij} + \bar{x}_j - \bar{x}$$

For the interaction of the factors, we adjust each data element as follows:

$$x'_{ijk} = x_{ijk} - \bar{x}_i - \bar{x}_j + \bar{x}$$

Now we conduct three separate ANOVA analyses, one for the rows main effect, one for the columns main effect and a third for the interaction effect. The data for these are the ranked values of the  $x'_{ijk}$ .

### Example 4.10: ART ANOVA Likert Scale Questionnaire

Four operators answered a three-survey with a Likert scale three different times over three days. The results were recorded and we want to analyze them for consistency. There are three hypotheses we will investigate. First, that there are no inconsistencies between the operators answering the questions. Second, there is no inconsistencies within each operator. Three, there are no interactions.

Implementing ART ANOVA in Excel requires running a two-factor ANOVA, followed by iterating through the ART process for the rows (operators), columns(questions), and interactions. The results are shown in Figure 4.35 through Figure 4.45. The formulas appear in **Table 4-30** through **Table 4-34**.

	S	T	U	V	W	X	Y	Z
3	ANOVA				Alpha	0.05		
4		SS	df	MS	F	p-value	p eta-sq	significant
5	Operators	64.056	1	64.0556	0.5087	0.48123	0.76815	No
6	Questions	25.542	2	12.7708	0.1079	0.89804	0.56917	No
7	Interactions	1477.389	2	738.6944	9.2678	0.00073	0.98708	Yes
8	Within	19.333	30	0.6444				
9	Total	34.306	35	0.9802				

Figure 4.35. Final ART ANOVA table

	A	B	C
1	ART ANOVA		
2			
3	Row (Subj)	Col (Ques)	Resp
4	1	1	5
5	1	1	4
6	1	1	3
7	1	2	2
8	1	2	1
9	1	2	3
10	1	3	4
11	1	3	3
12	1	3	4
13	2	1	3
14	2	1	2
15	2	1	4
16	2	2	4
17	2	2	4
18	2	2	2
19	2	3	3
20	2	3	3
21	2	3	2
22	3	1	4
23	3	1	5
24	3	1	3
25	3	2	2
26	3	2	3
27	3	2	4
28	3	3	2
29	3	3	3
30	3	3	2
31	4	1	3
32	4	1	3
33	4	1	2
34	4	2	3
35	4	2	4
36	4	2	5
37	4	3	2
38	4	3	3
39	4	3	4

Figure 4.36. Questionnaire data

	E	F	G	H	I
5	COUNT				
6		1	2	3	
7	1	3	3	3	9
8	2	3	3	3	9
9	3	3	3	3	9
10	4	3	3	3	9
11		12	12	12	36
12	MEAN				
13		1	2	3	
14	1	4	2	3.666667	3.222222
15	2	3	3.333333	2.666667	3
16	3	4	3	2.333333	3.111111
17	4	2.666667	4	3	3.222222
18		3.416667	3.083333	2.916667	3.138889
19	VARIANCE				
20		1	2	3	
21	1	1	1	0.333333	0.148148
22	2	1	1.333333	0.333333	0.259259
23	3	1	1	0.333333	0.148148
24	4	0.333333	1	1	0.148148
25		0.111111	0.027778	0.111111	0.130471

Figure 4.37. Summary statistics

	K	L	M	N	O	P	Q
5	ANOVA				Alpha	0.05	
6		SS	df	MS	F	p-value	p eta-sq
7	Operators	16.5278	3	5.50926	4.83740	0.00900	0.37682
8	Questions	4.0556	2	2.02778	1.78049	0.19011	0.12920
9	Interactions	3.0556	6	0.50926	0.44715	0.83976	0.10055
10	Within	27.3333	24	1.13889			
11	Total	50.9722	35	1.45635			

Figure 4.38. ANOVA results

	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
4	Means				Aligned Data				Adjusted Means				
5	Tot	Row	Col	Int	Row	Col	Int		Tot	Row	Col	Int	
6	3.028	3.667	3.500	4.000	1.639	1.472	0.861		3.028	3.667	3.500	4.306	
7	3.028	3.667	3.500	4.000	0.639	0.472	-0.139		3.028	3.667	3.500	4.306	
8	3.028	3.667	3.500	4.000	-0.361	-0.528	-1.139		3.028	3.667	3.500	4.306	
9	3.028	3.667	2.833	3.333	0.306	-0.528	-0.472		3.028	3.667	2.833	4.306	
10	3.028	3.667	2.833	3.333	-0.694	-1.528	-1.472		3.028	3.667	2.833	4.306	
11	3.028	3.667	2.833	3.333	2.306	1.472	1.528		3.028	3.667	2.833	4.306	
12	3.028	3.667	2.750	3.667	1.972	1.056	1.611		3.028	3.667	2.750	4.306	
13	3.028	3.667	2.750	3.667	-0.028	-0.944	-0.389		3.028	3.667	2.750	4.306	
14	3.028	3.667	2.750	3.667	-0.028	-0.944	-0.389		3.028	3.667	2.750	4.306	
15	3.028	3.333	3.500	4.333	0.972	1.139	1.194		3.028	3.333	3.500	3.639	
16	3.028	3.333	3.500	4.333	-0.028	0.139	0.194		3.028	3.333	3.500	3.639	
17	3.028	3.333	3.500	4.333	-0.028	0.139	0.194		3.028	3.333	3.500	3.639	
18	3.028	3.333	2.833	3.000	0.306	-0.194	-0.139		3.028	3.333	2.833	3.639	
19	3.028	3.333	2.833	3.000	-0.694	-1.194	-1.139		3.028	3.333	2.833	3.639	
20	3.028	3.333	2.833	3.000	1.306	0.806	0.861		3.028	3.333	2.833	3.639	
21	3.028	3.333	2.750	2.667	-1.361	-1.944	-2.056		3.028	3.333	2.750	3.639	
22	3.028	3.333	2.750	2.667	1.639	1.056	0.944		3.028	3.333	2.750	3.639	
23	3.028	3.333	2.750	2.667	0.639	0.056	-0.056		3.028	3.333	2.750	3.639	
24	3.028	1.889	3.500	2.000	-1.139	0.472	-0.361		3.028	1.889	3.500	0.750	
25	3.028	1.889	3.500	2.000	-2.139	-0.528	-1.361		3.028	1.889	3.500	0.750	
26	3.028	1.889	3.500	2.000	-0.139	1.472	0.639		3.028	1.889	3.500	0.750	
27	3.028	1.889	2.833	1.667	-0.806	0.139	0.306		3.028	1.889	2.833	0.750	
28	3.028	1.889	2.833	1.667	-1.806	-0.861	-0.694		3.028	1.889	2.833	0.750	
29	3.028	1.889	2.833	1.667	-0.806	0.139	0.306		3.028	1.889	2.833	0.750	
30	3.028	1.889	2.750	2.000	-1.139	-0.278	0.389		3.028	1.889	2.750	0.750	
31	3.028	1.889	2.750	2.000	-1.139	-0.278	0.389		3.028	1.889	2.750	0.750	
32	3.028	1.889	2.750	2.000	-1.139	-0.278	0.389		3.028	1.889	2.750	0.750	
33	3.028	3.222	3.500	3.667	1.528	1.806	1.306		3.028	3.222	3.500	3.417	
34	3.028	3.222	3.500	3.667	-1.472	-1.194	-1.694		3.028	3.222	3.500	3.417	
35	3.028	3.222	3.500	3.667	0.528	0.806	0.306		3.028	3.222	3.500	3.417	
36	3.028	3.222	2.833	3.333	0.861	0.472	0.972		3.028	3.222	2.833	3.417	
37	3.028	3.222	2.833	3.333	-0.139	-0.528	-0.028		3.028	3.222	2.833	3.417	
38	3.028	3.222	2.833	3.333	-0.139	-0.528	-0.028		3.028	3.222	2.833	3.417	
39	3.028	3.222	2.750	2.667	-0.472	-0.944	-0.944		3.028	3.222	2.750	3.417	
40	3.028	3.222	2.750	2.667	1.528	1.056	1.056		3.028	3.222	2.750	3.417	
41	3.028	3.222	2.750	2.667	-0.472	-0.944	-0.944		3.028	3.222	2.750	3.417	

Figure 4.39. Means, ART aligned, and adjusted means tables

	E	F	G	H	I
42	Descriptive Statistics				
43					
44	COUNT	balanced			
45		1	2	3	
46	1	3	3	3	9
47	2	3	3	3	9
48	3	3	3	3	9
49	4	3	3	3	9
50		12	12	12	36
51					
52	MEAN				
53		1	2	3	
54	1	24.6667	23.3333	24.6667	24.2222
55	2	24	22	21.6667	22.5556
56	3	8.16667	7	6.5	7.22222
57	4	19.8333	20.6667	19.5	20
58		19.1667	18.25	18.0833	18.5
59					
60	VARIANCE				
61		1	2	3	
62	1	82.3333	156.333	80.0833	80.1319
63	2	18.75	84	246.333	88.4653
64	3	66.0833	18.75	0	21.7569
65	4	223.083	40.3333	108	93.125
66		118.697	101.432	131.447	110.729

Figure 4.40. Row rank and descriptive statistics

	K	L	M	N	O	P	Q	R
44	ANOVA				Alpha	0.05		
45		SS	df	MS	F	p-value	p eta-sq	significant
46	Operators	1607.667	3	535.8889	5.72081	0.00422	0.98328	Yes
47	Questions	8.167	2	4.0833	0.04359	0.95742	0.23005	No
48	Interaction:	11.500	6	1.9167	0.02046	0.99995	0.29614	No
49	Within	2248.167	24	93.6736				
50	Total	3875.500	35	110.7286				

Figure 4.41. ANOVA for ART rows

A	B	C	D	E	F	G	H	I
Row	Col	Col Rank	Descriptive Statistics					
1	1	34						
1	1	25	COUNT balanced					
1	1	12		1	3	3	3	9
1	2	12		2	3	3	3	9
1	2	2		3	3	3	3	9
1	2	34		4	3	3	3	9
1	3	30			12	12	12	36
1	3	6.5						
1	3	6.5						
2	1	32	MEAN					
2	1	21.5		1	4.0000	3.3333	3.6667	3.6667
2	1	21.5		2	4.3333	3.0000	2.6667	3.3333
2	2	18		3	2.0000	1.6667	2.0000	1.8889
2	2	3.5		4	3.6667	3.3333	2.6667	3.2222
2	2	27.5			3.5000	2.8333	2.7500	18.5000
2	3	1						
2	3	30						
2	3	19	VARIANCE					
3	1	25			1.0000	2.0000	3.0000	
3	1	12		1	122.3333	268.0000	184.0833	162.1875
3	1	34		2	36.7500	146.0833	214.3333	117.3750
3	2	21.5		3	122.3333	52.0833	0.0000	56.1875
3	2	9		4	284.0833	56.3333	184.0833	144.1250
3	2	21.5			103.7879	95.2727	107.0606	110.1714
3	3	16						
3	3	16						
3	3	16						
4	1	36						
4	1	3.5						
4	1	27.5						
4	2	25						
4	2	12						
4	2	12						
4	3	6.5						
4	3	30						
4	3	6.5						

Figure 4.42. Column ranks and descriptive statistics

K	L	M	N	O	P	Q	R
ANOVA				Alpha	0.05		
	SS	df	MS	F	p-value	p eta-sq	significant
Operators	16.5278	3	5.50926	0.03958	0.98922	0.37682	No
Questions	4.05556	2	2.02778	0.01457	0.98555	0.1292	No
Interactions	494.417	6	82.4028	0.59194	0.73357	0.94761	No
Within	3341	24	139.208				
Total	3856	35	110.171				

Figure 4.43. ANOVA for ART columns

A	B	C	D	E	F	G	H	I
Row	Col	Int Rank		Descriptive Statistics				
120	1	1	29					
121	1	1	15	COUNT	balanced			
123	1	1	6		1	2	3	
124	1	2	10	1	3	3	3	9
125	1	2	3	2	3	3	3	9
126	1	2	35	3	3	3	3	9
127	1	3	36	4	3	3	3	9
128	1	3	11.5		12	12	12	36
129	1	3	11.5					
130	2	1	33	MEAN	1	2	3	
131	2	1	19.5		4.00000	3.33333	3.66667	3.66667
132	2	1	19.5	1	4.33333	3.00000	2.66667	3.33333
133	2	2	14	2	2.00000	1.66667	2.00000	1.88889
134	2	2	5	3	3.66667	3.33333	2.66667	3.22222
135	2	2	28	4	3.50000	2.83333	2.75000	18.50000
136	2	3	1					
137	2	3	30					
138	2	3	16	VARIANCE	1	2	3	
139	3	1	13		134.3333	283.0000	200.0833	157.2153
140	3	1	4	1	60.7500	134.3333	210.3333	118.7153
141	3	1	27	2	134.3333	56.3333	0.0000	68.8611
142	3	2	22	3	261.3333	60.7500	200.0833	138.1250
143	3	2	9	4	120.7424	104.1061	127.0000	110.8286
144	3	2	22					
145	3	3	25					
146	3	3	25					
147	3	3	25					
148	4	1	34					
149	4	1	2					
150	4	1	22					
151	4	2	31					
152	4	2	17.5					
153	4	2	17.5					
154	4	3	7.5					
155	4	3	32					
156	4	3	7.5					

Figure 4.44. Interactions ranks and descriptive statistics

K	L	M	N	O	P	Q	R
ANOVA				Alpha	0.05		
	SS	df	MS	F	p-value	p eta-sq	significant
Operators	16.5278	3	5.50926	0.03809	0.98981	0.37682	No
Questions	4.05556	2	2.02778	0.01402	0.98609	0.1292	No
Interactions	387.083	6	64.5139	0.44603	0.84054	0.93404	No
Within	3471.33	24	144.639				
Total	3879	35	110.829				

Figure 4.45. ANOVA for ART interactions

**Table 4-30. Formulas Summary Statistics Counts, Means, and Variances**

Function	Cell	Formula
COUNT(1,1) to COUNT(4,3)	F7	=COUNTIFS(\$A\$4:\$A\$39,\$E7,\$B\$4:\$B\$39,F\$6)
ROW1 SUM to ROW4 SUM	I7	=SUM(F7:H7)
COL1 SUM to COL3 SUM	F11	=SUM(F7:F10)
GRAND SUM	I11	=SUM(F7:H10)
MEAN(1,1) to MEAN(4,3)	F14	=COUNTIFS(\$A\$4:\$A\$39,\$E7,\$B\$4:\$B\$39,F\$6)
ROW1 MEAN to ROW4 MEAN	I14	=AVERAGE(F15:H15)
COL1 MEAN to COL3 MEAN	F18	=AVERAGE(F15:F18)
GRAND MEAN	I18	=AVERAGE(C4:C39)
VAR(1,1)	F21	=VAR.S(C4:C6)
VAR(2,1)	F22	=VAR.S(C13:C15)
VAR(3,1)	F23	=VAR.S(C22:C24)
VAR(4,1)	F24	=VAR.S(C31:C33)
VAR(1,1)	G21	=VAR.S(C7:C9)
VAR(1,2)	G22	=VAR.S(C16:C18)
VAR(2,2)	G23	=VAR.S(C25:C27)
VAR(4,2)	G24	=VAR.S(C34:C36)
VAR(1,3)	H21	=VAR.S(C10:C12)
VAR(2,3)	H22	=VAR.S(C19:C21)
VAR(2,2)	H23	=VAR.S(C28:C30)
VAR(4,3)	H24	=VAR.S(C37:C39)
ROW1 VAR to ROW4 VAR	I21	=VAR.S(C4:C12)
COL1 VAR to COL3 VAR	F25	=VAR.S(C4:C6,C13:C15,C22:C24,C31:C33)
GRAND VAR	I25	=VAR.S(C4:C39)
$SS_{Rows}$	L7	=DEVSQ(I15:I18)*I7
$SS_{Columns}$	L8	=DEVSQ(F19:H19)*F11
$SS_{Inter}$	L9	=L11-L7-L8-L10
$SS_{Within}$	L10	=SUM(F23:H26)*(F7-1)
$SS_{Total}$	L11	=DEVSQ(\$C\$4:\$C\$39)
$df_{Rows}$	M7	=COUNT(I15:I18)-1
$df_{Columns}$	M8	=COUNT(F19:H19)-1
$df_{Inter}$	M9	=M8*M7
$df_{within}$	M10	=M11-M7-M8-M9
$df_{Total}$	M11	=I11-1
$MS_{Rows}$	N7	=L7/M7
$MS_{Columns}$	N8	=L8/M8
$MS_{Inter}$	N9	=L9/M9

Function	Cell	Formula
$MS_{Within}$	N10	=L10/M10
$MS_{Total}$	N11	=L11/M11
$F_{Stat\ Rows}$	O7	=N7/N10
$F_{Stat\ Columns}$	O8	=N8/N10
$F_{Stat\ Inter}$	O9	=N9/N10
$p\text{-value}_{Rows}$	P7	=F.DIST.RT(O7,M7,M10)
$p\text{-value}_{Cols}$	P8	=F.DIST.RT(O8,M8,M10)
$p\text{-value}_{Inter}$	P9	=F.DIST.RT(O9,M9,M10)
$p\text{-eta-sq}_{Rows}$	Q7	=L7/(L7+L\$10)
$p\text{-eta-sq}_{columns}$	Q8	=L8/(L8+L\$10)
$p\text{-eta-sq}_{Inter}$	Q9	=L9/(L9+L\$10)

Table 4-31. Means, ART aligned, and adjusted means formulas

Function	Cell	Formula
Means Total copied down	S43	=AVERAGE(\$C\$4:\$C\$39)
Means Rows copied down	T43	=AVERAGEIF(\$A\$4:\$A\$39,A4,\$C\$4:\$C\$39)
Means Columns copied down	U43	=AVERAGEIF(\$B\$4:\$B\$39,B4,\$C\$4:\$C\$39)
Means Inter copied down	V43	=AVERAGEIFS(\$C\$4:\$C\$39,\$A\$4:\$A\$39,A4,\$B\$4:\$B\$39,B4)
Aligned Rows copied down	X43	=C4-V6+T6-S6
Aligned Columns copied down	Y43	=C4-V6+U6-S6
Aligned Inter copied down	Z43	=C4-T6-U6+S6
Adj Total Means copied down	AB43	=\$I\$19
Adj Row Means copied down	AC43	=INDEX(\$I\$15:\$I\$18,MATCH(A4,\$E\$15:\$E\$18),1)
Adj Col Means copied down	AD43	=INDEX(\$F\$19:\$H\$19,1,MATCH(B4,\$F\$14:\$H\$14))
Adj Inter Means copied down	AE44	=C4-Y6-Z6+X6

Table 4-32. ART ANOVA formulas, first iteration (rows)

Function	Cell	Formula
Row Rank copied down	C43	=RANK.AVG(X6,X\$6:X\$41,1)
COUNT(1,1) to COUNT(4,3)	F46	=COUNTIFS(\$A\$43:\$A\$78,\$E46,\$B\$43:\$B\$78,F\$6)
MEAN(1,1) to MEAN(4,3)	F54	=AVERAGEIFS(\$C\$43:\$C\$78,\$A\$43:\$A\$78,\$E54,\$B\$43:\$B\$78,F\$14)
VAR(1,1) like Cells F21:H24	F62:H65	=VAR.S(C43:C45)
$SS_{Rows}$	L46	=DEVSQ(I54:I57)*I46
$SS_{Columns}$	L47	=DEVSQ(F58:H58)*F50
$SS_{Inter}$	L48	=L50-L46-L47-L49
$SS_{Within}$	L49	=SUM(F62:H65)*(F46-1)

Function	Cell	Formula
$SS_{Total}$	L50	=DEVSQ(\$C\$43:\$C\$78)
$df_{Rows}$	M46	=COUNT(I54:I57)-1
$df_{Columns}$	M47	=COUNT(F58:H58)-1
$df_{Inter}$	M48	=M47*M46
$df_{Within}$	M49	=M50-M46-M47-M48
$df_{Total}$	M50	=I50-1
$MS_{Rows}$	N46	=L46/M46
$MS_{Columns}$	N47	=L47/M47
$MS_{Inter}$	N48	=L48/M48
$MS_{Within}$	N49	=L49/M49
$MS_{Total}$	N50	=L50/M50
$F_{Stat\ Rows}$	O46	=N46/N49
$F_{Stat\ Columns}$	O47	=N47/N49
$F_{Stat\ Inter}$	O48	=N48/N49
$p\text{-value}_{Rows}$	P46	=F.DIST.RT(O46,M46,M49)
$p\text{-value}_{Cols}$	P47	=F.DIST.RT(O47,M47,M49)
$p\text{-value}_{Inter}$	P48	=F.DIST.RT(O48,M48,M49)
$p\text{-eta-sq}_{Rows}$	Q46	=L46/(L46+L\$10)
$p\text{-eta-sq}_{Columns}$	Q47	=L47/(L47+L\$10)
$p\text{-eta-sq}_{Inter}$	Q48	=L48/(L48+L\$10)

Table 4-33. ART ANOVA formulas, second iteration (columns)

Function	Cell	Formula
$Col\ Rank\ copied\ down$	C82	=RANK.AVG(Y6,Y\$6:Y\$41,1)
$COUNT(1,1)\ to\ COUNT(4,3)$	F85	=RANK.AVG(Y7,Y\$6:Y\$41,1)
$MEAN(1,1)\ to\ MEAN(4,3)$	F93	=RANK.AVG(Y8,Y\$6:Y\$41,1)
$VAR(1,1)\ like\ Cells\ F21:H24$	F101:H104	=RANK.AVG(Y9,Y\$6:Y\$41,1)
$SS_{Rows}$	L85	=RANK.AVG(Y10,Y\$6:Y\$41,1)
$SS_{Columns}$	L86	=RANK.AVG(Y11,Y\$6:Y\$41,1)
$SS_{Inter}$	L87	=RANK.AVG(Y12,Y\$6:Y\$41,1)
$SS_{Within}$	L88	=RANK.AVG(Y13,Y\$6:Y\$41,1)
$SS_{Total}$	L89	=RANK.AVG(Y14,Y\$6:Y\$41,1)
$df_{Rows}$	M85	=RANK.AVG(Y15,Y\$6:Y\$41,1)
$df_{Columns}$	M86	=RANK.AVG(Y16,Y\$6:Y\$41,1)
$df_{Inter}$	M87	=RANK.AVG(Y17,Y\$6:Y\$41,1)
$df_{Within}$	M88	=RANK.AVG(Y18,Y\$6:Y\$41,1)
$df_{Total}$	M89	=RANK.AVG(Y19,Y\$6:Y\$41,1)
$MS_{Rows}$	N85	=RANK.AVG(Y20,Y\$6:Y\$41,1)
$MS_{Columns}$	N86	=RANK.AVG(Y21,Y\$6:Y\$41,1)

Function	Cell	Formula
$MS_{Inter}$	N87	=RANK.AVG(Y22,Y\$6:Y\$41,1)
$MS_{Within}$	N88	=RANK.AVG(Y23,Y\$6:Y\$41,1)
$MS_{Total}$	N89	=RANK.AVG(Y24,Y\$6:Y\$41,1)
F Stat Rows	O85	=RANK.AVG(Y25,Y\$6:Y\$41,1)
F Stat Columns	O86	=RANK.AVG(Y26,Y\$6:Y\$41,1)
F Stat Inter	O87	=RANK.AVG(Y27,Y\$6:Y\$41,1)
p-value Rows	P85	=RANK.AVG(Y28,Y\$6:Y\$41,1)
p-value Cols	P86	=RANK.AVG(Y29,Y\$6:Y\$41,1)
p-value Inter	P87	=RANK.AVG(Y30,Y\$6:Y\$41,1)
p eta-sq Rows	Q85	=RANK.AVG(Y31,Y\$6:Y\$41,1)
p eta-sq columns	Q86	=RANK.AVG(Y32,Y\$6:Y\$41,1)
p eta-sq Inter	Q87	=RANK.AVG(Y33,Y\$6:Y\$41,1)

Table 4-34. ART ANOVA formulas, third iteration (interactions)

Function	Cell	Formula
Int Rank copied down	C121	=RANK.AVG(Z6,Z\$6:Z\$41,1)
COUNT(1,1) to COUNT(4,3)	F124	=COUNTIFS(\$A\$4:\$A\$39,\$E124,\$B\$4:\$B\$39,F\$6)
MEAN(1,1) to MEAN(4,3)	F132	=AVERAGEIFS(\$C\$4:\$C\$39,\$A\$4:\$A\$39,\$E132,\$B\$4:\$B\$39,F\$14)
VAR(1,1) like Cells F175:H178	F140:H143	=VAR.S(C121:C123)
SS <sub>Rows</sub>	L124	=DEVSQ(I132:I135)*I124
SS <sub>columns</sub>	L125	=DEVSQ(F136:H136)*F128
SS <sub>Inter</sub>	L126	=L128-L124-L125-L127
SS <sub>Within</sub>	L127	=SUM(F140:H143)*(F124-1)
SS <sub>Total</sub>	L128	=DEVSQ(\$C\$121:\$C\$156)
df <sub>Rows</sub>	M124	=COUNT(I132:I135)-1
df <sub>Columns</sub>	M125	=COUNT(F136:H136)-1
df <sub>Inter</sub>	M126	=M125*M124
df <sub>Within</sub>	M127	=M128-M124-M125-M126
df <sub>Total</sub>	M128	=I128-1
MS <sub>Rows</sub>	N124	=L124/M124
MS <sub>Columns</sub>	N125	=L125/M125
MS <sub>Inter</sub>	N126	=L126/M126
MS <sub>Within</sub>	N127	=L127/M127
MS <sub>Total</sub>	N128	=L128/M128
F Stat Rows	O124	=N124/N127
F Stat Columns	O125	=N125/N127
F Stat Inter	O126	=N126/N127
p-value Rows	P124	=F.DIST.RT(O124,M124,M127)

Function	Cell	Formula
p-value Cols	P125	=F.DIST.RT(O125,M125,M127)
p-value Inter	P126	=F.DIST.RT(O126,M126,M127)
p eta-sq Rows	Q124	=L124/(L124+L\$10)
p eta-sq columns	Q125	=L125/(L125+L\$10)
p eta-sq Inter	Q126	=L126/(L126+L\$10)

## 2<sup>k</sup> Factorial Design Basic Concepts

### Basic Concepts

2<sup>k</sup> **factorial designs** consist of  $k$  factors, each of which has two levels. A key use of such designs is to identify which of many variables is most important and should be considered for further analysis in more detail.

We restrict our discussion to completely randomized designs with fixed factors. To illustrate the key concepts, we will consider the case with 3 factors, which we will label A, B, and C. The levels will be denoted + and – (or +1 and -1). We will suppose that each factor is replicated  $n$  times.

### Example 4.11: Factorial Designs

Suppose that three factors, A, B, and C, each at two levels, are of interest. The design is called a 2<sup>3</sup> factorial design. We'll determine the effect of machining factors on the strength of ceramic material used in satellite construction.

- Response Variable Y = Mean (over 15 reps) of Ceramic Strength
- Factor 1 = Table Speed (2 levels: slow (.025 m/s) and fast (.125 m/s))
- Factor 2 = Down Feed Rate (2 levels: slow (.05 mm), fast (.125 mm))
- Factor 3 = Wheel Grit (2 levels: 140/170 and 80/100)
- 4 replications
- There are eight treatment combinations written in standard order as: (1), a, b, ab, c, ac, bc, abc.
- There are seven degrees of freedom; one degree of freedom associated with each main effect and interaction: A, B, C, AB, AC, BC, ABC.

Create the  $2^3$  factorial design for the data in **Figure 4.46**.

	A	B	C	D	E	F	G
1	2 <sup>3</sup> factorial design						
2							
3	A	B	C	1	2	3	4
4	-	-	-	18	28	25	24
5	+	-	-	28	41	35	28
6	-	+	-	31	31	40	48
7	+	+	-	48	43	39	43
8	-	-	+	40	39	36	40
9	+	-	+	35	33	31	38
10	-	+	+	50	45	43	52
11	+	+	+	35	38	45	47
12							
13	k	3	n	4			

**Figure 4.46.**  $2^3$  design with 4 replications

In this example,  $k = 3$  and  $n = 4$ . Three factors result in  $2^k = 2^3 = 8$  rows in **Figure 4.47**. The average effect and SS value for each factor, including interactions, is shown on the left side of **Figure 4.48**.

	H	I	J	K	L	M	N	O	P
3		A	B	C	AB	AC	BC	ABC	Tot
4		-1	-1	-1	1	1	1	-1	95
5		1	-1	-1	-1	-1	1	1	132
6		-1	1	-1	-1	1	-1	1	150
7		1	1	-1	1	-1	-1	-1	173
8		-1	-1	1	1	-1	-1	1	155
9		1	-1	1	-1	1	-1	-1	137
10		-1	1	1	-1	-1	1	-1	190
11		1	1	1	1	1	1	1	165
12	effect	1.0625	9.9375	6.0625	-1.313	-6.438	-2.063	0.4375	
13	SS	9.0313	790.03	294.03	13.781	331.53	34.031	1.5313	

**Figure 4.47.** Average effect sizes and ANOVA

	R	S	T	U	V	W	X
3		df	SS	MS	F	p-value	sig
4	A	1	9.03125	9.03125	0.361401	0.553366	No
5	B	1	790.0313	790.0313	31.61442	8.68E-06	Yes
6	C	1	294.0313	294.0313	11.76615	0.002189	Yes
7	AB	1	13.78125	13.78125	0.55148	0.464921	No
8	AC	1	331.5313	331.5313	13.26678	0.001294	Yes
9	BC	1	34.03125	34.03125	1.361817	0.254687	No
10	ABC	1	1.53125	1.53125	0.061276	0.806597	No
11	Err	24	599.75	24.98958			
12	Tot	31	2073.719	66.89415			

Figure 4.48.  $2^3$  factorial analysis output

Range I4:K11 is derived from range A4:C11 and range L4:O11 is calculated from I4:K11 in the usual way (e.g., cell O4, corresponding to A\*B\*C for row 4, is calculated via the formula =I4\*J4\*K4). The values in column P are the sums of the replication values (e.g., cell P4 contains the formula =SUM(D4:G4)).

## Effect Sizes

In Figure 4.49. Effect and Sum of Squares table now calculate the effect size and SS for A (cells I12 and I13) using the formulas in Table 4-35

	H	I	J	K	L	M	N	O
12	effect	1.0625	9.9375	6.0625	-1.313	-6.438	-2.063	0.4375
13	SS	9.0313	790.03	294.03	13.781	331.53	34.031	1.5313

Figure 4.49. Effect and Sum of Squares table

If we highlight range I12:O13 and press Ctrl-R, the effect sizes and SS for the other factors (including interactions) will be filled in as shown.

Since we only have two levels,  $df$  for each factor = 1 and so  $MS = SS$  for each factor. Since  $df_T = 2^k * n - 1 = 2^3 * 4 - 1 = 31$ , it follows that  $df_{Err} = df_{Tot} - (2^k - 1) = 2^k * (n - 1) = 2^3 * (4 - 1) = 24$ .

All the other ANOVA values shown on the right side of Figure 2 are calculated in the usual way, making use of the fact that  $SS_{Tot} =$

$\text{DEVSQ}(\text{D4:G11})$  and  $SS_{Err} = SS_{Tot} - (SSA + SSB + SSC + SSAB + SSAC + SSBC + SSABC)$ .

We see from Figure 2 that factors B, C, and AC are significant. Note that since AB, BC and ABC are not significant, we could include them in the error term. Note that even though A is not significant, we should keep it in the model since AC is significant.

## Standard Error

The standard error for each effect size is equal to  $\sqrt{MSE/(2n)}$ , which in our example is  $\sqrt{24.9896/8} = 1.7674$ . We can thus find the 95% confidence interval for the effect sizes as shown in **Figure 4.50**. The formulas used to generate this figure are provided in **Table 4-37**. Note too that the  $p$ -values are identical to those shown in **Figure 4.48**.

	H	I	J	K	L	M	N
19		effect	s.e.	t-stat	p-val	lower	upper
20	A	1.0625	1.7674	0.6012	0.5534	-2.585	4.7102
21	B	9.9375	1.7674	5.6227	9E-06	6.2898	13.585
22	C	6.0625	1.7674	3.4302	0.0022	2.4148	9.7102
23	AB	-1.313	1.7674	-0.743	0.4649	-4.96	2.3352
24	AC	-6.438	1.7674	-3.642	0.0013	-10.09	-2.79
25	BC	-2.063	1.7674	-1.167	0.2547	-5.71	1.5852
26	ABC	0.4375	1.7674	0.2475	0.8066	-3.21	4.0852

**Figure 4.50. 95% confidence intervals of effect sizes**

## Example 4:12: Factorial Regression Model

We can also establish a regression model (using A, B, C, etc. as the names for the corresponding variables) as follows:

$$y = \beta_0 + \beta_A A + \beta_B B + \beta_C C + \beta_{AB} AB + \beta_{AC} AC + \beta_{BC} BC + \beta_{ABC} ABC + \varepsilon$$

It turns out that the intercept  $\beta_0 = \text{AVERAGE}(\text{D4:G11})$  and  $\beta_A = \text{half of the average effect size for factor A}$ . and similarly for all the other factors. Thus, for our example

$$y = 37.406 + 0.513A + 4.969B + 3.031C - 0.656AB - 3.219AC \\ - 1.031BC + 0.219ABC + \varepsilon$$

where each variable A, B and C takes the value +1 or -1 (and expressions such as AB should be interpreted as A times B). The standard error for each coefficient is equal to half of the standard error calculated previously, namely  $1.7807/2 = 0.8904$ . The regression analysis is shown in **Figure 4.51** and **Figure 4.52**.

	Z	AA
3	Regression Analysis	
4		
5	OVERALL FIT	
6	Multiple R	0.843081
7	R Square	0.710785
8	Adjusted R Sq	0.626431
9	Standard Error	4.998958
10	Observations	32

*Figure 4.51. Regression analysis output*

	Z	AA	AB	AC	AD	AE	AF
12	ANOVA				Alpha	0.05	
13		df	SS	MS	F	p-value	sig
14	Regression	7	1473.969	210.567	8.426189	3.32E-05	yes
15	Residual	24	599.75	24.98958			
16	Total	31	2073.719				
17							
18		coeff	std err	t stat	p-value	lower	upper
19	Intercept	37.40625	0.883699	42.32916	4.61E-24	40.50529	44.15303
20	A	0.53125	0.883699	0.601166	0.553366	-1.2227	2.425032
21	B	4.96875	0.883699	5.62267	8.68E-06	3.798805	7.446536
22	C	3.03125	0.883699	3.430183	0.002189	1.606317	5.254048
23	AB	-0.65625	0.883699	-0.74262	0.464921	-2.56648	1.081249
24	AC	-3.21875	0.883699	-3.64236	0.001294	-5.46622	-1.81849
25	BC	-1.03125	0.883699	-1.16697	0.254687	-2.99084	0.656896
26	ABC	0.21875	0.883699	0.247539	0.806597	-1.57633	2.071405

*Figure 4.52.  $2^3$  factorial design regression model*

The Excel formulas for each figure follows in **Table 4-35** through **Table 4-39**.

**Table 4-35. Excel formulas for Figure 4.47. Average effect sizes and ANOVA**

Name	Cell	Formula
effect	I12:O12	=SUMPRODUCT(I4:I11,\$P4:\$P11)/(2^(\$B\$13-1)*\$E\$13)
SS	I13:O13	=SUMPRODUCT(I4:I11,\$P4:\$P11)^2/(2^\$B\$13*\$E\$13)

**Table 4-36. Excel formulas for Figure 4.48**

Name	Cell	Formula
effect	I20:26	{=TRANSPOSE(I13:K13)}
MS-Factors	U4:U11	=T4/S4
F-Stats Factors	V4:V10	=U4/U\$11
p-value	W_4:W10	=FDIST(V4,S4,\$\$11)
Significant	X4:X10	=IF(W4<0.05,"Yes","No")

**Table 4-37. Excel formulas for Figure 4.50**

Name	Cell	Formula
effect	I20:26	{=TRANSPOSE(I12:O12)}
s.e.	J20	=SQRT(\$U\$11/(2*\$E\$13))
t-stat	K20	=I20/J20
p-val	L20	=TDIST(ABS(K20),\$\$11,2)
lower	M20	=I20-J20*TINV(0.05,\$\$11)
upper	N20	=I20+J20*TINV(0.05,\$\$11)

**Table 4-38. Excel formulas for**

**Figure 4.51**

Name	Cell	Formula
Multiple R	AA6	=SQRT(AA7)
R Square	AA7	=AB14/AB16
Adjusted R Sq	AA8	=1-(1-AA7)*(AA10-1)/(AA10-AA14-1)
Standard Error	AA9	=SQRT(AC15)
Observations	AA10	=COUNT(D4:G11)

**Table 4-39. Excel formulas for Figure 4.52**

Name	Cell	Formula
<i>Regression df</i>	AA14	=AA16-AA15
<i>Residual df</i>	AA15	=S11
<i>Total df</i>	AA16	=S12
<i>Regression SS</i>	AB14	=AB16-AB15
<i>Residual SS</i>	AB15	=T11
<i>Total SS</i>	AB16	=T12
<i>Regression MS</i>	AC14	=AB14/AA14
<i>Residual MS</i>	AC15	=AB15/AA15
<i>F-Stat</i>	AD14	=AC14/AC15
<i>p-value</i>	AE14	=FDIST(AD14,AA14,AA15)
<i>Significant</i>	AF14	=IF(AE14<AE12,"yes","no")
<i>Intercept coef</i>	AA19	=I16
<i>A coef</i>	AA20	{=TRANSPOSE(I17:O17)}
<i>B coef</i>	AA21	{=TRANSPOSE(I17:O17)}
<i>C coef</i>	AA22	{=TRANSPOSE(I17:O17)}
<i>AB coef</i>	AA23	{=TRANSPOSE(I17:O17)}
<i>AC coef</i>	AA24	{=TRANSPOSE(I17:O17)}
<i>BC coef</i>	AA25	{=TRANSPOSE(I17:O17)}
<i>ABC coef</i>	AA26	{=TRANSPOSE(I17:O17)}
<i>Intercept std err</i>	AB19	=AB20
<i>A std err</i>	AB20	=J20/2
<i>B std err</i>	AB21	=AB20
<i>C std err</i>	AB22	=AB21
<i>AB std err</i>	AB23	=AB22
<i>AC std err</i>	AB24	=AB23
<i>BC std err</i>	AB25	=AB24
<i>ABC std err</i>	AB26	=AB25
<i>Intercept t-stat</i>	AC19:AC16	=AA19/AB19
<i>Intercept p-value</i>	AD19:AD16	=TDIST(ABS(AC19),AA\$15,2)
<i>Intercept lower</i>	AE19:AE26	=AC19-TINV(\$AE\$12,AA\$15)*AB19
<i>Intercept upper</i>	AF19:AF26	=AC19+TINV(\$AE\$12,AA\$15)*AB19



## 5. Blocking Scenarios

**Blocking** is a technique for dealing with **nuisance factors**. A **nuisance factor** is a factor that has some effect on the response, but is of no interest to the experimenter; however, the variability it transmits to the response needs to be minimized or explained. We will talk about treatment factors, which we are interested in, and blocking factors, which we are not interested in. We will try to account for these nuisance factors in our model and analysis.

Typical nuisance factors include *batches* of raw material if you are in a production situation, different *operators*, nurses or subjects in studies, the *pieces* of test equipment, when studying a process, and *time* (shifts, days, etc.) where the time of the day or the shift can be a factor that influences the response.

**Many** industrial and human-subject experiments involve blocking, or when they do not, probably should in order to reduce the unexplained variation.

The original use of the term block for removing a source of variation comes from agriculture. Given that you have a plot of land and you want to do an experiment on crops, for instance perhaps testing different varieties or different levels of fertilizer, you would take a section of land and divide it into plots and assigned your treatments at random to these plots. If the section of land contains a large number of plots, they will tend to be very variable - heterogeneous.

A block is characterized by a set of homogeneous plots or a set of similar experimental units. In agriculture a typical block is a set of contiguous plots of land under the assumption that fertility, moisture, weather, will all be similar, and thus the plots are homogeneous.

Failure to block is a common flaw in designing an experiment. If the nuisance variable is **known** and **controllable**, we use **blocking** and control it by including a blocking factor in our experiment.

If we have a nuisance factor that is **known** but **uncontrollable**, sometimes we can use **analysis of covariance** (see Chapter 15) to

measure and remove the effect of the nuisance factor from the analysis. In that case we adjust statistically to account for a covariate, whereas in blocking, we design the experiment with a block factor as an essential component of the design. Which do you think is preferable?

Often, there are nuisance factors that are **unknown** and **uncontrollable** (sometimes called a “lurking” variable). We use **randomization** to balance out their impact. We always randomize so that every experimental unit has an equal chance of being assigned to a given treatment. Randomization is our insurance against a systematic bias due to a nuisance factor.

Sometimes several sources of variation are **combined** to define the block, so the block becomes an aggregate variable. Consider a scenario where we want to test various subjects with different treatments.

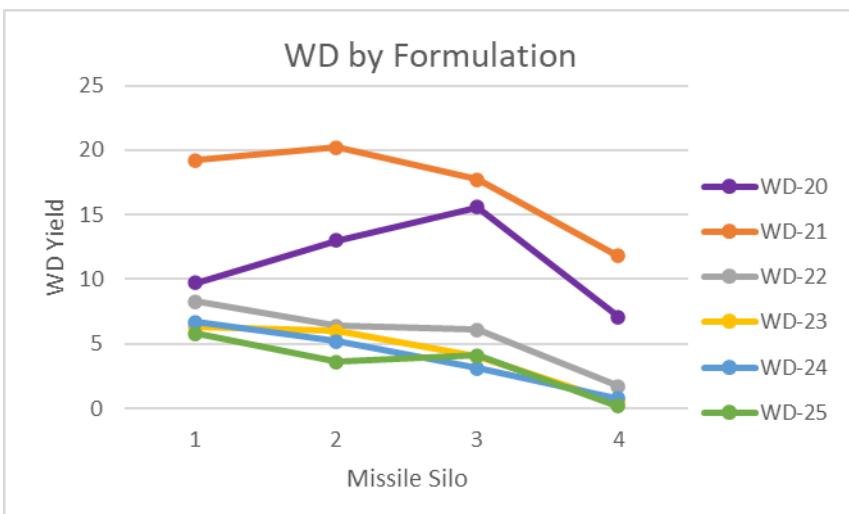
### Example 5.1: WD-40 Revisited

We'll return to our WD-40 example but add a few more formulations and missile silos. The yields (lower is better) are as shown in **Figure 5.1**.

	A	B	C	D	E	F	G	H
2	Randomized complete block design (RCB)							
3								
4	Formula	20	21	22	23	24	25	Mean
5	Silo 1	9.7	19.2	8.3	6.3	6.7	5.8	9.3333
6	Silo 2	13	20.2	6.4	6	5.2	3.6	9.0667
7	Silo 3	15.6	17.7	6.1	4	3.1	4.1	8.4333
8	Silo 4	7.1	11.8	1.7	0.5	0.8	0.2	3.6833
9	Mean	11.35	17.225	5.625	4.2	3.95	3.425	7.6292
10	Variance	13.8567	14.1358	7.7958	7.1267	6.5900	5.5092	33.7691

**Figure 5.1. Yield based on herbicide dosage per field**

We use a randomized complete block design, which can be implemented using Two Factor ANOVA without Replication. A key assumption for this test is that there is no interaction effect. We test this assumption by creating the chart of the yields by field as shown in **Figure 5.2**.



*Figure 5.2. Chart of the yield*

We see that the lines for the four fields are roughly parallel, which indicates that the interaction assumption is reasonable.

We now run the Real Statistics Two Factor ANOVA data analysis tool using the data in Figure 1 as input, selecting the Excel input format and inserting 1 in the Number of Rows per Sample field. The main part of the output is shown in **Figure 5.3**.

	J	K	L	M	N	O	P
2	Two Factor Anova						
3							
4	ANOVA				Alpha	0.05	
5		SS	df	MS	F	p-value	p eta-sq
6	Rows	127.1213	3	42.3738	16.7612	4.69E-05	0.7702
7	Columns	611.6471	5	122.329	48.3882	1E-08	0.9416
8	Error	37.9212	15	2.5281			
9							
10	Total	776.6896	23	33.7691			

*Figure 5.3. RCBD using ANOVA*

The rows correspond to the blocking factor and the columns correspond to the treatments. We are really only interested in the columns factor,

and see that there is a significant difference between the formulas ( $p$ -value = 1E-08).

## Example 5.2: RCBD Follow-up Testing

As we can see from cell O7 of **Figure 5.3** of Randomized Complete Block Design, there is a significant difference in crop yield between the various dosages of herbicide for Example 1 of Randomized Complete Block Design. We can now perform a number of follow-up tests as described in Planned Follow-up Tests and Unplanned Follow-up Tests. In particular, the Real Statistics Resource Pack supports Contrasts and Tukey's HSD.

Now, we'll use Contrasts to determine whether there is a significant difference in yield between 6 WD-25 formulations the RCBD and (the average of) the other formulations. We now fill in range AM5:AR5 with the contrast coefficient values shown in **Figure 5.4**. We see from cell V49 that there is a significant difference between the yield using a WD-25 and the average of the other formulations.

	R	S	T	U	V	W	X	Y
35	Contrast: RCBD Anova							
36								
37	Formula	WD-20	WD-21	WD-22	WD-23	WD-24	WD-25	
38	Contrast	-0.2	1	-0.2	-0.2	-0.2	-0.2	Diff
39	Silo 1	9.7	19.2	8.3	6.3	6.7	5.8	11.84
40	Silo 2	13	20.2	6.4	6	5.2	3.6	13.36
41	Silo 3	15.6	17.7	6.1	4	3.1	4.1	11.12
42	Silo 4	7.1	11.8	1.7	0.5	0.8	0.2	9.74
43						mean	11.515	
44						std dev	1.5073	
45						std err	0.7537	
46								
47	T TEST		Alpha	0.05				
48	mean	std err	t-stat	df	p-value	sig	effect d	effect r
49	11.515	0.7537	15.2785	3	0.00061	yes	7.6392	0.9889

**Figure 5.4. Contrasts for RCBD**

Now, we'll use Tukey's HSD to determine which WD formulas are significantly different as shown in **Figure 5.5**.

	R	S	T	U	V	W	X
2	TUKEY HSD: RCB Design Anova						
3	group	mean	std err	df	q-crit	mean-crit	
4	WD-20	11.35	0.7950	15	4.595	3.6530	
5	WD-21	17.225					
6	WD-22	5.625					
7	WD-23	4.2					
8	WD-24	3.95					
9	WD-25	3.425					
10							
11	Q Test Table Lookup						
12	$\alpha$	0.05					
13	df		15				
14	k		6				
15							
16	Q TEST				alpha	0.05	
17	group 1	group 2	mean	q-stat	lower	upper	Cohen d
18	WD-20	WD-21	5.875	7.3900	5.0800	9.5280	4.2666
19	WD-20	WD-22	5.725	7.2013	4.9300	9.3780	4.1577
20	WD-20	WD-23	7.15	8.9937	6.3550	10.8030	5.1925
21	WD-20	WD-24	7.4	9.3082	6.6050	11.0530	5.3741
22	WD-20	WD-25	7.925	9.9686	7.1300	11.5780	5.7554
23	WD-21	WD-22	11.6	14.5912	10.8050	15.2530	8.4243
24	WD-21	WD-23	13.025	16.3837	12.2300	16.6780	9.4591
25	WD-21	WD-24	13.275	16.6982	12.4800	16.9280	9.6407
26	WD-21	WD-25	13.8	17.3585	13.0050	17.4530	10.0220
27	WD-22	WD-23	1.425	1.7925	0.6300	5.0780	1.0349
28	WD-22	WD-24	1.675	2.1069	0.8800	5.3280	1.2164
29	WD-22	WD-25	2.2	2.7673	1.4050	5.8530	1.5977
30	WD-23	WD-24	0.25	0.3145	-0.5450	3.9030	0.1816
31	WD-23	WD-25	0.775	0.9748	-0.0200	4.4280	0.5628
32	WD-24	WD-25	0.525	0.6604	-0.2700	4.1780	0.3813

Figure 5.5. Tukey HSD RCB Design ANOVA and Q-Test

So, all of the pairings have significantly different means, except WW-22 and WD-23, WD-22 and WD-24, WD-22 and WD-25, WD-23 and WD-24, WD-23 and WD-25, WD-24 and WD-25.

The formulas for calculating our results are supplied in **Table 5-1** through **Table 5-6**.

*Table 5-1. Excel formulas for the RCB Design*

RCB Design		
Function	Cell	Formula
Col Mean copy across	B9	=AVERAGE(B5:B8)
Row Mean copy down	H5	=AVERAGE(B5:G5)
Col Var copy across	B10	=VAR(B5:B8)

*Table 5-2. Excel formulas for the TWO-factor ANOVA*

Two-Factor ANOVA		
Function	Cell	Formula
$SS_{Rows}$	K6	=DEVSQ(H24:H27)*H16
$SS_{Columns}$	K7	=DEVSQ(B28:G28)*B20
$SS_{Error}$	K8	=K10-K6-K7
$SS_{Total}$	K10	=M10*L10
$df_{Rows}$	L6	=COUNT(H24:H27)-1
$df_{Columns}$	L7	=COUNT(B28:G28)-1
$df_{Error}$	L8	=L7*L6
$df_{Total}$	L10	=H20-1
$MS_{Rows}$	M6	=K6/L6
$MS_{Columns}$	M7	=K7/L7
$MS_{Error}$	M8	=K8/L8
$MS_{Total}$	M10	=H36
$F_{Rows}$	N6	=M6/M8
$F_{Columns}$	N7	=M7/M8
p-value Rows	O6	=FDIST(N6,L6,L8)
p-value Columns	O7	=FDIST(N7,L7,L8)
$\eta^2$ Rows	P6	=K6/(K6+K8)
$\eta^2$ Columns	P7	=K7/(K7+K8)

**Table 5-3. Excel formulas for the Tukey HSD RCBD ANOVA**

Tukey HSD: RCBD ANOVA		
Function	Cell	Formula
Formulas (groups)	S4:S9	{=TRANSPOSE(B4:G4)}
WD-20 Mean copy down	S4	=AVERAGE(B5:B8)
Std Error	T4	=SQRT(M8/COUNT(B5:B8))
df	U4	=L8
q-critical	V4	=VLOOKUP(S13,'Stud. Q Table 2'!\$A\$65:\$W\$185,S14,TRUE)
mean-crit	W4	=T4*V4

**Table 5-4. Excel formulas for the Contrast RCBD ANOVA**

Contrast: RCBD ANOVA		
Function	Cell	Formula
Contrast Differences	Y39:Y42	=AVERAGE(Y39:Y42)
Mean of Differences	Y43	=SUMPRODUCT(\$S\$38:\$X\$38,S40:X40)
Std Dev of Differences		=SUMPRODUCT(\$S\$38:\$X\$38,S41:X41)
Std Err of Differences		=SUMPRODUCT(\$S\$38:\$X\$38,S42:X42)

**Table 5-5. Excel Formulas for the T-Test**

T-TEST		
Function	Cell	Formula
Mean	R48	=Y43
Std Err	S48	=Y45
t-stat	T47	=R49/S49
df	U48	=COUNTA(Y39:Y42)-1
p-value	V48	=T.DIST.2T(ABS(T49),U49)
sig	W48	=IF(V49<V47,"yes","no")
effect d	X48	=ABS(Y43)/Y44
effect r	Y48	=SQRT(R49^2/(R49^2+U49))

**Table 5-6. Excel Formulas for the Q-Tests**

Q-TEST		
Function	Cell	Formula
Mean	T18:T32	=ABS(S4-S5)
Q-stat	U18:U32	=T18/T\$4
Lower	V18:V32	=T18-T\$4
Upper	W18:W32	=T18+W\$4
Cohen d	X18:X32	=U18/SQRT(COUNT(B\$4:B\$7))

### Example 5.3: Hardness Testing

In this example we wish to determine whether 4 different tips (the treatment factor) produce different (mean) hardness readings on a Rockwell hardness tester. The treatment factor is the design of the tip for the machine that determines the hardness of metal. The tip is one component of the testing machine.

To conduct this experiment, we assign the tips to an **experimental unit**; that is, to a test specimen (called a coupon), which is a piece of metal on which the tip is tested.

If the structure were a completely randomized experiment (CRD) that we discussed in lesson 3, we would assign the tips to a random piece of metal for each test. In this case, the test specimens would be considered a source of **nuisance variability**. If we conduct this as a blocked experiment, we would assign all four tips to the same test specimen, randomly assigned to be tested on a different location on the specimen. Since each treatment occurs once in each block, the number of test specimens is the number of replicates.

Back to the hardness testing example, the experimenter may very well want to test the tips across specimens of various hardness levels. This shows the importance of blocking. To conduct this experiment as a RCBD, we assign all 4 tips to each specimen.

In this experiment, each specimen is called a “**block**”; thus, we have designed a more homogenous set of experimental units on which to test the tips.

Variability **between** blocks can be large, since we will remove this source of variability, whereas variability **within** a block should be relatively small. In general, a **block** is a specific level of the nuisance factor.

Another way to think about this is that a complete replicate of the basic experiment is conducted in each block. In this case, a block represents an experimental-wide **restriction on randomization**. However, experimental runs **within** a block are **randomized**.

Suppose that we use  $b = 4$  blocks as shown in Table 5-7:

**Table 5-7. Randomized Complete Block Design - Hardness Testing Experiment**

Test Coupon (Block)			
1	2	3	4
Tip 3	Tip 3	Tip 2	Tip 1
Tip 1	Tip 4	Tip 1	Tip 4
Tip 4	Tip 2	Tip 3	Tip 3
Tip 2	Tip 1	Tip 4	Tip 3

Notice the **two-way structure** of the experiment. Here we have four blocks and within each of these blocks is a random assignment of the tips within each block.

We are primarily interested in testing the equality of treatment means, but now we have the ability to remove the variability associated with the nuisance factor (the blocks) through the grouping of the experimental units prior to having assigned the treatments.

## The ANOVA for Randomized Complete Block Design (RCBD)

In the RCBD we have one run of each treatment in each block. In some disciplines, each block is called an experiment (because a copy of the entire experiment is in the block) but in statistics, we call the block to be a replicate. This is a matter of scientific jargon, the design and analysis of the study is an RCBD in both cases.

Suppose that there are  $a$  treatments (factor levels) and  $b$  blocks.

A **statistical model** (effects model) for the RCBD is:

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \end{cases}$$

This is just an extension of the model we had in the one-way case. We have for each observation  $Y_{ij}$  an additive model with an overall mean, plus an effect due to treatment, plus an effect due to block, plus error.

The relevant (fixed effects) hypothesis for the treatment effect is:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_a$$

where

$$\mu_i = -\left(\frac{1}{b}\right) \sum_{j=1}^b (\mu + \tau_i + \beta_j) = \mu + \tau_i \text{ if } \sum_{j=1}^b \beta_j = 0$$

We make the assumption that the errors are independent and normally distributed with constant variance.

The ANOVA is just a partitioning of the variation:

$$\begin{aligned} & \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{..})^2 \\ &= \sum_{i=1}^a \sum_{j=1}^b [(\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) \\ &\quad + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})] \\ &= b \sum_{i=1}^a a (\bar{y}_{i.} - \bar{y}_{..})^2 + a \sum_{j=1}^b b (\bar{y}_{.j} - \bar{y}_{..})^2 \\ &\quad + \sum_{i=1}^a \sum_{j=1}^b (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \end{aligned}$$

$$SS_T = SS_{Treatments} + SS_{Blocks} + SS_E$$

The algebra of the sum of squares falls out in this way. We can partition the effects into three parts: sum of squares due to treatments, sum of squares due to the blocks and the sum of squares due to error.

The degrees of freedom for the sums of squares in:

$$SS_T = SS_{Treatments} + SS_{Blocks} + SS_E$$

are as follows for  $a$  treatments and  $b$  blocks:

$$ab - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1)$$

The partitioning of the variation of the sum of squares and the corresponding partitioning of the degrees of freedom provides the basis for our orthogonal analysis of variance.

## ANOVA Display for the RCBD

**Table 5-8** provides the formulas for the RCBD ANOVA.

*Table 5-8. Analysis of Variance for a Randomized Complete Block Design*

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	$F_0$
Treatments	$SS_{Treatment}$	$a - 1$	$\frac{SS_{Treatment}}{a - 1}$	$\frac{MS_{Treatment}}{MS_g}$
Blocks	$SS_{Blocks}$	$b - 1$	$\frac{SS_{Blocks}}{(b - 1)}$	
Error	$SS_E$	$(a - 1)(b - 1)$	$\frac{SS_E}{(a - 1)(b - 1)}$	
Total	$SS_T$	$N - 1$		

In **Table 5-8** we have the sum of squares due to treatment, the sum of squares due to blocks, and the sum of squares due to error. The degrees of freedom add up to a total of  $N - 1$ , where  $N = ab$ . We obtain the Mean Square values by dividing the sum of squares by the degrees of freedom.

Then, under the null hypothesis of no treatment effect, the ratio of the mean square for treatments to the error mean square is an F statistic that is used to test the hypothesis of equal treatment means.

The text provides manual computing formulas; however, we will use Minitab to analyze the RCBD.

### Example 5.4: Tip Hardness continued

Remember, the hardness of specimens (coupons) is tested with 4 different tips. **NOTE:** Tips are the treatment factor levels, and the coupons are the block levels, composed of homogeneous specimens.

**Table 5-9** shows the data for this experiment ([tip hardness.csv](#)):

*Table 5-9. Tip harness data*

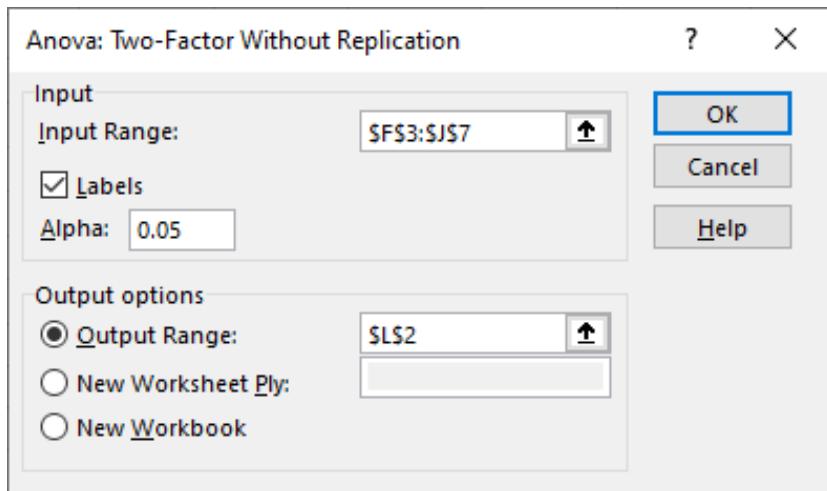
Obs	Tip	Hardness	Coupon
1	1	9.3	1
2	1	9.4	2
3	1	9.6	3
4	1	10.0	4
5	2	9.4	1
6	2	9.3	2
7	2	9.8	3
8	2	9.9	4
9	3	9.2	1
10	3	9.4	2
11	3	9.5	3
12	3	9.7	4
13	4	9.7	1
14	4	9.6	2
15	4	10.0	3
16	4	10.2	4

We now put the data in a block design in the Excel format.

	F	G	H	I	J
2			HARDNESS		
3	TIP	Coup 1	Coup 2	Coup 3	Coup 4
4	Tip 1	9.3	9.4	9.6	10
5	Tip 2	9.4	9.3	9.8	9.9
6	Tip 3	9.2	9.4	9.5	9.7
7	Tip 4	9.7	9.6	10	10.2

**Figure 5.6.** Blocking design in Excel format

We use Excels Data Analysis add-in tool to perform a two-way ANOVA without replication, as shown in **Figure 5.7**. The range of cells we selected for input include the column and row names or Labels, and selected an empty cell to the right for dumping the results.



**Figure 5.7.** Excel's Data Analysis tool dialog for Two-way ANOVA

**Figure 5.8** the output from Data Analysis Tool in Excel. We can see four levels of the Tip and four levels for Coupon:

	L	M	N	O	P	Q	R
2	Anova: Two-Factor Without Replication						
3							
4	SUMMARY	Count	Sum	Average	Variance		
5	Tip 1	4	38.3	9.575	0.09583		
6	Tip 2	4	38.4	9.6	0.08667		
7	Tip 3	4	37.8	9.45	0.04333		
8	Tip 4	4	39.5	9.875	0.07583		
9							
10	Coup 1	4	37.6	9.4	0.04667		
11	Coup 2	4	37.7	9.425	0.01583		
12	Coup 3	4	38.9	9.725	0.04917		
13	Coup 4	4	39.8	9.95	0.04333		
14							
15							
16	ANOVA						
17	Source of Variation	SS	df	MS	F	P-value	F crit
18	Rows	0.385	3	0.12833	14.4375	0.00087	3.8625
19	Columns	0.825	3	0.275	30.9375	4.5E-05	3.8625
20	Error	0.08	9	0.00889			
21							
22	Total	1.29	15				

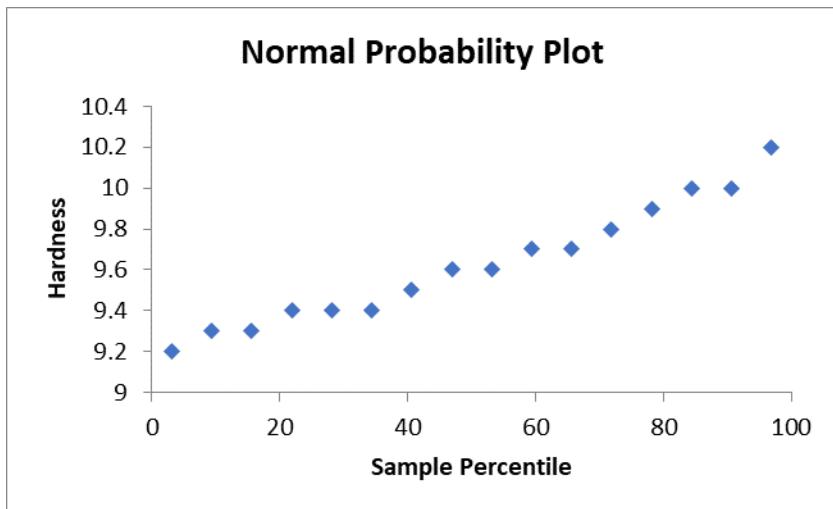
*Figure 5.8. Analysis of Variance for Hardness*

The Analysis of Variance table shows three degrees of freedom for Tip, three for Coupon, and the error degrees of freedom is nine. The ratio of mean squares of treatment over error gives us an *F* ratio that is equal to 14.44 which is highly significant since it is greater than the .001 percentile of the *F* distribution with three and nine degrees of freedom.

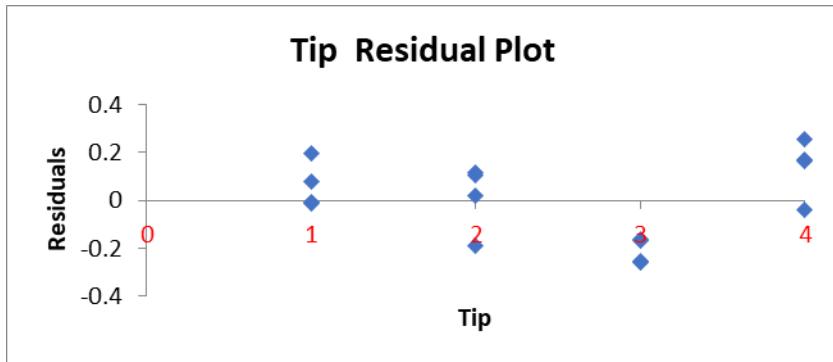
Our 2-way analysis also provides a test for the block factor, Coupon. The ANOVA shows that this factor is also significant with an *F* test = 30.94. So, there is a large amount of variation in hardness between the pieces of metal. This is why we used specimen (or coupon) as our blocking factor. We expected in advance that it would account for a large amount of variation. By including block in the model and in the analysis, we

removed this large portion of the variation, such that the residual error is quite small. By including a block factor in the model, the error variance is reduced, and the test on treatments is more powerful. The test on the block factor is typically not of interest except to confirm that you used a good blocking factor. The results are summarized by the table of means given below.

We confirm that our results are valid using a normal probability plot (see **Figure 5.9**) and plots of the residuals (see **Figure 5.10** and **Figure 5.11**). These can be produced by running a regression for the same data.



*Figure 5.9. Tip Hardness Normal Probability Plot*



*Figure 5.10. Tip Hardness factor residual plot*



**Figure 5.11. Tip Hardness and block residual plot**

### Other Aspects of the RCBD

The RCBD utilizes an **additive model** – one in which there is no interaction between treatments and blocks. *The error term in a randomized complete block model reflects how the treatment effect varies from one block to another.*

Both the treatments and blocks can be looked at as random effects rather than fixed effects, if the levels were selected at random from a population of possible treatments or blocks. We consider this case later, but it does not change the test for a treatment effect.

What are the **consequences of not blocking** if we should have? Generally the unexplained error in the model will be larger, and therefore the test of the treatment effect less powerful.

**How to determine the sample size** in the RCBD? The **OC curve** approach can be used to determine the number of blocks to run. The number of blocks,  $b$ , represents the number of replications. The power calculations that we looked at before would be the same, except that we use  $b$  rather than  $n$ , and we use the estimate of error,  $\sigma^2$ , that reflects the improved precision based on having used blocks in our experiment. So, the major benefit or power comes not from the number of replications but from the error variance which is much smaller because you removed the effects due to block.

# 6. Logistic Regression

Logistic regression, or logit regression, is a type of probabilistic (also called stochastic), statistical model used for classifying categorical responses (Bishop, 2006). It is also used to predict a binary response from a binary predictor, and used for predicting the outcome of a categorical dependent variable (i.e., a class label like “yes” or “no”), based on one or more predictor variables (features). In other words, it is used in estimating the parameters of a qualitative response model. The probabilities describing the possible outcomes of a single trial are modeled, as a function of the explanatory (predictor) variables, using a logistic function (Strickland, 2017). Frequently (and subsequently in this text) “logistic regression” is used to refer specifically to the problem in which the dependent variable is binary—that is, the number of available categories is two. Problems with more than two categories are referred to as *multinomial logistic regression* or, if the multiple categories are ordered, as *ordinal logistic regression*.

Formally, logistic regression measures the relationship between a categorical dependent variable  $Y$  and one or more independent variables  $X$  or  $X_i$ , with  $i = 1, 2, \dots, n$ , which are usually (but not necessarily) continuous, by using *probability scores* as the predicted values of the dependent variable (Bhandari & Joensson, 2008). That is, we extract the probability of  $A$  given the probability of  $X$ , or  $p(A|X)$ , or the probability of  $A$  given the probabilities of the  $X_i$ 's, with  $i = 1, 2, \dots, n$ . These are also called the *posterior probabilities*, as we do not know there distribution beforehand. As such it treats the same set of problems as does *probit regression* using similar techniques.

## Basics

Recall that logistic regression can be *binomial* or *multinomial*. Binomial or *binary logistic regression* deals with situations in which the observed outcome for a dependent variable can have only two possible types (for example, “click-through” vs. “opt-out” in a email campaign). *Multinomial logistic regression* deals with situations where the outcome can have three or more possible types (e.g., “disease A” vs. “disease B” vs. “disease C”). In *binary logistic regression*, the outcome is usually

coded as “0” or “1”, as this leads to the most direct explanation of the outcomes (Hosmer & Lemeshow, 2000). If a particular observed outcome for the dependent variable is the notable possible outcome, (referred to as a “success” or a “case”), like “will purchase” in a product marketing campaign, it is usually coded as “1”. The opposing outcome (referred to as a “failure” or a “noncase”) is coded as “0”, like “will not purchase.” Logistic regression is used to predict the odds of being a *case* based on the values of the independent variables (predictors). The odds are defined as the probability that a particular outcome is a *case* divided by the probability that it is a *noncase*,

$$\frac{p(\text{case})}{p(\text{noncase})} \text{ or } \frac{\text{odds}(\text{success})}{\text{odds}(\text{failure})}.$$

These are referred to as the *odds ratios*. More cases than noncases results in an odds ratio greater than one, and more noncases than cases results in an odds ratio less than one and greater than zero, i.e.

$$0 < \frac{p(\text{case})}{p(\text{noncase})} < 1.$$

Like other forms of regression analysis, logistic regression makes use of one or more predictor variables that may be either continuous or categorical data. Unlike conventional *linear regression*, however, logistic regression is used for predicting binary outcomes of the dependent variable (treating the dependent variable as the outcome of a Bernoulli trial) rather than continuous outcomes. Given this difference, it is necessary that logistic regression take the *natural logarithm* of the odds of the dependent variable being a case (referred to as the *logit* or *log-odds*) to create a continuous criterion as a transformed version of the dependent variable, for example

$$\text{logit} = \log_e \left( \frac{p(\text{case})}{p(\text{noncase})} \right) \quad (6.1)$$

lies between negative and positive infinity, ( $-\infty < \log \text{odds} < +\infty$ ). Hence, the *logit* transformation is referred to as the link function in logistic regression—although the dependent variable in logistic

regression is binomial, the logit is the continuous criterion upon which linear regression is conducted (Hosmer & Lemeshow, 2000).

The logit of success is then fit to the predictors using linear regression analysis. The predicted value of the logit is converted back into predicted odds via the inverse of the natural logarithm, namely the exponential function. Equivalent to Equation (6.1), the logit can be represented as the exponentiation of the log-odds

$$\text{logit} = e^{\beta_0 + \beta_1 x} \quad (4.2a)$$

Or for multiple predictors

$$\text{logit} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n} \quad (4.2b)$$

This exponential relationship provides an interpretation for  $\beta_1$ : the odds multiply by  $e^{\beta_1}$  for every one-unit increase in the predictor variable,

$$\log_e e^{\beta_0 + \beta_1 x} = \beta_0 + \beta_1 x \quad (4.3a)$$

with

$$\log_e \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \quad (4.3b)$$

where  $p = \text{case}$  and  $1 - p = \text{noncase}$  for multiple predictors.

Therefore, although the observed dependent variable in logistic regression is a zero-or-one variable, the logistic regression estimates the odds, as a continuous variable, that the dependent variable is a success (a case). In some applications, the odds are all that is needed. In others, a specific yes-or-no prediction is needed for whether the dependent variable is or is not a case; this categorical prediction can be based on the computed odds of a success, with predicted odds above some chosen cut-off value being translated into a prediction of a success.

### **Logistic function, odds ratio, and logit**

An explanation of logistic regression begins with an explanation of the logistic function, which always takes on values between zero and one (Hosmer & Lemeshow, 2000):

$$F(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}},$$
(4.4)

Observing  $t$  as a linear function of an explanatory variable  $x$  (or of a linear combination of explanatory variables), the logistic function can be written as:

$$F(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}.$$

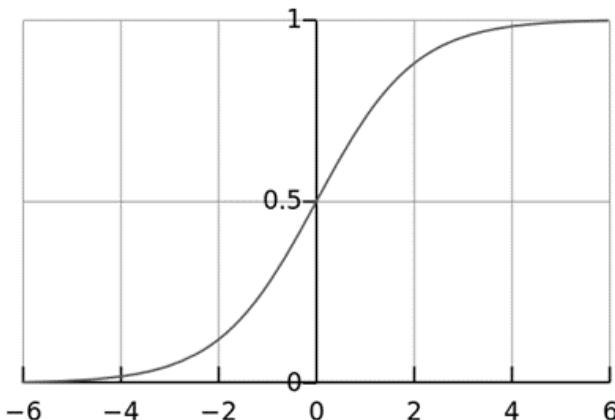
This will be interpreted as the probability of the dependent variable equaling a “success” or “case” rather than a failure or non-case. We also define the inverse of the logistic function, the *logit* (equivalent to (4.1)):

$$g(x) = \ln \frac{F(x)}{1 - F(x)} = \beta_0 + \beta_1 x$$
(4.5)

Equivalently:

$$\frac{F(x)}{1 - F(x)} = e^{(\beta_0 + \beta_1 x)}.$$

A graph of the logistic function is shown in **Figure 6.1**. The input is the value of  $x$  and the output is  $F(x)$ . The logistic function is useful because it can take an input with any value from negative infinity to positive infinity, whereas the output is confined to values between 0 and 1 and hence is interpretable as a probability. In the above equations,  $g(x)$  refers to the logit function of some given linear combination of the predictors,  $\ln$  denotes the natural logarithm,  $F(x)$  is the probability that the dependent variable equals a case,  $\beta_0$  is the intercept from the linear regression equation (the value of the criterion when the predictor is equal to zero),  $\beta_1$  is the regression coefficient multiplied by some value of the predictor  $x$ , and base  $e$  denotes the exponential function.



**Figure 6.1. The logistic function, with  $\beta_0 + \beta_1$  on the horizontal axis and  $F(x)$  on the vertical axis**

The formula for  $F(x)$  illustrates that the probability of the dependent variable equaling a case is equal to the value of the logistic function of the linear regression expression. This is important in that it shows that the value of the linear regression expression can vary from negative to positive infinity and yet, after transformation, the resulting expression for the probability  $F(x)$  ranges between 0 and 1. The equation for  $g(x)$  illustrates that the logit (i.e., log-odds or natural logarithm of the odds) is equivalent to the linear regression expression. Likewise, the next equation illustrates that the odds of the dependent variable equaling a case is equivalent to the exponential function of the linear regression expression. This illustrates how the logit serves as a link function between the probability and the linear regression expression. Given that the logit ranges between negative infinity and positive infinity, it provides an adequate criterion upon which to conduct linear regression and the logit is easily converted back into the odds (Hosmer & Lemeshow, 2000).

## Multiple explanatory variables

If there are multiple explanatory variables, then the above expression (4),  $\beta_0 + \beta_1 x$ , can be revised to  $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m$ . Then when this is used in the equation relating the logged odds of a success to the values of the predictors, the linear regression will be a multiple

regression with  $m$  explanators; the parameters  $\beta_j$  for all  $j = 0, 1, 2, \dots, m$  are all estimated.

Excel's Data Analysis add-in does not perform logistic regression. However, we can use what we know about the logit function to put our data into a format that we can operate on using another Excel add-in, the Solver.

Basically, we want to value that represents the sum of the Log Likelihoods and optimize it by changing the values three parameters derived from the data, using the logit value, the  $e^{\logit}$  function, and the probability function.

The logit (link-) function is given by the regression equation:

$$Y = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_n \cdot X_n$$

where  $b_0, b_1, b_2, \dots, b_n$  are coefficients of the intercept and the three independent variables. Y is the response variable. Now, the second function is  $e^{\logit}$ , which we calculate in Excel using its exponential or EXP() function.

Next, the probability is given by :

$$\frac{1}{1 + e^{\logit}}$$

Finally, we need to define a function for the log likelihood values that takes into account the value of the response, RSO=Yes and RSO>No.

$$\text{Log likelihood} = \begin{cases} \ln(1 - \text{Prob}(Y = 1)), & \text{for } y = 1 \\ \ln \text{Prob}(Y = 0), & \text{for } y = 0 \end{cases}$$

### **Example 6.1: Orbit Classification SW**

We're testing new software that will classify object in LEO orbit as resident space objects (RSOs). The dataset we'll use shows whether or not an object in a LEO orbit is identified as a RSO using 0 = No, 1 = Yes) based on three criteria, labeled Char1, Char2, and Char3, as shown in **Figure 6.2**. In reality, the actual identification is more complicated, but for the sake of space (book space) and easy understanding, we applied simplifying assumptions.

	A	B	C	D
1	RSO	Char1	Char2	Char3
2	0	12	3	6
3	1	13	4	4
4	0	13	4	6
5	1	12	9	9
6	1	14	4	5
7	0	14	4	4
8	0	17	2	2
9	1	17	6	5
10	1	21	5	7
11	0	21	9	3
12	1	24	11	11
13	0	24	4	5

Figure 6.2. Test data for RSO identification software.

So, our logit function is

$$Y = b_0 + b_1 \cdot Char_1 + b_2 \cdot Char_2 + b_3 \cdot Char_3$$

We set the coefficient with initial conditions,  $b_0 = b_1 = b_2 = b_3 = 0.001$

Referring to Figure 6.3, we implement these in Excel as follows:

In Cells A15 to A18, we enter **b0**, **b1**, **b2** and **b3**, respectively, and their corresponding values in Cells B15 to B18.

Then, in cell E2, we enter the logit function as

$$= \$B\$15 + \$B\$16 * B2 + \$B\$17 * C2 + \$B\$18 * D2$$

And copy it down from Cell E2 to Cell E13, with result seen in Figure 6.3.

Next, we enter the formula  $=EXP(E2)$  in Cell F2 and copy it down to Cell F13.

Next, we enter  $=1/(1+F2)$  in Cell G3 and copy it down to Cell G13.

Now, we enter the log likelihood function in Cell H2 as a conditional statement:

=IF(A2=0,LN(1-G2),LN(G2))

And copy it down to Cell G13, as seen in **Figure 6.3**. Finally, we sum the log likelihood values in Cell G14 as =SUM(H2:H13). Now we are ready to solve the logistic regression.

	A	B	C	D	E	F	G	H
1	RSO	Char1	Char2	Char3	logit	e <sup>logit</sup>	Prob	Log Likelihood
2	0	12	3	6	0.0220	1.0222	0.4945	-0.6822
3	1	13	4	4	0.0220	1.0222	0.4945	-0.7042
4	0	13	4	6	0.0240	1.0243	0.4940	-0.6812
5	1	12	9	9	0.0310	1.0315	0.4923	-0.7088
6	1	14	4	5	0.0240	1.0243	0.4940	-0.7052
7	0	14	4	4	0.0230	1.0233	0.4943	-0.6817
8	0	17	2	2	0.0220	1.0222	0.4945	-0.6822
9	1	17	6	5	0.0290	1.0294	0.4928	-0.7078
10	1	21	5	7	0.0340	1.0346	0.4915	-0.7103
11	0	21	9	3	0.0340	1.0346	0.4915	-0.6763
12	1	24	11	11	0.0470	1.0481	0.4883	-0.7169
13	0	24	4	5	0.0340	1.0346	0.4915	-0.6763
14						SUM		-8.3331
15	B0	0.001						
16	B1	0.001						
17	B2	0.001						
18	B3	0.001						

**Figure 6.3. Data with logistic regression model.**

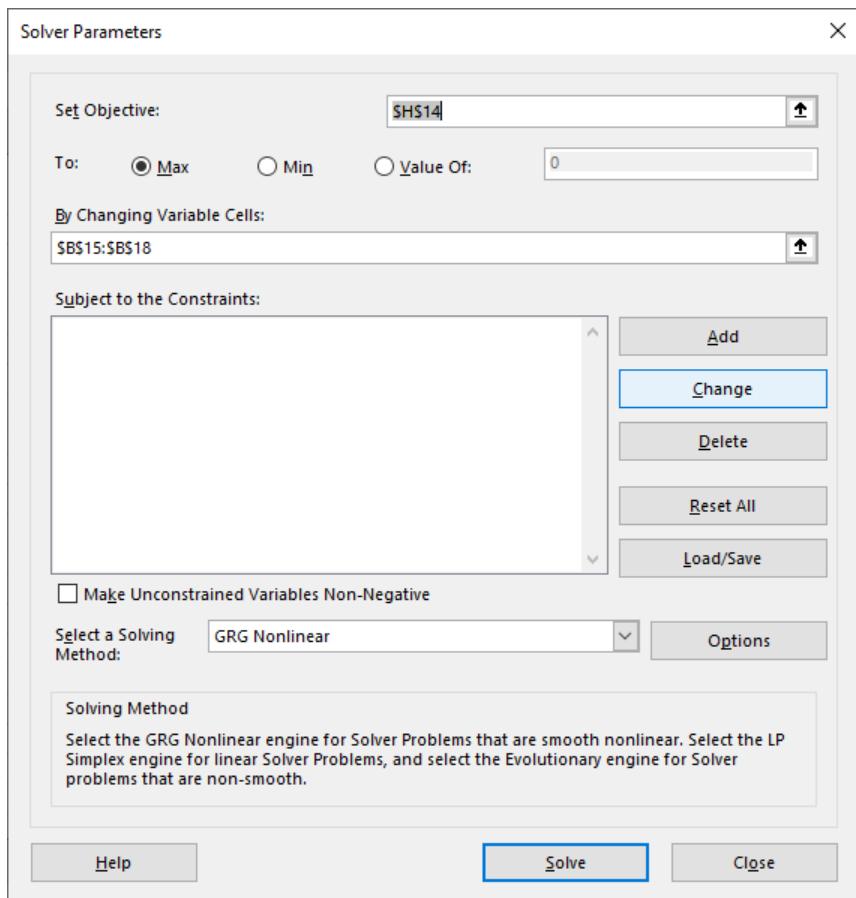
### Installing the Solver

If you haven't already installed the Solver in Excel, use the following steps to do so:

1. Click File.
2. Click Options.
3. Click Add-Ins.
4. Click Solver Add-In, then click Go.
5. In the new window that pops up, check the box next to Solver Add-In, then click Go.

Once the Solver is installed, go to the Analysis group on the Data tab and click Solver. Enter the following information, as shown in **Figure 6.4**:

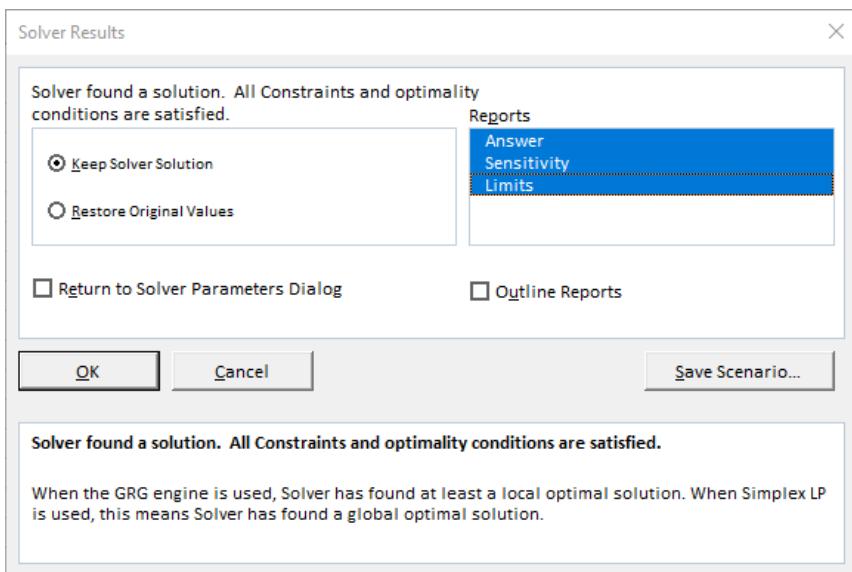
- Set Objective: Choose cell H14 that contains the sum of the log likelihoods.
- By Changing Variable Cells: Choose the cell range B15:B18 that contains the regression coefficients.
- Make Unconstrained Variables Non-Negative: Uncheck this box.
- Select a Solving Method: Choose GRG Nonlinear.



**Figure 6.4. Solver setup window**

Then press the Solve button.

Soon, the Solver Results window will appear as seen in **Figure 6.5**.



**Figure 6.5. Solver results window.**

We will select three reports to view, Answer, Sensitivity, and Limits. We do not have to see the reports( but they may be useful), as the coefficients and objective Log Likelihood sum shows in the spreadsheet as seen in **Figure 6.6**.

Also note that the values in Cells E2 through H13 have change. Since we told the solver to change Cells B15 to B18 to reach a maximum Log Likelihood sum, and they're linked to the logit function, all the cells linked in columns E to H also change. When the Log Likelihood value Cell H14, reaches a local maximum, the algorithm halts and the cells stop changing.

Now, using the coefficient values, our logistic regression function is:

$$Prob(Y = 1) = 1 - \frac{e^{3.6811 + 0.1128X_1 - 0.3957X_2 - 0.6795X_3}}{1 + e^{3.6811 + 0.1128X_1 - 0.3957X_2 - 0.6795X_3}}$$

So, so these calculations would yield the same probabilities as we calculated in column G, Cells G2 to G13. Note that we calculated

$$Prob(Y = 1) = 1 - \frac{e^X}{1 + e^X}$$

Rather than

$$Prob(Y = 0) = \frac{e^X}{1 + e^X}$$

By default, the regression coefficients can be used to find the probability that  $Prob(Y = 0)$ . However, typically in logistic regression we're interested in the probability that the response variable = 1. So, we can either change the signs of the coefficients or subtract  $P(Y = 0)$  from 1, giving us  $P(Y = 1)$ .

	A	B	C	D	E	F	G	H
1	RSO	Char1	Char2	Char3	logit	$e^{\text{logit}}$	Prob	Log Likelihood
2	0	12	3	6	-0.2292	0.7952	0.5570	-0.8143
3	1	13	4	4	0.8470	2.3327	0.3001	-1.2038
4	0	13	4	6	-0.5120	0.5993	0.6253	-0.9816
5	1	12	9	9	-4.6419	0.0096	0.9905	-0.0096
6	1	14	4	5	0.2803	1.3236	0.4304	-0.8431
7	0	14	4	4	0.9599	2.6113	0.2769	-0.3242
8	0	17	2	2	3.4488	31.4616	0.0308	-0.0313
9	1	17	6	5	-0.1726	0.8415	0.5430	-0.6106
10	1	21	5	7	-0.6846	0.5043	0.6648	-0.4083
11	0	21	9	3	0.4508	1.5695	0.3892	-0.4930
12	1	24	11	11	-5.4384	0.0043	0.9957	-0.0043
13	0	24	4	5	1.4086	4.0901	0.1965	-0.2187
14							SUM	-5.9428
15	B0	3.681166						
16	B1	0.112826						
17	B2	-0.39568						
18	B3	-0.67953						

Figure 6.6. Logistic regression with Solver solution shown.

Now, we want to do two things. First, we want to test the function to ensure we get the correct probabilities. We would normally hold out part of our dataset as a test set, but we did not have a lot of data to fit the model. So, we'll use the same data. Second, we run a larger test, get more data and validate our logistic regression model. The set we use for this is the validation set.

## Test the Logistic Regression Function

On the same spreadsheet, copy Cell B1 through Cell D13 (the original data) to Cells J1 through Cell L13 as seen in **Figure 6.7**. Then in Cell M1, enter “P(RSO=1)” and in Cell M2, enter:

=1-

$\text{EXP}(\$B\$15+\$B\$16*\text{J}2+\$B\$17*\text{K}2+\$B\$18*\text{L}2)/(1+\text{EXP}(\$B\$15+\$B\$16*\text{J}2+\$B\$17*\text{K}2+\$B\$18*\text{L}2))$

Copy this formula down to Cell M13, and compare the values in column M with the values in column G.

	G	H	I	J	K	L	M
1	Prob	Likelihood		Char1	Char2	Char3	P(RSO = 1)
2	0.5570	-0.8143		12	3	6	0.5570
3	0.3001	-1.2038		13	4	4	0.3001
4	0.6253	-0.9816		13	4	6	0.6253
5	0.9905	-0.0096		12	9	9	0.9905
6	0.4304	-0.8431		14	4	5	0.4304
7	0.2769	-0.2242		14	4	4	0.2769
8	0.0308	-0.0313		17	2	2	0.0308
9	0.5430	-0.6106		17	6	5	0.5430
10	0.6648	-0.4083		21	5	7	0.6648
11	0.3892	-0.4930		21	9	3	0.3892
12	0.9957	-0.0043		24	11	11	0.9957
13	0.1965	-0.2187		24	4	5	0.1965
14	SUM	-5.9428					

Figure 6.7. Test for the logistic regression model.

## Validate the Logistic Regression Function

To perform this task, we received a data set from a subsequent test, but this time there are 104 objects (more is better). The data was collected and the P(RSO=1) was determined using subject expert analysis. An object was determined to be an RSO if their data gave it 90% probability. In that case the RSO value is 1, otherwise it is 0.

The data is provided as a .CSV file and we imported it to Excel in a separate spreadsheet in the same Excel Workbook we've been using, as shown in **Figure 6.8**. We named the tab “Validation” and entered the data in cells A1 to D105.

	A	B	C	D	E	F	G
1	Actual(RSO)	Char1	Char2	Char3	P(RSO = 1)	PRED(RSO)	DIFF?
2	0	26	8	6	0.65204285	0	1
3	0	17	5	2	0.09434322	0	1
4	0	9	4	3	0.25440301	0	1
5	0	25	4	3	0.05312730	0	1
6	1	4	11	15	0.99996996	1	1
7	1	27	10	10	0.98244498	1	1
8	0	15	3	6	0.47270213	0	1
9	0	30	17	3	0.84543926	1	0
96	0	28	5	3	0.05607963	0	1
97	0	23	5	4	0.17085117	0	1
98	0	20	5	3	0.12778755	0	1
99	0	30	5	4	0.08553720	0	1
100	0	13	3	5	0.36281135	0	1
101	0	10	3	5	0.44405864	0	1
102	0	4	4	3	0.37492679	0	1
103	0	3	2	4	0.37515987	0	1
104	0	21	3	4	0.10476932	0	1
105	0	29	4	3	0.03449703	0	1

*Figure 6.8. Validation for the logistic regression model.*

In Cell E2, we enter the formula from our logistic regression model that calculate posterior probabilities, namely,

=1-

$\text{EXP}(\text{Data}!\$B\$15+\text{Data}!\$B\$16*\text{B2}+\text{Data}!\$B\$17*\text{C2}+\text{Data}!\$B\$18*\text{D2})/(1+\text{EXP}(\text{Data}!\$B\$15+\text{Data}!\$B\$16*\text{B2}+\text{Data}!\$B\$17*\text{C2}+\text{Data}!\$B\$18*\text{D2}))$

Note, our coefficients are in the **Data** tab.

We've been given guidance that our model should be at least 80% accurate or we'll have to rebuild the model with new data. Consequently, we use a conditional statement in Cell F2, that states the object is a RSO if its probability is greater than or equal to 80%. So, in Cell F2 we write, =IF(E2>=0.8,1,0).

We copy this formula down to Cell F105, and a column of 0s and 1s appear. We then compare these with the actual values, which we did using a conditional statement to annotate differences, in Cell G2, namely, =IF(F2=A2,1,0), copied down to Cell G105. So, a zero indicates a "mismatch."

Counting the entries, we have a total of 104, a matching count of 102, and 2 incorrect classifications, as shown in **Figure 6.9**. A simple check of accuracy yields 98.08%.

	I	J	K
1	TOT	104	=COUNT(G2:G105)
2	Correct	102	=SUM(G2:G105)
3	Incorrect	2	=J1-J2
4	Accuracy	0.980769	=J2/J1

**Figure 6.9. Validation results.**

Recall that we had three reports selected from the Solver solution. We will look at one of these, as they are more useful when performing mathematical optimization, like linear programming. Take a moment and view the information the report provides in addition to the coefficient values in **Figure 6.10**.

	A	B	C	D	E	F	G	H
1	Microsoft Excel 16.0 Answer Report							
2	Worksheet: [logistic regression.xlsx]Sheet3							
3	Report Created: 5/26/2022 2:48:30 PM							
4	Result: Solver found a solution. All Constraints and optimality conditions are satisfied.							
5	Solver Engine							
6	Engine: GRG Nonlinear							
7	Solution Time: 0.187 Seconds.							
8	Iterations: 17 Subproblems: 0							
9	Solver Options							
10	Max Time Unlimited, Iterations Unlimited, Precision 0.000001, Use Automatic Scalinh							
11	Convergence 0.0001, Population Size 100, Random Seed 123, Derivatives Central, Re							
12	Max Subproblems Unlimited, Max Integer Sols Unlimited, Integer Tolerance 1%							
13								
14	Objective Cell (Max)							
15	Cell	Name	Original Value	Final Value				
16	\$H\$1	SUM Log Likelihood	-8.3331	-5.9428				
17								
18								
19	Variable Cells							
20	Cell	Name	Original Value	Final Value	Integer			
21	\$B\$1	B0 pts	0.001	3.681166217	Contin			
22	\$B\$1	B1 pts	0.001	0.112825515	Contin			
23	\$B\$1	B2 pts	0.001	-0.395682349	Contin			
24	\$B\$1	B3 pts	0.001	-0.679533625	Contin			
25								
26								
27	Constraints							
	◀	▶	...	Answer Report 1	Sensitivity Report 1	Limits Report 1	Valida	...

Figure 6.10. Answer report for the Solver solution.



# Appendix A: Statistical Tables

## Student t-Distribution

Student t-Distribution						
df/α	α =0.1	α =0.05	α =0.025	α =0.01	α =0.005	α =0.0005
1	3.07768	6.31375	12.70620	31.82052	63.65674	636.61925
2	1.88562	2.91999	4.30265	6.96456	9.92484	31.59905
3	1.63774	2.35336	3.18245	4.54070	5.84091	12.92398
4	1.53321	2.13185	2.77645	3.74695	4.60409	8.61030
5	1.47588	2.01505	2.57058	3.36493	4.03214	6.86883
6	1.43976	1.94318	2.44691	3.14267	3.70743	5.95882
7	1.41492	1.89458	2.36462	2.99795	3.49948	5.40788
8	1.39682	1.85955	2.30600	2.89646	3.35539	5.04131
9	1.38303	1.83311	2.26216	2.82144	3.24984	4.78091
10	1.37218	1.81246	2.22814	2.76377	3.16927	4.58689
11	1.36343	1.79588	2.20099	2.71808	3.10581	4.43698
12	1.35622	1.78229	2.17881	2.68100	3.05454	4.31779
13	1.35017	1.77093	2.16037	2.65031	3.01228	4.22083
14	1.34503	1.76131	2.14479	2.62449	2.97684	4.14045
15	1.34061	1.75305	2.13145	2.60248	2.94671	4.07277
16	1.33676	1.74588	2.11991	2.58349	2.92078	4.01500
17	1.33338	1.73961	2.10982	2.56693	2.89823	3.96513
18	1.33039	1.73406	2.10092	2.55238	2.87844	3.92165
19	1.32773	1.72913	2.09302	2.53948	2.86093	3.88341
20	1.32534	1.72472	2.08596	2.52798	2.84534	3.84952
21	1.32319	1.72074	2.07961	2.51765	2.83136	3.81928
22	1.32124	1.71714	2.07387	2.50832	2.81876	3.79213
23	1.31946	1.71387	2.06866	2.49987	2.80734	3.76763
24	1.31784	1.71088	2.06390	2.49216	2.79694	3.74540
25	1.31635	1.70814	2.05954	2.48511	2.78744	3.72514
26	1.31497	1.70562	2.05553	2.47863	2.77871	3.70661
27	1.31370	1.70329	2.05183	2.47266	2.77068	3.68959
28	1.31253	1.70113	2.04841	2.46714	2.76326	3.67391
29	1.31143	1.69913	2.04523	2.46202	2.75639	3.65941

Student t-Distribution						
df/α	α =0.1	α =0.05	α =0.025	α =0.01	α =0.005	α =0.0005
30	1.31042	1.69726	2.04227	2.45726	2.75000	3.64596
31	1.30946	1.69552	2.03951	2.45282	2.74404	3.63346
32	1.30857	1.69389	2.03693	2.44868	2.73848	3.62180
33	1.30774	1.69236	2.03452	2.44479	2.73328	3.61091
34	1.30695	1.69092	2.03224	2.44115	2.72839	3.60072
35	1.30621	1.68957	2.03011	2.43772	2.72381	3.59115
36	1.30551	1.68830	2.02809	2.43449	2.71948	3.58215
37	1.30485	1.68709	2.02619	2.43145	2.71541	3.57367
38	1.30423	1.68595	2.02439	2.42857	2.71156	3.56568
39	1.30364	1.68488	2.02269	2.42584	2.70791	3.55812
40	1.30308	1.68385	2.02108	2.42326	2.70446	3.55097
41	1.30254	1.68288	2.01954	2.42080	2.70118	3.54418
42	1.30204	1.68195	2.01808	2.41847	2.69807	3.53775
43	1.30155	1.68107	2.01669	2.41625	2.69510	3.53163
44	1.30109	1.68023	2.01537	2.41413	2.69228	3.52580
45	1.30065	1.67943	2.01410	2.41212	2.68959	3.52025
46	1.30023	1.67866	2.01290	2.41019	2.68701	3.51496
47	1.29982	1.67793	2.01174	2.40835	2.68456	3.50990
48	1.29944	1.67722	2.01063	2.40658	2.68220	3.50507
49	1.29907	1.67655	2.00958	2.40489	2.67995	3.50044
50	1.29871	1.67591	2.00856	2.40327	2.67779	3.49601
51	1.29837	1.67528	2.00758	2.40172	2.67572	3.49177
52	1.29805	1.67469	2.00665	2.40022	2.67373	3.48769
53	1.29773	1.67412	2.00575	2.39879	2.67182	3.48378
54	1.29743	1.67356	2.00488	2.39741	2.66998	3.48002
55	1.29713	1.67303	2.00404	2.39608	2.66822	3.47640
56	1.29685	1.67252	2.00324	2.39480	2.66651	3.47292
57	1.29658	1.67203	2.00247	2.39357	2.66487	3.46956
58	1.29632	1.67155	2.00172	2.39238	2.66329	3.46633
59	1.29607	1.67109	2.00100	2.39123	2.66176	3.46321
60	1.29582	1.67065	2.00030	2.39012	2.66028	3.46020
61	1.29558	1.67022	1.99962	2.38905	2.65886	3.45729

Student t-Distribution						
df/α	α =0.1	α =0.05	α =0.025	α =0.01	α =0.005	α =0.0005
62	1.29536	1.66980	1.99897	2.38801	2.65748	3.45448
63	1.29513	1.66940	1.99834	2.38701	2.65615	3.45177
64	1.29492	1.66901	1.99773	2.38604	2.65485	3.44914
65	1.29471	1.66864	1.99714	2.38510	2.65360	3.44660
66	1.29451	1.66827	1.99656	2.38419	2.65239	3.44414
67	1.29432	1.66792	1.99601	2.38330	2.65122	3.44175
68	1.29413	1.66757	1.99547	2.38245	2.65008	3.43944
69	1.29394	1.66724	1.99495	2.38161	2.64898	3.43719
70	1.29376	1.66691	1.99444	2.38081	2.64790	3.43501
71	1.29359	1.66660	1.99394	2.38002	2.64686	3.43290
72	1.29342	1.66629	1.99346	2.37926	2.64585	3.43085
73	1.29326	1.66600	1.99300	2.37852	2.64487	3.42885
74	1.29310	1.66571	1.99254	2.37780	2.64391	3.42692
75	1.29294	1.66543	1.99210	2.37710	2.64298	3.42503
76	1.29279	1.66515	1.99167	2.37642	2.64208	3.42320
77	1.29264	1.66488	1.99125	2.37576	2.64120	3.42141
78	1.29250	1.66462	1.99085	2.37511	2.64034	3.41968
79	1.29236	1.66437	1.99045	2.37448	2.63950	3.41799
80	1.29222	1.66412	1.99006	2.37387	2.63869	3.41634
81	1.29209	1.66388	1.98969	2.37327	2.63790	3.41473
82	1.29196	1.66365	1.98932	2.37269	2.63712	3.41317
83	1.29183	1.66342	1.98896	2.37212	2.63637	3.41164
84	1.29171	1.66320	1.98861	2.37156	2.63563	3.41015
85	1.29159	1.66298	1.98827	2.37102	2.63491	3.40870
86	1.29147	1.66277	1.98793	2.37049	2.63421	3.40728
87	1.29136	1.66256	1.98761	2.36998	2.63353	3.40590
88	1.29125	1.66235	1.98729	2.36947	2.63286	3.40455
89	1.29114	1.66216	1.98698	2.36898	2.63220	3.40323
90	1.29103	1.66196	1.98667	2.36850	2.63157	3.40194
91	1.29092	1.66177	1.98638	2.36803	2.63094	3.40067
92	1.29082	1.66159	1.98609	2.36757	2.63033	3.39944
93	1.29072	1.66140	1.98580	2.36712	2.62973	3.39824

Student t-Distribution						
df/α	α =0.1	α =0.05	α =0.025	α =0.01	α =0.005	α =0.0005
94	1.29062	1.66123	1.98552	2.36667	2.62915	3.39706
95	1.29053	1.66105	1.98525	2.36624	2.62858	3.39590
96	1.29043	1.66088	1.98498	2.36582	2.62802	3.39477
97	1.29034	1.66071	1.98472	2.36541	2.62747	3.39367
98	1.29025	1.66055	1.98447	2.36500	2.62693	3.39259
99	1.29016	1.66039	1.98422	2.36461	2.62641	3.39153
100	1.29007	1.66023	1.98397	2.36422	2.62589	3.39049

## Wilcoxon Rank-Sum Table

1-tail		$\alpha = 0.025$				$\alpha = 0.05$			
2-tail		$\alpha = 0.05$				$\alpha = 0.10$			
m	n	W	d	P	W	d	P		
3	3				6	15	1	0.05	
3	4				6	18	1	0.0286	
3	5	6	21	1	0.0179	7	20	2	0.0357
3	6	7	23	2	0.0238	8	22	3	0.0476
3	7	7	26	2	0.0167	8	25	3	0.0333
3	8	8	28	3	0.0242	9	27	4	0.0424
3	9	8	31	3	0.0182	10	29	5	0.05
3	10	9	33	4	0.0245	10	32	5	0.0385
3	11	9	36	4	0.0192	11	34	6	0.044
3	12	10	38	5	0.0242	11	37	6	0.0352
3	13	10	41	5	0.0196	12	39	7	0.0411
3	14	11	43	6	0.0235	13	41	8	0.0456
3	15	11	46	6	0.0196	13	44	8	0.038
3	16	12	48	7	0.0237	14	46	9	0.0423
3	17	12	51	7	0.0202	15	48	10	0.0465
3	18	13	53	8	0.0233	15	51	10	0.0398
3	19	13	56	8	0.0201	16	53	11	0.0435
3	20	14	58	9	0.0232	17	55	12	0.0469
3	21	14	61	9	0.0203	17	58	12	0.041
3	22	15	63	10	0.023	18	60	13	0.0443
3	23	15	66	10	0.0204	19	62	14	0.0473
3	24	16	68	11	0.0229	19	65	14	0.0421
3	25	16	71	11	0.0205	20	67	15	0.0449
4	4	10	26	1	0.0143	11	25	2	0.0286
4	5	11	29	2	0.0159	12	28	3	0.0317
4	6	12	32	3	0.019	13	31	4	0.0333
4	7	13	35	4	0.0212	14	34	5	0.0364
4	8	14	38	5	0.0242	15	37	6	0.0364
4	9	14	42	5	0.0168	16	40	7	0.0378
4	10	15	45	6	0.018	17	43	8	0.038

1-tail		$\alpha = 0.025$				$\alpha = 0.05$			
2-tail		$\alpha = 0.05$				$\alpha = 0.10$			
m	n	W	d	P	W	d	P		
4	11	16	48	7	0.0198	18	46	9	0.0388
4	12	17	51	8	0.0209	19	49	10	0.039
4	13	18	54	9	0.0223	20	52	11	0.0395
4	14	19	57	10	0.0232	21	55	12	0.0395
4	15	20	60	11	0.0243	22	58	13	0.04
4	16	21	63	12	0.025	24	60	15	0.0497
4	17	21	67	12	0.0202	25	63	16	0.0493
4	18	22	70	13	0.0212	26	66	17	0.0491
4	19	23	73	14	0.0219	27	69	18	0.0487
4	20	24	76	15	0.0227	28	72	19	0.0485
4	21	25	79	16	0.0233	29	75	20	0.0481
4	22	26	82	17	0.024	30	78	21	0.048
4	23	27	85	18	0.0246	31	81	22	0.0477
4	24	27	89	18	0.0211	32	84	23	0.0475
4	25	28	92	19	0.0217	33	87	24	0.0473
5	5	17	38	3	0.0159	19	36	5	0.0476
5	6	18	42	4	0.0152	20	40	6	0.0411
5	7	20	45	6	0.024	21	44	7	0.0366
5	8	21	49	7	0.0225	23	47	9	0.0466
5	9	22	53	8	0.021	24	51	10	0.0415
5	10	23	57	9	0.02	26	54	12	0.0496
5	11	24	61	10	0.019	27	58	13	0.0449
5	12	26	64	12	0.0242	28	62	14	0.0409
5	13	27	68	13	0.023	30	65	16	0.0473
5	14	28	72	14	0.0218	31	69	17	0.0435
5	15	29	76	15	0.0209	33	72	19	0.0491
5	16	30	80	16	0.0201	34	76	20	0.0455
5	17	32	83	18	0.0238	35	80	21	0.0425
5	18	33	87	19	0.0229	37	83	23	0.0472
5	19	34	91	20	0.022	38	87	24	0.0442
5	20	35	95	21	0.0212	40	90	26	0.0485

1-tail		$\alpha = 0.025$				$\alpha = 0.05$			
2-tail		$\alpha = 0.05$				$\alpha = 0.10$			
m	n	W	d	P	W	d	P		
5	21	37	98	23	0.0243	41	94	27	0.0457
5	22	38	102	24	0.0234	43	97	29	0.0496
5	23	39	106	25	0.0226	44	101	30	0.0469
5	24	40	110	26	0.0219	45	105	31	0.0445
5	25	42	113	28	0.0246	47	108	33	0.048
6	6	26	52	6	0.0206	28	50	8	0.0465
6	7	27	57	7	0.0175	29	55	9	0.0367
6	8	29	61	9	0.0213	31	59	11	0.0406
6	9	31	65	11	0.0248	33	63	13	0.044
6	10	32	70	12	0.021	35	67	15	0.0467
6	11	34	74	14	0.0238	37	71	17	0.0491
6	12	35	79	15	0.0207	38	76	18	0.0415
6	13	37	83	17	0.0231	40	80	20	0.0437
6	14	38	88	18	0.0204	42	84	22	0.0457
6	15	40	92	20	0.0224	44	88	24	0.0474
6	16	42	96	22	0.0244	46	92	26	0.049
6	17	43	101	23	0.0219	47	97	27	0.0433
6	18	45	105	25	0.0236	49	101	29	0.0448
6	19	46	110	26	0.0214	51	105	31	0.0462
6	20	48	114	28	0.0229	53	109	33	0.0475
6	21	50	118	30	0.0244	55	113	35	0.0487
6	22	51	123	31	0.0224	57	117	37	0.0498
6	23	53	127	33	0.0237	58	122	38	0.0452
6	24	54	132	34	0.0219	60	126	40	0.0463
6	25	56	136	36	0.0231	62	130	42	0.0473
7	7	36	69	9	0.0189	39	66	12	0.0487
7	8	38	74	11	0.02	41	71	14	0.0469
7	9	40	79	13	0.0209	43	76	16	0.0454
7	10	42	84	15	0.0215	45	81	18	0.0439
7	11	44	89	17	0.0221	47	86	20	0.0427
7	12	46	94	19	0.0225	49	91	22	0.0416

1-tail		$\alpha = 0.025$				$\alpha = 0.05$			
2-tail		$\alpha = 0.05$				$\alpha = 0.10$			
m	n	W	d	P	W	d	P		
7	13	48	99	21	0.0228	52	95	25	0.0484
7	14	50	104	23	0.023	54	100	27	0.0469
7	15	52	109	25	0.0233	56	105	29	0.0455
7	16	54	114	27	0.0234	58	110	31	0.0443
7	17	56	119	29	0.0236	61	114	34	0.0497
7	18	58	124	31	0.0237	63	119	36	0.0484
7	19	60	129	33	0.0238	65	124	38	0.0471
7	20	62	134	35	0.0239	67	129	40	0.046
7	21	64	139	37	0.024	69	134	42	0.0449
7	22	66	144	39	0.024	72	138	45	0.0492
7	23	68	149	41	0.0241	74	143	47	0.0481
7	24	70	154	43	0.0241	76	148	49	0.047
7	25	72	159	45	0.0242	78	153	51	0.0461
8	8	49	87	14	0.0249	51	85	16	0.0415
8	9	51	93	16	0.0232	54	90	19	0.0464
8	10	53	99	18	0.0217	56	96	21	0.0416
8	11	55	105	20	0.0204	59	101	24	0.0454
8	12	58	110	23	0.0237	62	106	27	0.0489
8	13	60	116	25	0.0223	64	112	29	0.0445
8	14	62	122	27	0.0211	67	117	32	0.0475
8	15	65	127	30	0.0237	69	123	34	0.0437
8	16	67	133	32	0.0224	72	128	37	0.0463
8	17	70	138	35	0.0247	75	133	40	0.0487
8	18	72	144	37	0.0235	77	139	42	0.0452
8	19	74	150	39	0.0224	80	144	45	0.0475
8	20	77	155	42	0.0244	83	149	48	0.0495
8	21	79	161	44	0.0233	85	155	50	0.0464
8	22	81	167	46	0.0223	88	160	53	0.0483
8	23	84	172	49	0.024	90	166	55	0.0454
8	24	86	178	51	0.0231	93	171	58	0.0472
8	25	89	183	54	0.0247	96	176	61	0.0488

1-tail		$\alpha = 0.025$				$\alpha = 0.05$			
2-tail		$\alpha = 0.05$				$\alpha = 0.10$			
m	n	W	d	P	W	d	P		
9	9	62	109	18	0.02	66	105	22	0.047
9	10	65	115	21	0.0217	69	111	25	0.0474
9	11	68	121	24	0.0232	72	117	28	0.0476
9	12	71	127	27	0.0245	75	123	31	0.0477
9	13	73	134	29	0.0217	78	129	34	0.0478
9	14	76	140	32	0.0228	81	135	37	0.0478
9	15	79	146	35	0.0238	84	141	40	0.0478
9	16	82	152	38	0.0247	87	147	43	0.0477
9	17	84	159	40	0.0223	90	153	46	0.0476
9	18	87	165	43	0.0231	93	159	49	0.0475
9	19	90	171	46	0.0239	96	165	52	0.0474
9	20	93	177	49	0.0245	99	171	55	0.0473
9	21	95	184	51	0.0225	102	177	58	0.0472
9	22	98	190	54	0.0231	105	183	61	0.0471
9	23	101	196	57	0.0237	108	189	64	0.047
9	24	104	202	60	0.0243	111	195	67	0.0469
9	25	107	208	63	0.0249	114	201	70	0.0468
10	10	78	132	24	0.0216	82	128	28	0.0446
10	11	81	139	27	0.0215	86	134	32	0.0493
10	12	84	146	30	0.0213	89	141	35	0.0465
10	13	88	152	34	0.0247	92	148	38	0.0441
10	14	91	159	37	0.0242	96	154	42	0.0478
10	15	94	166	40	0.0238	99	161	45	0.0455
10	16	97	173	43	0.0234	103	167	49	0.0487
10	17	100	180	46	0.023	106	174	52	0.0465
10	18	103	187	49	0.0226	110	180	56	0.0493
10	19	107	193	53	0.025	113	187	59	0.0472
10	20	110	200	56	0.0245	117	193	62	0.0498
10	21	113	207	59	0.0241	120	200	65	0.0478
10	22	116	214	62	0.0237	123	207	68	0.0459
10	23	119	221	65	0.0233	127	213	72	0.0482

1-tail		$\alpha = 0.025$				$\alpha = 0.05$			
2-tail		$\alpha = 0.05$				$\alpha = 0.10$			
m	n	W	d	P	W	d	P		
10	24	122	228	68	0.023	130	220	75	0.0465
10	25	126	234	72	0.0248	134	226	79	0.0486
11	11	96	157	31	0.0237	100	153	34	0.044
11	12	99	165	34	0.0219	104	160	38	0.0454
11	13	103	172	38	0.0237	108	167	42	0.0467
11	14	106	180	41	0.0221	112	174	46	0.0477
11	15	110	187	45	0.0236	116	181	50	0.0486
11	16	113	195	48	0.0221	120	188	54	0.0494
11	17	117	202	52	0.0235	123	196	57	0.0453
11	18	121	209	56	0.0247	127	203	61	0.0461
11	19	124	217	59	0.0233	131	210	65	0.0468
11	20	128	224	63	0.0244	135	217	69	0.0474
11	21	131	232	66	0.023	139	224	73	0.048
11	22	135	239	70	0.024	143	231	77	0.0486
11	23	139	246	74	0.025	147	238	81	0.049
11	24	142	254	77	0.0237	151	245	85	0.0495
11	25	146	261	81	0.0246	155	252	89	0.0499
12	12	115	185	38	0.0225	120	180	42	0.0444
12	13	119	193	42	0.0229	125	187	47	0.0488
12	14	123	201	46	0.0232	129	195	51	0.0475
12	15	127	209	50	0.0234	133	203	55	0.0463
12	16	131	217	54	0.0236	138	210	60	0.05
12	17	135	225	58	0.0238	142	218	64	0.0486
12	18	139	233	62	0.0239	146	226	68	0.0474
12	19	143	241	66	0.024	150	234	72	0.0463
12	20	147	249	70	0.0241	155	241	77	0.0493
12	21	151	257	74	0.0242	159	249	81	0.0481
12	22	155	265	78	0.0242	163	257	85	0.0471
12	23	159	273	82	0.0243	168	264	90	0.0496
12	24	163	281	86	0.0243	172	272	94	0.0486
12	25	167	289	90	0.0243	176	280	98	0.0475

1-tail		$\alpha = 0.025$				$\alpha = 0.05$			
2-tail		$\alpha = 0.05$				$\alpha = 0.10$			
m	n	W	d	P	W	d	P		
13	13	136	215	46	0.0221	142	209	51	0.0454
13	14	141	223	51	0.0241	147	217	56	0.0472
13	15	145	232	55	0.0232	152	225	61	0.0489
13	16	150	240	60	0.025	156	234	65	0.0458
13	17	154	249	64	0.024	161	242	70	0.0472
13	18	158	258	68	0.0232	166	250	75	0.0485
13	19	163	266	73	0.0247	171	258	80	0.0497
13	20	167	275	77	0.0238	175	267	84	0.047
13	21	171	284	81	0.0231	180	275	89	0.0481
13	22	176	292	86	0.0243	185	283	94	0.0491
13	23	180	301	90	0.0236	189	292	98	0.0467
13	24	185	309	95	0.0247	194	300	103	0.0476
13	25	189	318	99	0.024	199	308	108	0.0485
14	14	160	246	56	0.0249	166	240	61	0.0469
14	15	164	256	60	0.0229	171	249	66	0.0466
14	16	169	265	65	0.0236	176	258	72	0.0463
14	17	174	274	70	0.0242	182	266	78	0.05
14	18	179	283	75	0.0247	187	275	83	0.0495
14	19	183	293	79	0.023	192	284	88	0.0489
14	20	188	302	84	0.0235	197	293	93	0.0484
14	21	193	311	89	0.0239	202	302	98	0.048
14	22	198	320	94	0.0243	207	311	103	0.0475
14	23	203	329	99	0.0247	212	320	108	0.0471
14	24	207	339	103	0.0233	218	328	114	0.0498
14	25	212	348	108	0.0236	223	337	119	0.0492
15	15	184	281	65	0.0227	192	273	73	0.0488
15	16	190	290	71	0.0247	197	283	78	0.0466
15	17	195	300	76	0.0243	203	292	84	0.0485
15	18	200	310	81	0.0239	208	302	89	0.0465
15	19	205	320	86	0.0235	214	311	95	0.0482
15	20	210	330	91	0.0232	220	320	101	0.0497

1-tail		$\alpha = 0.025$				$\alpha = 0.05$			
2-tail		$\alpha = 0.05$				$\alpha = 0.10$			
m	n	W	d	P	W	d	P		
15	21	216	339	97	0.0247	225	330	106	0.0478
15	22	221	349	102	0.0243	231	339	112	0.0492
15	23	226	359	107	0.0239	236	349	117	0.0474
15	24	231	369	112	0.0235	242	358	123	0.0486
15	25	237	378	118	0.0248	248	367	129	0.0499
16	16	211	317	76	0.0234	219	309	84	0.0469
16	17	217	327	82	0.0243	225	319	90	0.0471
16	18	222	338	87	0.0231	231	329	96	0.0473
16	19	228	348	93	0.0239	237	339	102	0.0474
16	20	234	358	99	0.0247	243	349	108	0.0475
16	21	239	369	104	0.0235	249	359	114	0.0475
16	22	245	379	110	0.0242	255	369	120	0.0476
16	23	251	389	116	0.0248	261	379	126	0.0476
16	24	256	400	121	0.0238	267	389	132	0.0476
16	25	262	410	127	0.0243	273	399	138	0.0476
17	17	240	355	88	0.0243	249	346	97	0.0493
17	18	246	366	94	0.0243	255	357	103	0.0479
17	19	252	377	100	0.0243	262	367	110	0.0499
17	20	258	388	106	0.0242	268	378	116	0.0485
17	21	264	399	112	0.0242	274	389	122	0.0473
17	22	270	410	118	0.0241	281	399	129	0.049
17	23	276	421	124	0.024	287	410	135	0.0477
17	24	282	432	130	0.0239	294	420	141	0.0492
17	25	288	443	136	0.0238	300	431	147	0.048
18	18	270	396	100	0.0235	280	386	109	0.0485
18	19	277	407	107	0.0246	287	397	116	0.049
18	20	283	419	113	0.0238	294	408	123	0.0495
18	21	290	430	120	0.0247	301	419	130	0.0499
18	22	296	442	126	0.024	307	431	136	0.0474
18	23	303	453	133	0.0248	314	442	143	0.0478
18	24	309	465	139	0.024	321	453	150	0.0481

1-tail		$\alpha = 0.025$				$\alpha = 0.05$			
2-tail		$\alpha = 0.05$				$\alpha = 0.10$			
m	n	W	d	P	W	d	P		
18	25	316	476	146	0.0248	328	464	157	0.0484
19	19	303	438	114	0.0248	313	428	123	0.0482
19	20	309	451	120	0.0234	320	440	130	0.0474
19	21	316	463	127	0.0236	328	451	138	0.0494
19	22	323	475	134	0.0238	335	463	145	0.0486
19	23	330	487	141	0.024	342	475	152	0.0478
19	24	337	499	148	0.0241	350	486	160	0.0496
19	25	344	511	155	0.0243	357	498	167	0.0488
20	20	337	483	128	0.0245	348	472	138	0.0482
20	21	344	496	135	0.0241	356	484	146	0.049
20	22	351	509	142	0.0236	364	496	154	0.0497
20	23	359	521	150	0.0246	371	509	161	0.0478
20	24	366	534	157	0.0242	379	521	169	0.0484
20	25	373	547	164	0.0237	387	533	177	0.049
21	21	373	530	143	0.0245	385	518	154	0.0486
21	22	381	543	151	0.0249	393	531	162	0.0482
21	23	388	557	158	0.0238	401	544	170	0.0478
21	24	396	570	166	0.0242	410	556	179	0.0497
21	25	404	583	174	0.0245	418	569	187	0.0492
22	22	411	579	159	0.0247	424	566	171	0.0491
22	23	419	593	167	0.0244	432	580	179	0.0477
22	24	427	607	175	0.0242	441	593	188	0.0486
22	25	435	621	183	0.024	450	606	197	0.0494
23	23	451	630	176	0.0249	465	616	189	0.0499
23	24	459	645	184	0.0242	474	630	198	0.0497
23	25	468	659	193	0.0246	483	644	207	0.0495
24	24	492	684	193	0.0241	507	669	207	0.0486
24	25	501	699	202	0.0241	517	683	217	0.0496
25	25	536	739	212	0.0247	552	723	227	0.0497

## Student Q Table

Alpha	0.05						
df	k=2	k=3	k=4	k=5	k=6	k=7	k=8
1	17.97	26.98	32.82	37.08	40.41	43.12	45.40
2	6.085	8.331	9.798	10.88	11.74	12.44	13.03
3	4.501	5.910	6.825	7.502	8.037	8.478	8.853
4	3.927	5.040	5.757	6.287	6.707	7.053	7.347
5	3.635	4.602	5.218	5.673	6.033	6.330	6.582
6	3.461	4.339	4.896	5.305	5.628	5.895	6.122
7	3.344	4.165	4.681	5.060	5.359	5.606	5.815
8	3.261	4.041	4.529	4.886	5.167	5.399	5.597
9	3.199	3.949	4.415	4.756	5.024	5.244	5.432
10	3.151	3.877	4.327	4.654	4.912	5.124	5.305
11	3.113	3.820	4.256	4.574	4.823	5.028	5.202
12	3.082	3.773	4.199	4.508	4.751	4.950	5.119
13	3.055	3.735	4.151	4.453	4.690	4.885	5.049
14	3.033	3.702	4.111	4.407	4.639	4.829	4.990
15	3.014	3.674	4.076	4.367	4.595	4.782	4.940
16	2.998	3.649	4.046	4.333	4.557	4.741	4.897
17	2.984	3.628	4.020	4.303	4.524	4.705	4.858
18	2.971	3.609	3.997	4.277	4.495	4.673	4.824
19	2.960	3.593	3.977	4.253	4.469	4.645	4.794
20	2.950	3.578	3.958	4.232	4.445	4.620	4.768
21	2.941	3.565	3.942	4.213	4.424	4.597	4.744
22	2.933	3.553	3.927	4.196	4.406	4.577	4.722
23	2.926	3.542	3.914	4.181	4.388	4.558	4.702
24	2.919	3.532	3.901	4.166	4.373	4.541	4.684
25	2.9126	3.5226	3.89	4.1534	4.3583	4.5258	4.6672
26	2.907	3.5142	3.8796	4.1415	4.3451	4.5115	4.6519
27	2.9017	3.5064	3.8701	4.1305	4.3329	4.4983	4.6378
28	2.8969	3.4992	3.8612	4.1203	4.3217	4.4861	4.6248
29	2.8924	3.4926	3.853	4.1109	4.3112	4.4747	4.6127
30	2.8882	3.4865	3.8454	4.1021	4.3015	4.4642	4.6014
31	2.8843	3.4806	3.8383	4.0939	4.2924	4.4543	4.5909

Alpha	0.05							
	df	k=2	k=3	k=4	k=5	k=6	k=7	k=8
32	2.8807	3.4752	3.8316	4.0862	4.2839	4.4451	4.5811	
33	2.8773	3.4702	3.8254	4.079	4.2759	4.4365	4.5718	
34	2.874	3.4654	3.8195	4.0723	4.2684	4.4284	4.5632	
35	2.871	3.461	3.814	4.066	4.2614	4.4207	4.555	
36	2.8682	3.4568	3.8088	4.06	4.2548	4.4136	4.5473	
37	2.8655	3.4528	3.8039	4.0543	4.2485	4.4068	4.5401	
38	2.863	3.449	3.7993	4.049	4.2426	4.4003	4.5332	
39	2.8605	3.4455	3.7949	4.0439	4.237	4.3942	4.5267	
40	2.8583	3.4421	3.7907	4.0391	4.2317	4.3885	4.5205	
41	2.8561	3.4389	3.7867	4.0346	4.2266	4.383	4.5147	
42	2.854	3.4358	3.783	4.0302	4.2218	4.3778	4.5091	
43	2.8521	3.4329	3.7794	4.0261	4.2172	4.3728	4.5038	
44	2.8502	3.4302	3.776	4.0222	4.2128	4.3681	4.4987	
45	2.8484	3.4275	3.7727	4.0184	4.2087	4.3635	4.4939	
46	2.8467	3.425	3.7696	4.0148	4.2047	4.3592	4.4893	
47	2.8451	3.4226	3.7666	4.0114	4.2009	4.3551	4.4849	
48	2.8435	3.4203	3.7637	4.0081	4.1972	4.3511	4.4806	
49	2.842	3.418	3.761	4.005	4.1938	4.3474	4.4766	
50	2.8406	3.4159	3.7584	4.002	4.1904	4.3437	4.4727	
51	2.8392	3.4139	3.7559	3.9991	4.1872	4.3402	4.469	
52	2.8379	3.4119	3.7535	3.9963	4.1841	4.3369	4.4654	
53	2.8366	3.41	3.751	3.9936	4.1811	4.3337	4.4619	
54	2.8354	3.4082	3.7489	3.991	4.1783	4.3305	4.4586	
55	2.8342	3.4065	3.7468	3.9885	4.1755	4.3276	4.4554	
56	2.8331	3.4048	3.7448	3.9862	4.1729	4.3247	4.4524	
57	2.832	3.4032	3.7427	3.9839	4.1703	4.3219	4.4494	
58	2.8309	3.4016	3.7408	3.9816	4.1678	4.3192	4.4465	
59	2.8299	3.4001	3.7389	3.9795	4.1655	4.3166	4.4438	
60	2.8289	3.3987	3.7371	3.9774	4.1632	4.3142	4.4411	
61	2.828	3.3973	3.7354	3.9754	4.1609	4.3117	4.4385	
62	2.8271	3.3959	3.7337	3.9735	4.1588	4.3094	4.436	
63	2.8262	3.3946	3.732	3.9716	4.1567	4.3071	4.4336	

Alpha	0.05							
	df	k=2	k=3	k=4	k=5	k=6	k=7	k=8
64	2.8253	3.3933	3.7305	3.9698	4.1547	4.305	4.4313	
65	2.8245	3.3921	3.7289	3.968	4.1527	4.3028	4.429	
66	2.8237	3.3909	3.7275	3.9663	4.1509	4.3008	4.4268	
67	2.8229	3.3897	3.726	3.9647	4.149	4.2988	4.4247	
68	2.8221	3.3886	3.7246	3.9631	4.1472	4.2969	4.4226	
69	2.8214	3.3875	3.7233	3.9615	4.1455	4.295	4.4206	
70	2.8207	3.3864	3.722	3.96	4.1438	4.2932	4.4186	
71	2.82	3.3854	3.7207	3.9585	4.1422	4.2914	4.4168	
72	2.8193	3.3844	3.7195	3.9571	4.1406	4.2897	4.4149	
73	2.8186	3.3834	3.7183	3.9557	4.1391	4.288	4.4131	
74	2.818	3.3825	3.7171	3.9544	4.1376	4.2864	4.4114	
75	2.8174	3.3816	3.716	3.9531	4.1361	4.2848	4.4097	
76	2.8167	3.3807	3.7149	3.9518	4.1347	4.2833	4.4081	
77	2.8161	3.3798	3.7138	3.9506	4.1333	4.2818	4.4064	
78	2.8155	3.3789	3.7127	3.9494	4.132	4.2803	4.4049	
79	2.8149	3.3781	3.7117	3.9482	4.1307	4.2789	4.4034	
80	2.8144	3.3773	3.7107	3.947	4.1294	4.2775	4.4019	
81	2.8138	3.3765	3.7097	3.9459	4.1282	4.2761	4.4004	
82	2.8133	3.3758	3.7088	3.9448	4.127	4.2748	4.399	
83	2.8128	3.375	3.7079	3.9437	4.1258	4.2735	4.3977	
84	2.8123	3.3743	3.707	3.9427	4.1246	4.2723	4.3963	
85	2.8118	3.3736	3.7061	3.9417	4.1235	4.2711	4.395	
86	2.8114	3.3729	3.7052	3.9407	4.1224	4.2699	4.3937	
87	2.8109	3.3722	3.7044	3.9397	4.1213	4.2687	4.3925	
88	2.8105	3.3715	3.7036	3.9388	4.1203	4.2676	4.3913	
89	2.81	3.3709	3.7028	3.9378	4.1192	4.2664	4.3901	
90	2.8096	3.3702	3.702	3.9369	4.1182	4.2654	4.3889	
91	2.8092	3.3696	3.7012	3.9361	4.1173	4.2643	4.3878	
92	2.8088	3.369	3.7005	3.9352	4.1163	4.2633	4.3866	
93	2.8083	3.3684	3.6997	3.9344	4.1154	4.2622	4.3855	
94	2.808	3.3678	3.699	3.9336	4.1144	4.2612	4.3845	
95	2.8076	3.3673	3.6983	3.9327	4.1135	4.2603	4.3834	

Alpha	0.05							
	df	k=2	k=3	k=4	k=5	k=6	k=7	k=8
96	2.8072	3.3667	3.6976	3.9319	4.1127	4.2593	4.3824	
97	2.8068	3.3662	3.6969	3.9312	4.1118	4.2584	4.3814	
98	2.8065	3.3656	3.6963	3.9304	4.111	4.2574	4.3804	
99	2.8061	3.3651	3.6956	3.9297	4.1101	4.2566	4.3795	
100	2.8058	3.3646	3.695	3.9289	4.1093	4.2557	4.3785	
101	2.8054	3.3641	3.6944	3.9282	4.1085	4.2548	4.3776	
102	2.8051	3.3636	3.6938	3.9275	4.1077	4.254	4.3767	
103	2.8048	3.3631	3.6932	3.9268	4.107	4.2531	4.3758	
104	2.8044	3.3627	3.6926	3.9262	4.1062	4.2523	4.3749	
105	2.8041	3.3622	3.692	3.9255	4.1055	4.2515	4.3741	
106	2.8038	3.3617	3.6915	3.9249	4.1048	4.2507	4.3733	
107	2.8035	3.3613	3.6909	3.9242	4.1041	4.25	4.3724	
108	2.8032	3.3609	3.6904	3.9236	4.1034	4.2492	4.3716	
109	2.8029	3.3604	3.6898	3.923	4.1027	4.2485	4.3708	
110	2.8026	3.36	3.6893	3.9224	4.102	4.2478	4.3701	
111	2.8024	3.3596	3.6888	3.9218	4.1014	4.2471	4.3693	
112	2.8021	3.3592	3.6883	3.9212	4.1007	4.2464	4.3686	
113	2.8018	3.3588	3.6878	3.9206	4.1001	4.2457	4.3678	
114	2.8015	3.3584	3.6873	3.9201	4.0995	4.245	4.3671	
115	2.8013	3.358	3.6869	3.9195	4.0989	4.2443	4.3664	
116	2.801	3.3576	3.6864	3.919	4.0983	4.2437	4.3657	
117	2.8008	3.3573	3.6859	3.9185	4.0977	4.243	4.365	
118	2.8005	3.3569	3.6855	3.918	4.0971	4.2424	4.3643	
119	2.8003	3.3565	3.685	3.9174	4.0965	4.2418	4.3637	
120	2.8	3.3562	3.6846	3.9169	4.096	4.2412	4.363	

Alpha	0.05						
df	k=9	k=10	k=11	k=12	k=13	k=14	k=15
1	47.36	49.07	50.59	51.96	53.20	54.33	55.36
2	13.54	13.99	14.39	14.75	15.08	15.38	15.65
3	9.177	9.462	9.717	9.946	10.15	10.35	10.53
4	7.602	7.826	8.027	8.208	8.373	8.525	8.664
5	6.802	6.995	7.168	7.324	7.466	7.596	7.717
6	6.319	6.493	6.649	6.789	6.917	7.034	7.143
7	5.998	6.158	6.302	6.431	6.550	6.658	6.759
8	5.767	5.918	6.054	6.175	6.287	6.389	6.483
9	5.595	5.739	5.867	5.983	6.089	6.186	6.276
10	5.461	5.599	5.722	5.833	5.935	6.028	6.114
11	5.353	5.487	5.605	5.713	5.811	5.901	5.984
12	5.265	5.395	5.511	5.615	5.710	5.798	5.878
13	5.192	5.318	5.431	5.533	5.625	5.711	5.789
14	5.131	5.254	5.364	5.463	5.554	5.637	5.714
15	5.077	5.198	5.306	5.404	5.493	5.574	5.649
16	5.031	5.150	5.256	5.352	5.439	5.520	5.593
17	4.991	5.108	5.212	5.307	5.392	5.471	5.544
18	4.956	5.071	5.174	5.267	5.352	5.429	5.501
19	4.924	5.038	5.140	5.231	5.315	5.391	5.462
20	4.895	5.008	5.1083	5.199	5.2815	5.3573	5.4273
21	4.870	4.981	5.0806	5.1703	5.252	5.3269	5.3961
22	4.847	4.957	5.0556	5.1443	5.2252	5.2993	5.3678
23	4.826	4.935	5.0328	5.1207	5.2008	5.2743	5.3421
24	4.807	4.915	5.0119	5.0991	5.1785	5.2514	5.3186
25	4.7894	4.8969	4.9928	5.0793	5.1581	5.2303	5.297
26	4.7733	4.88	4.9753	5.0611	5.1393	5.211	5.2772
27	4.7584	4.8645	4.959	5.0443	5.122	5.1931	5.2589
28	4.7446	4.85	4.944	5.0287	5.1059	5.1766	5.2419
29	4.7319	4.8366	4.93	5.0143	5.0909	5.1612	5.2261
30	4.7199	4.8241	4.917	5.0008	5.077	5.1469	5.2114
31	4.7088	4.8125	4.9049	4.9882	5.064	5.1335	5.1977
32	4.6984	4.8016	4.8936	4.9765	5.0519	5.121	5.1848

Alpha	0.05						
df	k=9	k=10	k=11	k=12	k=13	k=14	k=15
33	4.6887	4.7914	4.8829	4.9654	5.0405	5.1093	5.1728
34	4.6795	4.7818	4.8729	4.955	5.0298	5.0982	5.1614
35	4.6709	4.7728	4.8635	4.9453	5.0197	5.0879	5.1508
36	4.6628	4.7642	4.8546	4.9361	5.0102	5.0781	5.1407
37	4.6551	4.7562	4.8462	4.9274	5.0012	5.0688	5.1312
38	4.6479	4.7486	4.8383	4.9191	4.9927	5.06	5.1222
39	4.641	4.7414	4.8308	4.9113	4.9846	5.0517	5.1137
40	4.6345	4.7345	4.8237	4.9039	4.977	5.0439	5.1056
41	4.6283	4.728	4.8169	4.8969	4.9697	5.0364	5.0979
42	4.6224	4.7218	4.8104	4.8902	4.9628	5.0293	5.0906
43	4.6168	4.716	4.8043	4.8839	4.9562	5.0225	5.0836
44	4.6114	4.7103	4.7984	4.8778	4.9499	5.016	5.077
45	4.6063	4.705	4.7928	4.872	4.9439	5.0098	5.0706
46	4.6014	4.6998	4.7875	4.8664	4.9382	5.0039	5.0646
47	4.5967	4.6949	4.7824	4.8611	4.9327	4.9983	5.0587
48	4.5923	4.6903	4.7775	4.856	4.9275	4.9929	5.0532
49	4.588	4.6858	4.7728	4.8512	4.9224	4.9877	5.0479
50	4.5839	4.6814	4.7683	4.8465	4.9176	4.9827	5.0427
51	4.5799	4.6773	4.764	4.842	4.913	4.9779	5.0378
52	4.5761	4.6733	4.7598	4.8377	4.9085	4.9733	5.0331
53	4.5725	4.6695	4.7558	4.8336	4.9042	4.9689	5.0286
54	4.569	4.6658	4.752	4.8296	4.9001	4.9647	5.0242
55	4.5656	4.6623	4.7483	4.8257	4.8961	4.9606	5.02
56	4.5623	4.6589	4.7447	4.822	4.8923	4.9566	5.016
57	4.5592	4.6556	4.7413	4.8185	4.8886	4.9528	5.0121
58	4.5562	4.6524	4.738	4.815	4.885	4.9492	5.0083
59	4.5533	4.6493	4.7348	4.8117	4.8816	4.9456	5.0046
60	4.5504	4.6463	4.7317	4.8085	4.8783	4.9422	5.0011
61	4.5477	4.6435	4.7287	4.8054	4.8751	4.9389	4.9977
62	4.5451	4.6407	4.7258	4.8024	4.872	4.9357	4.9944
63	4.5425	4.638	4.723	4.7995	4.869	4.9326	4.9912
64	4.54	4.6354	4.7203	4.7967	4.8661	4.9296	4.9881

Alpha	0.05	df	k=9	k=10	k=11	k=12	k=13	k=14	k=15
65		4.5376	4.6329	4.7177	4.7939	4.8632	4.9267	4.9852	
66		4.5353	4.6305	4.7151	4.7913	4.8605	4.9239	4.9823	
67		4.533	4.6281	4.7126	4.7887	4.8578	4.9211	4.9794	
68		4.5309	4.6258	4.7102	4.7862	4.8553	4.9185	4.9767	
69		4.5287	4.6236	4.7079	4.7838	4.8528	4.9159	4.9741	
70		4.5267	4.6214	4.7057	4.7815	4.8503	4.9134	4.9715	
71		4.5247	4.6193	4.7035	4.7792	4.848	4.911	4.969	
72		4.5227	4.6173	4.7013	4.777	4.8457	4.9086	4.9666	
73		4.5208	4.6153	4.6993	4.7748	4.8435	4.9063	4.9642	
74		4.519	4.6133	4.6972	4.7727	4.8413	4.9041	4.9619	
75		4.5172	4.6115	4.6953	4.7707	4.8392	4.9019	4.9597	
76		4.5155	4.6096	4.6934	4.7687	4.8371	4.8998	4.9575	
77		4.5138	4.6078	4.6915	4.7668	4.8351	4.8977	4.9554	
78		4.5121	4.6061	4.6897	4.7649	4.8332	4.8957	4.9533	
79		4.5105	4.6044	4.6879	4.7631	4.8313	4.8938	4.9513	
80		4.5089	4.6028	4.6862	4.7613	4.8295	4.8919	4.9494	
81		4.5074	4.6012	4.6845	4.7595	4.8277	4.89	4.9474	
82		4.5059	4.5996	4.6829	4.7578	4.8259	4.8882	4.9456	
83		4.5045	4.5981	4.6813	4.7562	4.8242	4.8864	4.9438	
84		4.503	4.5966	4.6798	4.7546	4.8225	4.8847	4.942	
85		4.5016	4.5951	4.6782	4.753	4.8209	4.883	4.9403	
86		4.5003	4.5937	4.6767	4.7515	4.8193	4.8814	4.9386	
87		4.499	4.5923	4.6753	4.7499	4.8177	4.8798	4.9369	
88		4.4977	4.591	4.6739	4.7485	4.8162	4.8782	4.9353	
89		4.4964	4.5896	4.6725	4.747	4.8147	4.8767	4.9337	
90		4.4952	4.5883	4.6711	4.7456	4.8133	4.8752	4.9322	
91		4.494	4.5871	4.6698	4.7443	4.8118	4.8737	4.9307	
92		4.4928	4.5858	4.6685	4.7429	4.8104	4.8722	4.9292	
93		4.4916	4.5846	4.6673	4.7416	4.8091	4.8708	4.9277	
94		4.4905	4.5834	4.666	4.7403	4.8077	4.8695	4.9263	
95		4.4894	4.5823	4.6648	4.739	4.8064	4.8681	4.925	
96		4.4883	4.5811	4.6636	4.7378	4.8052	4.8668	4.9236	

Alpha	0.05						
df	k=9	k=10	k=11	k=12	k=13	k=14	k=15
97	4.4873	4.58	4.6625	4.7366	4.8039	4.8655	4.9223
98	4.4862	4.5789	4.6613	4.7354	4.8027	4.8643	4.921
99	4.4852	4.5778	4.6602	4.7343	4.8015	4.863	4.9197
100	4.4842	4.5768	4.6591	4.7331	4.8003	4.8618	4.9185
101	4.4832	4.5758	4.658	4.732	4.7992	4.8606	4.9172
102	4.4823	4.5748	4.657	4.7309	4.798	4.8595	4.916
103	4.4813	4.5738	4.656	4.7299	4.7969	4.8583	4.9149
104	4.4804	4.5728	4.6549	4.7288	4.7958	4.8572	4.9137
105	4.4795	4.5719	4.654	4.7278	4.7948	4.8561	4.9126
106	4.4786	4.5709	4.653	4.7268	4.7937	4.855	4.9115
107	4.4778	4.57	4.652	4.7258	4.7927	4.854	4.9104
108	4.4769	4.5691	4.6511	4.7248	4.7917	4.8529	4.9093
109	4.4761	4.5682	4.6502	4.7238	4.7907	4.8519	4.9083
110	4.4752	4.5674	4.6493	4.7229	4.7898	4.8509	4.9072
111	4.4744	4.5665	4.6484	4.722	4.7888	4.8499	4.9062
112	4.4736	4.5657	4.6475	4.7211	4.7879	4.849	4.9052
113	4.4729	4.5649	4.6467	4.7202	4.7869	4.848	4.9043
114	4.4721	4.5641	4.6458	4.7193	4.786	4.8471	4.9033
115	4.4713	4.5633	4.645	4.7185	4.7852	4.8462	4.9024
116	4.4706	4.5625	4.6442	4.7176	4.7843	4.8453	4.9015
117	4.4699	4.5618	4.6434	4.7168	4.7834	4.8444	4.9005
118	4.4692	4.561	4.6426	4.716	4.7826	4.8435	4.8997
119	4.4685	4.5603	4.6419	4.7152	4.7818	4.8427	4.8988
120	4.4678	4.5596	4.6411	4.7144	4.781	4.8418	4.8979

Alpha	0.05				
df	k=16	k=18	k=20	k=24	k=28
1	56.32	58.04	59.56	62.12	64.23
2	15.91	16.37	16.77	17.45	18.02
3	10.69	10.98	11.24	11.68	12.05
4	8.794	9.028	9.233	9.584	9.875
5	7.828	8.030	8.208	8.512	8.764
6	7.244	7.426	7.587	7.861	8.088
7	6.852	7.020	7.170	7.423	7.634
8	6.571	6.729	6.870	7.109	7.307
9	6.359	6.510	6.644	6.871	7.061
10	6.194	6.339	6.467	6.686	6.868
11	6.062	6.202	6.326	6.536	6.712
12	5.953	6.089	6.209	6.414	6.585
13	5.862	5.995	6.112	6.312	6.478
14	5.786	5.915	6.029	6.224	6.387
15	5.720	5.846	5.958	6.149	6.309
16	5.662	5.786	5.897	6.084	6.241
17	5.612	5.734	5.842	6.027	6.181
18	5.568	5.688	5.794	5.977	6.128
19	5.528	5.647	5.752	5.932	6.081
20	5.4923	5.5529	5.6097	5.6632	5.7136
21	5.4603	5.5203	5.5765	5.6293	5.6792
22	5.4314	5.4908	5.5464	5.5987	5.648
23	5.4051	5.4639	5.5189	5.5707	5.6196
24	5.381	5.4393	5.4939	5.5452	5.5936
25	5.359	5.4167	5.4709	5.5218	5.5698
26	5.3387	5.396	5.4497	5.5002	5.5478
27	5.3199	5.3768	5.4301	5.4802	5.5275
28	5.3025	5.359	5.412	5.4618	5.5087
29	5.2863	5.3425	5.3951	5.4446	5.4913
30	5.2713	5.3271	5.3794	5.4286	5.475
31	5.2572	5.3127	5.3647	5.4136	5.4597
32	5.244	5.2993	5.351	5.3996	5.4455

Alpha	0.05				
df	k=16	k=18	k=20	k=24	k=28
33	5.2317	5.2866	5.3381	5.3865	5.4321
34	5.2201	5.2748	5.326	5.3741	5.4196
35	5.2091	5.2636	5.3146	5.3625	5.4077
36	5.1988	5.253	5.3038	5.3515	5.3965
37	5.1891	5.2431	5.2936	5.3412	5.386
38	5.1799	5.2337	5.284	5.3313	5.376
39	5.1711	5.2247	5.2749	5.322	5.3665
40	5.1629	5.2162	5.2662	5.3132	5.3575
41	5.155	5.2082	5.258	5.3048	5.349
42	5.1475	5.2005	5.2502	5.2968	5.3408
43	5.1403	5.1932	5.2427	5.2892	5.3331
44	5.1335	5.1862	5.2356	5.282	5.3257
45	5.127	5.1796	5.2288	5.275	5.3186
46	5.1208	5.1732	5.2223	5.2684	5.3119
47	5.1148	5.1671	5.2161	5.2621	5.3054
48	5.1091	5.1613	5.2101	5.256	5.2992
49	5.1037	5.1557	5.2044	5.2502	5.2933
50	5.0984	5.1503	5.1989	5.2446	5.2876
51	5.0934	5.1452	5.1937	5.2392	5.2822
52	5.0886	5.1402	5.1886	5.234	5.2769
53	5.0839	5.1355	5.1837	5.2291	5.2718
54	5.0794	5.1309	5.1791	5.2243	5.267
55	5.0751	5.1265	5.1745	5.2197	5.2623
56	5.071	5.1222	5.1702	5.2153	5.2578
57	5.067	5.1181	5.166	5.211	5.2534
58	5.0631	5.1142	5.162	5.2069	5.2492
59	5.0594	5.1103	5.1581	5.2029	5.2452
60	5.0557	5.1066	5.1543	5.199	5.2412
61	5.0523	5.1031	5.1506	5.1953	5.2374
62	5.0489	5.0996	5.1471	5.1917	5.2338
63	5.0456	5.0963	5.1437	5.1882	5.2302
64	5.0424	5.093	5.1404	5.1848	5.2268

Alpha	0.05				
df	k=16	k=18	k=20	k=24	k=28
65	5.0394	5.0899	5.1372	5.1816	5.2234
66	5.0364	5.0868	5.1341	5.1784	5.2202
67	5.0335	5.0839	5.131	5.1753	5.2171
68	5.0307	5.081	5.1281	5.1723	5.214
69	5.028	5.0783	5.1253	5.1694	5.2111
70	5.0254	5.0756	5.1225	5.1666	5.2082
71	5.0228	5.0729	5.1198	5.1639	5.2054
72	5.0203	5.0704	5.1172	5.1612	5.2027
73	5.0179	5.0679	5.1147	5.1586	5.2001
74	5.0155	5.0655	5.1122	5.1561	5.1975
75	5.0132	5.0631	5.1098	5.1537	5.195
76	5.011	5.0609	5.1075	5.1513	5.1926
77	5.0088	5.0586	5.1052	5.149	5.1902
78	5.0067	5.0565	5.103	5.1467	5.1879
79	5.0047	5.0544	5.1008	5.1445	5.1857
80	5.0027	5.0523	5.0987	5.1424	5.1835
81	5.0007	5.0503	5.0967	5.1403	5.1813
82	4.9988	5.0483	5.0947	5.1382	5.1793
83	4.9969	5.0464	5.0927	5.1362	5.1772
84	4.9951	5.0446	5.0908	5.1343	5.1753
85	4.9933	5.0427	5.089	5.1324	5.1733
86	4.9916	5.041	5.0872	5.1305	5.1714
87	4.9899	5.0392	5.0854	5.1287	5.1696
88	4.9882	5.0375	5.0836	5.127	5.1678
89	4.9866	5.0359	5.0819	5.1252	5.166
90	4.985	5.0342	5.0803	5.1235	5.1643
91	4.9835	5.0327	5.0787	5.1219	5.1626
92	4.982	5.0311	5.0771	5.1203	5.161
93	4.9805	5.0296	5.0755	5.1187	5.1593
94	4.979	5.0281	5.074	5.1171	5.1578
95	4.9776	5.0267	5.0725	5.1156	5.1562
96	4.9762	5.0252	5.0711	5.1141	5.1547

Alpha	0.05				
df	k=16	k=18	k=20	k=24	k=28
97	4.9749	5.0238	5.0697	5.1127	5.1532
98	4.9735	5.0225	5.0683	5.1113	5.1518
99	4.9722	5.0211	5.0669	5.1099	5.1503
100	4.9709	5.0198	5.0656	5.1085	5.149
101	4.9697	5.0185	5.0642	5.1072	5.1476
102	4.9685	5.0173	5.063	5.1058	5.1462
103	4.9673	5.0161	5.0617	5.1045	5.1449
104	4.9661	5.0148	5.0605	5.1033	5.1436
105	4.9649	5.0137	5.0592	5.102	5.1424
106	4.9638	5.0125	5.0581	5.1008	5.1411
107	4.9627	5.0113	5.0569	5.0996	5.1399
108	4.9616	5.0102	5.0557	5.0985	5.1387
109	4.9605	5.0091	5.0546	5.0973	5.1376
110	4.9594	5.008	5.0535	5.0962	5.1364
111	4.9584	5.007	5.0524	5.0951	5.1353
112	4.9574	5.0059	5.0514	5.094	5.1342
113	4.9564	5.0049	5.0503	5.0929	5.1331
114	4.9554	5.0039	5.0493	5.0919	5.132
115	4.9544	5.0029	5.0483	5.0909	5.131
116	4.9535	5.002	5.0473	5.0898	5.1299
117	4.9526	5.001	5.0463	5.0888	5.1289
118	4.9516	5.0001	5.0453	5.0879	5.1279
119	4.9507	4.9991	5.0444	5.0869	5.1269
120	4.9499	4.9982	5.0435	5.086	5.126

## Dunnet's Critical Values

alpha= 0.01

df\k	2	3	4	5	6	7	8	9	10
2	9.925	12.388	13.826	14.825	15.582	16.189	16.692	17.121	17.494
3	5.841	6.974	7.639	8.104	8.460	8.746	8.985	9.189	9.367
4	4.604	5.364	5.809	6.121	6.361	6.554	6.716	6.854	6.975
5	4.032	4.627	4.975	5.219	5.406	5.557	5.683	5.792	5.887
6	3.707	4.212	4.506	4.711	4.869	4.997	5.104	5.196	5.276
7	3.499	3.948	4.208	4.389	4.529	4.642	4.736	4.817	4.888
8	3.355	3.766	4.002	4.168	4.295	4.397	4.483	4.557	4.621
9	3.250	3.633	3.853	4.006	4.124	4.219	4.299	4.367	4.427
10	3.169	3.531	3.739	3.883	3.994	4.084	4.159	4.223	4.279
11	3.106	3.452	3.649	3.787	3.892	3.978	4.049	4.110	4.164
12	3.055	3.387	3.577	3.709	3.811	3.892	3.960	4.019	4.070
13	3.012	3.335	3.518	3.646	3.743	3.822	3.888	3.944	3.994
14	2.977	3.290	3.468	3.592	3.687	3.763	3.827	3.882	3.930
15	2.947	3.253	3.426	3.547	3.639	3.713	3.776	3.829	3.875
16	2.921	3.220	3.390	3.508	3.598	3.671	3.731	3.783	3.829
17	2.898	3.192	3.359	3.474	3.563	3.634	3.693	3.744	3.788
18	2.878	3.168	3.331	3.445	3.531	3.601	3.659	3.709	3.753
19	2.861	3.146	3.307	3.419	3.504	3.572	3.630	3.679	3.722
20	2.845	3.127	3.285	3.395	3.479	3.547	3.603	3.651	3.694
21	2.831	3.109	3.266	3.375	3.457	3.524	3.579	3.627	3.669
22	2.819	3.094	3.249	3.356	3.437	3.503	3.558	3.605	3.646
23	2.807	3.080	3.233	3.339	3.420	3.485	3.539	3.585	3.626
24	2.797	3.067	3.218	3.323	3.403	3.468	3.521	3.567	3.608
25	2.787	3.055	3.205	3.309	3.388	3.452	3.505	3.551	3.591
26	2.779	3.044	3.193	3.296	3.375	3.438	3.491	3.536	3.575
27	2.771	3.034	3.182	3.284	3.362	3.425	3.477	3.522	3.561
28	2.763	3.025	3.172	3.273	3.351	3.413	3.465	3.509	3.548
29	2.756	3.017	3.162	3.263	3.340	3.402	3.453	3.497	3.536
30	2.750	3.009	3.154	3.254	3.330	3.391	3.442	3.486	3.524
31	2.744	3.001	3.145	3.245	3.321	3.381	3.432	3.476	3.514

32	2.738	2.994	3.138	3.237	3.312	3.372	3.423	3.466	3.504
33	2.733	2.988	3.131	3.229	3.304	3.364	3.414	3.457	3.495
34	2.728	2.982	3.124	3.222	3.296	3.356	3.406	3.449	3.486
35	2.724	2.976	3.118	3.215	3.289	3.349	3.398	3.441	3.478
36	2.719	2.971	3.112	3.209	3.282	3.342	3.391	3.433	3.470
37	2.715	2.966	3.106	3.203	3.276	3.335	3.384	3.426	3.463
38	2.712	2.961	3.101	3.197	3.270	3.329	3.378	3.420	3.456
39	2.708	2.957	3.096	3.192	3.264	3.323	3.372	3.413	3.450
40	2.704	2.952	3.091	3.186	3.259	3.317	3.366	3.408	3.444
48	2.682	2.925	3.060	3.154	3.224	3.281	3.329	3.369	3.405
60	2.660	2.898	3.030	3.121	3.190	3.246	3.292	3.332	3.366
80	2.639	2.871	3.001	3.090	3.157	3.211	3.256	3.295	3.328
120	2.617	2.845	2.972	3.059	3.124	3.177	3.221	3.259	3.291
240	2.596	2.820	2.943	3.028	3.092	3.143	3.186	3.223	3.255
Inf	2.576	2.794	2.915	2.998	3.060	3.110	3.152	3.188	3.219

**alpha= 0.01**

df\k	11	12	13	14	15	16	17	18	19	20
2	17.823	18.117	18.383	18.624	18.846	19.051	19.241	19.418	19.583	19.739
3	9.525	9.666	9.794	9.910	10.017	10.116	10.208	10.294	10.375	10.450
4	7.082	7.179	7.266	7.346	7.419	7.487	7.550	7.609	7.664	7.716
5	5.971	6.047	6.116	6.178	6.236	6.289	6.339	6.386	6.429	6.470
6	5.347	5.411	5.469	5.523	5.572	5.617	5.659	5.699	5.736	5.770
7	4.951	5.008	5.059	5.106	5.150	5.190	5.227	5.262	5.295	5.326
8	4.679	4.730	4.777	4.820	4.859	4.896	4.930	4.961	4.991	5.019
9	4.480	4.528	4.571	4.611	4.647	4.681	4.713	4.742	4.770	4.796
10	4.329	4.374	4.415	4.452	4.487	4.519	4.548	4.576	4.602	4.627
11	4.211	4.254	4.293	4.328	4.361	4.391	4.419	4.446	4.470	4.494
12	4.116	4.157	4.194	4.228	4.259	4.288	4.315	4.340	4.364	4.387
13	4.038	4.077	4.113	4.146	4.176	4.204	4.230	4.254	4.277	4.299
14	3.972	4.011	4.045	4.077	4.106	4.133	4.158	4.182	4.204	4.225
15	3.917	3.954	3.988	4.019	4.047	4.073	4.098	4.121	4.142	4.163
16	3.869	3.905	3.938	3.968	3.996	4.022	4.046	4.068	4.089	4.109
17	3.828	3.863	3.896	3.925	3.952	3.977	4.001	4.023	4.043	4.062
18	3.792	3.826	3.858	3.887	3.914	3.938	3.961	3.983	4.003	4.022

19	3.760	3.794	3.825	3.853	3.879	3.904	3.926	3.947	3.967	3.986
20	3.731	3.765	3.795	3.823	3.849	3.873	3.895	3.916	3.935	3.954
21	3.706	3.739	3.769	3.797	3.822	3.845	3.867	3.888	3.907	3.925
22	3.683	3.716	3.745	3.773	3.798	3.821	3.842	3.862	3.881	3.899
23	3.662	3.694	3.724	3.751	3.775	3.798	3.820	3.840	3.858	3.876
24	3.643	3.675	3.704	3.731	3.755	3.778	3.799	3.819	3.837	3.855
25	3.626	3.658	3.686	3.713	3.737	3.759	3.780	3.800	3.818	3.835
26	3.610	3.642	3.670	3.696	3.720	3.742	3.763	3.782	3.800	3.817
27	3.596	3.627	3.655	3.681	3.705	3.727	3.747	3.766	3.784	3.801
28	3.582	3.613	3.641	3.667	3.690	3.712	3.732	3.751	3.769	3.786
29	3.570	3.600	3.628	3.654	3.677	3.699	3.719	3.738	3.755	3.772
30	3.558	3.589	3.616	3.641	3.665	3.686	3.706	3.725	3.743	3.759
31	3.547	3.578	3.605	3.630	3.653	3.675	3.695	3.713	3.731	3.747
32	3.537	3.567	3.595	3.620	3.643	3.664	3.684	3.702	3.719	3.736
33	3.528	3.558	3.585	3.610	3.633	3.654	3.673	3.692	3.709	3.725
34	3.519	3.549	3.576	3.601	3.623	3.644	3.664	3.682	3.699	3.715
35	3.511	3.541	3.567	3.592	3.614	3.635	3.655	3.673	3.690	3.706
36	3.503	3.533	3.559	3.584	3.606	3.627	3.646	3.664	3.681	3.697
37	3.496	3.525	3.552	3.576	3.598	3.619	3.638	3.656	3.673	3.689
38	3.489	3.518	3.544	3.569	3.591	3.611	3.631	3.649	3.665	3.681
39	3.482	3.511	3.538	3.562	3.584	3.604	3.623	3.641	3.658	3.674
40	3.476	3.505	3.531	3.555	3.577	3.598	3.617	3.634	3.651	3.667
48	3.436	3.464	3.490	3.513	3.535	3.555	3.573	3.590	3.607	3.622
60	3.397	3.424	3.449	3.472	3.493	3.512	3.530	3.547	3.563	3.578
80	3.358	3.385	3.409	3.431	3.452	3.471	3.488	3.504	3.520	3.534
120	3.320	3.346	3.370	3.392	3.411	3.430	3.447	3.463	3.478	3.492
240	3.283	3.308	3.331	3.352	3.372	3.389	3.406	3.422	3.436	3.450
Inf	3.246	3.271	3.293	3.314	3.333	3.350	3.366	3.381	3.395	3.409

alpha= 0.05

df\k	2	3	4	5	6	7	8	9	10
1	12.706	17.369	20.034	21.850	23.209	24.285	25.171	25.922	26.570
2	4.303	5.418	6.065	6.513	6.852	7.124	7.349	7.540	7.707
3	3.182	3.867	4.263	4.538	4.748	4.916	5.056	5.176	5.280
4	2.776	3.310	3.618	3.832	3.994	4.125	4.235	4.328	4.410

5	2.571	3.030	3.293	3.476	3.615	3.727	3.821	3.900	3.970
6	2.447	2.863	3.099	3.263	3.388	3.489	3.573	3.644	3.707
7	2.365	2.752	2.971	3.123	3.238	3.331	3.408	3.475	3.533
8	2.306	2.673	2.880	3.023	3.131	3.219	3.292	3.354	3.408
9	2.262	2.614	2.812	2.948	3.052	3.135	3.205	3.264	3.316
10	2.228	2.568	2.759	2.890	2.990	3.070	3.137	3.194	3.244
11	2.201	2.532	2.717	2.845	2.941	3.019	3.084	3.139	3.187
12	2.179	2.502	2.683	2.807	2.901	2.977	3.040	3.094	3.140
13	2.160	2.478	2.654	2.776	2.868	2.942	3.003	3.056	3.102
14	2.145	2.457	2.631	2.750	2.840	2.912	2.973	3.024	3.069
15	2.131	2.439	2.610	2.727	2.816	2.887	2.946	2.997	3.041
16	2.120	2.424	2.592	2.708	2.795	2.865	2.924	2.974	3.017
17	2.110	2.410	2.577	2.691	2.777	2.846	2.904	2.953	2.996
18	2.101	2.399	2.563	2.676	2.761	2.830	2.887	2.935	2.977
19	2.093	2.388	2.551	2.663	2.747	2.815	2.871	2.919	2.961
20	2.086	2.379	2.540	2.651	2.735	2.802	2.857	2.905	2.946
21	2.080	2.370	2.531	2.640	2.723	2.790	2.845	2.892	2.933
22	2.074	2.363	2.522	2.631	2.713	2.779	2.834	2.880	2.921
23	2.069	2.356	2.514	2.622	2.704	2.769	2.824	2.870	2.910
24	2.064	2.349	2.507	2.614	2.695	2.760	2.814	2.860	2.900
25	2.060	2.344	2.500	2.607	2.688	2.752	2.806	2.851	2.891
26	2.056	2.338	2.494	2.600	2.680	2.745	2.798	2.843	2.883
27	2.052	2.333	2.488	2.594	2.674	2.738	2.791	2.836	2.875
28	2.048	2.329	2.483	2.588	2.668	2.731	2.784	2.829	2.868
29	2.045	2.325	2.478	2.583	2.662	2.725	2.778	2.823	2.862
30	2.042	2.321	2.474	2.578	2.657	2.720	2.772	2.817	2.856
31	2.040	2.317	2.469	2.574	2.652	2.715	2.767	2.811	2.850
32	2.037	2.314	2.466	2.569	2.647	2.710	2.762	2.806	2.845
33	2.035	2.310	2.462	2.565	2.643	2.705	2.757	2.801	2.840
34	2.032	2.307	2.458	2.561	2.639	2.701	2.753	2.797	2.835
35	2.030	2.305	2.455	2.558	2.635	2.697	2.748	2.792	2.830
36	2.028	2.302	2.452	2.555	2.632	2.693	2.745	2.788	2.826
37	2.026	2.299	2.449	2.551	2.628	2.690	2.741	2.784	2.822
38	2.024	2.297	2.447	2.548	2.625	2.686	2.737	2.781	2.819

39	2.023	2.295	2.444	2.546	2.622	2.683	2.734	2.777	2.815
40	2.021	2.293	2.441	2.543	2.619	2.680	2.731	2.774	2.812
48	2.011	2.279	2.426	2.526	2.601	2.661	2.711	2.753	2.790
60	2.000	2.265	2.410	2.508	2.582	2.642	2.691	2.733	2.769
80	1.990	2.252	2.394	2.491	2.564	2.623	2.671	2.712	2.748
120	1.980	2.238	2.379	2.475	2.547	2.604	2.651	2.692	2.727
240	1.970	2.225	2.364	2.458	2.529	2.585	2.632	2.672	2.707
Inf	1.960	2.212	2.349	2.442	2.511	2.567	2.613	2.652	2.686

alpha= 0.05

df\k	11	12	13	14	15	16	17	18	19	20
1	27.141	27.649	28.106	28.521	28.901	29.251	29.575	29.876	30.158	30.422
2	7.853	7.985	8.103	8.211	8.310	8.401	8.485	8.564	8.638	8.707
3	5.372	5.455	5.529	5.597	5.660	5.717	5.771	5.821	5.868	5.912
4	4.482	4.546	4.605	4.658	4.707	4.752	4.794	4.834	4.870	4.905
5	4.032	4.087	4.137	4.183	4.225	4.264	4.300	4.334	4.366	4.395
6	3.763	3.812	3.857	3.898	3.936	3.971	4.004	4.034	4.063	4.089
7	3.584	3.630	3.671	3.709	3.744	3.777	3.807	3.835	3.861	3.886
8	3.457	3.500	3.539	3.575	3.608	3.638	3.666	3.693	3.718	3.741
9	3.362	3.403	3.440	3.474	3.506	3.535	3.562	3.587	3.610	3.633
10	3.288	3.328	3.364	3.396	3.427	3.454	3.480	3.504	3.527	3.549
11	3.230	3.268	3.303	3.334	3.363	3.390	3.415	3.439	3.461	3.481
12	3.182	3.219	3.253	3.284	3.312	3.338	3.363	3.385	3.407	3.427
13	3.142	3.179	3.212	3.242	3.269	3.295	3.319	3.341	3.362	3.381
14	3.109	3.144	3.177	3.206	3.233	3.258	3.282	3.303	3.324	3.343
15	3.080	3.115	3.147	3.176	3.202	3.227	3.250	3.271	3.291	3.310
16	3.056	3.090	3.121	3.150	3.176	3.200	3.222	3.243	3.263	3.282
17	3.034	3.068	3.099	3.127	3.152	3.176	3.199	3.219	3.239	3.257
18	3.015	3.048	3.079	3.107	3.132	3.156	3.177	3.198	3.217	3.235
19	2.998	3.031	3.061	3.089	3.114	3.137	3.159	3.179	3.198	3.216
20	2.983	3.016	3.045	3.073	3.098	3.121	3.142	3.162	3.181	3.198
21	2.969	3.002	3.031	3.058	3.083	3.106	3.127	3.147	3.165	3.183
22	2.957	2.989	3.019	3.045	3.070	3.092	3.113	3.133	3.151	3.169
23	2.946	2.978	3.007	3.033	3.058	3.080	3.101	3.121	3.139	3.156
24	2.936	2.968	2.996	3.023	3.047	3.069	3.090	3.109	3.127	3.144

25	2.927	2.958	2.987	3.013	3.037	3.059	3.080	3.099	3.117	3.134
26	2.918	2.949	2.978	3.004	3.028	3.050	3.070	3.089	3.107	3.124
27	2.910	2.941	2.970	2.995	3.019	3.041	3.061	3.080	3.098	3.115
28	2.903	2.934	2.962	2.988	3.011	3.033	3.053	3.072	3.090	3.107
29	2.896	2.927	2.955	2.981	3.004	3.026	3.046	3.065	3.082	3.099
30	2.890	2.921	2.949	2.974	2.997	3.019	3.039	3.058	3.075	3.092
31	2.884	2.915	2.942	2.968	2.991	3.012	3.032	3.051	3.068	3.085
32	2.879	2.909	2.937	2.962	2.985	3.006	3.026	3.045	3.062	3.079
33	2.873	2.904	2.931	2.956	2.979	3.001	3.021	3.039	3.056	3.073
34	2.869	2.899	2.926	2.951	2.974	2.996	3.015	3.034	3.051	3.067
35	2.864	2.894	2.922	2.947	2.969	2.991	3.010	3.029	3.046	3.062
36	2.860	2.890	2.917	2.942	2.965	2.986	3.005	3.024	3.041	3.057
37	2.856	2.886	2.913	2.938	2.960	2.981	3.001	3.019	3.036	3.052
38	2.852	2.882	2.909	2.934	2.956	2.977	2.997	3.015	3.032	3.048
39	2.848	2.878	2.905	2.930	2.952	2.973	2.993	3.011	3.028	3.044
40	2.845	2.875	2.902	2.926	2.949	2.970	2.989	3.007	3.024	3.040
48	2.823	2.852	2.879	2.903	2.925	2.945	2.964	2.982	2.999	3.015
60	2.801	2.830	2.856	2.880	2.901	2.922	2.940	2.958	2.974	2.989
80	2.780	2.808	2.833	2.857	2.878	2.898	2.916	2.933	2.950	2.965
120	2.758	2.786	2.811	2.834	2.855	2.875	2.893	2.910	2.925	2.940
240	2.737	2.764	2.789	2.812	2.832	2.852	2.869	2.886	2.901	2.916
Inf	2.716	2.743	2.767	2.790	2.810	2.829	2.846	2.862	2.878	2.892

## Z-Table (upper-tail)

<b>z</b>	<b>0</b>	<b>0.01</b>	<b>0.02</b>	<b>0.03</b>	<b>0.04</b>	<b>0.05</b>	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>
<b>0</b>	0.500	0.504	0.508	0.512	0.516	0.520	0.524	0.528	0.532	0.536
<b>0.1</b>	0.540	0.544	0.548	0.552	0.556	0.560	0.564	0.567	0.571	0.575
<b>0.2</b>	0.579	0.583	0.587	0.591	0.595	0.599	0.603	0.606	0.610	0.614
<b>0.3</b>	0.618	0.622	0.626	0.629	0.633	0.637	0.641	0.644	0.648	0.652
<b>0.4</b>	0.655	0.659	0.663	0.666	0.670	0.674	0.677	0.681	0.684	0.688
<b>0.5</b>	0.691	0.695	0.698	0.702	0.705	0.709	0.712	0.716	0.719	0.722
<b>0.6</b>	0.726	0.729	0.732	0.736	0.739	0.742	0.745	0.749	0.752	0.755
<b>0.7</b>	0.758	0.761	0.764	0.767	0.770	0.773	0.776	0.779	0.782	0.785
<b>0.8</b>	0.788	0.791	0.794	0.797	0.800	0.802	0.805	0.808	0.811	0.813
<b>0.9</b>	0.816	0.819	0.821	0.824	0.826	0.829	0.831	0.834	0.836	0.839
<b>1</b>	0.841	0.844	0.846	0.848	0.851	0.853	0.855	0.858	0.860	0.862
<b>1.1</b>	0.864	0.867	0.869	0.871	0.873	0.875	0.877	0.879	0.881	0.883
<b>1.2</b>	0.885	0.887	0.889	0.891	0.893	0.894	0.896	0.898	0.900	0.901
<b>1.3</b>	0.903	0.905	0.907	0.908	0.910	0.911	0.913	0.915	0.916	0.918
<b>1.4</b>	0.919	0.921	0.922	0.924	0.925	0.926	0.928	0.929	0.931	0.932
<b>1.5</b>	0.933	0.934	0.936	0.937	0.938	0.939	0.941	0.942	0.943	0.944
<b>1.6</b>	0.945	0.946	0.947	0.948	0.949	0.951	0.952	0.953	0.954	0.954
<b>1.7</b>	0.955	0.956	0.957	0.958	0.959	0.960	0.961	0.962	0.962	0.963
<b>1.8</b>	0.964	0.965	0.966	0.966	0.967	0.968	0.969	0.969	0.970	0.971
<b>1.9</b>	0.971	0.972	0.973	0.973	0.974	0.974	0.975	0.976	0.976	0.977
<b>2</b>	0.977	0.978	0.978	0.979	0.979	0.980	0.980	0.981	0.981	0.982
<b>2.1</b>	0.982	0.983	0.983	0.983	0.984	0.984	0.985	0.985	0.985	0.986
<b>2.2</b>	0.986	0.986	0.987	0.987	0.987	0.988	0.988	0.988	0.989	0.989
<b>2.3</b>	0.989	0.990	0.990	0.990	0.990	0.991	0.991	0.991	0.991	0.992
<b>2.4</b>	0.992	0.992	0.992	0.992	0.993	0.993	0.993	0.993	0.993	0.994
<b>2.5</b>	0.994	0.994	0.994	0.994	0.994	0.995	0.995	0.995	0.995	0.995
<b>2.6</b>	0.995	0.995	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.996
<b>2.7</b>	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997
<b>2.8</b>	0.997	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998
<b>2.9</b>	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.999	0.999	0.999
<b>3</b>	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999

## Standard Normal Margin of Error (E) Table

E = 0.1 Low, E = 0.3 Medium, E = 0.5 High

Z <sub>1-α</sub> Margin of Error (MOE) Table						Proportion = 0.5
n/1-α	0.95	0.975	0.99	0.999	0.9999	0.99999
10	0.2601	0.3099	0.3678	0.4886	0.5880	0.6743
20	0.1839	0.2191	0.2601	0.3455	0.4158	0.4768
30	0.1502	0.1789	0.2124	0.2821	0.3395	0.3893
40	0.1300	0.1549	0.1839	0.2443	0.2940	0.3372
50	0.1163	0.1386	0.1645	0.2185	0.2630	0.3016
60	0.1062	0.1265	0.1502	0.1995	0.2401	0.2753
70	0.0983	0.1171	0.1390	0.1847	0.2223	0.2549
80	0.0920	0.1096	0.1300	0.1727	0.2079	0.2384
90	0.0867	0.1033	0.1226	0.1629	0.1960	0.2248
100	0.0822	0.0980	0.1163	0.1545	0.1860	0.2132
200	0.0582	0.0693	0.0822	0.1093	0.1315	0.1508
300	0.0475	0.0566	0.0672	0.0892	0.1074	0.1231
400	0.0411	0.0490	0.0582	0.0773	0.0930	0.1066
500	0.0368	0.0438	0.0520	0.0691	0.0832	0.0954
600	0.0336	0.0400	0.0475	0.0631	0.0759	0.0871
700	0.0311	0.0370	0.0440	0.0584	0.0703	0.0806
800	0.0291	0.0346	0.0411	0.0546	0.0657	0.0754
900	0.0274	0.0327	0.0388	0.0515	0.0620	0.0711
1000	0.0260	0.0310	0.0368	0.0489	0.0588	0.0674
2000	0.0184	0.0219	0.0260	0.0346	0.0416	0.0477
3000	0.0150	0.0179	0.0212	0.0282	0.0340	0.0389
4000	0.0130	0.0155	0.0184	0.0244	0.0294	0.0337
5000	0.0116	0.0139	0.0165	0.0219	0.0263	0.0302
6000	0.0106	0.0127	0.0150	0.0199	0.0240	0.0275
7000	0.0098	0.0117	0.0139	0.0185	0.0222	0.0255
8000	0.0092	0.0110	0.0130	0.0173	0.0208	0.0238
9000	0.0087	0.0103	0.0123	0.0163	0.0196	0.0225
10000	0.0082	0.0098	0.0116	0.0155	0.0186	0.0213

Kolmogorov-Smirnov Critical Values							
KS-Value	1.073	1.14	1.223	1.358	1.518	1.629	
n	$\alpha = 0.20$	$\alpha = 0.15$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.02$	$\alpha = 0.01$	
10	0.324	0.344	0.369	0.409	0.458	0.491	
11	0.310	0.329	0.353	0.392	0.438	0.470	
12	0.298	0.316	0.339	0.377	0.421	0.452	
13	0.287	0.305	0.327	0.363	0.406	0.435	
14	0.277	0.294	0.316	0.351	0.392	0.421	
15	0.268	0.285	0.306	0.340	0.380	0.407	
16	0.260	0.276	0.297	0.329	0.368	0.395	
17	0.253	0.269	0.288	0.320	0.358	0.384	
18	0.246	0.262	0.281	0.312	0.348	0.374	
19	0.240	0.255	0.273	0.304	0.339	0.364	
20	0.234	0.249	0.267	0.296	0.331	0.355	
21	0.229	0.243	0.261	0.290	0.324	0.347	
22	0.224	0.238	0.255	0.283	0.317	0.340	
23	0.219	0.233	0.250	0.277	0.310	0.333	
24	0.215	0.228	0.245	0.272	0.304	0.326	
25	0.210	0.224	0.240	0.266	0.298	0.319	
26	0.206	0.219	0.235	0.261	0.292	0.314	
27	0.203	0.215	0.231	0.257	0.287	0.308	
28	0.199	0.212	0.227	0.252	0.282	0.302	
29	0.196	0.208	0.223	0.248	0.277	0.297	
30	0.193	0.205	0.220	0.244	0.273	0.293	
31	0.190	0.202	0.216	0.240	0.268	0.288	
32	0.187	0.198	0.213	0.236	0.264	0.284	
33	0.184	0.196	0.210	0.233	0.260	0.279	
34	0.181	0.193	0.207	0.230	0.257	0.275	
35	0.179	0.190	0.204	0.226	0.253	0.272	
36	0.176	0.187	0.201	0.223	0.250	0.268	
37	0.174	0.185	0.198	0.220	0.246	0.264	
38	0.172	0.183	0.196	0.217	0.243	0.261	
39	0.170	0.180	0.193	0.215	0.240	0.258	
40	0.168	0.178	0.191	0.212	0.237	0.254	
41	0.166	0.176	0.189	0.210	0.234	0.251	
42	0.164	0.174	0.187	0.207	0.231	0.248	
43	0.162	0.172	0.184	0.205	0.229	0.246	
44	0.160	0.170	0.182	0.202	0.226	0.243	
45	0.158	0.168	0.180	0.200	0.224	0.240	
46	0.157	0.166	0.178	0.198	0.221	0.238	
47	0.155	0.165	0.177	0.196	0.219	0.235	
48	0.153	0.163	0.175	0.194	0.217	0.233	

Kolmogorov-Smirnov Critical Values							
KS-Value	1.073	1.14	1.223	1.358	1.518	1.629	
n	$\alpha = 0.20$	$\alpha = 0.15$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.02$	$\alpha = 0.01$	
49	0.152	0.161	0.173	0.192	0.215	0.230	
50	0.150	0.160	0.171	0.190	0.213	0.228	
60	0.137	0.146	0.157	0.174	0.194	0.209	
70	0.127	0.135	0.145	0.161	0.180	0.193	
80	0.119	0.127	0.136	0.151	0.169	0.181	
90	0.112	0.120	0.128	0.142	0.159	0.171	
100	0.107	0.113	0.122	0.135	0.151	0.162	
Over 50	$\frac{ks\text{-value}}{\sqrt{n}}$						

## Chi-Square Table (upper-tail)

df/α	Chi-Square Table (one-tailed) T.S. < $\chi^2_{(1-\alpha, N-1)}$					
	α =0.1	α =0.05	α =0.025	α =0.01	α =0.005	α =0.001
1	2.7055	3.8415	5.0239	6.6349	7.8794	10.8276
2	4.6052	5.9915	7.3778	9.2103	10.5966	13.8155
3	6.2514	7.8147	9.3484	11.3449	12.8382	16.2662
4	7.7794	9.4877	11.1433	13.2767	14.8603	18.4668
5	9.2364	11.0705	12.8325	15.0863	16.7496	20.5150
6	10.6446	12.5916	14.4494	16.8119	18.5476	22.4577
7	12.0170	14.0671	16.0128	18.4753	20.2777	24.3219
8	13.3616	15.5073	17.5345	20.0902	21.9550	26.1245
9	14.6837	16.9190	19.0228	21.6660	23.5894	27.8772
10	15.9872	18.3070	20.4832	23.2093	25.1882	29.5883
11	17.2750	19.6751	21.9200	24.7250	26.7568	31.2641
12	18.5493	21.0261	23.3367	26.2170	28.2995	32.9095
13	19.8119	22.3620	24.7356	27.6882	29.8195	34.5282
14	21.0641	23.6848	26.1189	29.1412	31.3193	36.1233
15	22.3071	24.9958	27.4884	30.5779	32.8013	37.6973
16	23.5418	26.2962	28.8454	31.9999	34.2672	39.2524
17	24.7690	27.5871	30.1910	33.4087	35.7185	40.7902
18	25.9894	28.8693	31.5264	34.8053	37.1565	42.3124
19	27.2036	30.1435	32.8523	36.1909	38.5823	43.8202
20	28.4120	31.4104	34.1696	37.5662	39.9968	45.3147
21	29.6151	32.6706	35.4789	38.9322	41.4011	46.7970
22	30.8133	33.9244	36.7807	40.2894	42.7957	48.2679
23	32.0069	35.1725	38.0756	41.6384	44.1813	49.7282
24	33.1962	36.4150	39.3641	42.9798	45.5585	51.1786
25	34.3816	37.6525	40.6465	44.3141	46.9279	52.6197
26	35.5632	38.8851	41.9232	45.6417	48.2899	54.0520
27	36.7412	40.1133	43.1945	46.9629	49.6449	55.4760
28	37.9159	41.3371	44.4608	48.2782	50.9934	56.8923
29	39.0875	42.5570	45.7223	49.5879	52.3356	58.3012
30	40.2560	43.7730	46.9792	50.8922	53.6720	59.7031
31	41.4217	44.9853	48.2319	52.1914	55.0027	61.0983

df/ $\alpha$	Chi-Square Table (one-tailed) T.S. < $\chi^2_{(1-\alpha, N-1)}$					
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.001$
32	42.5847	46.1943	49.4804	53.4858	56.3281	62.4872
33	43.7452	47.3999	50.7251	54.7755	57.6484	63.8701
34	44.9032	48.6024	51.9660	56.0609	58.9639	65.2472
35	46.0588	49.8018	53.2033	57.3421	60.2748	66.6188
36	47.2122	50.9985	54.4373	58.6192	61.5812	67.9852
37	48.3634	52.1923	55.6680	59.8925	62.8833	69.3465
38	49.5126	53.3835	56.8955	61.1621	64.1814	70.7029
39	50.6598	54.5722	58.1201	62.4281	65.4756	72.0547
40	51.8051	55.7585	59.3417	63.6907	66.7660	73.4020
41	52.9485	56.9424	60.5606	64.9501	68.0527	74.7449
42	54.0902	58.1240	61.7768	66.2062	69.3360	76.0838
43	55.2302	59.3035	62.9904	67.4593	70.6159	77.4186
44	56.3685	60.4809	64.2015	68.7095	71.8926	78.7495
45	57.5053	61.6562	65.4102	69.9568	73.1661	80.0767
46	58.6405	62.8296	66.6165	71.2014	74.4365	81.4003
47	59.7743	64.0011	67.8206	72.4433	75.7041	82.7204
48	60.9066	65.1708	69.0226	73.6826	76.9688	84.0371
49	62.0375	66.3386	70.2224	74.9195	78.2307	85.3506
50	63.1671	67.5048	71.4202	76.1539	79.4900	86.6608
51	64.2954	68.6693	72.6160	77.3860	80.7467	87.9680
52	65.4224	69.8322	73.8099	78.6158	82.0008	89.2722
53	66.5482	70.9935	75.0019	79.8433	83.2526	90.5734
54	67.6728	72.1532	76.1920	81.0688	84.5019	91.8718
55	68.7962	73.3115	77.3805	82.2921	85.7490	93.1675
56	69.9185	74.4683	78.5672	83.5134	86.9938	94.4605
57	71.0397	75.6237	79.7522	84.7328	88.2364	95.7510
58	72.1598	76.7778	80.9356	85.9502	89.4769	97.0388
59	73.2789	77.9305	82.1174	87.1657	90.7153	98.3242
60	74.3970	79.0819	83.2977	88.3794	91.9517	99.6072
61	75.5141	80.2321	84.4764	89.5913	93.1861	100.8879
62	76.6302	81.3810	85.6537	90.8015	94.4187	102.1662
63	77.7454	82.5287	86.8296	92.0100	95.6493	103.4424

df/α	Chi-Square Table (one-tailed) T.S. < $\chi^2_{(1-\alpha, N-1)}$					
	α =0.1	α =0.05	α =0.025	α =0.01	α =0.005	α =0.001
64	78.8596	83.6753	88.0041	93.2169	96.8781	104.7163
65	79.9730	84.8206	89.1771	94.4221	98.1051	105.9881
66	81.0855	85.9649	90.3489	95.6257	99.3304	107.2579
67	82.1971	87.1081	91.5194	96.8278	100.5540	108.5256
68	83.3079	88.2502	92.6885	98.0284	101.7759	109.7913
69	84.4179	89.3912	93.8565	99.2275	102.9962	111.0551
70	85.5270	90.5312	95.0232	100.4252	104.2149	112.3169
71	86.6354	91.6702	96.1887	101.6214	105.4320	113.5769
72	87.7430	92.8083	97.3531	102.8163	106.6476	114.8351
73	88.8499	93.9453	98.5163	104.0098	107.8617	116.0915
74	89.9560	95.0815	99.6783	105.2020	109.0744	117.3462
75	91.0615	96.2167	100.8393	106.3929	110.2856	118.5991
76	92.1662	97.3510	101.9993	107.5825	111.4954	119.8503
77	93.2702	98.4844	103.1581	108.7709	112.7038	121.1000
78	94.3735	99.6169	104.3159	109.9581	113.9109	122.3480
79	95.4762	100.7486	105.4728	111.1440	115.1166	123.5944
80	96.5782	101.8795	106.6286	112.3288	116.3211	124.8392
81	97.6796	103.0095	107.7834	113.5124	117.5242	126.0826
82	98.7803	104.1387	108.9373	114.6949	118.7261	127.3244
83	99.8805	105.2672	110.0902	115.8763	119.9268	128.5648
84	100.9800	106.3948	111.2423	117.0565	121.1263	129.8037
85	102.0789	107.5217	112.3934	118.2357	122.3246	131.0412
86	103.1773	108.6479	113.5436	119.4139	123.5217	132.2773
87	104.2750	109.7733	114.6929	120.5910	124.7177	133.5121
88	105.3722	110.8980	115.8414	121.7671	125.9125	134.7455
89	106.4689	112.0220	116.9891	122.9422	127.1063	135.9776
90	107.5650	113.1453	118.1359	124.1163	128.2989	137.2084
91	108.6606	114.2679	119.2819	125.2895	129.4905	138.4379
92	109.7556	115.3898	120.4271	126.4617	130.6811	139.6661
93	110.8502	116.5110	121.5715	127.6329	131.8706	140.8931
94	111.9442	117.6317	122.7151	128.8032	133.0591	142.1189
95	113.0377	118.7516	123.8580	129.9727	134.2465	143.3435

df/ $\alpha$	Chi-Square Table (one-tailed) T.S. < $\chi^2_{(1-\alpha, N-1)}$					
	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.001$
96	114.1307	119.8709	125.0001	131.1412	135.4330	144.5670
97	115.2232	120.9896	126.1414	132.3089	136.6186	145.7892
98	116.3153	122.1077	127.2821	133.4757	137.8032	147.0104
99	117.4069	123.2252	128.4220	134.6416	138.9868	148.2304
100	118.4980	124.3421	129.5612	135.8067	140.1695	149.4493

## F Critical Value Table, $F_{1-\alpha, df_1, df_2}$

		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
df1	df2	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
		0.05	0.025	0.01	0.005	0.0025	0.001
10	10	2.9782	3.7168	4.8491	5.8467	6.9875	8.7539
10	11	2.9430	3.6649	4.7715	5.7462	6.8608	8.5864
10	12	2.9130	3.6209	4.7059	5.6613	6.7538	8.4452
10	13	2.8872	3.5832	4.6496	5.5887	6.6623	8.3245
10	14	2.8647	3.5504	4.6008	5.5257	6.5831	8.2200
10	15	2.8450	3.5217	4.5581	5.4707	6.5139	8.1288
10	16	2.8276	3.4963	4.5204	5.4221	6.4528	8.0484
10	17	2.8120	3.4737	4.4869	5.3789	6.3986	7.9770
10	18	2.7980	3.4534	4.4569	5.3403	6.3501	7.9131
10	19	2.7854	3.4351	4.4299	5.3055	6.3064	7.8557
10	20	2.7740	3.4185	4.4054	5.2740	6.2669	7.8037
10	21	2.7636	3.4035	4.3831	5.2454	6.2310	7.7565
10	22	2.7541	3.3897	4.3628	5.2192	6.1982	7.7134
10	23	2.7453	3.3770	4.3441	5.1953	6.1682	7.6739
10	24	2.7372	3.3654	4.3269	5.1732	6.1405	7.6376
10	25	2.7298	3.3546	4.3111	5.1528	6.1149	7.6041
10	26	2.7229	3.3446	4.2963	5.1339	6.0913	7.5730
10	27	2.7164	3.3353	4.2827	5.1164	6.0693	7.5442
10	28	2.7104	3.3267	4.2700	5.1001	6.0489	7.5173
10	29	2.7048	3.3186	4.2581	5.0848	6.0298	7.4922
10	30	2.6996	3.3110	4.2469	5.0706	6.0119	7.4688
11	10	2.8536	3.5257	4.5393	5.4183	6.4101	7.9224
11	11	2.8179	3.4737	4.4624	5.3197	6.2869	7.7614
11	12	2.7876	3.4296	4.3974	5.2363	6.1828	7.6256
11	13	2.7614	3.3917	4.3416	5.1649	6.0937	7.5094
11	14	2.7386	3.3588	4.2932	5.1031	6.0165	7.4089
11	15	2.7186	3.3299	4.2509	5.0489	5.9490	7.3210
11	16	2.7009	3.3044	4.2134	5.0011	5.8895	7.2435
11	17	2.6851	3.2816	4.1801	4.9586	5.8366	7.1747
11	18	2.6709	3.2612	4.1503	4.9205	5.7892	7.1131

		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
11	19	2.6581	3.2428	4.1234	4.8863	5.7466	7.0577
11	20	2.6464	3.2261	4.0990	4.8552	5.7080	7.0076
11	21	2.6358	3.2109	4.0769	4.8270	5.6729	6.9620
11	22	2.6261	3.1970	4.0566	4.8012	5.6408	6.9204
11	23	2.6172	3.1843	4.0380	4.7775	5.6115	6.8823
11	24	2.6090	3.1725	4.0209	4.7557	5.5844	6.8471
11	25	2.6014	3.1616	4.0051	4.7356	5.5594	6.8147
11	26	2.5943	3.1516	3.9904	4.7170	5.5363	6.7847
11	27	2.5877	3.1422	3.9768	4.6997	5.5148	6.7568
11	28	2.5816	3.1334	3.9641	4.6835	5.4947	6.7309
11	29	2.5759	3.1253	3.9522	4.6684	5.4760	6.7066
11	30	2.5705	3.1176	3.9411	4.6543	5.4585	6.6839
12	10	2.7534	3.3736	4.2961	5.0855	5.9663	7.2920
12	11	2.7173	3.3215	4.2198	4.9884	5.8458	7.1362
12	12	2.6866	3.2773	4.1553	4.9062	5.7440	7.0046
12	13	2.6602	3.2393	4.0999	4.8358	5.6568	6.8920
12	14	2.6371	3.2062	4.0518	4.7748	5.5812	6.7945
12	15	2.6169	3.1772	4.0096	4.7213	5.5151	6.7092
12	16	2.5989	3.1515	3.9724	4.6741	5.4567	6.6340
12	17	2.5828	3.1286	3.9392	4.6321	5.4048	6.5672
12	18	2.5684	3.1081	3.9095	4.5945	5.3583	6.5074
12	19	2.5554	3.0896	3.8827	4.5606	5.3165	6.4535
12	20	2.5436	3.0728	3.8584	4.5299	5.2786	6.4048
12	21	2.5328	3.0575	3.8363	4.5020	5.2442	6.3605
12	22	2.5229	3.0434	3.8161	4.4765	5.2127	6.3200
12	23	2.5139	3.0306	3.7976	4.4530	5.1838	6.2829
12	24	2.5055	3.0187	3.7805	4.4314	5.1572	6.2488
12	25	2.4977	3.0077	3.7647	4.4115	5.1326	6.2172
12	26	2.4905	2.9976	3.7500	4.3930	5.1099	6.1880
12	27	2.4838	2.9881	3.7364	4.3759	5.0887	6.1608
12	28	2.4776	2.9793	3.7237	4.3599	5.0690	6.1355
12	29	2.4718	2.9710	3.7119	4.3449	5.0506	6.1119

		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
12	30	2.4663	2.9633	3.7008	4.3309	5.0334	6.0898
13	10	2.6710	3.2497	4.1003	4.8199	5.6153	6.7992
13	11	2.6347	3.1975	4.0245	4.7240	5.4970	6.6474
13	12	2.6037	3.1532	3.9603	4.6429	5.3970	6.5192
13	13	2.5769	3.1150	3.9052	4.5733	5.3113	6.4094
13	14	2.5536	3.0819	3.8573	4.5129	5.2370	6.3144
13	15	2.5331	3.0527	3.8154	4.4600	5.1719	6.2312
13	16	2.5149	3.0269	3.7783	4.4132	5.1145	6.1578
13	17	2.4987	3.0039	3.7452	4.3716	5.0634	6.0926
13	18	2.4841	2.9832	3.7156	4.3344	5.0177	6.0342
13	19	2.4709	2.9646	3.6888	4.3008	4.9765	5.9816
13	20	2.4589	2.9477	3.6646	4.2703	4.9391	5.9340
13	21	2.4479	2.9322	3.6425	4.2426	4.9052	5.8907
13	22	2.4379	2.9181	3.6224	4.2173	4.8741	5.8511
13	23	2.4287	2.9052	3.6038	4.1940	4.8456	5.8148
13	24	2.4202	2.8932	3.5868	4.1726	4.8194	5.7814
13	25	2.4123	2.8821	3.5710	4.1528	4.7951	5.7505
13	26	2.4050	2.8719	3.5563	4.1344	4.7727	5.7219
13	27	2.3982	2.8623	3.5427	4.1174	4.7518	5.6954
13	28	2.3918	2.8534	3.5300	4.1015	4.7323	5.6706
13	29	2.3859	2.8451	3.5182	4.0866	4.7142	5.6475
13	30	2.3803	2.8372	3.5070	4.0727	4.6971	5.6258
14	10	2.6022	3.1469	3.9394	4.6034	5.3311	6.4041
14	11	2.5655	3.0946	3.8640	4.5085	5.2146	6.2556
14	12	2.5342	3.0502	3.8001	4.4281	5.1161	6.1302
14	13	2.5073	3.0119	3.7452	4.3591	5.0316	6.0228
14	14	2.4837	2.9786	3.6975	4.2993	4.9584	5.9297
14	15	2.4630	2.9493	3.6557	4.2468	4.8942	5.8483
14	16	2.4446	2.9234	3.6187	4.2005	4.8376	5.7764
14	17	2.4282	2.9003	3.5857	4.1592	4.7871	5.7124
14	18	2.4134	2.8795	3.5561	4.1221	4.7419	5.6551
14	19	2.4000	2.8607	3.5294	4.0888	4.7012	5.6035

		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
14	20	2.3879	2.8437	3.5052	4.0585	4.6643	5.5568
14	21	2.3768	2.8282	3.4832	4.0310	4.6308	5.5143
14	22	2.3667	2.8139	3.4630	4.0058	4.6001	5.4755
14	23	2.3573	2.8009	3.4445	3.9827	4.5719	5.4399
14	24	2.3487	2.7888	3.4274	3.9614	4.5459	5.4070
14	25	2.3407	2.7777	3.4116	3.9417	4.5220	5.3767
14	26	2.3333	2.7673	3.3969	3.9234	4.4997	5.3486
14	27	2.3264	2.7577	3.3833	3.9064	4.4791	5.3225
14	28	2.3199	2.7487	3.3706	3.8906	4.4598	5.2982
14	29	2.3139	2.7403	3.3587	3.8758	4.4418	5.2754
14	30	2.3082	2.7324	3.3476	3.8619	4.4249	5.2542
15	10	2.5437	3.0602	3.8049	4.4235	5.0966	6.0808
15	11	2.5068	3.0078	3.7299	4.3295	4.9817	5.9352
15	12	2.4753	2.9633	3.6662	4.2497	4.8844	5.8121
15	13	2.4481	2.9249	3.6115	4.1813	4.8009	5.7066
15	14	2.4244	2.8915	3.5639	4.1219	4.7286	5.6151
15	15	2.4034	2.8621	3.5222	4.0698	4.6651	5.5351
15	16	2.3849	2.8360	3.4852	4.0237	4.6091	5.4644
15	17	2.3683	2.8128	3.4523	3.9827	4.5592	5.4015
15	18	2.3533	2.7919	3.4228	3.9459	4.5145	5.3452
15	19	2.3398	2.7730	3.3961	3.9127	4.4741	5.2944
15	20	2.3275	2.7559	3.3719	3.8826	4.4376	5.2484
15	21	2.3163	2.7403	3.3498	3.8552	4.4044	5.2066
15	22	2.3060	2.7260	3.3297	3.8301	4.3739	5.1683
15	23	2.2966	2.7128	3.3111	3.8071	4.3460	5.1332
15	24	2.2878	2.7006	3.2940	3.7859	4.3203	5.1009
15	25	2.2797	2.6894	3.2782	3.7662	4.2965	5.0710
15	26	2.2722	2.6790	3.2635	3.7480	4.2745	5.0433
15	27	2.2652	2.6692	3.2499	3.7311	4.2540	5.0176
15	28	2.2587	2.6602	3.2372	3.7153	4.2348	4.9936
15	29	2.2525	2.6517	3.2253	3.7006	4.2170	4.9712
15	30	2.2468	2.6437	3.2141	3.6867	4.2002	4.9502

		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
16	10	2.4935	2.9862	3.6909	4.2719	4.9000	5.8117
16	11	2.4564	2.9337	3.6162	4.1785	4.7863	5.6684
16	12	2.4247	2.8890	3.5527	4.0994	4.6901	5.5473
16	13	2.3973	2.8506	3.4981	4.0314	4.6075	5.4434
16	14	2.3733	2.8170	3.4506	3.9723	4.5358	5.3533
16	15	2.3522	2.7875	3.4089	3.9205	4.4730	5.2745
16	16	2.3335	2.7614	3.3720	3.8747	4.4175	5.2048
16	17	2.3167	2.7380	3.3391	3.8338	4.3680	5.1428
16	18	2.3016	2.7170	3.3096	3.7972	4.3237	5.0872
16	19	2.2880	2.6980	3.2829	3.7641	4.2837	5.0372
16	20	2.2756	2.6808	3.2587	3.7342	4.2475	4.9918
16	21	2.2642	2.6651	3.2367	3.7069	4.2145	4.9505
16	22	2.2538	2.6507	3.2165	3.6819	4.1843	4.9128
16	23	2.2443	2.6374	3.1979	3.6589	4.1566	4.8781
16	24	2.2354	2.6252	3.1808	3.6378	4.1310	4.8462
16	25	2.2272	2.6138	3.1650	3.6182	4.1074	4.8167
16	26	2.2196	2.6033	3.1503	3.6000	4.0855	4.7893
16	27	2.2125	2.5935	3.1366	3.5831	4.0651	4.7638
16	28	2.2059	2.5844	3.1238	3.5674	4.0461	4.7401
16	29	2.1997	2.5758	3.1119	3.5527	4.0284	4.7180
16	30	2.1938	2.5678	3.1007	3.5389	4.0118	4.6972
17	10	2.4499	2.9222	3.5931	4.1424	4.7328	5.5844
17	11	2.4126	2.8696	3.5185	4.0496	4.6203	5.4431
17	12	2.3807	2.8249	3.4552	3.9709	4.5249	5.3237
17	13	2.3531	2.7863	3.4007	3.9033	4.4431	5.2212
17	14	2.3290	2.7526	3.3533	3.8445	4.3720	5.1323
17	15	2.3077	2.7230	3.3117	3.7929	4.3097	5.0544
17	16	2.2888	2.6968	3.2748	3.7473	4.2546	4.9856
17	17	2.2719	2.6733	3.2419	3.7066	4.2056	4.9244
17	18	2.2567	2.6522	3.2124	3.6701	4.1615	4.8695
17	19	2.2429	2.6331	3.1857	3.6372	4.1218	4.8200
17	20	2.2304	2.6158	3.1615	3.6073	4.0859	4.7751

		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
17	21	2.2189	2.6000	3.1394	3.5801	4.0531	4.7343
17	22	2.2084	2.5855	3.1192	3.5552	4.0231	4.6969
17	23	2.1987	2.5721	3.1006	3.5323	3.9955	4.6627
17	24	2.1898	2.5598	3.0835	3.5112	3.9701	4.6311
17	25	2.1815	2.5484	3.0676	3.4916	3.9466	4.6019
17	26	2.1738	2.5378	3.0529	3.4735	3.9248	4.5748
17	27	2.1666	2.5280	3.0392	3.4566	3.9046	4.5496
17	28	2.1599	2.5187	3.0264	3.4409	3.8857	4.5261
17	29	2.1536	2.5101	3.0145	3.4262	3.8680	4.5042
17	30	2.1477	2.5020	3.0032	3.4124	3.8514	4.4836
18	10	2.4117	2.8664	3.5082	4.0305	4.5891	5.3900
18	11	2.3742	2.8137	3.4338	3.9382	4.4775	5.2505
18	12	2.3421	2.7689	3.3706	3.8599	4.3829	5.1324
18	13	2.3143	2.7302	3.3162	3.7926	4.3017	5.0312
18	14	2.2900	2.6964	3.2689	3.7341	4.2312	4.9433
18	15	2.2686	2.6667	3.2273	3.6827	4.1693	4.8663
18	16	2.2496	2.6404	3.1904	3.6373	4.1146	4.7982
18	17	2.2325	2.6168	3.1575	3.5967	4.0658	4.7376
18	18	2.2172	2.5956	3.1280	3.5603	4.0221	4.6833
18	19	2.2033	2.5764	3.1013	3.5275	3.9826	4.6343
18	20	2.1906	2.5590	3.0771	3.4977	3.9468	4.5899
18	21	2.1791	2.5431	3.0550	3.4705	3.9142	4.5494
18	22	2.1685	2.5285	3.0348	3.4456	3.8843	4.5124
18	23	2.1587	2.5151	3.0161	3.4228	3.8569	4.4784
18	24	2.1497	2.5027	2.9990	3.4017	3.8316	4.4471
18	25	2.1413	2.4912	2.9831	3.3822	3.8083	4.4182
18	26	2.1335	2.4806	2.9683	3.3641	3.7866	4.3913
18	27	2.1262	2.4706	2.9546	3.3472	3.7664	4.3663
18	28	2.1195	2.4613	2.9418	3.3315	3.7476	4.3430
18	29	2.1131	2.4527	2.9298	3.3168	3.7300	4.3213
18	30	2.1071	2.4445	2.9185	3.3030	3.7135	4.3009
19	10	2.3779	2.8172	3.4338	3.9329	4.4641	5.2219

		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
19	11	2.3402	2.7645	3.3596	3.8410	4.3534	5.0840
19	12	2.3080	2.7196	3.2965	3.7631	4.2595	4.9672
19	13	2.2800	2.6808	3.2422	3.6961	4.1788	4.8669
19	14	2.2556	2.6469	3.1949	3.6378	4.1087	4.7799
19	15	2.2341	2.6171	3.1533	3.5866	4.0473	4.7037
19	16	2.2149	2.5907	3.1165	3.5412	3.9928	4.6362
19	17	2.1977	2.5670	3.0836	3.5008	3.9443	4.5762
19	18	2.1823	2.5457	3.0541	3.4645	3.9008	4.5223
19	19	2.1683	2.5265	3.0274	3.4318	3.8616	4.4738
19	20	2.1555	2.5089	3.0031	3.4020	3.8259	4.4297
19	21	2.1438	2.4930	2.9810	3.3749	3.7935	4.3896
19	22	2.1331	2.4783	2.9607	3.3500	3.7638	4.3529
19	23	2.1233	2.4648	2.9421	3.3272	3.7365	4.3192
19	24	2.1141	2.4523	2.9249	3.3062	3.7113	4.2881
19	25	2.1057	2.4408	2.9089	3.2867	3.6880	4.2594
19	26	2.0978	2.4300	2.8941	3.2686	3.6664	4.2327
19	27	2.0905	2.4200	2.8804	3.2517	3.6462	4.2079
19	28	2.0836	2.4107	2.8675	3.2360	3.6275	4.1848
19	29	2.0772	2.4019	2.8555	3.2213	3.6099	4.1632
19	30	2.0712	2.3937	2.8442	3.2075	3.5935	4.1429
20	10	2.3479	2.7737	3.3682	3.8470	4.3546	5.0752
20	11	2.3100	2.7209	3.2941	3.7555	4.2446	4.9386
20	12	2.2776	2.6758	3.2311	3.6779	4.1513	4.8229
20	13	2.2495	2.6369	3.1769	3.6111	4.0711	4.7236
20	14	2.2250	2.6030	3.1296	3.5530	4.0014	4.6374
20	15	2.2033	2.5731	3.0880	3.5020	3.9403	4.5618
20	16	2.1840	2.5465	3.0512	3.4568	3.8861	4.4949
20	17	2.1667	2.5228	3.0183	3.4164	3.8379	4.4353
20	18	2.1511	2.5014	2.9887	3.3802	3.7945	4.3819
20	19	2.1370	2.4821	2.9620	3.3475	3.7554	4.3337
20	20	2.1242	2.4645	2.9377	3.3178	3.7200	4.2900
20	21	2.1124	2.4484	2.9156	3.2907	3.6876	4.2501

		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
20	22	2.1016	2.4337	2.8953	3.2659	3.6580	4.2137
20	23	2.0917	2.4201	2.8766	3.2431	3.6308	4.1802
20	24	2.0825	2.4076	2.8594	3.2220	3.6057	4.1493
20	25	2.0739	2.3959	2.8434	3.2025	3.5825	4.1208
20	26	2.0660	2.3851	2.8286	3.1845	3.5609	4.0943
20	27	2.0586	2.3751	2.8148	3.1676	3.5409	4.0696
20	28	2.0517	2.3657	2.8019	3.1519	3.5221	4.0466
20	29	2.0452	2.3569	2.7898	3.1372	3.5046	4.0251
20	30	2.0391	2.3486	2.7785	3.1234	3.4882	4.0050
21	10	2.3210	2.7348	3.3098	3.7709	4.2579	4.9462
21	11	2.2829	2.6819	3.2359	3.6798	4.1485	4.8107
21	12	2.2504	2.6368	3.1730	3.6024	4.0557	4.6960
21	13	2.2222	2.5978	3.1187	3.5358	3.9759	4.5975
21	14	2.1975	2.5638	3.0715	3.4779	3.9066	4.5119
21	15	2.1757	2.5338	3.0300	3.4270	3.8457	4.4369
21	16	2.1563	2.5071	2.9931	3.3818	3.7918	4.3705
21	17	2.1389	2.4833	2.9602	3.3416	3.7437	4.3114
21	18	2.1232	2.4618	2.9306	3.3054	3.7006	4.2583
21	19	2.1090	2.4424	2.9039	3.2728	3.6616	4.2104
21	20	2.0960	2.4247	2.8796	3.2431	3.6263	4.1670
21	21	2.0842	2.4086	2.8574	3.2160	3.5941	4.1274
21	22	2.0733	2.3938	2.8370	3.1912	3.5645	4.0912
21	23	2.0633	2.3801	2.8183	3.1684	3.5374	4.0579
21	24	2.0540	2.3675	2.8010	3.1474	3.5124	4.0272
21	25	2.0454	2.3558	2.7850	3.1279	3.4892	3.9988
21	26	2.0374	2.3450	2.7702	3.1098	3.4677	3.9724
21	27	2.0299	2.3348	2.7563	3.0930	3.4477	3.9479
21	28	2.0229	2.3254	2.7434	3.0773	3.4290	3.9250
21	29	2.0164	2.3165	2.7313	3.0625	3.4115	3.9036
21	30	2.0102	2.3082	2.7200	3.0488	3.3951	3.8836
22	10	2.2967	2.6998	3.2576	3.7030	4.1717	4.8317
22	11	2.2585	2.6469	3.1837	3.6122	4.0629	4.6973

		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
22	12	2.2258	2.6017	3.1209	3.5350	3.9706	4.5835
22	13	2.1975	2.5626	3.0667	3.4686	3.8912	4.4857
22	14	2.1727	2.5285	3.0195	3.4108	3.8222	4.4007
22	15	2.1508	2.4984	2.9779	3.3600	3.7616	4.3262
22	16	2.1313	2.4717	2.9411	3.3150	3.7079	4.2602
22	17	2.1138	2.4478	2.9082	3.2748	3.6600	4.2015
22	18	2.0980	2.4262	2.8786	3.2387	3.6170	4.1487
22	19	2.0837	2.4067	2.8518	3.2060	3.5782	4.1011
22	20	2.0707	2.3890	2.8274	3.1764	3.5429	4.0579
22	21	2.0587	2.3728	2.8052	3.1494	3.5108	4.0186
22	22	2.0478	2.3579	2.7849	3.1246	3.4813	3.9825
22	23	2.0377	2.3442	2.7661	3.1018	3.4542	3.9494
22	24	2.0283	2.3315	2.7488	3.0807	3.4293	3.9189
22	25	2.0196	2.3198	2.7328	3.0613	3.4061	3.8906
22	26	2.0116	2.3088	2.7179	3.0432	3.3847	3.8644
22	27	2.0040	2.2986	2.7040	3.0263	3.3647	3.8400
22	28	1.9970	2.2891	2.6910	3.0106	3.3460	3.8172
22	29	1.9904	2.2802	2.6789	2.9959	3.3286	3.7959
22	30	1.9842	2.2718	2.6675	2.9821	3.3122	3.7759
23	10	2.2747	2.6682	3.2106	3.6420	4.0946	4.7296
23	11	2.2364	2.6152	3.1368	3.5515	3.9863	4.5962
23	12	2.2036	2.5699	3.0740	3.4745	3.8944	4.4831
23	13	2.1752	2.5308	3.0199	3.4083	3.8154	4.3859
23	14	2.1502	2.4966	2.9727	3.3506	3.7466	4.3015
23	15	2.1282	2.4665	2.9311	3.2999	3.6862	4.2274
23	16	2.1086	2.4396	2.8943	3.2549	3.6327	4.1618
23	17	2.0910	2.4157	2.8613	3.2148	3.5850	4.1034
23	18	2.0751	2.3940	2.8317	3.1787	3.5421	4.0509
23	19	2.0608	2.3745	2.8049	3.1461	3.5034	4.0036
23	20	2.0476	2.3567	2.7805	3.1165	3.4682	3.9606
23	21	2.0356	2.3404	2.7583	3.0895	3.4361	3.9214
23	22	2.0246	2.3254	2.7378	3.0647	3.4068	3.8856

		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
23	23	2.0144	2.3116	2.7191	3.0419	3.3797	3.8526
23	24	2.0050	2.2989	2.7017	3.0208	3.3548	3.8222
23	25	1.9963	2.2871	2.6856	3.0014	3.3317	3.7940
23	26	1.9881	2.2761	2.6707	2.9833	3.3103	3.7679
23	27	1.9805	2.2659	2.6568	2.9664	3.2903	3.7436
23	28	1.9734	2.2563	2.6438	2.9507	3.2717	3.7209
23	29	1.9668	2.2473	2.6316	2.9359	3.2543	3.6997
23	30	1.9605	2.2389	2.6202	2.9221	3.2379	3.6798
24	10	2.2547	2.6396	3.1681	3.5870	4.0252	4.6379
24	11	2.2163	2.5865	3.0944	3.4967	3.9174	4.5053
24	12	2.1834	2.5411	3.0316	3.4199	3.8258	4.3929
24	13	2.1548	2.5019	2.9775	3.3538	3.7471	4.2963
24	14	2.1298	2.4677	2.9303	3.2962	3.6785	4.2124
24	15	2.1077	2.4374	2.8887	3.2456	3.6183	4.1387
24	16	2.0880	2.4105	2.8519	3.2007	3.5650	4.0735
24	17	2.0703	2.3865	2.8189	3.1606	3.5174	4.0154
24	18	2.0543	2.3648	2.7892	3.1246	3.4746	3.9631
24	19	2.0399	2.3452	2.7624	3.0920	3.4360	3.9160
24	20	2.0267	2.3273	2.7380	3.0624	3.4009	3.8732
24	21	2.0146	2.3109	2.7157	3.0354	3.3689	3.8342
24	22	2.0035	2.2959	2.6953	3.0106	3.3396	3.7985
24	23	1.9932	2.2821	2.6765	2.9878	3.3126	3.7657
24	24	1.9838	2.2693	2.6591	2.9667	3.2877	3.7354
24	25	1.9750	2.2574	2.6430	2.9472	3.2647	3.7073
24	26	1.9668	2.2464	2.6280	2.9291	3.2433	3.6813
24	27	1.9591	2.2361	2.6140	2.9123	3.2233	3.6571
24	28	1.9520	2.2265	2.6010	2.8965	3.2047	3.6345
24	29	1.9453	2.2174	2.5888	2.8818	3.1873	3.6133
24	30	1.9390	2.2090	2.5773	2.8679	3.1710	3.5935
25	10	2.2365	2.6135	3.1294	3.5370	3.9623	4.5551
25	11	2.1979	2.5603	3.0558	3.4470	3.8549	4.4233
25	12	2.1649	2.5149	2.9931	3.3704	3.7637	4.3116

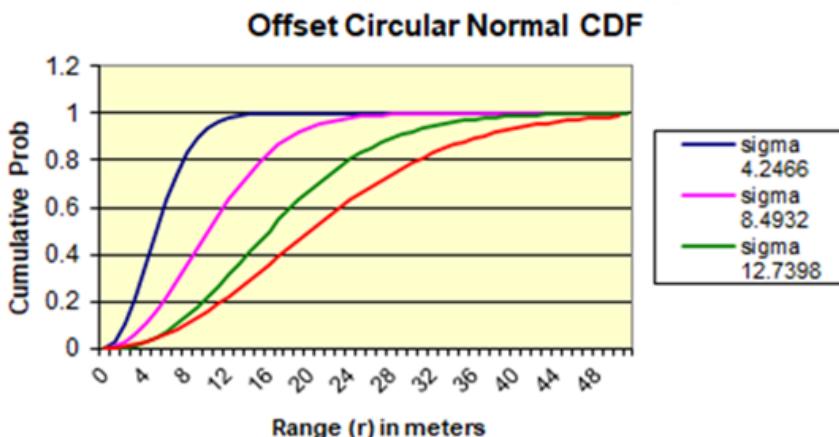
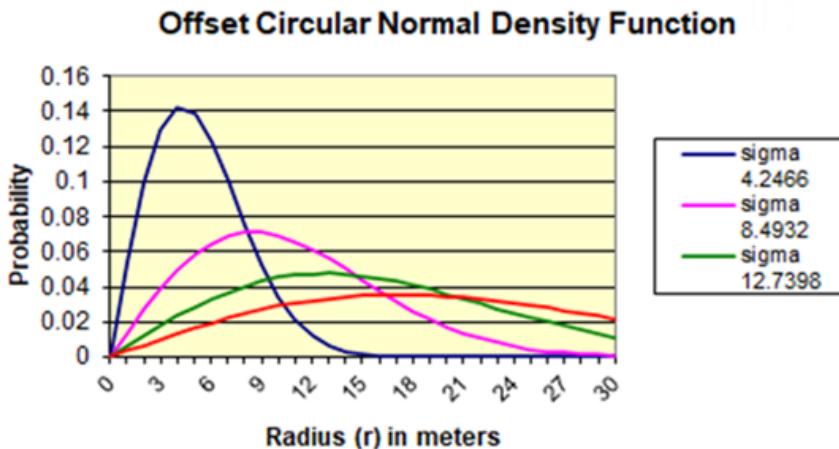
		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
25	13	2.1362	2.4756	2.9389	3.3044	3.6852	4.2155
25	14	2.1111	2.4413	2.8917	3.2469	3.6169	4.1320
25	15	2.0889	2.4110	2.8502	3.1963	3.5569	4.0587
25	16	2.0691	2.3840	2.8133	3.1515	3.5037	3.9938
25	17	2.0513	2.3599	2.7803	3.1114	3.4562	3.9359
25	18	2.0353	2.3381	2.7506	3.0754	3.4136	3.8839
25	19	2.0207	2.3184	2.7238	3.0429	3.3750	3.8370
25	20	2.0075	2.3005	2.6993	3.0133	3.3400	3.7944
25	21	1.9953	2.2840	2.6770	2.9862	3.3080	3.7555
25	22	1.9842	2.2690	2.6565	2.9615	3.2788	3.7199
25	23	1.9738	2.2551	2.6377	2.9387	3.2518	3.6872
25	24	1.9643	2.2422	2.6203	2.9176	3.2270	3.6570
25	25	1.9554	2.2303	2.6041	2.8981	3.2039	3.6291
25	26	1.9472	2.2192	2.5891	2.8800	3.1825	3.6031
25	27	1.9395	2.2089	2.5751	2.8631	3.1626	3.5789
25	28	1.9323	2.1992	2.5620	2.8473	3.1440	3.5564
25	29	1.9255	2.1901	2.5498	2.8326	3.1266	3.5353
25	30	1.9192	2.1816	2.5383	2.8187	3.1103	3.5155
26	10	2.2197	2.5896	3.0941	3.4916	3.9052	4.4801
26	11	2.1811	2.5363	3.0205	3.4017	3.7981	4.3490
26	12	2.1479	2.4908	2.9578	3.3252	3.7072	4.2378
26	13	2.1192	2.4515	2.9038	3.2594	3.6290	4.1422
26	14	2.0939	2.4171	2.8566	3.2020	3.5609	4.0591
26	15	2.0716	2.3867	2.8150	3.1515	3.5010	3.9861
26	16	2.0518	2.3597	2.7781	3.1067	3.4479	3.9215
26	17	2.0339	2.3355	2.7451	3.0666	3.4006	3.8638
26	18	2.0178	2.3137	2.7153	3.0306	3.3580	3.8120
26	19	2.0032	2.2939	2.6885	2.9981	3.3195	3.7653
26	20	1.9898	2.2759	2.6640	2.9685	3.2846	3.7228
26	21	1.9776	2.2594	2.6416	2.9415	3.2527	3.6841
26	22	1.9664	2.2443	2.6211	2.9167	3.2234	3.6486
26	23	1.9560	2.2303	2.6022	2.8939	3.1965	3.6160

		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
26	24	1.9464	2.2174	2.5848	2.8728	3.1717	3.5859
26	25	1.9375	2.2054	2.5686	2.8533	3.1487	3.5580
26	26	1.9292	2.1943	2.5536	2.8352	3.1273	3.5321
26	27	1.9215	2.1839	2.5395	2.8183	3.1074	3.5080
26	28	1.9142	2.1742	2.5264	2.8025	3.0888	3.4856
26	29	1.9074	2.1651	2.5141	2.7877	3.0714	3.4645
26	30	1.9010	2.1565	2.5026	2.7738	3.0551	3.4448
27	10	2.2043	2.5676	3.0618	3.4499	3.8530	4.4117
27	11	2.1655	2.5143	2.9882	3.3602	3.7463	4.2812
27	12	2.1323	2.4688	2.9256	3.2839	3.6557	4.1706
27	13	2.1035	2.4293	2.8715	3.2182	3.5776	4.0754
27	14	2.0781	2.3949	2.8243	3.1608	3.5097	3.9926
27	15	2.0558	2.3644	2.7827	3.1104	3.4500	3.9200
27	16	2.0358	2.3373	2.7458	3.0656	3.3970	3.8556
27	17	2.0179	2.3131	2.7127	3.0256	3.3497	3.7981
27	18	2.0017	2.2912	2.6830	2.9896	3.3073	3.7466
27	19	1.9870	2.2713	2.6561	2.9571	3.2688	3.7000
27	20	1.9736	2.2533	2.6316	2.9275	3.2339	3.6576
27	21	1.9613	2.2367	2.6092	2.9005	3.2021	3.6190
27	22	1.9500	2.2216	2.5887	2.8757	3.1729	3.5837
27	23	1.9396	2.2076	2.5697	2.8529	3.1460	3.5511
27	24	1.9299	2.1946	2.5522	2.8318	3.1212	3.5211
27	25	1.9210	2.1826	2.5360	2.8123	3.0982	3.4933
27	26	1.9126	2.1714	2.5209	2.7941	3.0768	3.4675
27	27	1.9048	2.1609	2.5069	2.7772	3.0569	3.4434
27	28	1.8975	2.1512	2.4937	2.7614	3.0383	3.4210
27	29	1.8907	2.1420	2.4814	2.7466	3.0209	3.4000
27	30	1.8842	2.1334	2.4699	2.7327	3.0046	3.3803
28	10	2.1900	2.5473	3.0320	3.4117	3.8051	4.3491
28	11	2.1512	2.4940	2.9585	3.3222	3.6988	4.2193
28	12	2.1179	2.4484	2.8959	3.2460	3.6084	4.1091
28	13	2.0889	2.4089	2.8418	3.1803	3.5306	4.0143

		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
28	14	2.0635	2.3743	2.7946	3.1231	3.4628	3.9319
28	15	2.0411	2.3438	2.7530	3.0727	3.4032	3.8595
28	16	2.0210	2.3167	2.7160	3.0279	3.3503	3.7953
28	17	2.0030	2.2924	2.6830	2.9879	3.3031	3.7381
28	18	1.9868	2.2704	2.6532	2.9520	3.2607	3.6867
28	19	1.9720	2.2505	2.6263	2.9194	3.2224	3.6402
28	20	1.9586	2.2324	2.6017	2.8899	3.1875	3.5980
28	21	1.9462	2.2158	2.5793	2.8628	3.1557	3.5595
28	22	1.9349	2.2006	2.5587	2.8380	3.1265	3.5242
28	23	1.9244	2.1865	2.5398	2.8152	3.0996	3.4918
28	24	1.9147	2.1735	2.5223	2.7941	3.0748	3.4618
28	25	1.9057	2.1615	2.5060	2.7746	3.0519	3.4341
28	26	1.8973	2.1502	2.4909	2.7564	3.0305	3.4083
28	27	1.8894	2.1397	2.4768	2.7395	3.0106	3.3843
28	28	1.8821	2.1299	2.4636	2.7236	2.9920	3.3619
28	29	1.8752	2.1207	2.4513	2.7088	2.9746	3.3409
28	30	1.8687	2.1121	2.4397	2.6949	2.9583	3.3213
29	10	2.1768	2.5286	3.0045	3.3765	3.7611	4.2917
29	11	2.1379	2.4752	2.9311	3.2871	3.6550	4.1624
29	12	2.1045	2.4295	2.8685	3.2110	3.5649	4.0526
29	13	2.0755	2.3900	2.8144	3.1454	3.4872	3.9582
29	14	2.0500	2.3554	2.7672	3.0882	3.4196	3.8761
29	15	2.0275	2.3248	2.7256	3.0379	3.3601	3.8039
29	16	2.0073	2.2976	2.6886	2.9932	3.3074	3.7400
29	17	1.9893	2.2732	2.6555	2.9532	3.2603	3.6829
29	18	1.9730	2.2512	2.6257	2.9173	3.2179	3.6316
29	19	1.9581	2.2313	2.5987	2.8847	3.1796	3.5853
29	20	1.9446	2.2131	2.5742	2.8551	3.1448	3.5432
29	21	1.9322	2.1965	2.5517	2.8281	3.1130	3.5048
29	22	1.9208	2.1812	2.5311	2.8033	3.0838	3.4696
29	23	1.9103	2.1671	2.5121	2.7805	3.0570	3.4373
29	24	1.9005	2.1540	2.4946	2.7594	3.0322	3.4074

		F Critical Values, $F_{1-\alpha, df_1, df_2}$					
		$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	$\alpha = 0.0025$	$\alpha = 0.001$
29	25	1.8915	2.1419	2.4783	2.7398	3.0092	3.3797
29	26	1.8830	2.1306	2.4631	2.7216	2.9879	3.3540
29	27	1.8751	2.1201	2.4490	2.7047	2.9680	3.3300
29	28	1.8677	2.1102	2.4358	2.6888	2.9494	3.3076
29	29	1.8608	2.1010	2.4234	2.6740	2.9320	3.2867
29	30	1.8543	2.0923	2.4118	2.6600	2.9157	3.2671
30	10	2.1646	2.5112	2.9791	3.3440	3.7205	4.2388
30	11	2.1256	2.4577	2.9057	3.2547	3.6147	4.1100
30	12	2.0921	2.4120	2.8431	3.1787	3.5247	4.0006
30	13	2.0630	2.3724	2.7890	3.1132	3.4473	3.9065
30	14	2.0374	2.3378	2.7418	3.0560	3.3798	3.8247
30	15	2.0148	2.3072	2.7002	3.0057	3.3204	3.7527
30	16	1.9946	2.2799	2.6632	2.9611	3.2677	3.6890
30	17	1.9765	2.2554	2.6301	2.9211	3.2207	3.6321
30	18	1.9601	2.2334	2.6003	2.8852	3.1784	3.5810
30	19	1.9452	2.2134	2.5732	2.8526	3.1401	3.5348
30	20	1.9317	2.1952	2.5487	2.8230	3.1053	3.4928
30	21	1.9192	2.1785	2.5262	2.7960	3.0735	3.4545
30	22	1.9077	2.1631	2.5055	2.7712	3.0444	3.4194
30	23	1.8972	2.1490	2.4865	2.7483	3.0176	3.3870
30	24	1.8874	2.1359	2.4689	2.7272	2.9928	3.3572
30	25	1.8782	2.1237	2.4526	2.7076	2.9698	3.3296
30	26	1.8698	2.1124	2.4374	2.6894	2.9485	3.3039
30	27	1.8618	2.1018	2.4233	2.6725	2.9286	3.2800
30	28	1.8544	2.0919	2.4100	2.6566	2.9100	3.2576
30	29	1.8474	2.0827	2.3976	2.6417	2.8926	3.2367
30	30	1.8409	2.0739	2.3860	2.6278	2.8763	3.2171

## Offset Circular Normal Density Function



**Offset Circular Normal Tables**

r	CEP 5		CEP 10		CEP 15	
	sigma	$f_R(r,s)$	sigma	$f_R(r,s)$	sigma	$f_R(r,s)$
0	4.2466	0	8.4932	0	12.7398	0
1	0.05394	0.02735	0.01377	0.00691	0.00614	0.00308
2	0.09926	0.10497	0.02697	0.02735	0.01217	0.01225
3	0.12962	0.22084	0.03907	0.06048	0.01798	0.02735
4	0.14234	0.35829	0.04963	0.10497	0.02346	0.04810
5	0.13863	0.50000	0.05829	0.15910	0.02852	0.07413
6	0.12263	0.63143	0.06481	0.22084	0.03309	0.10497
7	0.09977	0.74297	0.06910	0.28797	0.03709	0.14011

	CEP	5	CEP	10	CEP	15
	sigma	4.2466	sigma	8.4932	sigma	12.7398
r	fR(r,s)	P(R<=r)	fR(r,s)	P(R<=r)	fR(r,s)	P(R<=r)
8	0.07523	0.83042	0.07117	0.35829	0.04047	0.17894
9	0.05282	0.89416	0.07116	0.42962	0.04321	0.22084
10	0.03466	0.93750	0.06931	0.50000	0.04528	0.26513
11	0.02130	0.96508	0.06592	0.56773	0.04669	0.31117
12	0.01228	0.98155	0.06131	0.63143	0.04745	0.35829
13	0.00665	0.99077	0.05585	0.69007	0.04759	0.40585
14	0.00339	0.99564	0.04988	0.74297	0.04716	0.45327
15	0.00162	0.99805	0.04371	0.78978	0.04621	0.50000
16	0.00073	0.99917	0.03761	0.83042	0.04480	0.54554
17	0.00031	0.99967	0.03179	0.86510	0.04300	0.58947
18	0.00013	0.99987	0.02641	0.89416	0.04088	0.63143
19	0.00005	0.99996	0.02157	0.91810	0.03850	0.67114
20	0.00002	0.99998	0.01733	0.93750	0.03594	0.70837
21	0.00001	1.00000	0.01369	0.95296	0.03326	0.74297
22	0.00000	1.00000	0.01065	0.96508	0.03052	0.77486
23	0.00000	1.00000	0.00815	0.97444	0.02777	0.80400
24	0.00000	1.00000	0.00614	0.98155	0.02508	0.83042
25	0.00000	1.00000	0.00455	0.98686	0.02246	0.85418
26	0.00000	1.00000	0.00333	0.99077	0.01996	0.87538
27	0.00000	1.00000	0.00239	0.99361	0.01761	0.89416
28	0.00000	1.00000	0.00169	0.99564	0.01541	0.91065
29	0.00000	1.00000	0.00118	0.99706	0.01339	0.92504
30	0.00000	1.00000	0.00081	0.99805	0.01155	0.93750
31	0.00000	1.00000	0.00055	0.99872	0.00989	0.94821
32	0.00000	1.00000	0.00037	0.99917	0.00841	0.95734
33	0.00000	1.00000	0.00024	0.99947	0.00710	0.96508
34	0.00000	1.00000	0.00016	0.99967	0.00595	0.97160
35	0.00000	1.00000	0.00010	0.99979	0.00495	0.97704
36	0.00000	1.00000	0.00006	0.99987	0.00409	0.98155
37	0.00000	1.00000	0.00004	0.99992	0.00336	0.98526
38	0.00000	1.00000	0.00002	0.99996	0.00274	0.98830
39	0.00000	1.00000	0.00001	0.99997	0.00222	0.99077
40	0.00000	1.00000	0.00001	0.99998	0.00178	0.99277
41	0.00000	1.00000	0.00000	0.99999	0.00142	0.99436
42	0.00000	1.00000	0.00000	1.00000	0.00113	0.99564
43	0.00000	1.00000	0.00000	1.00000	0.00089	0.99664
44	0.00000	1.00000	0.00000	1.00000	0.00070	0.99743
45	0.00000	1.00000	0.00000	1.00000	0.00054	0.99805
46	0.00000	1.00000	0.00000	1.00000	0.00042	0.99852

47	0.00000	1.00000	0.00000	1.00000	0.00032	0.99889
48	0.00000	1.00000	0.00000	1.00000	0.00024	0.99917
49	0.00000	1.00000	0.00000	1.00000	0.00019	0.99939
50	0.00000	1.00000	0.00000	1.00000	0.00014	0.99955





# References

- Amerise, I. L., & Tarsitano, A. (2015). Correction methods for ties in rank correlations. *Journal of Applied Statistics*, 42(12), 2584-2596. doi:10.1080/02664763.2015.1043870
- Bhandari, M., & Joensson, A. (2008). *Clinical Research for Surgeons*. Thieme.
- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York: Springer. Retrieved from <http://www.hua.edu.vn/khoa/fita/wp-content/uploads/2013/08/Pattern-Recognition-and-Machine-Learning-Christophe-M-Bishop.pdf>
- Conroy, R. (2012). What Hypotheses do "Nonparametric" Two-Group Tests Actually Test? *Stata Journal*, 12(2), 182–190. doi:10.1177/1536867X1201200202
- Glen, S. (2022, May 10). *Scales of Measurement / Level of Measurement*. Retrieved from StatisticsHowTo.com: Elementary Statistics for the rest of us! : <https://www.statisticshowto.com/probability-and-statistics/descriptive-statistics/scales-of-measurement/>
- Global. (2022, April 3). *Comparative Studies Definition and Statistical Analysis*. Retrieved from Global Stats Academic: <https://www.en.globalstatistik.com/comparative-studies-definition-and-statistical-analysis/>
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). New York: Wiley-Interscience Publication.
- Nelson, S. L., & Nelson, E. C. (2021). How to use t-distributions in Excel. *www.dummies*. Retrieved from <https://www.dummies.com/software/microsoft-office/excel/how-to-use-t-distributions-in-excel/>
- Snedecor, G. W., & Cochran, W. G. (1989). *Statistical Methods* (Eighth ed.). Ames, Iowa: Iowa State University Press.

Strickland, J. (2017). *Logistic Regression Inside-Out*. Lulu, Inc. Retrieved from [https://www.lulu.com/spotlight/strickland\\_jeffrey](https://www.lulu.com/spotlight/strickland_jeffrey)

# Index

## A

- absolute value ..... 42, 44, 45, 63, 66  
accuracy ..... 23, 111  
alternative hypothesis ..... 32, 33, 34, 35, 45, 52, 58, 63, 79  
ANOVA ..... 56, 112, 113, 138, 139, 140

## B

- Bernoulli ..... 225  
binary ..... 224, 225  
binary predictor ..... 224  
binomial ..... 224, 226  
binomial distribution ..... 48  
**block** ..... 67, 73, 75, 111, 112  
blocking ..... 67, 111  
blood pressure ..... 53  
boxplot ..... 70, 82, 84, 86

## C

- case ..... 225, 226, 227, 228  
    non-case ..... 227  
categorical distributions ..... 224, 225, 226  
Central Limit Theorem ..... 15, 19, 32  
classification ..... xi  
confidence coefficient ..... 20, 21  
confidence interval . 15, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 36, 42, 47, 49,  
    58, 61, 79, 80, 81  
    two-sided ..... 22, 23  
confidence level ..... 26, 27, 28, 30, 36, 48  
confidence limits ..... See confidence interval  
continuous variable ..... 226  
**critical region** ..... 22, 60  
critical value ..... 32, 33, 35, 57, 60, 62, 63, 64, 65, 66, 70, 88, 138  
cross-sectional study ..... 42

## D

- DaaS ..... 24  
data analytics ..... xi  
Data as a Service ..... See DaaS

dataset	
AUTO83B.DAT .....	61, 64
Globalstar.csv.....	89
ZARR13.DAT .....	23
decision rule .....	32
degrees of freedom.....	17, 18, 19, 20, 21, 54, 58, 60
dependent variable .....	224, 225, 226, 227, 228
target variable.....	225
Dunnett's critical value .....	138
Dunnett's test.....	138, 139, 141
<b>E</b>	
effect size .....	31, 35, 36, 42, 44, 45, 47
effectiveness .....	88
equivalence interval.....	79
error variance.....	24
Excel function	
ABS.....	43, 45, 141
AVERAGE.....	141
COUNT .....	59, 88, 140, 141
COUNTA .....	140
COUNTIF.....	82
DEVSQ.....	140, 141
F.DIST .....	140
F.INV.RT .....	140
IF 88, 141	
MIN .....	71, 84, 88
RANK.AVG .....	.87, 88
SQRT.....	.45, 141
SUM .....	.88, 140, 141
T.DIST .....	141
VLOOKUP .....	.88, 141
experimental design.....	23
randomized block design .....	111, 112
randomized complete block design .....	67
explanatory variables .....	227, 228
exponential distribution.....	226, 227, 228
<b>F</b>	
factors .....	21, 111, 113
F-statistic.....	.57, 62, 65

F-test .....	57, 90
<b>G</b>	
Goal Seek .....	50, 51, 52
<b>H</b>	
hypothesis test.....	15, 22, 24, 32, 36, 41, 43, 44, 45, 52, 53, 58, 62, 65
two-sided.....	32, 58
<b>I</b>	
independence .....	45, 79
independent variables .....	224, 225
explanatory variables .....	228
predictors .....	226
intercept .....	227
interval estimates .....	<i>See</i> confidence intervals
<b>L</b>	
latency.....	24, 25, 82, 83, 84, 86, 88
levels .....	30, 44
linear combination.....	227
link function .....	225, 228
logistic function.....	224, 226, 227, 228
logistic regression .....	xiii, 224, 225, 226
logit .....	224, 225, 226, 227, 228
logit function.....	227, 228
logit regression .....	<i>See</i> logistic regression
log-odds .....	225, 226, 228
<b>M</b>	
machine learning .....	x
margin of error.....	24, 25, 26, 27, 28, 30, 31, 47, 48
marketing.....	225
mean .....	26
estimated mean.....	24
population mean .....	22, 24, 41, 49, 52, 57, 59, 62, 63, 65, 66, 79, 80
sample mean .....	21, 24, 32, 33, 34, 35, 56, 57
true mean .....	20, 21, 33, 34, 35
mean square between .....	138
missile .....	46, 47
model	

parameters .....	224, 229
Multinomial .....	224
multinomial logistic regression .....	224

## N

natural logarithm .....	225, 226, 227, 228
noise .....	<i>See variance</i>
non-parametric test .....	82, 85
normal distribution .....	15, 18, 24, 26, 27, 28, 30, 41, 44, 45, 82
normal probability plot .....	61, 64, 82, 83, 86
null hypothesis ..	22, 31, 32, 33, 34, 35, 36, 42, 44, 52, 53, 54, 56, 58, 59, 60, 62,
	63, 65, 66, 68, 70, 79, 81, 82, 83, 85, 88, 138, 139, 140

## O

observations.....	23, 53, 54
Odds Ratio.....	225
ordinal distribution .....	85
outcomes.....	224, 225
outliers .....	56

## P

parameter .....	24, 31, 35, 36
placebo.....	46
point estimate .....	24
population .....	15, 17, 19, 24, 26, 27, 28, 41, 42, 43, 49, 50, 57, 79, 85
posterior probability .....	224
power .....	35, 36, 41, 44, 52, 53
predictors .....	225, 226, 227, 228
probability .....	224, 225, 227, 228
probability density function .....	18, 19
<i>probit regression</i> .....	224
propagator .....	80, 81
proportion.....	21, 28, 29, 43, 44, 48
p-value.....	58, 79, 81, 83, 84, 113, 141

## Q

QQ plot.....	<i>See normal probability plot</i>
quartile .....	71

## R

random assignment .....	47
-------------------------	----

random sampling .....	26
random variables .....	18
regression coefficient.....	227
rejection region.....	63, 64
replication .....	24
residuals.....	24
response.....	22, 111
response variable.....	224

## S

sample size..	15, 18, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 36, 41, 42, 43, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 56, 83
sample standard deviation .....	<i>See</i> standard deviation
satellite .....	27, 42, 80, 111, 112
scenario.....	42, 52
sensitivity .....	23, 24
shape.....	18, 56, 86
sign test.....	82
signal-to-noise-ratio.....	53, 57
significance level .....	21, 22, 31, 36, 41, 43, 44, 45, 46, 51, 60, 63, 66
standard deviation..	15, 17, 18, 19, 21, 24, 26, 27, 30, 32, 42, 45, 46, 47, 49, 53, 54, 56, 57, 76
pooled standard deviation .....	45, 48, 57
sample standard deviation .....	19, 21, 26
standard error.....	19
standard normal distribution.....	18, 41, 44, 45
statistical control .....	22
statistical difference.....	81, 88, 113
statistical inference .....	26, 56
statistical power.....	31, 49
statistical significance .....	70, 141
Student t test .....	<i>See</i> t-test
sum of squares between.....	138
summary statistics .....	56, 61, 64

## T

t-distribution .....	15, 17, 18, 19, 20, 21, 24, 57, 58
test statistic.....	18, 32, 57, 58, 63, 66, 70, 88
TOST .....	79
transformation.....	225, 228
treatment.....	46, 59, 113, 138

trial .....	46, 83
t-statistic .....	57, 58
<b>t-test</b> 22, 53, 54, 56, 57, 58, 59, 60, 63, 64, 65, 67, 68, 79, 80, 81, 82, 85, 86, 92, 140, 141	
one-sample .....	49, 63
paired.....	54, 67
two-sample .....	53, 56, 59
two-sided .....	79
two-sample equivalence test .....	79
two-sided test .....	33, 42, 44, 46
Type I error.....	31, 33, 52
Type II error.....	31, 35
<b>U</b>	
uncertainty .....	20
<b>V</b>	
variance .....	xi, 54, 56, 57, 58, 59, 60, 61, 62, 63, 65
<b>W</b>	
<i>W critical value</i> .....	88
<i>W statistic</i> .....	88
Wilcoxon Rank-Sum test .....	85, 87, 88
<b>Z</b>	
z-statistic .....	57



