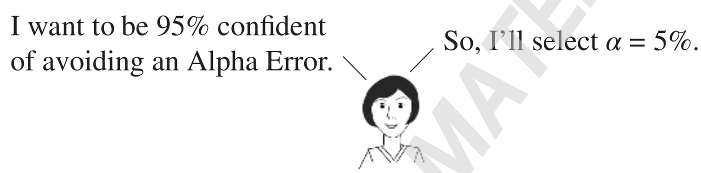
JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

ALPHA, α

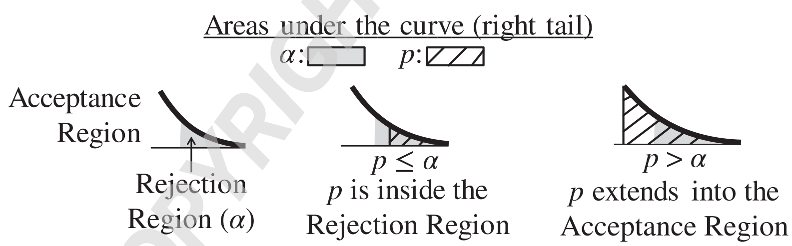
Summary of Keys to Understanding

1. In Inferential Statistics, *p* is the Probability of an Alpha (“False Positive”) **Error.**

2. Alpha is the highest value of *p* that we are willing to tolerate and still say that a difference, change, or effect observed in the Sample is “Statistically Significant.”



3. Alpha is a Cumulative Probability, represented as an area under the curve, at one or both tails of a Probability Distribution. *p* is also a Cumulative Probability.



4. In Hypothesis Testing, if $p \leq \alpha$, Reject the Null Hypothesis. If $p > \alpha$, Accept (Fail to Reject) the Null Hypothesis.

5. Alpha defines the Critical Value(s) of Test Statistics, such as *z*, *t*, *F*, or Chi-Square. The Critical Value or Values, in turn, define the Confidence Interval.

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

Explanation

1. In Inferential Statistics, *p* is the Probability of an Alpha ("False Positive") Error.

In Inferential Statistics, we use data from a Sample to estimate a property (say, the Mean) of the Population or Process from which the Sample was taken. **Being an estimate, there is a risk of error.**

One type of error is the Alpha Error (also known as "Type I Error" or "False Positive").

An Alpha Error is the error of seeing something which is not there, that is, **concluding that there is a Statistically Significant difference, change, or effect, when in fact there is not.** For example,

- Erroneously concluding that there is a difference in the Means of two Populations, when there is not, or
- Erroneously concluding that there has been a change in the Standard Deviation of a Process, when there has not, or
- Erroneously concluding that a medical treatment has an effect, when it does not.

In Hypothesis Testing, the Null Hypothesis states that there is no difference, change, or effect. All these are examples of **Rejecting the Null Hypothesis when the Null Hypothesis is true.**

***p* is the Probability of an Alpha Error, a "False Positive."**

It is calculated as part of the Inferential Statistical analysis, for example, in a *t*-test or ANOVA.

How does an Alpha Error happen? **An Alpha Error occurs when data in our Sample are not representative of the overall Population or Process from which the Sample was taken.**

If the Sample Size is large enough, the great majority of Samples of that size will do a good job of representing the Population or Process. However, some won't. ***p* tells us how probable it is that our Sample is unrepresentative enough to produce an Alpha Error.**

2. Alpha is the highest value of *p* that we are willing to tolerate and still say that a difference, change, or effect observed in the Sample is "Statistically Significant."

In this article, we use Alpha both as an adjective and as a noun. This might cause some confusion, so let's explain.

"Alpha," as an adjective, describes a type of error, the Alpha Error. Alpha as a noun is something related, but different.

First of all, what it is not: **Alpha**, as a noun, **is not**

- a Statistic or a Parameter, which describes a property (e.g., the Mean) of a Sample or Population
- a Constant, like those shown in some statistical tables.

Second, what it is: Alpha, as a noun, **is**

- **a value of *p* which defines the boundary of the values of *p* which we are willing to tolerate from those which we are not.**

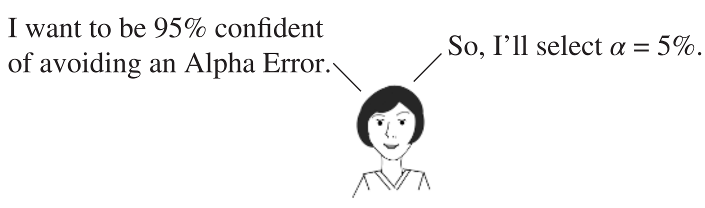
For example, if we are willing to tolerate a 5% risk of a False Positive, then we would select $\alpha = 5\%$. That would mean that we are willing to tolerate $p \leq 5\%$, but not $p > 5\%$.

Alpha must be selected prior to collecting the Sample data. This is to help ensure the integrity of the test or experiment. If we have a look at the data first, that might influence our selection of a value for Alpha.

Rather than starting with Alpha, it's probably more natural to think in terms of a Level of Confidence first. Then we subtract it from 1 (100%) to get Alpha.

If we want to be 95% sure, then we want a 95% Level of Confidence (aka "Confidence Level").

By definition, $\boldsymbol{\alpha}$ **= 100% – Confidence Level.** (And, so Confidence Level = 100% – α.)

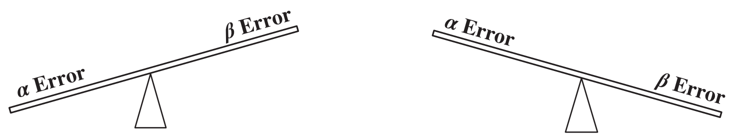


Alpha is called the "Level or Significance" or "Significance Level."

- **If *p* is calculated to be less than or equal to the Significance Level, $\boldsymbol{\alpha}$, then any observed difference, change, or effect calculated from our Sample data is said to be "Statistically Significant."**

- **If $p > \alpha$, then it is <u>not</u> Statistically Significant.**

Popular choices for Alpha are 10% (0.1), 5% (0.05), 1% (0.01), 0.5% (0.005), and 0.1% (0.001). But, why wouldn't we always select as low a level of Alpha as possible? Because, **the choice of Alpha is a tradeoff** between Alpha (Type I) Error and Beta (Type 2) Error – or put another way – between a False Positive and a False Negative. If you reduce the chance (Probability) of one, you increase the chance of the other.



Choosing $\alpha = 0.05$ (5%) is generally accepted as a good balance for most uses. The pros and cons of various choices for Alpha (and Beta) in different situations are covered in the article, *Alpha and Beta Errors.*

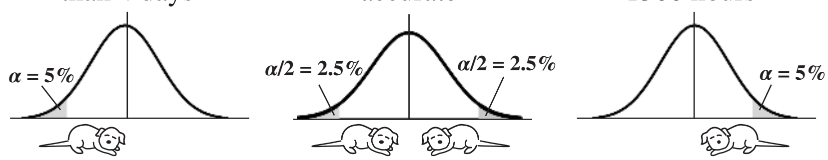
3. Alpha is a Cumulative Probability, represented by an area under the curve, at one or both tails of a Probability Distribution. p is also a Cumulative Probability.

Below are diagrams of the Standard Normal Distribution. The Variable on its horizontal axis is the Test Statistic, z. Any point on the curve is the Probability of the value of z directly below that point.

Probabilities of individual points are usually less useful in statistics than Probabilities of ranges of values. The latter are called Cumulative Probabilities. **The Cumulative Probability of a range of values is calculated as the area under the curve above that range of values.** The Cumulative Probability of all values under the curve is 100%.

We start by selecting a value for Alpha, most commonly 5%, which tells us how big the shaded area under the curve will be. **Depending on the type of problem we're trying to solve, we position the shaded area (α) under the left tail, the right tail, or both tails.**

left-tailed: our orders ship in less than 4 days	2-tailed: our estimate of the Mean is accurate	right-tailed: our light bulbs last longer than 1300 hours



If it's one tail only, the analysis is called "1-tailed" or "1-sided" (or "left-tailed or "right-tailed"), and Alpha is entirely under one side of the curve. If it's both tails, it's called a "2-tailed" or **"2-sided" analysis. In that case, we divide Alpha by two, and put half under each tail.** For more on tails, see the article *Alternative Hypothesis*.

There are two main methods in Inferential Statistics – Hypothesis Testing and Confidence Intervals. Alpha plays a key role in both. First, let's take a look at Hypothesis Testing:

4. In Hypothesis Testing, if $p \leq \alpha$, Reject the Null Hypothesis. If $p > \alpha$, Accept (Fail to Reject) the Null Hypothesis.

In Hypothesis testing, *p* is compared to Alpha, in order to determine what we can conclude from the test.

Hypothesis Testing starts with a **Null Hypothesis** – a statement that there is no (Statistically Significant) difference, change, or effect.

We select a value for Alpha (say 5%) and then collect a Sample of data. Next, a statistical test (like a *t*-test or *F*-test) is performed. The test output includes a value for *p*.

***p* is** the Probability of an Alpha Error, a False Positive, that is, **the Probability that any difference, effect, or change shown by the Sample data is not Statistically Significant.**

If *p* is small enough, then we can be confident that there really is a difference, change, or effect. How small is small enough? **Less than or equal to Alpha.** Remember, we picked Alpha as the upper boundary for the values of *p* which indicate a tolerable Probability of an Alpha Error. So, $p > \alpha$ is an unacceptably high Probability of an Alpha Error.

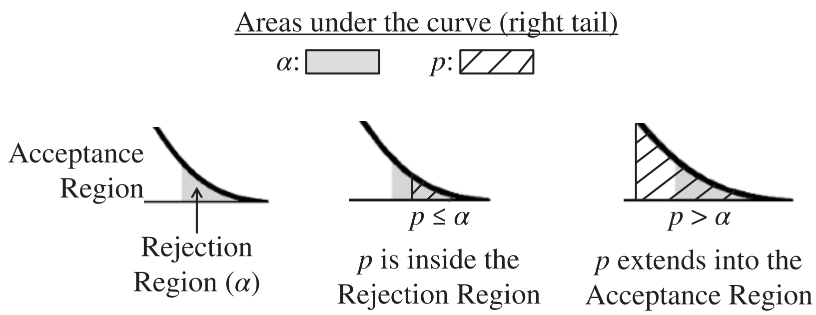
How confident can we be? As confident as the Level of Confidence. For example, with a 5% Alpha (Significance Level), we have a 100% – 5% = 95% Confidence Level. So, …

If $p \leq \alpha$, then we conclude that:

– the Probability of an Alpha Error is within the range we said we would tolerate, so **the observed difference, change, or effect we are testing is Statistically Significant.**
– **in a Hypothesis test, we would Reject the Null Hypothesis.**
– **the smaller the *p*-value, the stronger the evidence for this conclusion.**

How does this look graphically? Below are three close-ups of the right tail of a Distribution. This is for a 1-tailed test, in which the shaded area represents Alpha and the hatched areas represent p. (In a 2-tailed test, the left and right tails would each have $\alpha/2$ as the shaded areas.)

- Left graph below: in Hypothesis Testing, some use the term "Acceptance Region" or "Non-critical Region" for the unshaded white area under the Distribution curve, and "Rejection Region" or "Critical Region" for the shaded area representing Alpha.



- Center graph: **if** the hatched area representing p is entirely in the shaded Rejection Region (because $\boldsymbol{p \leq \alpha}$) **we Reject the Null Hypothesis.**
- Right graph: **If** p extends into the white Acceptance Region (because $\boldsymbol{p > \alpha}$), **we Accept** (or **"Fail to Reject"**) **the Null Hypothesis.**

For example, here is a portion of the output from an analysis which includes an F-test. $\alpha = 0.05$.

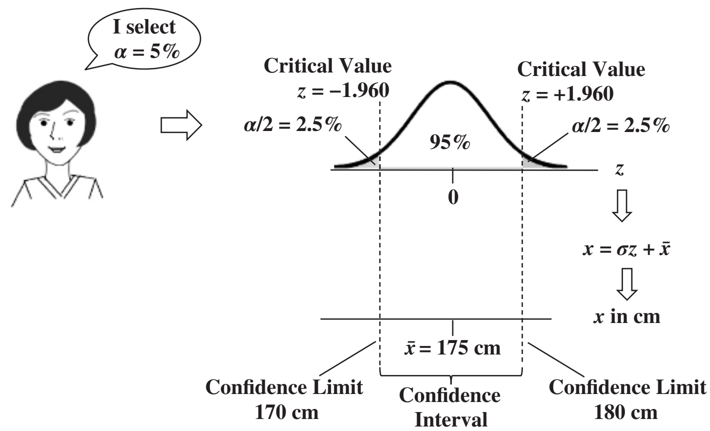
Factors	F	Effect Size	p
A: Detergent	729	6.75	0.02
B: Water Temp.	225	3.74	0.04
C: Washing Machine	49	1.75	0.09

- We see that $p < \alpha$ for both Factor A and Factor B. So, we can say that A and B do have a Statistically Significant effect. (We Reject the Null Hypothesis.)
- The p-value for A is considerably smaller than that for B, so the evidence is stronger that A has an effect.
- $p > \alpha$ for Factor C, so we conclude that C does not have a Statistically Significant effect. (We Accept/Fail to Reject the Null Hypothesis.)

5. Alpha defines the Critical Value(s) of Test Statistics, such as z, t, F, or Chi-Square. The Critical Value or Values, in turn, define the Confidence Interval.

We explained how Alpha plays a key role in the Hypothesis Testing method of Inferential Statistics. It is also an integral part of the other main method – Confidence Intervals. This is explained in detail in the article, *Confidence Intervals – Part 1*. It is also illustrated in the following concept flow diagram (follow the arrows):

Here's how it works. Let's say we want a Confidence Interval around the Mean height of males.



Top part of the diagram:

- The person performing the analysis selects a value for Alpha.
- Alpha – split into two halves – is shown as the shaded areas under the two tails of the curve of a Test Statistic, like z.
- Tables or calculations provide the values of the Test Statistic which form the boundaries of these shaded $\alpha/2$ areas. In this example, $z = -1.960$ and $z = +1.960$.
- These values are the Critical Values of the Test Statistic for $\alpha = 5\%$. They are in the units of the Test Statistic (z is in units of Standard Deviations).

Bottom part of the diagram:

- A Sample of data is collected and a Statistic (e.g., the Sample Mean, $\bar{x}$) is calculated (175 cm in this example).
- To make use of the Critical Values in the real world, we need to convert the Test Statistic Values into real-world values – like centimeters in the example above.

There are different conversion formulas for different Test Statistics and different tests. In this illustration, z is the Test Statistic and it is defined as $z = (x - \bar{x})/\sigma$. So $x = \sigma z + \bar{x}$. We multiply σ (the Population Standard

Deviation), by each critical value of z (−1.960 and +1.960), and we add those to the Sample Mean (175 cm).

- That converts the Critical Values −1.960 and +1.960 into the Confidence Limits of 170 and 180 cm.
- These Confidence Limits define the lower and upper boundaries of the Confidence Interval.

To further your understanding of how Alpha is used, it would be a good idea to next read the article *Alpha, p, Critical Value, and Test Statistic – How they Work Together.*

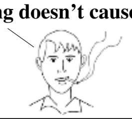
Related Articles in This Book: *Alpha and Beta Errors*; *p, p-Value*; *Statistically Significant*; *Alpha, p, Critical Value, and Test Statistic – How They Work Together*; *Test Statistic*; *p, t, and F: “>” or “<”?*; *Hypothesis Testing – Part 1: Overview*; *Critical Value*; *Confidence Intervals – Parts 1 and 2*; *z*

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

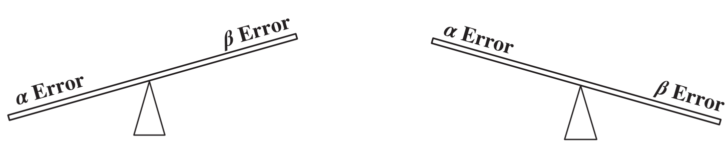
ALPHA AND BETA ERRORS

Summary of Keys to Understanding

1. There is a risk of an Alpha (aka Type I) **Error or a Beta** (aka Type II) **Error in any Inferential Statistical analysis.**

2.	**Alpha Error, "False Positive"**	**Beta Error, "False Negative"**
	I saw a unicorn. 	**Smoking doesn't cause cancer.** 
What it is	**The error of concluding that there is something** – a difference, or a change, or an effect – **when, in reality, there is not.**	**The error of concluding that there is nothing** – no difference, or no change, or no effect, **when, in reality, there is.**
In Hypothesis Testing	The error of Rejecting the Null Hypothesis when it is true.	The error of Failing to Reject the Null Hypothesis when it is false.
Found in:	Hypothesis Testing and Confidence Levels, *t*-tests, ANOVA, ANOM, etc.	

3. There is a tradeoff between Alpha and Beta Errors.



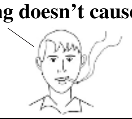
The subject being analyzed determines which type is more troublesome.

4. To reduce both Alpha and Beta Errors, increase the Sample Size.

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

Explanation

1. There is a risk of an Alpha (aka Type I) **Error or a Beta** (aka Type II) **Error in any Inferential Statistical analysis.**

2.	**Alpha Error (False Positive)**	**Beta Error (False Negative)**
	I saw a unicorn. 	**Smoking doesn't cause cancer.** 
What it is	**The error of concluding that there is something** – a difference, or a change, or an effect – **when, in reality, there is not.**	**The error of concluding that there is nothing** – no difference, or no change, or no effect – **when, in reality, there is.**
In Hypothesis Testing	The error of Rejecting the Null Hypothesis when it is true.	The error of Failing to Reject the Null Hypothesis when it is false.
Also known as	Type I Error, Error of the First Kind Colloquially: **False Positive**, False Alarm, Crying Wolf	Type II Error, Error of the Second Kind, **False Negative**
Found in:	Hypothesis Testing and Confidence Levels, *t*-tests, ANOVA, ANOM, etc.	
Example: in blood tests	Indicate a disease in a healthy person.	Fail to find a disease that exists.
Probability of the error	p	β (Beta)

In Descriptive Statistics, we have complete data on the entire universe we wish to observe. So we can just directly calculate various properties like the Mean or Standard Deviation.

On the other hand, in **Inferential Statistics** methods like Hypothesis Testing and Confidence Intervals, **we don't have the complete data**. The Population or Process is too big or it is always changing, so we can never be 100% sure about it. **We can collect a Sample of data and make an**

estimate from that. As a result, **there will always be a chance for error**. There are two types of this kind of Sampling Error; they are like mirror images of each other.

It may be easiest to think in terms of "False Positive" and "False Negative."

False Positive (Alpha Error) – **is the error of concluding that there is a difference, change, or effect, when, in reality there is no difference, change, or effect.**

"False Negative" is the opposite – **the error of concluding there is nothing happening, when, in fact, something is.** For example, the statistical analysis of a Process Mean concluded that it has not changed over time, when, in reality the Process Mean has "drifted."

In this context **"positive" does not mean "beneficial," and "negative" does not mean "undesirable."** In fact, for medical diagnostic tests, a "positive" result indicates that a disease was found. And a "negative" result is no disease found.

Alpha, α (*see the article by that name*) **is selected by the tester as the maximum Probability of an Alpha** (aka Type 1 aka False Positive) **Error they will accept** and still be able to call the results "Statistically Significant." That's why Alpha is called the "Significance Level" or "Level of Significance."

Beta, β, is the Probability of a Beta Error. Unlike Alpha, which is selected by us, Beta is calculated by the analysis. $1 - \beta$ is the Probability of there not being a Beta Error. So, if we call Beta the Probability of a False Negative, we might think of $1 - \beta$ as the Probability of a "true negative." $1 - \beta$ is called the "Power" of the test, and it is used in Design of Experiments to determine the required Sample Size.

You may have noticed a lack of symmetry in the terminology. This can be confusing; hopefully the following table will help:

p is the Probability of an Alpha Error	β is the Probability of a Beta Error
α is the maximum acceptable Probability for an Alpha Error	
$1 - \alpha$ is called the Confidence Level	$1 - \beta$ is called the Power of the test

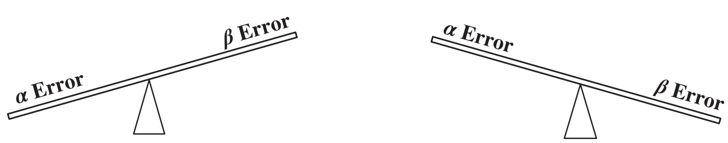
In Hypothesis Testing

Let's say we're testing the effect of a new medicine compared to a placebo. **The Null Hypothesis (H_0) says that there is no difference** between the new medicine and the placebo.

- **If the reality is that there is no difference (H_0 is true), and if …**
  - our testing concludes that there is no difference, then there is no error.
  - **our testing concludes that there is a difference, then there is an Alpha Error.**
- **If the reality is that there is a difference (H_0 is false), and if …**
  - **our testing concludes that there is no difference, then there is a Beta Error**
  - our testing concludes that there is a difference, then there is no error.

Conclusion from our testing	Reality: No difference, H_0 is True	Reality: There is a difference, H_0 is False
Accept (Fail to Reject) H_0	No error	Beta Error
Reject H_0	Alpha Error	No error

3. There is a tradeoff between Alpha and Beta Errors.



This makes sense. Consider the situation of airport security scanning. We want to detect metal weapons. We don't adjust the scanner to detect only metallic objects which are the size of an average gun or knife or larger. That would **reduce the risk of Alpha Errors** (e.g., identifying coins as possible weapons), but it **would increase the risk of Beta Errors** (not detecting small guns and knives).

This is the reason why we don't select an Alpha (maximum tolerable Probability of an Alpha Error) which is much smaller than the usual 0.05. There is a price to pay for making α extremely small. And the price is making the Probability of a Beta Error larger.

So, we need to select a value for Alpha which balances the need to avoid both types of error. The consensus seems to be that 0.05 is good for most uses.

How to make the tradeoff between Alpha and Beta depends on the situation being analyzed. In some cases, the effect of an Alpha Error is

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

relatively benign and you don't want to risk a False Negative. In other cases, the opposite is true. Some examples:

Situation	Consequence of an Alpha Error (False Positive)	Consequence of a Beta Error (False Negative)	Wise choice for level of risk	
			Alpha Error (risk of False Positive)	Beta Error (risk of False Negative)
Airport Security	Detain an innocent person as a terrorist	Let a terrorist on board	higher	lower
Inspect critical components for jet engine	Reject a good component	Engine failure	higher	lower
Inspect painting on the underside of a wheelbarrow	Cost of a reject	Customer will probably not notice or care	lower	higher

4. To reduce both Alpha and Beta Errors, increase the Sample Size.

There are other factors involved, but increasing the Sample Size will reduce both Alpha and Beta Errors. If the Sample Size is relatively large, increasing it further will yield diminishing returns in error reduction. (See the articles on Sample Size.)

Related Articles in This Book: *Alpha, α*; *Alpha, p-Value, Critical Value, and Test Statistic – How They Work Together*; *p, p-Value*; *Inferential Statistics*; *Power*; *Sample Size – Parts 1 and 2*

ALPHA, *p*, CRITICAL VALUE, AND TEST STATISTIC – HOW THEY WORK TOGETHER

Summary of Keys to Understanding

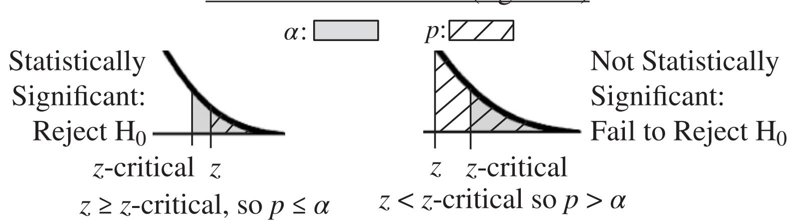
1. **Alpha and *p* are Cumulative Probabilities.** They are represented as **areas under the curve** of the Test Statistic Distribution.

2. **The Critical Value** (e.g., z-critical) **and the value of the Test Statistic** (e.g., z) **are point values** on the horizontal axis of the Test Statistic Distribution. **They mark the inner boundaries of the areas representing Alpha and *p*, respectively.**

3. **The person performing the analysis selects the value of Alpha, α.**
   Alpha and the Distribution are then used to calculate the Critical Value of the Test Statistic (e.g., z-critical). It is the value which forms the inner boundary of Alpha.

4. **Sample data are used to calculate the value of the Test Statistic** (e.g., z).
   The value of the Test Statistic and the Distribution are then used to calculate the value of *p*. p is the area under the curve outward from this calculated value of the Test Statistic.

Areas under the curve (right tail)



5. To determine Statistical Significance, compare p to Alpha, or (equivalently) compare the value of the Test Statistic to its Critical value.

   If $p \leq \alpha$ or (same thing) **$z \geq z$-critical**,

   then there is a Statistically Significant difference, change, or effect. Reject the Null Hypothesis, H_0.

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

Explanation

Much of statistics involves taking a Sample of data and using it to infer something about the Population or Process from which the Sample was collected. This is called Inferential Statistics.

There are 4 key concepts at the heart of Inferential Statistics:

- Alpha, the Level of Significance
- p, the Probability of an Alpha (False Positive) Error
- a Test Statistic, such as z, t, F, or χ^2 (and its associated Distribution)
- Critical Value, the value of the Test Statistic corresponding to Alpha

This article describes how these 4 concepts work together in Inferential Statistics. It assumes you are familiar with the individual concepts. If you are not, it's easy enough to get familiar with them by reading the individual articles for each of them.

<table>
<tr><th></th><th>Alpha, α</th><th>p</th><th></th><th>Critical Value of Test Statistic</th><th>Test Statistic Value</th></tr>
<tr><td rowspan="2">What is it?
How is it pictured?</td><td colspan="2">a Cumulative Probability</td><td></td><td colspan="2">a value of the Test Statistic</td></tr>
<tr><td colspan="2">an area under the curve of the Distribution of the Test Statistic</td><td></td><td colspan="2">a point on the horizontal axis of the Distribution of the Test Statistic</td></tr>
<tr><td>Boundary</td><td>Critical Value marks its boundary</td><td>Test Statistic Value marks its boundary</td><td></td><td>Forms the boundary for Alpha</td><td>Forms the boundary for p</td></tr>
<tr><td>How is its value determined?</td><td>Selected by the tester</td><td>area bounded by the Test Statistic value</td><td></td><td>boundary of the Alpha area</td><td>calculated from Sample Data</td></tr>
<tr><td>Compared with</td><td>p</td><td>α</td><td></td><td>Test Statistic Value</td><td>Critical Value of Test Statistic</td></tr>
<tr><td>Statistically Significant/ Reject the Null Hypothesis if</td><td colspan="2">$p \leq \alpha$</td><td></td><td colspan="2">Test Statistic $\geq$ Critical Value
e.g., $z \geq z$-critical</td></tr>
</table>

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

The preceding compare-and-contrast table is a visual summary of the 5 Keys to Understanding from the previous page and the interrelationships among the 4 concepts. This article will cover its content in detail. At the end of the article is a concept flow visual which explains the same things as this table, but using a different format. Use whichever one works better for you.

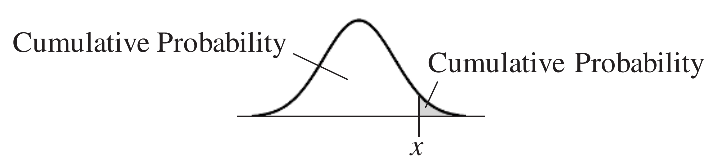
1. Alpha and *p* are Cumulative Probabilities. They are represented as **areas under the curve** of the Test Statistic Distribution.

A Test Statistic is calculated using Sample data. But, unlike other Statistics (e.g., the Mean or Standard Deviation), Test Statistics have an associated Probability Distribution (or family of such Distributions). Common Test Statistics are z, t, F, and χ^2 (Chi-Square).

The Distribution is plotted as a curve over a horizontal axis. The Test Statistic values are along the horizontal axis. The Point Probability of any value of a Test Statistic is the height of the curve above that Test Statistic value. But, we're really interested in Cumulative Probabilities.

A Cumulative Probability is the total Probability of all values in a range. Pictorially, **it is shown as the area under the part of curve of the Distribution which is above the range.**

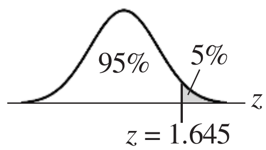
In the diagram below, the curve of the Probability Distribution is divided by x into two ranges: negative infinity to x and x to infinity. Above these two ranges are two areas (unshaded and shaded) representing two Cumulative Probabilities. The total area of the two is 100%.



In calculus-speak, the area under a curve is calculated as the integral of the curve over the range. Fortunately, when we use Test Statistics, we don't have to worry about calculus and integrals. The areas for specific values of the Test Statistic are shown in tables in books and websites, or they can be calculated with software, spreadsheets, or calculators on websites.

For example, if we select Alpha to be 5% (0.05), and we are using the Test Statistic z, then the value of z which corresponds to that value of

Alpha is $z = 1.645$



> **2. <u>The Critical Value</u>** (e.g., z-critical) **<u>and the value of the Test Statistic</u>** (e.g., z) **<u>are point values</u>** on the horizontal axis of the Test Statistic Distribution. **They mark the inner boundaries of the areas representing Alpha and p, respectively.**

- **The Critical Value is determined from the Distribution of the Test Statistic and the selected value of Alpha**. For example, as we showed earlier, if we select $\alpha = 5\%$ and we use z as our Test Statistic, then z-critical $= 1.645$.
- **The Sample data are used to calculate a value of the test Statistic.** For example, the following formula is used to calculate the value of z from Sample data:

$$\boldsymbol{z = (\mu - \bar{x})/s}$$

where $\bar{\boldsymbol{x}}$ is the Sample Mean, $\boldsymbol{s}$ is the Sample Standard Variation, and $\boldsymbol{\mu}$ is a specified value, for example, a target or historical value for the Mean.

The following tables illustrate some values for a 1-tailed/right-tailed situation (only shading under the right tail. See the article "*Alpha, α*" for more on 1-tailed and 2-tailed analyses.) Notice that **the larger the value of the boundary**, the farther out it is in the direction of the tail, and so **the smaller the area under the curve.**

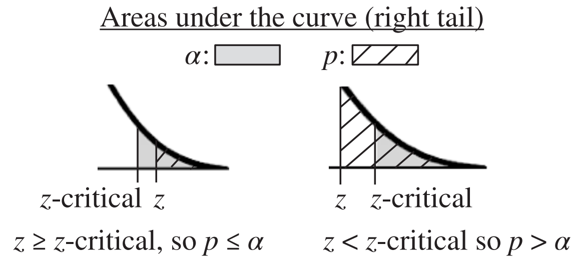
As the boundary point value grows larger ——————————>

Boundary: z or z-critical	0	0.675	1.282	1.645	2.327
Area: p or α	0.50	0.25	0.1	0.05	0.01

<————————— the Cumulative Probability area grows smaller

The graphs below are close-ups of the right tail of the z Distribution. The shaded area represents the Cumulative Probability, Alpha. The hatched area represents the Cumulative Probability, p. As explained in the tables above, **the larger the point value (z or z-critical), the smaller the value for its corresponding Cumulative Probability (p or α, respectively).**

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

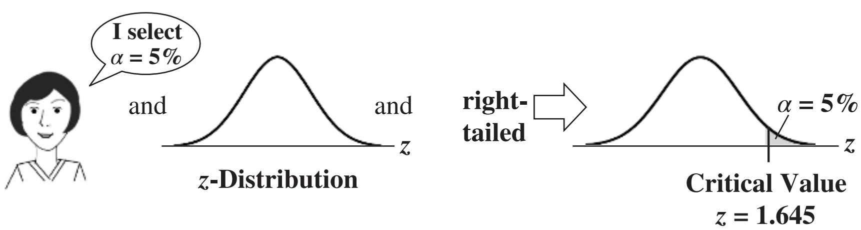


The left diagram above shows a value of z which is greater (farther from the Mean) than the Critical Value. So, p, the area under the curve bounded by z, is smaller than the area for Alpha, which is bounded by the Critical Value. The right diagram shows the opposite.

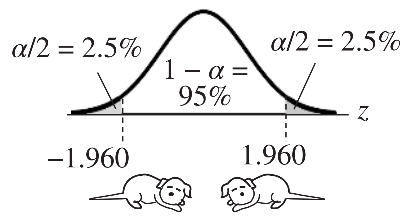
3. The person performing the analysis selects the value of Alpha.
Alpha and the Distribution are then used to calculate the Critical Value of the Test Statistic (e.g., z-critical). It is the value which forms the inner boundary of Alpha.

Alpha is called the Level of Significance. **Alpha is the upper limit for the Probability of an Alpha/"False-Positive" Error below which any observed difference, change, or effect is deemed Statistically Significant.** This is the only one of the four concepts featured in this article which is not calculated. It is selected by the person doing the analysis. Most commonly, $\alpha = 5\%$ (0.05) is selected. This gives a Level of Confidence of $1 - \alpha = 95\%$.

If we then plot this as a shaded area under the curve, the boundary can be calculated from it.



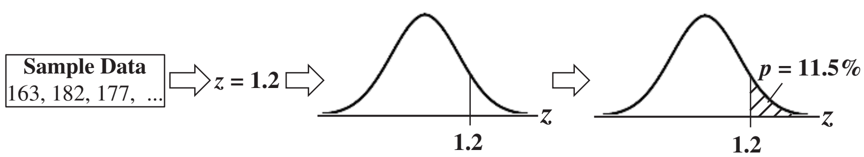
Note: for a 2-tailed analysis, half of Alpha (2.5%) would be placed under each tail. (left and right. The Critical Value would be 1.96 on the right and –1.96 on the left).



> **4. Sample data are used to calculate the value of the Test Statistic** (e.g., z).
>
> **The value of the Test Statistic and the Distribution are then used to calculate the value of *p*.** p is the area under the curve outward from this calculated value of the Test Statistic.

We saw how we use a Cumulative Probability (α) to get a point value (the Critical Value). We'll now go in the opposite direction. We use a point value for the Test Statistic, z, to get a Cumulative Probability (p).

***p* is the actual Probability of an Alpha Error** for a particular Sample of data.



Let's say that z is calculated from the Sample data to be 1.2. This gives us a value of $p = 0.115$ (11.5%).

> **5. To determine Statistical Significance, compare *p* to Alpha, or** (equivalently) **compare the value of the Test Statistic to its Critical value.**
>
> **If** $p \leq \alpha$ or (same thing) $z \geq z$**-critical**,
>
> **then there is a Statistically Significant difference, change, or effect. Reject the Null Hypothesis (H_0).**

- We selected Alpha as the Level of Significance – the maximum Probability of an Alpha/"False-Positive" Error) which we are willing to tolerate.
- We calculated p as the actual Probability of an Alpha Error for our Sample.

- **So if $p \leq \alpha$, then any difference, change, or effect observed in the Sample data is Statistically Significant.**

Note that:

- **The Critical Value is determined from Alpha. The two contain the same information**. Given a value for one, we could determine the other from the Distribution.
- Similarly, **the value of the Test Statistic (*z* in our example) contains the same information as *p*.**
- **So, comparing the Test Statistic to the Critical Value of the Test Statistic is statistically identical to comparing *p* to Alpha.**

Therefore:

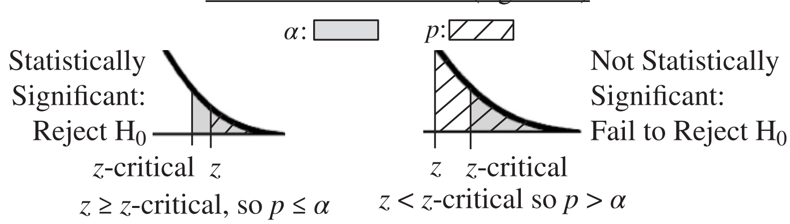
If $p \leq \alpha$ or **$z \geq z$-critical**,

then there <u>is</u> a Statistically Significant difference, **change**, **or effect**.

(**Reject the Null Hypothesis**).

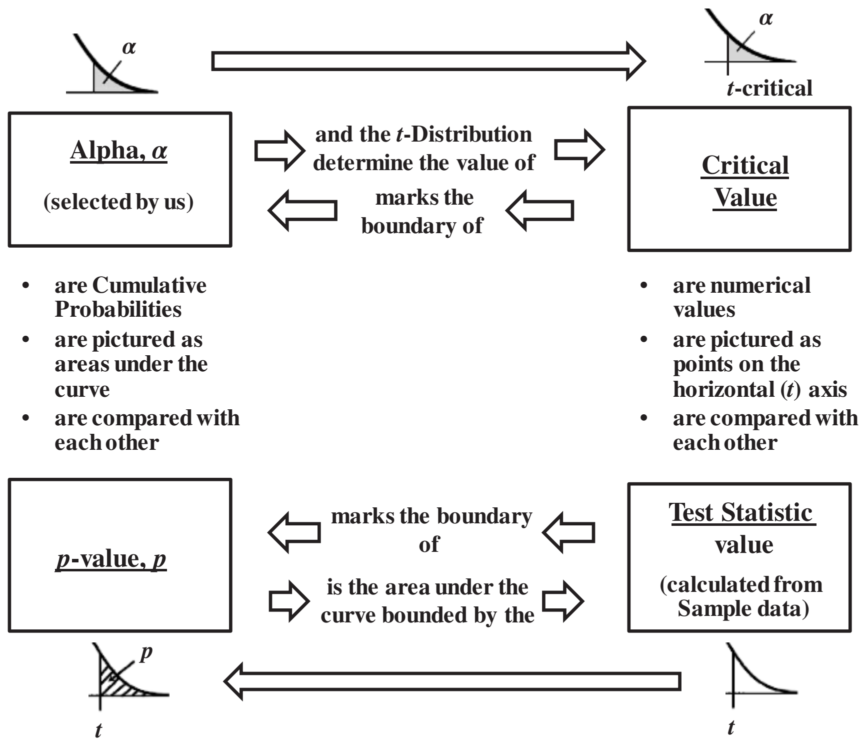
Depicted graphically:

<u>Areas under the curve (right tail)</u>



The table at the beginning of this article summarized the whole article in one visual. On the next page is the same information presented in another way. Use whichever one works best for you.

t is the Test Statistic in this illustration.



Related Articles in This Book: *Alpha, α*; *p-Value, p*; *Critical Value*; *Test Statistic*; *Distributions – Part 1: What They Are*; *Inferential Statistics*; *Hypothesis Testing – Parts 1–3*; *Confidence Intervals – Parts 1 and 2*; *p, t, and F: “<” or “>”?*

ALTERNATIVE HYPOTHESIS

Recommendation: read the article "Null Hypothesis" before reading this article.

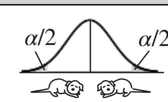
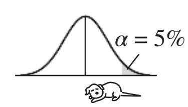
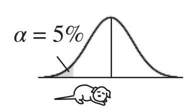
Symbols for the Alternative Hypothesis: H_A, H_1, or H_a

Summary of Keys to Understanding

1. Stating a **Null Hypothesis** (H_0) **and an Alternative Hypothesis** (H_A) **is the first step in our 5-step method for Hypothesis Testing.**

2. The Alternative Hypothesis is the opposite of the Null Hypothesis – and vice versa.

3. Stating the Alternative Hypothesis as a comparison formula, rather than in words, **can make things easier to understand. The formula must include an inequivalence in the comparison operator**, using one of these: "≠", ">", or "<".

Comparison Operator		Tails of the Test	
H_A	H_0		
$\neq$	$=$	2-tailed	
$>$	$\leq$	Right-tailed	
$<$	$\geq$	Left-tailed	

4. In a 1-tailed test, the Alternative Hypothesis (aka the "Research Hypothesis" or Maintained Hypothesis") **tells you in which direction** (right or left) **the tail points.**

Explanation

1. Stating a **Null Hypothesis** (H_0) **and an Alternative Hypothesis** (H_A) **is the first step in our 5-step method for Hypothesis Testing.**

Hypothesis Testing is one of two common methods for Inferential Statistics. Confidence Intervals is the other. In Inferential Statistics, we estimate a statistical property (e.g., the Mean or Standard Deviation) of a Population or Process by taking a Sample of data and calculating the property in the Sample.

In the article, "*Hypothesis Testing – Part 2: How To*" we describe a 5-step method of Hypothesis Testing:

1. State the problem or question in the form of a Null Hypothesis (H_0) and Alternative Hypothesis (H_A).
2. Select a Level of Significance (α).
3. Collect a Sample of data for analysis.
4. Perform a statistical analysis on the Sample data.
5. Come to a conclusion about the Null Hypothesis (Reject or Fail to Reject).

Hypothesis Testing can be very confusing, mainly because the language in steps 1 and 5 can be confusing. This article and the *Null Hypothesis* article are written to clear up the confusion in step 1.

Experts disagree on whether an Alternative Hypothesis should be used. It is included here, because, as we'll explain later, it is useful in 1-tailed tests.

2. The Alternative Hypothesis is the opposite of the Null Hypothesis – and vice versa.

What exactly does that mean? It means that:

- **If the Null Hypothesis is true, then the Alternative Hypothesis is false.**
- **If the Null Hypothesis is false, then the Alternative Hypothesis is true.**

These two statements imply that:

H_0 and H_A are

- **mutually exclusive and**
- **collectively exhaustive.**

This means that either H_0 or H_A must be true; you can't have neither being true. And you can't have both being true.

Here are some examples:

Example 1

H_0: There is no difference between the Standard Deviations of Population A and Population B.

H_A: There is a difference between the Standard Deviations of Population A and Population B.

Example 2

H_A: Our school's average test scores are better than the national average.

H_0: Our school's average test scores are less than or equal to the national average.

Example 3

H_A: Our orders ship in less than 4 days.

H_0: Our orders ship in 4 days or more

In addition to being mutually exclusive and collectively exhaustive, these three examples include a couple of other concepts:

- **Statistically Significant:** In Example 1, our two Samples of data will no doubt show some difference in the two Standard Deviations. The Inferential Statistical test will determine whether that difference is Statistically Significant. Likewise, the "better than" and "less than" in Examples 2 and 3 are implicitly modified by "to a Statistically Significant extent."
- **2-tailed or 1-tailed:** As we'll explain later, Example 1 will use a 2-tailed analysis. Example 2 (right-tailed) and Example 3 (left-tailed) are 1-tailed.

Note also that for Examples 2 and 3, we list H_A first and H_0 second. The reason for this is explained below, under Keys to Understanding #4.

3. Stating the Alternative Hypothesis as a comparison formula, rather than in words, **can make things easier to understand. The formula must include an inequivalence in the comparison operator**, using one of these: "$\neq$", "$>$", or "$<$".

Null and Alternative Hypotheses involve comparisons (equations or inequalities) between values of Parameters (properties) of Populations or

Processes. A Parameter could be a Mean (μ), a Standard Deviation (σ), or other descriptive statistical property.

In a Hypothesis, a Parameter from one Population or Process could be compared with that of another, for example,

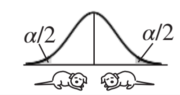
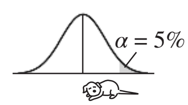
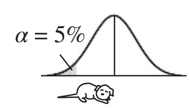
$$\sigma_A = \sigma_B$$

Or it could be compared with a numerical value, like a target or historical value:

$$\sigma < 1.5$$

There are 3 basic comparison symbols: equal "=", greater than ">", and less than "<".

There are also compound symbols: not equal ≠, greater than or equal to "≥", and less than or equal to "≤".

Comparison Operator		Tails of the Test	
H_A	H_0		
≠	=	2-tailed	
>	≤	Right-tailed	
<	≥	Left-tailed	

4. In a 1-tailed test, the Alternative Hypothesis (aka the "Research Hypothesis" or Maintained Hypothesis") **tells you in which direction** (right or left) **the tail points.**

If H_0 can be stated with an equal sign, "=", the situation is relatively straightforward. We are only interested in whether there is a Statistically Significant difference, change, or effect. There is no direction involved. When we tell our statistical tool what type of test it is, we say "2-tailed." The common wisdom is to state a Null Hypothesis, and then the Alternative Hypothesis is the opposite.

But, for 1-tailed tests, when "<" or ">" is involved, it gets more complicated. Once we determine which is the Null and which is the Alternative Hypothesis, it's easy to assign a comparison operator to each comparison formula. **But how do we decide which is which?**

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

It may help to know that **the Alternative Hypothesis is also known as the Research Hypothesis or the Maintained Hypothesis.** And that **the Alternative Hypothesis is the one that the researcher maintains and aims to prove.**

In Example 2 above, our school's average test scores are somewhat better than the national average, and we would like to prove that this is a Statistically Significant difference. So we select as our Alternative (Maintained) Hypothesis:

$$\mathrm{H_A}\colon \mu_{\text{school}} > \mu_{\text{national}}$$

The Null Hypothesis then becomes:

$$\mathrm{H_0}\colon \mu_{\text{school}} \leq \mu_{\text{national}}$$

Note that, **for 1-tailed tests, it is better to start with a statement of the Alternative Hypothesis and then derive the Null Hypothesis as the opposite.** This is because we know what we maintain and would like to prove.

Furthermore, **the "<" or ">" in the Alternative Hypothesis points in the direction of the tail.** "<" in the Alternative Hypothesis means that the test is Left-Tailed. ">" tells us that it is Right-Tailed.

Related Articles in This Book: *Null Hypothesis*; *Hypothesis Testing – Part 1: Overview*; *Hypothesis Testing – Part 2: How To*; *Reject the Null Hypothesis*; *Fail to Reject the Null Hypothesis*

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

ANALYSIS OF MEANS (ANOM)

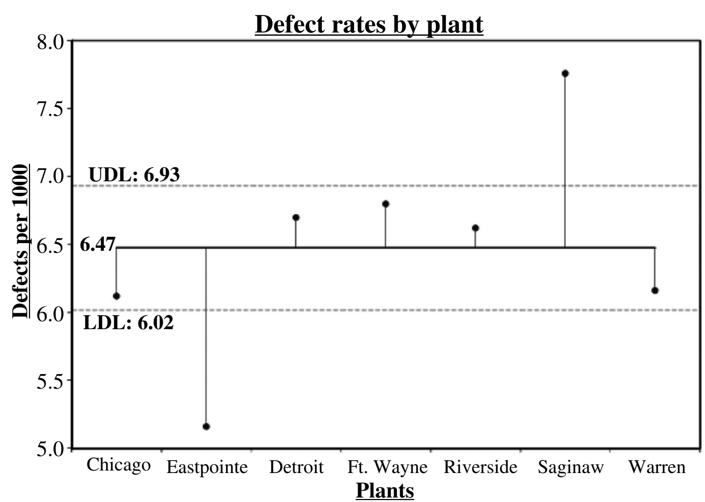
Summary of Keys to Understanding

1. Analysis of Means (ANOM) tells us whether the Means from several Samples are statistically the same as the Overall Mean.

2. ANOM has some similarities to, and some differences from, ANOVA

	ANOM	ANOVA
Assumptions	Approximately Normal data	
Analyzes Variation of several Means	Yes	
1-Way or 2-Way	Yes	
Variation	around the overall Mean	among each other
Identifies which Means are not statistically the same	Yes	No
Output	Graphical	Statistical: ANOVA Table

3. The graphical ANOM output is similar to a Control Chart.



JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

Explanation

1. Analysis of Means (ANOM) tells us whether the Means from Samples from several different Populations or Processes are statistically the same as the Overall Mean.

The different Populations or Processes are represented by different values of a Categorical/Nominal Variable. As such, they are names, for example,

Call center reps: John, Jane, Robert, Melissa, Judith, Mike

Vendors: Company A, Company B, Company C, Company D

Plants: Chicago, Eastpointe, Detroit, Fort Wayne, Riverside, Toledo, Warren

The Means here are the Means of an Independent Variable, *y*. *y* is numerical, such as the number of calls successfully handled, delivery times, and defect rates.

For each name, there will be a Sample of data – for example, for each call center rep, the number of calls handled each day for a number of days.

The Overall Mean, sometimes called the Grand Mean, is the average of all the *y*-Variable values from all the Samples.

ANOM has been most frequently used in industrial and process-improvement analyses, but it is applicable generally.

The underlying calculations for ANOM are more complicated than those for ANOVA, and explaining them is beyond the scope of this book.

2. ANOM has some similarities to, and some differences from, ANOVA

	ANOM	ANOVA
Assumptions	Approximately Normal data	
Analyzes Variation of several Means	Yes	
1-Way or 2-Way	Yes	

First, the similarities: In order to produce valid results, **both ANOM and ANOVA require that the data be approximately Normal.** "Approximately" Normal is not strictly defined, but the data should not be obviously non-Normal. That is, it should have one discernable peak and not be strongly skewed.

Second, they both analyze Variation in Means. ANOVA is "Analysis of Variation," but it analyzes Variation among Means. Both are usually used with 3 or more Means. For 2 Means, there is the 2-Sample *t*-test.

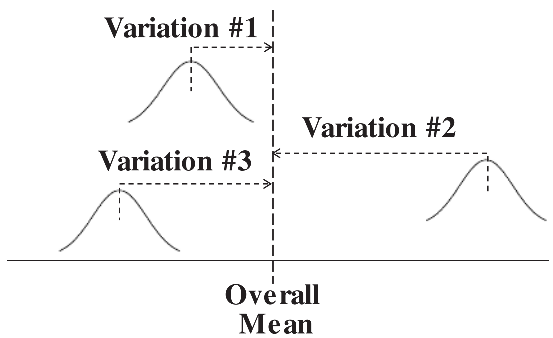
And both can perform 1-Way (aka Single Factor, i.e., one *x* Variable) or 2-Way (Two Factor, two *x* Variables) analyses.

	ANOM	ANOVA
Variation	around the overall Mean	among each other
Identifies which Means are not statistically the same	Yes	No

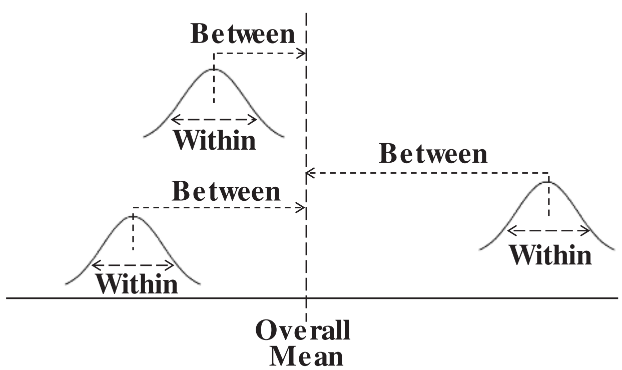
ANOM calculates the Overall Mean, and then it measures the Variation of each Mean from that. In the conceptual diagram below, each Sample is depicted by a Normal curve. The distance between each Sample Mean and the Overall Mean is identified as a "Variation."

ANOM retains the identity of the source of each of these Variations (#1, #2, and #3), and it displays this graphically in the ANOM chart (shown later in this article).

ANOM



ANOVA



ANOVA takes a more holistic approach, in which the identity of the individual Sample Variations is lost. This is explained in detail in the articles, *ANOVA, Parts 1, 2, and 3.* But briefly, …

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

ANOVA starts out like ANOM, calculating the Variation between each Sample Mean and the Overall Mean. But it then consolidates this information into one Statistic for all the Samples, the Mean Sum of Squares Between, MSB.

Next it calculates Variation within each Sample and then consolidates that into one Statistic, the Mean Sum of Squares Within, MSW*. So **any information about individual Sample Means and Variances is lost.** That is why **ANOVA can only tell us if there is a Statistically Significant difference somewhere among the Means, not which one(s) are Significantly different. However, ANOM can.**

*(ANOVA goes on to divide MSB by MSW, yielding the *F*-statistic, which is then compared to *F*-critical to determine Statistical Significance.)

3. The graphical ANOM output is similar to a Control Chart.

The output from ANOVA is a table of Statistics. **The output from ANOM is graphical.**

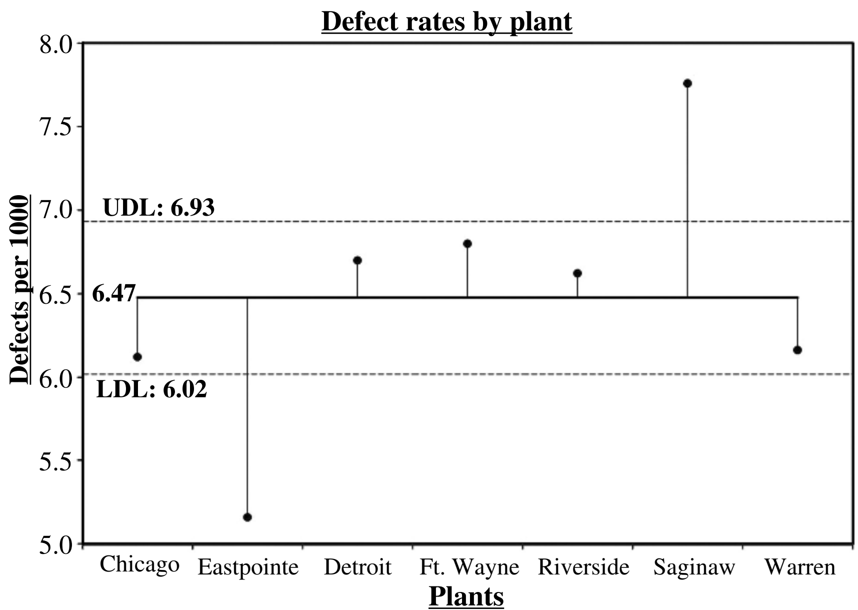
Example: Let's say we have 7 plants mass-producing the same product, and we want to determine whether any have a defect rate per thousand which is (Statistically) Significantly better or worse than the others. We collect data for 5 days.

	Chicago	Eastpointe	Detroit	Ft. Wayne	Riverside	Saginaw	Warren
	6.0	5.2	6.8	7.1	6.8	7.4	6.2
	6.5	4.3	7.0	6.7	6.0	7.9	6.9
	6.1	5.1	6.7	6.5	6.4	8.2	5.9
	6.2	5.3	6.4	6.9	7.3	7.7	5.7
	5.8	5.9	6.6	6.8	6.6	7.6	6.1
Means:	6.1	5.2	6.7	6.8	6.6	7.8	6.2

In the ANOM chart below, the dotted horizontal lines, the **Upper Decision Line (UDL) and Lower Decision Line (LDL) define a Confidence Interval**, in this case, for $\alpha = 0.05$. Our conclusion is that only Eastpointe (on the low side) and Saginaw (on the high side) exhibit a Statistically Significant difference in their Mean defect rates. So **ANOM tells us not only whether any plants are Significantly different, but also which ones are.**

ANOM Output The dots show the Means of the 5 days of data for each plant.

The Overall Mean for all plants is 6.47.



Related Articles in This Book: *ANOVA, Parts 1–4*; *Variation/Variability/ Dispersion/Spread*; *Confidence Intervals – Parts 1 and 2*; *Alpha, a*; *Control Charts – Part 1: General Concepts and Principles*

ANOVA – PART 1 (OF 4): WHAT IT DOES

Summary of Keys to Understanding

1. "ANOVA" is an acronym for ANalysis Of VAriance. However, its objective is to determine if one or more of the Means of several Groups is different from the others.

2. Assumptions (test requirements) are
- The groups being compared have a roughly Normal Distribution
- The groups have similar Variances

3. There are 3 types of ANOVA
- 1-Way aka Single Factor
- 2-Way without Replication
- 2-Way with Replication

4. ANOVA is often used in Designed Experiments. An ANOVA Table is often an output in Multiple Linear Regression analysis.

5. ANOVA Does	ANOVA Does Not		Do this instead
compare several Means with each other	compare several Means with the overall Mean		ANOM
say whether or not there is a difference among Means	say which Means differ		ANOM or Confidence Intervals
require Continuous data	handle Discrete data		Chi-square Test of Variance
require roughly Normal Distributions	handle very Non-Normal Distributions		Kruskal–Wallis
require somewhat equal Sample Variances	handle very unequal Sample Sizes and Variances		Ensure equal Sample Sizes when Sample Variances are unequal

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

Explanation

There are 4 articles in this series about ANOVA

Part 1: What it Does

Part 2: How it Does It

The underlying 7-step method which involves Sums of Squares and an *F*-test. Students may need to understand this for their exams. But if you just want the answer to the ANOVA analysis, spreadsheets or software can give it to you if you just provide the data.

Part 3: 1-Way

The method used when there is a single Factor affecting the outcome we are measuring. For example, the single Factor would be the drug used in a test. ANOVA would be used to determine whether any stood out from the rest.

Part 4: 2-Way

Used when 2 Factors affect the outcome. For example, in a laundry process, measuring the effect on cleanliness of the Factors, water temperature, and detergent type. Interactions between Factors are an important component of 2-Way ANOVA.

1. "ANOVA" is an acronym for ANalysis Of VAriance. However, its objective is to determine if one or more of the Means of several Groups are different from the others.

ANOVA is an acronym for "Analysis of Variance." But **analyzing Variances is not its objective**. **Its objective is to determine whether one or more of several Means are different** from the others by a Statistically Significant amount. **It does this by analyzing Variances.**

"Group" here is a generic term which can refer to:

– a Population or Process for which we have complete data.
– a Sample taken from a Population or Process, for example, the annual incomes of 30 people who live in particular neighborhood. In the case of a Sample, the Sum of the Squares of the Sample is an estimate of the Sum of the Squares for the Population.

2. Assumptions (test requirements) are
- Groups being compared have a roughly Normal Distribution
- Groups have similar Variances

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

As we will see in the Part 2 article, the Variances which are analyzed are not the Variances of the individual groups whose Means we are comparing. The Variances are the Mean Sum of Squares Between groups (MSB) and Mean Sum of Squares Within groups (MSW). These are two numbers, each of which summarizes different information about all the groups.

For ANOVA, the groups should be roughly Normal in their Distributions and their Variances should be roughly similar. ANOVA is fairly tolerant in terms of what is considered Normal enough or having similar enough Variances. If these assumptions are not roughly met, then the Kruskal–Wallis test can be used instead.

3. There are 3 types of ANOVA

- 1-Way, aka Single Factor
- 2-Way without Replication
- 2-Way with Replication

1-Way, also known as Single Factor, is covered in the Part 3 article. There is one Factor – the *x* Variable – which affects the outcome, or *y* Variable. For example, the single Factor could be blood pressure drug. There could be several different drugs being compared. The *y* Variable would be a measure of reduction in blood pressure.

The 2-Way types of ANOVA are covered in the Part 4 article. In both cases, there are two Factors, or *x* Variables. For example, water temperature and detergent type would be the two Factors, and a cleanliness measure would be the outcome or *y* Variables.

If the data show that the two Factors interact, then the 2-Way with Replication (repeated measurements) must be used.

4. ANOVA is often used in Designed Experiments. An ANOVA Table is often an output in a Multiple Linear Regression analysis.

ANOVA Table					
	df	SS	MS	*F*	*p*-value
Regression	−4.000	48,877.931	−12,219.483	32.727	0.009
Residual	3.000	1493.498	497.833		
Total	−1.000	50,371.429			

5. ANOVA Does	**ANOVA Does Not**		**Do this instead**
compare several Means with each other	compare several Means with the overall Mean		ANOM
say whether or not there is a difference among Means	say which Means differ		ANOM or Confidence Intervals
require Continuous data	handle Discrete data		Chi-square Test of Variance
require roughly Normal Distributions	handle very Non-Normal Distributions		Kruskal–Wallis
require somewhat equal Sample Variances	handle very unequal Sample Sizes and Variances		Ensure equal Sample Sizes when Sample Variances are unequal

Related Articles in This Book: *Part 2: How It Does It*; *Part 3: 1-Way*; *Part 4: 2-Way*; *Sums of Squares*; *ANOVA vs. Regression*; *Design of Experiments (DOE) – Part 3*; *Regression – Part 4: Multiple Linear*

ANOVA – PART 2 (OF 4): HOW IT DOES IT

Summary of Keys to Understanding

1. **Sum of Squares Within** (SSW) is **the sum of the Variations** (as expressed by the Sums of Squares, SS's) **within each of several Groups.**

$$\mathbf{SSW = SS_1 + SS_2 + \cdots + SS_n}$$

2. **Sum of Squares Between (SSB) measures Variation between** (among) **Groups,**

$$\mathbf{SSB = \sum n(\overline{X} - \overline{\overline{X}})^2}$$

and **Sums of Squares Total (SST) is the Total of both types of Variation.**

$$\mathbf{SST = SSW + SSB}$$

3. The Mean Sums (of Squares), MSB and MSW, are averages of SSB and SSW, respectively. **With MSB and MSW, we have only 2 Statistics which summarize the Variation in 3 or more groups.**

4. **Mean Sums of Squares are similar to the Variance. As such, they can be used to calculate the Test Statistic, *F***, which is the ratio of two Variances.

5. **Mean Sums of Squares are used in the *F*-tests** in **ANOVA.**
   A large value of MSB, compared with MSW, **indicates that the Sample Means are not close to each other.** This makes for a large value for *F*, **which makes it more likely that $F \geq F$-critical.**

SSB ⇨ MSB
SSW ⇨ MSW ⇨ $\dfrac{\mathbf{MSB}}{\mathbf{MSW}} = F$ ⇨
If $F \geq F$-critical, ⇨ there is a difference.
If $F < F$-critical, ⇨ there is no difference.

Explanation

This article is about what goes on behind the scenes in an ANOVA. Spreadsheets or software will do all the calculations for you

The generic **Sum of Squares (SS)** is the **sum of the Squared Deviations** of all the data values in a single Group (e.g., a Sample). **SS is one measure of Variation** (it also happens to be the numerator in the formula for Variance).

$$\mathbf{SS} = \sum (\boldsymbol{x} - \bar{\boldsymbol{x}})^2$$

MSB and MSW are special types of Sums of Squares. In this article, we will show how MSB and MSW are derived from the data, starting with the most basic kind of Sum of Squares.

The **Deviation** (of a single data value, x) **is $\boldsymbol{x} - \bar{\boldsymbol{x}}$.**

"Deviation" here means **distance from the Mean**: $\boldsymbol{x} - \bar{\boldsymbol{x}}$, where x is an individual data value in a Group, and $\bar{\boldsymbol{x}}$ is the Mean of the Group. It could just as easily be $\bar{\boldsymbol{x}} - x$ as $x - \bar{\boldsymbol{x}}$. For our purposes, we don't care whether a value is less than or greater than the Mean. We just want to know by how much it deviates from the Mean. **So we square it, to ensure we always get a positive number.** (Another article in this book, *Variance,* explains why we don't just use the absolute value instead of squaring).

A **Squared Deviation is just the square of a Deviation.**

If we want to find a measure of Variation for the Group we can total up all the Squared Deviations of all data values in the Sample. That gives us the Sum of the Squared Deviations, aka the **Sum of Squares.**

$$\mathbf{SS} = \sum (\boldsymbol{x} - \bar{\boldsymbol{x}})^2$$

So, it is easy to see that – like Variance and Standard Deviation – **Sum of Squares (SS) is a measure of Variation.** In fact, the **Sum of Squares is the numerator in the formula for Variance (s^2).**

$$s^2 = \frac{\sum (\boldsymbol{x} - \bar{\boldsymbol{x}})^2}{\boldsymbol{n} - \mathbf{1}} = \frac{\mathbf{SS}}{\boldsymbol{n} - \mathbf{1}}$$

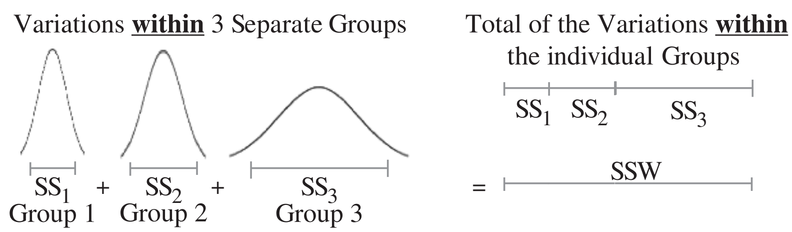
Variance is, for most purposes, a better measure of Variation than the generic SS, because it takes into account the Sample Size, and it approximates the square of the average Deviation. But there is more to the SS story. **ANOVA uses 3 particular types of Sums of Squares: Within, Between, and Total (SSW, SSB, and SST).** Whereas the generic SS is only about a single Group, **these three each measure different kinds of Variation involving multiple Groups.**

1. Sum of Squares Within (SSW) is **the sum of the Variations** (as expressed by the Sums of Squares, SS's) **within each of several Groups.**

$$\mathbf{SSW = SS_1 + SS_2 + \cdots + SS_n}$$

Sums of Squares Within (SSW) summarizes how much Variation there is within each of several Groups (usually Samples) – by giving the sum of all such **Variations**.

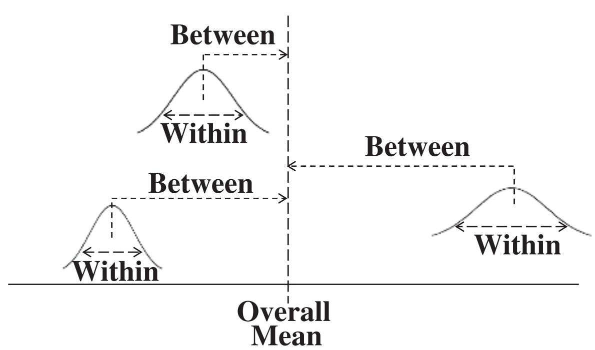
This is not numerically precise, but conceptually, one might picture SS as the width of the "meaty" part of a Distribution curve – the part without the skinny tails on either side.



A comparatively small SSW indicates that the data within the individual Groups are tightly clustered about their respective Means. If the data in each Group represent the effects of a particular treatment, for example, this **is indicative of consistent results** (good or bad) within each individual treatment.

"Small" is a relative term, so the word **"comparatively" is key here.** We'll need to compare SSW with another type of Sum of Squares (SSB) before being able to make a final determination.

A comparatively large SSW shows that the **data** within the individual Groups **are widely dispersed. This would indicate inconsistent results** within each individual treatment.



2. Sum of Squares Between (SSB) measures Variation <u>between</u> (among) **Groups,**

$$\mathbf{SSB} = \sum n(\overline{X} - \overline{\overline{X}})^2$$

and **Sums of Squares Total (SST) is the <u>Total</u> of both types of Variation.**

$$\mathbf{SST = SSW + SSB}$$

(According to the dictionary, "between" is about 2 things, so the word is used ungrammatically here; it should be "among." However, "between" is much more commonly used in this context, so we'll go with that in this book.)

<u>To calculate Sum of Squares between, SSB:</u>

$$\mathbf{SSB} = \sum n(\overline{X} - \overline{\overline{X}})^2$$

where $\overline{X}$ is a Group Mean and $\overline{\overline{X}}$ is the Overall Mean and n is the number of values in that Group. The Overall Mean (also called the Grand Mean) is the Mean of all the data values from all Groups.

– First, calculate the Overall Mean, (symbols $\overline{\overline{X}}$). You can forget the individual groupings, just add up the data values from all Groups and divide by the total number of values.

  In the form of a formula – Overall Mean: $\overline{\overline{X}} = \sum x_{ij}/N$

  where i represents the individual values in one Sample

  and j represents the individual Samples

  and N is the total of Sample Sizes of all Samples

  For example, Sample #1 has values 40, 45, 45, 50, and a Mean of 45; Sample #2 has values 25, 35, 35, 45, and a Mean of 35; Sample #3 has values 40, 55, 55, 70, and a Mean of 55.

$$\overline{\overline{X}} = \frac{40 + 45 + 45 + 50 + 25 + 35 + 35 + 45 + 40 + 55 + 55 + 70}{12}$$
$$= \frac{540}{12} = 45$$

– Next, subtract the Overall Mean, $\overline{\overline{X}}$, from each Group Mean $\overline{X}_j$

$$\overline{X}_j - \overline{\overline{X}}$$

  Sample #1: 45 − 45 = 0; Sample #2: 35 – 45 = −10; Sample #3: 55 − 45 = 10

– Then, square each of these deviations

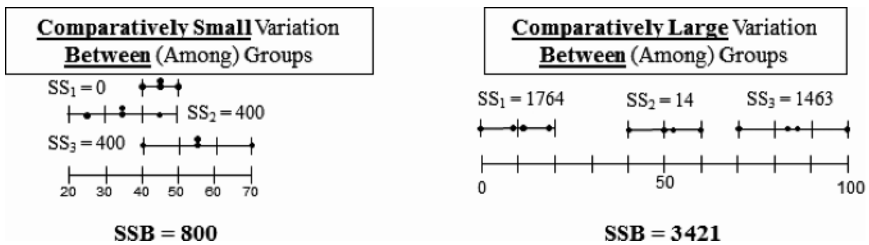
$$(0)^2 = 0 \qquad (-10)^2 = 100 \qquad (10)^2 = 100$$

– Multiply each squared Deviation by the Group size

$$0 \times 4 = 0 \qquad 100 \times 4 = 400 \qquad 100 \times 4 = 400$$

– Sum these: SSB = 0 + 400 + 400 = 800

These numbers are graphed below left and are indicative of a comparatively small Variation between the Groups. Notice that the ranges overlap.



On the right above is a graph of comparatively large Variation between the Groups. There is no overlap in the ranges. We keep saying "comparatively" because, as mentioned earlier, we need to consider both SSW and SSB together in order to come to a definitive conclusion.

If we add SSW and SSB, we get a measure of the total Variation, Sum of Squares Total, SST.

$$\mathbf{SST = SSW + SSB}$$

Notation Alert: some authors use the term "Sum of Squares Treatment" (SST) instead of Sum of Squares Between. That introduces a potential source of confusion, since SST is usually used for Sum of Squares Total.

3. The Mean Sums (of Squares), MSB and MSW, are averages of SSB and SSW, respectively. **With MSB and MSW, we have only 2 Statistics which summarize the Variation in 3 or more Groups.**

Sums of differences (like SSW and SSB) provide a gross measure of Variation, somewhat analogous to a Range. Averages (Means) are generally more meaningful than sums. (That is why Variances or Standard Deviations are generally more useful than Ranges.) So we calculate the Mean equivalents of SSW and SSB: MSW and MSB.

MSB and MSW are Statistics which each distill information about a type of **Variation involving multiple Groups into a single number**. We can then use these 2 Statistics in a single *F*-test to accomplish the same thing that multiple t-tests would accomplish. Thus, we avoid the compounding of Alpha Error which would occur with multiple t-tests.

The downside is that, in calculating the MS's, we lose specific information about the individual Groups. **This is why ANOVA will tell us whether there is a Statistically Significant difference among several Groups, but it will not tell us which one(s) are different.**

> **4. Mean Sums of Squares are similar to the Variance. As such, they can be used to calculate the Test Statistic, *F*,** which is the ratio of two Variances.

Earlier in this article, we said that the generic Sum of Squares is the numerator in the formula for Variance. The denominator in that formula is $n - 1$. As described in the Part 3 article, MSB and MSW are calculated by dividing the Sums of Squares, SSB and SSW, by terms similar to $n - 1$. So MSB and MSW are similar to the Variance.

The Test Statistic F is the ratio of two Variances. So, ANOVA is able to use the ratio of MSB and MSW in an F-test to determine if there is a Statistical Significant difference among the Means of the Groups.

> **5. Mean Sums of Squares are used in the *F*-tests** in **ANOVA.**
> **A large value of MSB**, compared with MSW, **indicates that the Sample Means are not close to each other.** This makes for a large value for F, **which makes it more likely that $F > F$-critical.**

SSB ⇨ MSB
SSW ⇨ MSW ⇨ $\dfrac{\textbf{MSB}}{\textbf{MSW}} = F$ ⇨
If $F \geq F$-critical, ⇨ **there <u>is</u> a difference.**
If $F < F$-critical, ⇨ **there is <u>no difference</u>.**

- The formulas for MSB and MSW are specific implementations of the generic formula for Variance.
- So, MSB divided by MSW is the ratio of two Variances.
- The Test Statistic F is the ratio of two Variances.
- ANOVA uses an F-Test (F = MSB/MSW) to come to a conclusion.
- If $F \geq F$-Critical, then we conclude that the Mean(s) of one or more Groups have a Statistically Significant difference from the others.

Related Articles in This Book: *ANOVA – Part 1: What It Does*; *ANOVA – Part 3: 1-Way*; *ANOVA – Part 4: 2-Way*; *Sums of Squares*; *Variation/Variability/Dispersion/Spread*; *Variance*

ANOVA – PART 3 (OF 4): 1-WAY (AKA SINGLE FACTOR)

Summary of Keys to Understanding

Builds on the content of the ANOVA Part 1 and Part 2 articles.

1. **In 1-Way ANOVA, we study the effect of one Nominal** (named) **Variable, *x*, on the Dependent Numerical Variable, *y*.**

2. **Objective:** Determine whether there is a **Statistically Significant difference among the Means** of 3 or more groups. **Do one or more group Means stand out from the rest?**

x:Script	*y*: sales in first 100 calls											Mean
A	175	50	225	60	180	170	230	45	90	190		141.5
B	95	150	160	75	120	140	250	70	85	180		132.5
C	80	120	95	225	60	110	160	90	120	140		126.5

3. **A 7-step method** (summarized graphically below) **performs the analysis.** Spreadsheets or software will do all this, you just provide the data.

SSB ⇨ MSB
SSW ⇨ MSW
⇨ $\frac{\mathbf{MSB}}{\mathbf{MSW}} = F$ ⇨
If $F \geq F$-critical, ⇨ there is a difference.
If $F < F$-critical, ⇨ there is no difference.

4. **The output includes an ANOVA Table like this:**

ANOVA	Cannot Reject Null Hypothesis because $p > 0.05$ (Means are the same.)					
Source of Variation	SS	df	MS	*F*	*p*-Value	*F*-crit
Between Groups	2686.67	2	1343.33	0.376	0.690	3.354
Within Groups	96567.50	27	3576.57			
Total	99254.17	29				

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

Explanation

Prerequisite articles: ANOVA Part 1 and ANOVA Part 2.

1. In 1-Way ANOVA, we study the effect of one Nominal (named) **Variable, *x*, on the Numerical Variable, *y*.**

A Nominal (aka Categorical) Variable is one whose values are names.

x is the Independent Variable, also called the Factor. *y* is the Dependent Variable, since its value depends on the value of *x*. We might say $y = f(x)$, but in ANOVA (unlike in Regression) we are not interested in determining what the function f is.

Three Examples of Variables in 1-Way ANOVA

Nominal Independent Variable, *x*	values of the *x* Variable	Numerical Dependent Variable, *y*
Script used in call center sales calls	“A”, “B”, “C”	Sales in dollars
Level of Training	Beginner, Intermediate, Advanced	A worker productivity measurement
School District	names of the 6 school districts	Test scores

ANOVA is frequently used in Designed Experiments. (See the articles on *Design of Experiments*.)

2. Objective: Determine whether there is a **Statistically Significant difference among the Means** of 3 or more groups. **Do one or more group Means stand out from the rest?**

A Sample of data is taken for each of the values of the *x* Variable, and the Means of the *y* measurements for each Sample is calculated.

For example, let’s say we’re starting up a call center to sell a new product. We hire 30 callers of similar background and divide them into 3 groups of 10. Each group was given a different script to use for their opening sales pitches. We recorded their sales in dollars for the first 100 calls. The *x* Variable is the name of the script, and the *y* Variable is the sales amount.

x:script	*y*: sales in first 100 calls											Mean
A	175	50	225	60	180	170	230	45	90	190		141.5
B	95	150	160	75	120	140	250	70	85	180		132.5
C	80	120	95	225	60	110	160	90	120	140		126.5

There are 3 Samples (groups), A, B, and C. Each has 10 data values, for a total of 30.

Script A appears to give the best results and Script C the worst. But are the differences in the 3 Means Statistically Significant? That's what 1-Way ANOVA can tell us.

3. A 7-step method performs the analysis. Spreadsheets or software will do all this, you just provide the data.

Before collecting data, select a value for Alpha. Most commonly $\alpha = 0.05$ is selected.

Step 1. Calculate the Sum of Squares (SS) for each Sample.

$$\mathbf{SS} = \sum (x_i - \bar{x})^2$$

SS is a measure of Variation within one Sample. In fact, it is the numerator in the formula for Variance.

Step 2. Add all these up for all Samples to get the Sum of Squares Within

$$\mathbf{SSW} = \mathbf{SS}_1 + \mathbf{SS}_2 + \cdots + \mathbf{SS}_n$$

SSW is a measure of Variation within all the Samples.

Step 3. Calculate the Overall Mean, ($\overline{\overline{X}}$), of all the data values in all Samples.

Forget which data values go with which Samples, just put them all in one bucket and calculate the Mean.

Step 4: Sum up the differences between each Sample Mean and the Overall Mean to get Sum of Squares Between.

$$\mathbf{SSB} = \sum n(\overline{X} - \overline{\overline{X}})^2$$

SSB is a measure of how much the Sample Means differ from the Overall Mean. It also contains information on how much the Sample Means differ from each other.

Step 5: Calculate the Mean Sum of Squares Within (MSW) and Between (MSB).

Sums of differences (like SSW and SSB) provide a gross measure of Variation, somewhat analogous to a Range. But it is often not meaningful to compare sums of different numbers of things. Averages (Means) are generally more meaningful than totals. (That is why Variances or Standard Deviations are generally more useful than Ranges.) So we calculate MSW and MSB.

$$\mathbf{MSW} = \frac{\mathbf{SSW}}{\boldsymbol{N-k}} \quad \text{and} \quad \mathbf{MSB} = \frac{\mathbf{SSB}}{\boldsymbol{k-1}}$$

where N is the overall number of data values in all groups, and k is the number of groups. In our example $N = 30$ and $k = 3$.

SSW and SSB are specific types of the generic Sum of Squares, SS. And the formula for SS is the numerator for the formula for Variance, s^2.

$$s^2 = \frac{\sum(\boldsymbol{x_i} - \bar{\boldsymbol{x}})}{\boldsymbol{n-1}} = \frac{\mathbf{SS}}{\boldsymbol{N-1}}$$

So, if we divide the two special types of Sums of Squares, SSW and SSB, by a Degrees-of-Freedom term (like N–k or k–1), it is easy to see that **MSW and MSB are Variances.**

Step 6: Perform an *F*-test

The crux of ANOVA is comparing the Variation Within groups to the Variation Between (Among) groups. The best way to do a comparison is to calculate a ratio. **The *F*-statistic is a ratio of** two Variances, **MSB and MSW.**

$$\boldsymbol{F} = \frac{\mathbf{MSB}}{\mathbf{MSW}}$$

Note that **this is a different concept from the usual *F*-test comparing Variances of two Samples**. In that case, the Null Hypothesis would be that there is not a Statistically Significant difference between the Variances of two Samples. Although MSB and MSW have formulas like Variances, **MSB and MSW contain information about the differences between the Means of the several groups.** They contain no information about the Variances of the groups.

In the *F*-Test within ANOVA, **the ANOVA Null Hypothesis is that** there is not a Statistically Significant difference between MSB and MSW – that is, **there is not a Statistically Significant difference among the Means of the several Groups.**

Step 7:

As described in the article on the *F*-test, Alpha determines the value of *F*-critical, and the *F*-statistic (calculated from the Sample data) determines the value of the Probability p. Comparing p to α is identical to comparing F and F-critical

If $F \geq F$-critical (equivalently**, $p \leq \alpha$**), then **there is a Statistically Significant difference between the Means of the groups**. (Reject the ANOVA Null Hypothesis.)

If $F < F$-critical ($p > \alpha$), then **there is not Statistically Significant difference between the Means of the groups**. (Accept/Fail to Reject the ANOVA Null Hypothesis.)

<u>**The 7–Step ANOVA Process**</u> summarized in a concept flow diagram:

SSB ⇨ MSB
SSW ⇨ MSW ⇨ $\frac{\textbf{MSB}}{\textbf{MSW}} = F$ ⇨
If $F \geq F$-critical, ⇨ there <u>is</u> a difference.
If $F < F$-critical, ⇨ there is <u>no difference</u>.

4. The output includes an ANOVA Table like this:

ANOVA	Cannot Reject Null Hypothesis because $p > 0.05$ (Means are the same.)					
Source of Variation	SS	df	MS	*F*	*p*-Value	*F*-crit
Between Groups	2686.67	2	1343.33	0.376	0.690	3.354
Within Groups	96567.50	27	3576.57			
Total	99254.17	29				

The conclusion of this ANOVA is stated at the top. Prior to the test, Alpha (α) was selected to be 0.05. We see that the p-Value (p) is 0.690, which is greater than Alpha (0.05). So, we do not reject the Null Hypothesis.

Details are given in the table beneath the conclusion about the Null Hypothesis:

"SS" stands for Sum of Squares, and values are given for Between Groups (SSB) and Within Groups (SSW).

"df" is Degrees of Freedom. For Between Groups, df = $k - 1$, where k is the number of groups. In our example, k is 3, so df = 3 – 1 = 2. For Within Groups df = $N - k$, where N is the total number (30) of y measurements, so df = 30 – 3 = 27.

"MS" is Mean Sum of Squares, and values are given for MSB and MSW. You can see that their ratio gives us F.

$F < F$-critical, which is statistically equivalent to $p > \alpha$.

You might remember that the Part 1 article said that an ANOVA assumption was Continuous, not Discrete data. And the data in this example appear to be Discrete, being in increments of dollars. However, Discrete data in money, which tend to have a large number of possible values, are effectively Continuous.

Related Articles in This Book: *ANOVA – Part 1: What It Does*; *ANOVA – Part 2: How It Does It*; *ANOVA – Part 4: 2-Way*; *Variation/Variability/Dispersion/Spread*; *Variance*; *F*; *Sums of Squares*; *Critical Values*; *Alpha(α)*; *p-Value*; *ANOVA vs. Regression*; *p, t, and F: ">" or "<"?*

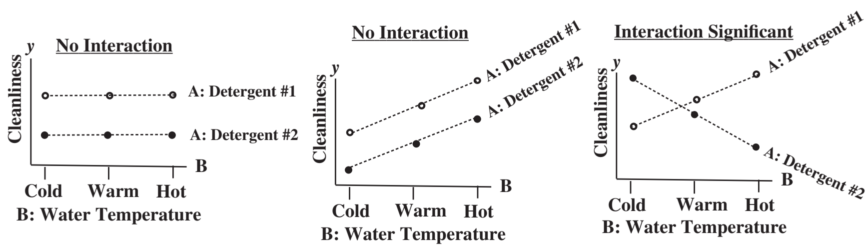
ANOVA – PART 4 (OF 4): 2-WAY (AKA 2-FACTOR)

Summary of Keys to Understanding

Builds on information in the article ANOVA: Part 3 – 1-Way.

1. In 2-Way ANOVA, we study the effect of 2 Nominal (named) Variables, A and B, on the Dependent Numerical Variable, *y*.
A and B are Factors influencing the value of *y*. "AB" – the Interaction between and A and B – can be the 3rd Factor.

2. There are 2 Methods for 2-way ANOVA
- **The WITHOUT Replication method can be used if there is no Interaction which is Statistically Significant.**
- **Otherwise, you must use the WITH Replication method.**



3. 2-Way ANOVA WITH Replication simply repeats (replicates) **the experiment several times** for each combination of A and B values **in order to obtain sufficient data to identify and quantify any Interaction, AB.**

4. In an ANOVA Table, $p \leq \alpha$ indicates Statistical Significance. If the Interaction, AB, is Statistically Significant, then p-values for A and B are not usable.

Explanation

1. In 2-Way ANOVA, we study the effect of 2 Nominal (named) Variables, A and B, on the Dependent Numerical Variable, *y*.

A and B are Factors influencing the value of *y*. "AB" – the Interaction between and A and B – can be the 3rd Factor

2-Way ANOVA is more complicated – and potentially much more confusing – than 1-Way ANOVA. So, we're going to proceed slowly and deliberately with descriptions of the individual elements involved.

First of all, the names used for different types of Variables can be confusing.

We're familiar with equations of the type

$$y = f(x) \text{ or } y = f(x_1, x_2, \ldots , x_n).$$

The value of the Variable y is a function of one or more x Variables. In other words, the value of y is dependent on the value of one or more x's. So, y is called the Dependent Variable. The x's can vary independently and are called Independent Variables.

In 2-Way ANOVA, the equation is of the type

$$\boldsymbol{y = \mathbf{f}(\mathbf{A}, \mathbf{B}, \mathbf{AB})}$$

- ***y* is the Dependent Variable** (also known as the **"Outcome Variable"**).

  ***y* is a Numerical Variable.** That is, its value is **a Number**, like 5, not a Name, like "Detergent #1."
- **A and B are Nominal (named) Variables.** That is, their values are Names (hence "nominal") within a Category. (Nominal Variables are also known as Categorical Variables.)
  - For example, if the Category A is type of detergent, the values of A would be names or labels for two detergents, say "Detergent #1" and "Detergent #2."
  - B, the second Category could be water temperature. It may have values of "Cold," "Warm," and "Hot." Note, that although these names may have corresponding numerical temperatures (say 40, 80, and 120 degrees Fahrenheit) we do no calculations with those numbers. We are naming 3 levels of temperature, but the numbers behind these names are not used.

- **A, B, and AB are Factors.** We don't use the term Independent Variable, in this context, because AB is not independent of A and B. **A and B** are also called **"Main Effects,"** to distinguish them from Interaction Factors like AB.
- **AB is the Interaction** of A and B. It has an effect on the Outcome Variable different from the effects of A or B separately. As we'll see later**, if the Interaction term is Statistically Significant, then the individual effects of A and B cannot be separately measured.**

Interaction:

Sometimes Factors interact synergistically, that is, the effect of the two of them together is more than just the sum of the effects of each individually. For example, some detergents work much better in hot water than in cold water.

Interacting Factors can also cancel each other out – as in two cleaners, one an acid and the other a base.

2. There are 2 Methods for 2-Way ANOVA
- **The WITHOUT Replication method can be used if there is no Interaction which is Statistically Significant.**
- **Otherwise, you must use the WITH Replication method**

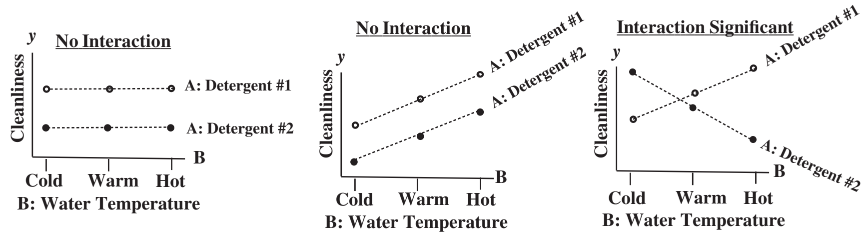
In 1-Way ANOVA, we worked with Samples of data in a Population. In 2-Way ANOVA, we **design experiments** to ensure that we get the kind of data that can be analyzed the way we need. For example, we select 2 different detergents and 3 levels of temperature. The numerical Outcome, *y*, is "Cleanliness," measured on a scale of 0 to 50.

There are 2 methods that can be used for 2-Way ANOVA. **The WITH Replication method is usually better**, because it uses more data and provides more information. However if the experimental budget and time are constraints, **the WITHOUT Replication method can be used, but only if there is no Interaction between the 2 Factors.**

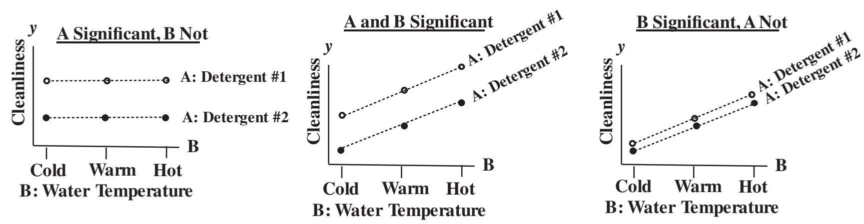
How do we know if there is no Interaction? **Plot the data. If the lines don't intersect there is not a Statistically Significant Interaction.**

**Parallel or roughly parallel lines imply no Interaction.
Crossed lines imply Interaction.**

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in



<u>Separated</u> Lines show that <u>Factor A</u> has an effect on the Outcome Variable *y*.
<u>Slanted</u> Lines show that <u>Factor B</u> has an effect.



If Factor A has i number of values and Factor B has j, then there are $i \times j$ pairs of combinations to test. In this example there are 2 values for A: Detergent and 3 values for B: Water Temperature, so there are $2 \times 3 = 6$ pairs of combinations to test – yielding 6 values of y (the numbers in the table above).

The WITHOUT Replication method measures only one value of y for each of these combinations. Here is the data we would enter into a spreadsheet or software.

	Cold	Warm	Hot
Detergent #1	30	36	45
Detergent #2	20	29	35

Here is the ANOVA Table produced. (The format and labels will vary somewhat by the tool.)

ANOVA Table: 2-Way WITHOUT Replication (Alpha = 0.05)						
Source of Variation	SS	df	MS	*F*	*p*-value	*F*-crit
Rows (A)	121.5	1	121.5	81	0.012	18.51282
Columns (B)	225	2	112.5	75	0.013	19
Error	3	2	1.5			
Total	349.5	5				

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

The key items in the ANOVA table are the *p*-values. In the above example, *p*-values for both Rows (Factor A) and Columns (Factor B) are less than 0.05 (the value selected for Alpha), so both have Statistically Significant effects.

Error is the Variation left over after totaling up the Variations caused by A and by B.

Sum of Squares (SS) is the measure of Variation shown. That column shows how much of the total Variation in y is caused by Factors A (Rows) and B (Columns), and how much is left over as Error.

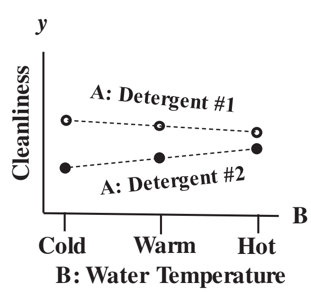
Degrees of Freedom (df), and Mean Sums of Squares (MS) are provided for your information. They are used in interim calculations in producing values for the p-value and F.

And, as is explained in the article, *Critical Values*, $F \geq F$-critical is statistically identical to $p \leq \alpha$. So that information is redundant.

The WITHOUT Replication method has lower experimental costs, but it is limited – **it does not identify or quantify Interactions.**

> **3. 2-Way ANOVA WITH Replication simply repeats** (replicates) **the experiment several times** for each combination of A and B values **in order to obtain sufficient data to identify and quantify any Interaction, AB.**

Suppose we collected data which produced the graph below.



- The lines are separated. But are they separated enough for us to say that Factor B has a Statistically Significant effect?
- The lines are slanted, indicating that Temperature has an effect. But is it a Statistically Significant effect?
- The two lines don't cross, but, if extended, they would. Does this indicate a Statistically Significant Interaction?

The WITHOUT Replication method could answer the first two. But the graph is ambiguous enough that we may want the greater accuracy to be achieved by using more data points, as with the WITH Replication method. That could also answer the question of whether or not there is an Interaction.

The WITH Replication method repeats (Replicates) the experiment several times for each combination of A and B values. That can provide sufficient data to identify and quantify an Interaction. The number of Replications required to achieve a specified level of accuracy is determined by the methods of **Design of Experiments, DOE.** (This book has a 3-part series of articles on DOE.)

Here's the data.

Data for WITH Replication method				
		Factor B		
		Cold	**Warm**	**Hot**
Factor A	**Detergent #1**	40, 42, 39	35, 33, 36	30, 29, 31
	Detergent #2	20, 18, 21	24, 26, 23	28, 27, 29

4. In an ANOVA Table, $p \leq \alpha$ indicates Statistical Significance. If – in the WITH Replication method – the Interaction, AB, is Statistically Significant, then p-values for A and B are not usable.

Here's the ANOVA Table, which is calculated from WITH Replication data above:

ANOVA Table: 2-Way With Replication (Alpha = 0.05)							
Source of Variation	SS	df	MS	F	p-value	F-crit	
Sample (A)	93.4	1	93.4	5.4	0.038	4.7	Reject Null Hypothesis
Columns (B)	310.3	2	155.2	9.0	0.004	3.9	Reject Null Hypothesis
Interaction (AB)	14.8	2	7.4	0.4	0.660	3.9	Do Not Reject Null Hypothesis
Within	206.0	12	17.2				
Total	624.5	17					

The first thing we do is check the *p*-value for the Interaction AB.

If that *p*-value indicates a Statistically Significant Interaction ($p \leq$ Alpha), then the *p*-values calculated for A and B separately would be meaningless. The effects of A and B would be too intertwined to separate them

In this example, the $p > \alpha$, so we do not reject the Null Hypothesis. The Null Hypothesis of No Statistically Significant effect for the Interaction is supported by the analysis.

(Note that there are 3 different Null Hypotheses here: one each for Factor A, Factor B, and the Interaction AB.)

Since there is no Statistically Significant Interaction, **we can check the *p*-values for the two Factors, A and B. If $p \leq \alpha$, then that Factor does have a Statistically Significant Effect.** The Null Hypothesis of No Statistically Significant effect for the Factor is Rejected.

In this example, the *p*-values for both Factors A and B are less than Alpha. So, we Reject the Null Hypothesis and conclude that both the Factors A and B have a Statistically Significant effect on the outcome Variable *y*.

Related Articles in This Book: *ANOVA: Part 1 – What It Is*; *ANOVA: Part 2 – What It Does*; *ANOVA: Part 3 – 1-Way*; *ANOVA vs. Regression*; *Design of Experiments – Parts 1–3*; *F*; *Variation/Variability/Dispersion/Spread*; *p, p-Value*; *Alpha (α)*; *Critical Value*

ANOVA vs. REGRESSION

The purpose of this article is to give you a more intuitive understanding of both ANOVA and Regression by exploring how they are similar and how they differ.

Summary of Keys to Understanding

	ANOVA	Regression
1. Purpose	Determine whether the Means of 2 or more Populations are statistically the same.	Model Cause and Effect; Predict y value from x value(s).
2. Type of Question	Is there a Statistically Significant difference between drugs A, B, and placebo?	How much do house prices increase as the number of bedrooms increases?
3. Variable Types	x: Categorical, y: Numerical	x and y both Numerical
4. Groups Being Compared	Individual Populations (or Samples of each)	data values for the y Variable vs. corresponding y values on the Regression Line
5. Focuses on Variation	Yes	Yes
6. Uses Sums of Squares to Partition Variation	Yes	Yes
7. Variation of …	Means of Different Populations	Dependent Variable (y) vs. Independent Variable(s) (x's)
8. Involves Correlation	No	Yes
9. Sum of Squares Total (SST) =	SSW + SSB	SSR + SSE
10. Key Sum of Squares Ratio	F = MSB/MSW	R^2 = SSR/SST
11. Analysis Output Includes ANOVA Table	Yes	Yes
12. Used Primarily In	Designed Experiments	Inferential Statistics, but validated via Designed Experiments

Explanation

ANOVA and Regression have a number of similarities and differences. The purpose of this article is to give you a more intuitive understanding of both ANOVA and Regression by exploring both how they overlap and how they differ. Let's start with some key differences.

	ANOVA	**Regression**
1. Purpose	Determine whether the Means of 2 or more Populations are statistically the same.	Model Cause and Effect; Predict *y* value from *x* value(s).
2. Type of Question	Is there a Statistically Significant difference between Drug A, Drug B, and Placebo?	How much do house prices increase as the number of bedrooms increase?

ANOVA and Regression differ in their purposes and in the type of question they answer.

ANOVA:

ANOVA is actually more similar to the *t*-test than to Regression. ANOVA and the 2-Sample *t*-test do the same thing if there are only 2 Populations – they **determine whether the Means** of the 2 Populations **are statistically the same or different**.

This, then, becomes **a way of determining whether the** 2 **Populations are the same or different – relative to the question being asked.** ANOVA can also answer the question for 3 or more Populations.

The answer to the question is Yes or No.

Regression:

The purpose of Regression is very different. It attempts to produce a Model (an equation for a Regression Line or Curve) which can be used to **predict the values of the *y* (Dependent) Variable given values of one or more *x* (Independent) Variables.**

Regression goes beyond mere Correlation (which does not imply Causation) to attempt to **establish a Cause and Effect relationship** between the *x* Variable(s) and the values of *y*.

The answer to the question is the equation, for the best-fit Regression Line, e.g., House Price = \$200,000 + (\$50,000 × Number of Bedrooms).

	ANOVA	Regression
3. Variable Types	x: Categorical, y: Numerical	x and y both Numerical

ANOVA

The Independent Variables (x) Must be Categorical (Nominal). That is, the different values of x in the category (e.g., drug) must be names (e.g., Drug A, Drug B, Drug C, Placebo), rather than numbers.

The Dependent Variable (y) must be Numerical, e.g., a blood pressure measurement.

Regression

Both the Independent and Dependent Variables must be Numerical. For example, x is Number of Bathrooms and y is House Price. As mentioned earlier, Regression attempts to establish a Cause and Effect relationship, that is, increasing the number of Bathrooms results in an increase in House Price.

	ANOVA	Regression
4. Groups Being Compared	Individual Populations (or Samples of each)	data values for the y Variable vs. corresponding y values on the Regression Line

Regression really doesn't compare groups as such. But if one wants to explore this similarity between Regression and ANOVA, one would describe Regression concepts in terms used by ANOVA.

We can consider the Sample of paired (x, y) data to represent one group. And the other group consists of corresponding paired (x, y) points on the Regression Line. By "corresponding" we mean having the same x values.

Illustration: 7 pairs of (x, y) data and their corresponding points on the Regression Line.

The Regression Line is $y = 2x$. We take the value of x from a data point, and calculate the y value for the Regression Population using $y = 2x$

Group 1	Data Points (x, y)	(1, 2.5)	(2, 1.9)	(4.7)	(5, 9)	(7, 15)	(8, 18)	(11, 22)
Group 2	Corresponding Regression Points	(1, 2)	(2, 2)	(4, 8)	(5, 10)	(7, 14)	(8, 16)	(11, 20)

ANOVA will compare the Means of the y values of these groups.

	ANOVA	Regression
5. Focuses on Variation	Yes	Yes
6. Uses Sums of Squares to Partition Variation	Yes	Yes

The main conceptual similarity between **ANOVA and Regression** is that they **both analyze Variation to come to their conclusions.**

"Partitioning" Variation Means dividing up the Total Variation – as measured by **Sum of Squares Total (SST)** – into components or portions of the total Variation.

	ANOVA	Regression
7. Variation of …	Means of Different Populations	Dependent Variable (y) vs. Independent Variable(s) (x's)
8. Involves Correlation	No	Yes

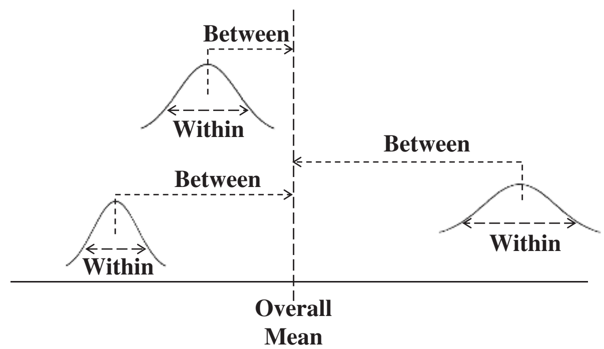
Both ANOVA and Regression use Variation as a tool. For Regression, we know that the Variables x and y vary – that is, all their values in a Sample will not be identical. That is, a Sample will not be something like (2, 3); (2, 3); (2, 3); (2, 3); (2, 3); (2, 3); (2, 3). **The first question for Regression is, do *x* and *y* vary together** – either increasing together, or moving in opposite directions. That is, **is there a Correlation between the *x* and *y* Variables?** If there is not a Correlation, then we will not even consider doing a Regression analysis.

For ANOVA, there is no question of "varying together," because the values of the x Variable – being a Categorical Variable. They don't increase or decrease.

	ANOVA	Regression
9. SST =	SSW + SSB	SSR + SSE

Since ANOVA and Regression measure very different types of Variation, one would expect that the components of their total Variations are very different.

ANOVA: SST = SSW + SSB where SST is **Sum of Squares Total,** SSW is **Sum of Squares Within,** and SSB is **Sum of Squares Between**

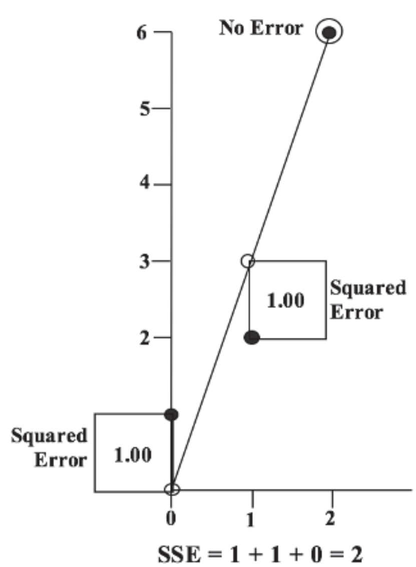


The total Variation (SST) is the sum of all the Variations **within** each of the individual Populations plus the sum of the Variations **between** each Population Mean and the Overall Mean.

Regression:

For Regression, the two components of SST are Sum of Squares Error (SSE) and Sum of Squares Regression (SSR). We use the data to calulate one component, SSE, and to calculate the total, SST. Then, we calculate the other component, SSR from SST and SSE:

In this very simple example, there are only 3 data points in our sample. These are illustrated by the 3 black dots. The 3 data points have *x, y* values of (2,6), (1,2), and (0,1). The Regression line is defined by formula $y = 3x$.

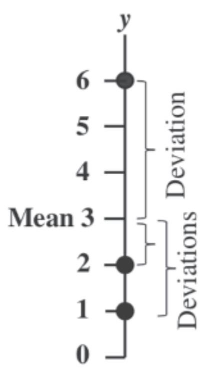


There is no error for the point at the top 2, 6. It is on the Regression line of y = 3*x*.

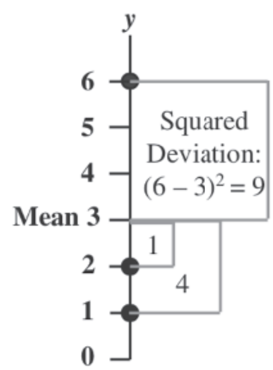
The black dots of the other two points, (1,2) and (0,1) are each one unit away from the Regression line. So, their error is 1 and their squared error is also 1.

And the Sum of these Squared Errors, SSE, is 0 + 1 + 1, which equals 2.

Now, let's look at Sum of Squares Total, SST.



Mean $\bar{y} = 3$		
y	Deviation $y - \bar{y}$	Squared Deviation $(y - \bar{y})^2$
1	1 – 3 = –2	4
2	2 – 3 = –1	1
6	6 – 3 = 3	9
Sum of Squared Deviations $\Sigma(y - \bar{y})^2$		14



SST, is the sum of the squared deviations of the data values of the Variable y to the Mean of *y*.

As shown as black dots in the vertical graph on the left, our 3 data points had y values of 1, 2, and 6.

They are also shown in the first column of the table in the middle.

1 + 2 + 6 = 9, divided by 3 gives us a Mean value of 3 for the y Variable, as stated in the top row of the table.

The middle column of the table calculates the 3 deviations from this Mean, -2, -1 and 3.

And the right column of the table shows the squared deviations, 4, 1, and 9. This is also illustrated in the diagram to the right of the table.

JWST737-A JWST737-Jawlik August 22, 2016 9:59 Printer Name: Trim: 9in × 6in

The sum of the Squared deviations is 4 + 1+ 9 = 14. This is SST, the Sum of Squares Total. Given SST and SSE, we can calculate SSR, the Sum of Squares Regression.

$$\text{SSR} = \text{SST} - \text{SSE} = 14 - 2 = 12$$

	ANOVA	Regression
10. Key Sum of Squares Ratio	$F = \text{MSB/MSW}$	$R^2 = \text{SSR/SST}$

ANOVA: *F* = MSB/MSW,

where MSW is the **Mean Sum of Squares Within** and MSB is the **Mean Sum of Squares Between.** These are calculated by dividing SSW and SSB, respectively, by their Degrees of Freedom. **MSW and MSB are different types of the Statistic, Variance.**

The F-statistic is a ratio of two Variances – MSW and MSB, in this case. Comparing F to its Critical Value tells us whether there is a Statistically Significant difference among the (Means of) the Groups being compared.

SSB ⇨ MSB
SSW ⇨ MSW ⇨ $\dfrac{\textbf{MSB}}{\textbf{MSW}} = F$ ⇨
If $F \geq F$-critical, ⇨ there is a difference.
If $F < F$-critical, ⇨ there is no difference.

Regression: R^2 = SSR/SST,

where **SSR is the Sum of Squares Regression**. SSR is the component of the Variation in the **Total Variation in the *y* Variable (SST)** which is explained by the Regression Line. SSR/SST is the proportion.

R^2 is a measure of the Goodness of Fit of the Regression Line. If R^2 is greater than a predetermined clip level, then the Regression Model is considered good enough, and its predictions can then be subjected to validation via Designed Experiments.

	ANOVA	Regression
11. Analysis Output Includes ANOVA Table	Yes	Yes

Spreadsheets and statistical software often include an ANOVA table in their outputs for both ANOVA and for Regression:

ANOVA Table from a Regression analysis					
	df	SS	MS	*F*	*p*-value
Regression	2	48,845.938	24,422.969	64.040	0.001
Residual	4	1525.490	381.373		
Total	6	50,371.429			

SS for Regression is SSR, SS for Residual is SSE ("Residual" is another name for Error) and SS for Total is SST.

Divide the SS's by the df's (Degrees of Freedom) to get the MS's (Mean Sums of Squares for Regression and Error). The *F*-statistic is MS Regression/MS Residual.

This particular table doesn't show the Critical Value of *F* with which to compare the value of *F*. But it does show the *p*-value, which can be compared to the value we selected for the Significance Level, Alpha (α).

So, in this example ***p*** **is much less than Alpha, so we can conclude that** the results are Statistically Significant. That's another way of saying **the Regression Line is a good fit for the data**. This was confirmed by the value (not shown in the ANOVA table) of $R^2 = 0.893$

	ANOVA	Regression
12. Used Primarily In	Designed Experiments	Inferential Statistics, but validated in Designed Experiments

One of the most significant differences between ANOVA and Regression is in how they are used. ANOVA has a wide variety of uses. It is well-suited for Designed Experiments, in which levels of the *x* Variable can be controlled – for example, testing the effects of specific dosages of drugs.

Regression can be used to draw conclusions about a Population, based on Sample data (Inferential Statistics). The purpose of Regression is to provide a Cause and Effect Model – an equation for a Best Fit Regression line or curve – which predicts a value for the *y* Variable from a value of the *x* Variable(s). Subsequent to that, data can be collected in Designed Experiments to prove or disprove the validity of the Model.

Related Articles in This Book: *ANOVA – Parts 1–4*; *Regression – Parts 1–5*; *r, Multiple R, R^2, R Square, Adjusted R^2*; *Sum of Squares*