The Data Driven Manager

4

Foundations of Hypothesis Testing

Learning Objectives

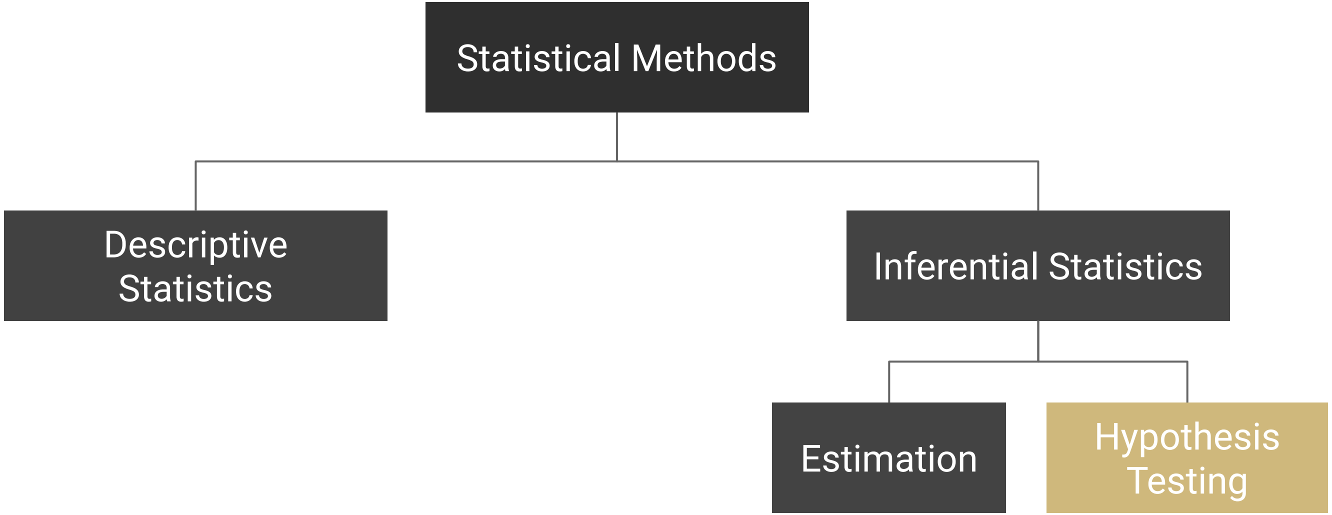
- Recall the basic assumptions and concepts related to hypothesis testing
- Interpret significance level and risk
- Discriminate between one-tail and two-tailed tests
- Differentiate between Type I and Type II Error

Learning Objectives

- Interpret the relationship between Beta and Power
- Calculate Power for Changes in Means and Variance
- Calculate Power for Changes in Proportions and Rates
- Calculate sample size for means, variance and Proportions

Introduction to Hypothesis Testing

Statistical Methods



What is a Hypothesis?

H_0 H_1

- An assumption related to a process or population

Statistical Hypotheses

H_0 H_1

- A **statistical hypothesis**, as opposed to a scientific hypothesis, is a statement related to the value of process or population parameter

Hypothesis Testing

H_0 H_1

- **Hypothesis testing** is a procedure which uses sample statistic(s) to make inferences about a population

Hypothesis Testing Process

H_0 H_1

(Hypothesis)

"The average work efficiency of employees is at least 90 percent."

Population

Parameters

Hypothesis Testing Process

H_0 H_1

(Hypothesis)

"The average work efficiency of employees is at least 90 percent."

Population

Parameters

Statistics

Representative Sample

Hypothesis Testing Process

H_0 H_1

(Hypothesis)

"The average work efficiency of employees is at least 90 percent."

Population

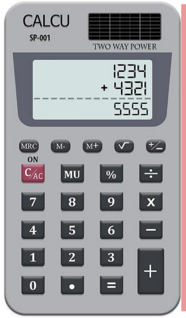
Parameters

Is 46 percent "*close enough*" to 90 percent to support the Hypothesis?

The Sample Data reveals a value of 46%

Reject or Fail to Reject

Probability of the Test Statistics



Test Statistics

Representative Sample

Statistics

Hypothesis Test

Statistical Significance

- When we use the term **“statistically significant,”** we are saying that the observed difference or association/phenomenon represents a significant departure from what might be expected by **chance alone**.

Example

- A compression molding process has been yielding sheets 0.5500" thick, on the average.
- This dimension must be reduced to meet capability requirements.
- The process is modified by the process engineers, tested, and a sample average from 50 randomly selected sheets yields a mean (X-bar) of 0.5300".

Example

- In this example, note that we are essentially asking whether a sample with a mean of 0.530" could have been drawn from a process with a mean (μ) of 0.550"

Example

- Given that the population standard deviation, $\boldsymbol{\sigma}$, was 0.04 and n was 50, we may calculate the following

$$\sigma_{\bar{X}} = \frac{.04}{\sqrt{50}} = 0.00566$$

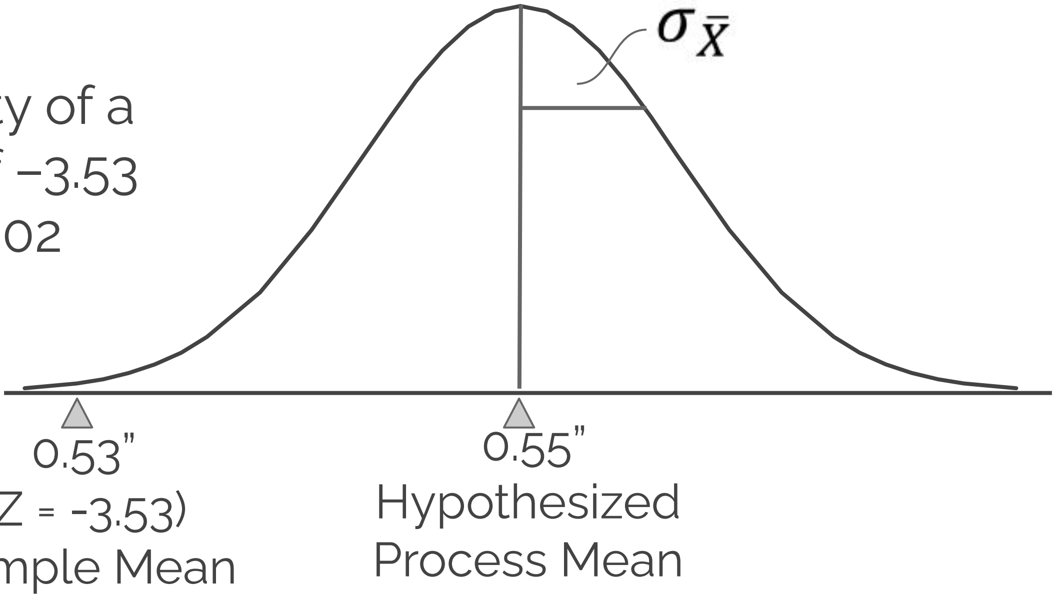
$$Z_{\bar{X}} = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \frac{0.530 - 0.550}{0.00566} = -3.53$$

Example

H_0 H_1

Analogy to
Individual scores

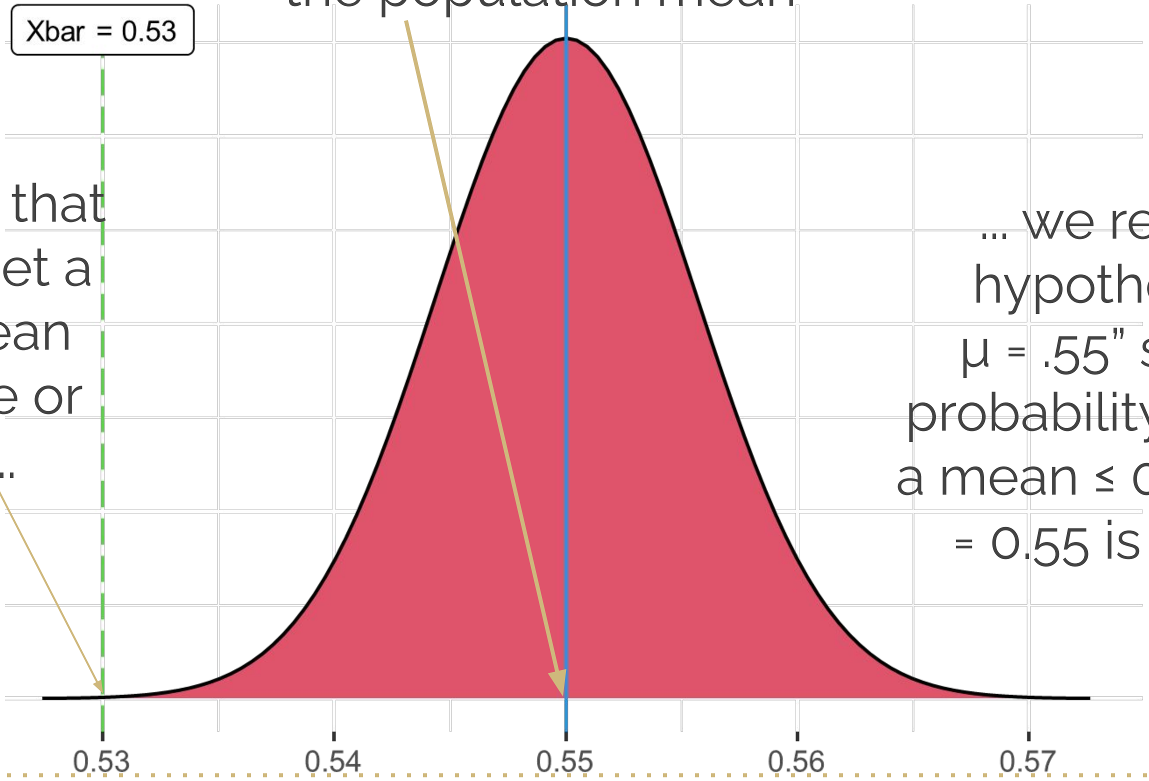
Probability of a
z value of −3.53
is 0.0002



RSD_{Mean}

It is unlikely that we would get a sample mean of this value or smaller …

… if in fact this were the population mean



… we reject the hypothesis that $\mu = .55$" since the probability of getting a mean ≤ 0.53 when $\mu = 0.55$ is so small.

Example

- What conclusion would we make if we observed this sample mean and the sample size was 100? Size 10? Why?

Business World	Data World
Research question driven by business needs	Develop statistical hypotheses
	Design a study and plan what data to collect
	Execute the study and collect the data
	Calculate appropriate sample statistics
	Conduct the appropriate statistical test
	Determine the probability associated with the results of the test statistic
Interpret the results for others	Make a statistical decision

Testing Hypotheses

Testing Statistical Hypotheses

H_0 H_1

- Statistical hypotheses allow researchers to **connect research questions** to the appropriate statistical tests used to answer those questions.

- Statistical hypotheses are generated in pairs, representing all possible outcomes
    - Null hypothesis
    - Alternative hypothesis

Null Hypothesis

H_0 H_1

- The **null hypothesis** always contains a statement of equality and indicates that no difference, no effect, or no relationship exists.
- The symbol for the null hypothesis is $\mathbf{H_0}$
- All hypothesis statements, both null and alternative, use **parametric** symbols

Null Hypothesis

- Presumed to be true unless **sufficient (significant) contradictory evidence** is presented
- Expression of equality: e.g. $=$ or $\geq$ or $\leq$
- H_0 = Status Quo
- Examples:
  - H_0: $\pi = 0.90$
  - H_0: $\mu = 25.75$
  - H_0: $\sigma^2 = 3.85$

H_0 H_1

Alternative Hypothesis

H_0 H_1

- The **alternative hypothesis** is what the researcher wants to detect if it is happening, good or bad
- The alternative hypothesis is a statement that includes all of the conditions NOT represented by the null hypothesis
- When the null hypothesis, H_0, is rejected, the alternative is accepted
- The alternative hypothesis is represented by the symbol H_1

Alternative Hypothesis

H_0 H_1

- Opposite of Null Hypothesis
- Always contains an inequality sign: ≠, <, or >
- Designated H_1 or $H_{A,}$ stated H_1: $\pi \neq 0.90$
- The Null and Alternative Hypotheses as a group must constitute a mutually exclusive, collectively exhaustive set

Directional Hypotheses

H_0 H_1

- Hypotheses may be either **directional** or **non-directional**

- A **directional** hypothesis states that if the null hypothesis is not true, the difference occurs in a specific direction

$$H_0: \quad \mu \leq 50$$
$$H_1: \quad \mu > 50$$

Accepting or Rejecting **Hypotheses**

H_0 H_1

- When testing the null hypothesis, our acceptance of H_0 is based on whether or not we have sufficient statistical evidence to reject it
- We can say we either have or do not have sufficient statistical evidence to reject the null hypothesis
- WHSSETIT - *"We have sufficient statistical evidence to infer that..."*

Accepting or Rejecting Hypotheses

H_0 H_1

- We cannot (will NOT) accept or reject a hypothesis based solely on intuition!
- Rather, we decide objectively, based on representative sample information/data and appropriate statistical analysis, whether to Reject or Not Reject the Null hypothesis

Observations and Cautions

H_0 H_1

- We either accept or reject the H_0; we never have ***proven*** that a difference either does or does not exist
- In essence, we have found that we do or do not have sufficient statistical evidence to accept or reject a hypothesis, respectively
- Development of the hypotheses takes place **before** the collection of the data

Observations and Cautions

H_0 H_1

- **Non-directional** hypotheses will most often lead to a **two-tailed** hypothesis test

- A **directional** hypothesis will lead to a one-tailed hypothesis test

Significance Level and Risk

Alpha Level and Risk

H_0 H_1

- Alpha, **α**, is a selection of risk that you are willing to take
- Given a true null hypothesis, **α** is the probability the null hypothesis could be rejected
- The smaller the selected level of **α** , the smaller the probability of rejecting a true null hypothesis
- Typical choices are 0.10, 0.05, and 0.01

p value

- The significance level, or p-value, is the probability that an observed statistic, or one that is more extreme, could have occurred by chance and chance alone, given a true null hypothesis
- The p-value is generated from calculation in statistical tests and is directly compared to **α**
- We will reject a null hypothesis if the p-value is less than or equal to the selected level of **α**

Test Statistics

- Hypothesis Testing
  - Draw samples
  - Calculate sample statistics
  - Calculate test statistics
  - Calculate probabilities (significance) using the test statistics

Test Statistics

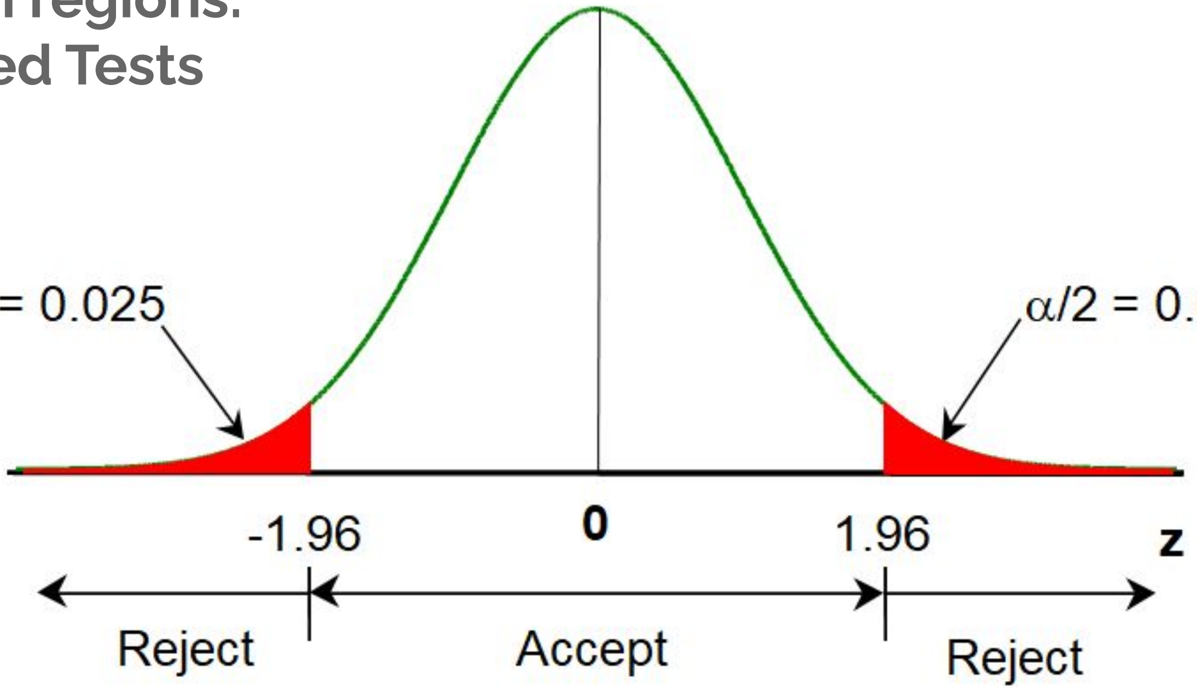
- Some important statistics:
  - z, t, □2 , and F
- Probabilities of test statistics can be approximate or in some cases exact, and require underlying assumptions are met

One and Two Tailed Tests

Two Tailed Tests

- In many instances, the researcher will not be able to make a prediction as to the direction of a possible change.
- In this case, an alternative H structure is appropriate, and the test will have 2 rejection areas.
- For example, if alpha = 0.05, then:

Rejection regions: Two-tailed Tests



What percent of the time would you expect z scores to fall into one of the rejection regions if the null hypothesis is true?

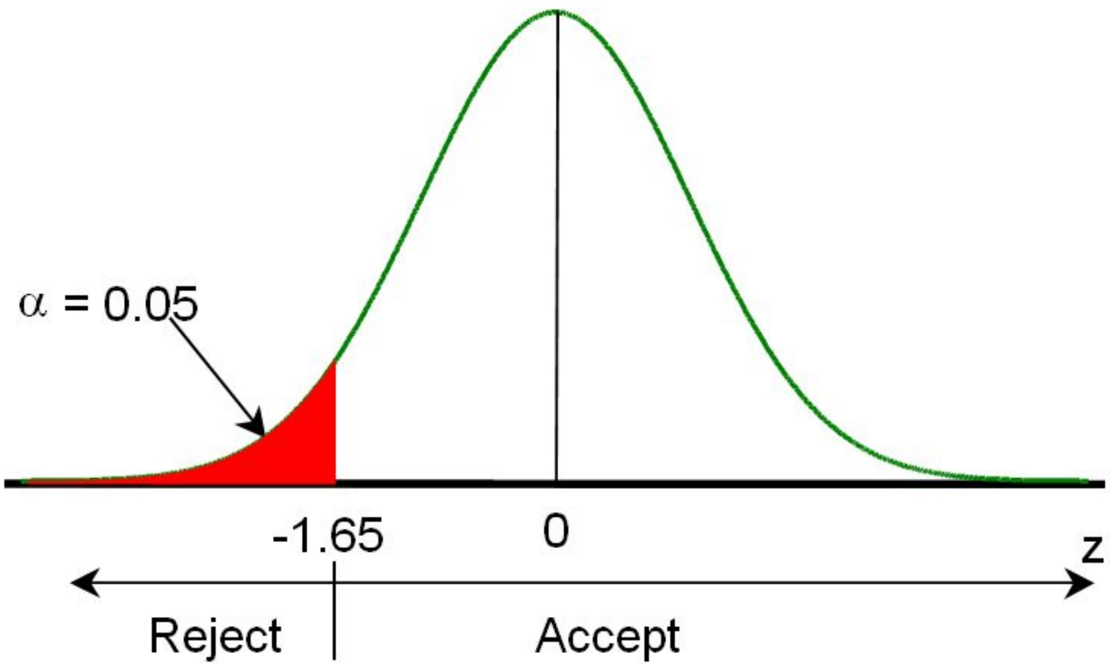
One Tailed Tests

H_0 H_1

- In some cases, an investigator will be able (or forced) to make a prediction based upon a theoretical rationale or prior research.
- In this case, a one-tailed test with a directional hypothesis may be appropriate.
- This choice would result in a single rejection region.

One Tailed Test Example

- Suppose that we do not wish to ship a production lot of armatures unless, based upon a randomly drawn sample from the lot and spin tested, the mean (μ) of the lot may be reasonably assumed to be at least 14,000 rpm to failure
- The hypotheses to be tested are:
  - H_0: $\mu \geq 14{,}000$
  - H_1: $\mu < 14{,}000$



What percent of the time would you expect z scores to fall into one of the rejection regions if the null hypothesis is true?

Cautions for Directional Hypotheses

H_0 H_1

- The one-tailed test will cause the rejection of the H_0 more often than a two-tailed test if the direction predicted is the same as the direction of the difference or association
- Note: there is no possibility of rejecting the null hypothesis if there is a difference in the "wrong direction"
- A directional test should be employed *only when thoroughly justified*

Summary Observations

H_0 H_1

- All decisions regarding maximum risk levels and hypotheses should be made before any data are collected
- When accepting an H_0, we are not saying it is true. We are noting that there is insufficient statistical evidence to reject it.
- We are also, in essence, rejecting the research hypothesis when we accept the null hypothesis.

Setting Up Null and Alternate Hypotheses

H_0 H_1

Non-Directional

- "The average life of light bulbs is 1,000 hours."
  - $H_0 : \mu = 1000$
  - $H_1 : \mu \neq 1000$

NOTE : This is almost always the case.

Setting Up Null and Alternate Hypotheses

H_0 H_1

Directional

- "The average life of light bulbs is equal to or less than 1,000 hours."
  - $H_0 : \mu \leq 1000$
  - $H_1 : \mu > 1000$
- "The average life of light bulbs is equal to or greater than 1,000 hours."
  - $H_0 : \mu \geq 1000$
  - $H_1 : \mu < 1000$

NOTE: There must be sound justification for running a 1-tailed test or testing directional hypotheses!

Error and Power in Hypothesis Testing

H_0 H_1

Definitions

- Type I Error (α):
  - A false signal (false positive)
  - The probability of rejecting a true null hypothesis
- Type II Error (□):
  - Something missed (false negative)
  - The probability of accepting a false null hypothesis

Definitions

- Power (1–□):
  - The ability to detect something
  - The probability of rejecting a false null hypothesis

- Confidence (1–α):
  - Chance of no false signals
  - The probability of accepting a true null hypothesis

Experimental Outcomes

H_0 H_1

Decision	Actual Situation or Reality - H_0	
	H_0 is True	H_0 is False
Accept H_0 as True	Correct Decision Confidence $1 - \alpha$	Type II Error: β
Reject H_0 as False	Type I Error: α	Correct Decision Power $p = 1 - \beta$

Example

Decision	Actual Situation or Reality - H_0	
	No Police with Radar	Police with Radar
Find No Police	😊 Confidence (No False Signal)	😢 Type II Error: (Something Missed)
Find Police	😠 Type I Error: (False Signal)	💪 Real Power Ability to Detect

Observations

- α + □ will almost never equal 1, they are conditional probabilities

- Specifically, α is based upon the premise that H_0 is true, β is predicated on the assumption that H_0 is false

H_0 H_1

Observations

- Both α + □ represent risk. They are an expression of the researcher's willingness to commit an error in their inference.

- Power, the ability of the test to correctly reject a false H_0, must be "purchased" with sample size and with the selection of an appropriate experimental design.

Observations

- α is not "always more important" than β. For example:
- A drug company wishes to test the safety of a new drug formulation. The hypotheses tested are:
  - H_0: The drug is safe
  - H_1: The drug is not safe

- In this case, which type of error is of most concern?

Beta and Power

Beta and Power

H_0 H_1

- Power is the probability of correctly rejecting a False H_0
- Correct decision when H_0 is false
- Power is designated as $1 - \beta$
- Power is used in determining how well a test is working and likely to detect a true effect (or difference)

Beta and Power

H_0 H_1

- Affected by
  - True value of the population parameter
  - Significance level, α
  - Standard deviation, s (or σ)
  - Sample size, n

Considering Power in Experiments

H_0 H_1

- **Strategy I:** Consider α and □ and calculate an appropriate sample size

- **Strategy II:** Select α and then select an economical sample size, with knowledge of the associated power

- When would you use Strategy I?
- When would you use Strategy II?

Calculating Beta and Power for Means

When **σ** is known

Calculating □ and Power for Means

H_0 H_1

- **Assumptions:**
  - The Central Limit Theorem is applicable
  - The RSD's employed will be approximately normal

Calculating □ and Power for Means

H_0 H_1

- Based on the results of an experiment, an engineer has been studying the effects of modifying a part.
- They want to know if the change significantly **reduces** a particular output measure.
- Past data show that the process has been “running” at a mean value of **60**, with a standard deviation of **16.9**.

Calculating □ and Power for Means

- Further, let us assume that a reduction in the mean of at least 5 increments (units) is necessary before the modification becomes cost-effective.
- This is referred to as the **effect size** (Δ) of an experiment.

Calculating □ and Power for Means

H_0 H_1

- Assume further that α was selected at a 0.05 level
- Note also that a one-tailed test has been employed for the hypotheses to be tested, which corresponds to a z of −1.645.

- We will assume that n = 100.
- What would be the power to detect a difference of five units?

Calculating □ and Power for Means

H_0 H_1

- https://casertamarco.shinyapps.io/power/

Calculating □ and Power for Means

Step 1:

- Determine the critical value in the H_0 RSD of means corresponding to the z value for the given value of α

$$Z_{\bar{X}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$-1.645 = \frac{\bar{X} - 60}{\frac{16.9}{\sqrt{100}}} = \frac{\bar{X} - 60}{1.69}$$

$$\bar{X} = 57.22$$

Calculating □ and Power for Means

H_0 H_1

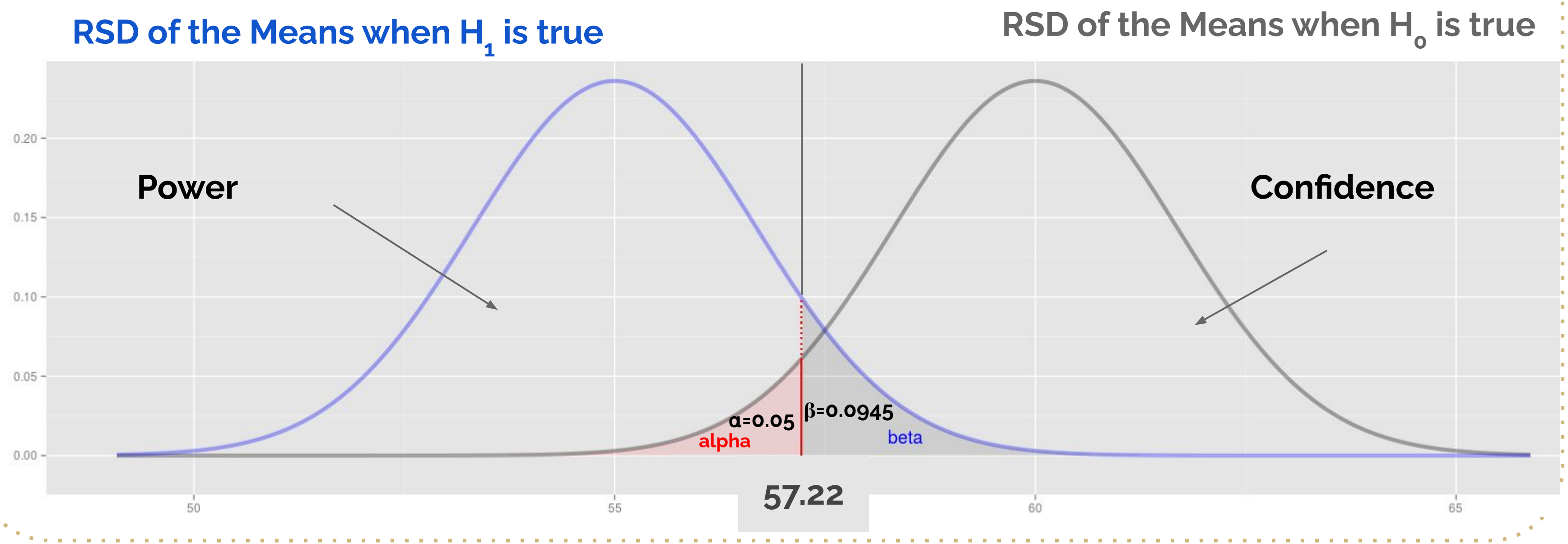
Step 2:

- Calculate the z value and area corresponding to the calculated on the "H_1 is true" curve

$$Z = \frac{57.22 - 55}{1.69} = 1.314$$

Calculating □ and Power for Means

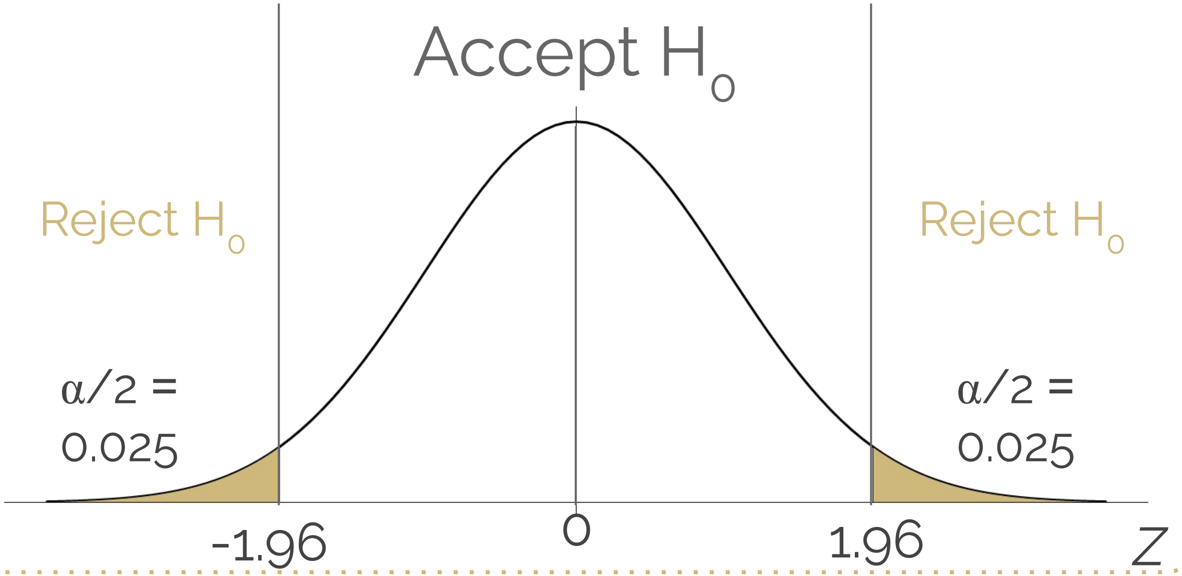
H0 H1



Accept/ Reject H_0?

Do we reject the null hypothesis if the *z*-test statistic is

- 2.00?
- 1.60?
- 1.65?
- -1.80?



Summary Observations

- The larger the value of $\mu_0 - \mu_1$ (delta), the larger power (1 –□) will become
- Generally, both α and □ should be small. In industry, studies planned without an initial regard for □ generally result in low power or high □ values
- It is not possible to commit Type I and Type II errors at the same time

Summary Observations

- Increasing α will generally reduce □
- Increasing n will generally increase the power of the test
- Increasing the power of the test can be accomplished by reducing the standard error through design modifications (for example, matched groups and stratified sampling)

In RStudio

- `power.mean.z.onesample( )`

In ROIStat

- Go to Sample Size, Means
- Select One-sample Mean z
- Toggle the power button 

Calculating **Beta** and Power for **Means**

When **σ** is unknown

Calculating □ and Power for Means

H_0 H_1

- **Assumptions:**
  - The Central Limit Theorem is applicable
  - The RSD's employed will be approximately normal

Calculating □ and Power for Means

H_0 H_1

- You are a manager attempting to determine whether or not a carefully considered improvement activity really increased performance of an assembly component.

- You have no interest in a decrease, regardless of how large it might be.

Calculating □ and Power for Means

H_0 H_1

- As part of these efforts you have designed a small experiment (with a sample size of n=15), executed it, measured the criterion measure, computed the sample mean and standard deviation.
  - X-bar = 4.25
  - s = 1.02
- After testing, the assumption of normality was accepted.

Calculating □ and Power for **Means**

- You then proceeded to use a t-test to test this hypothesis:
  - H_0: $\mu \leq 3.85$
  - H_1: $\mu > 3.85$
- And the test Failed to Reject H_0 at the 0.05 significance level
- What was the **power** (or probability) of detecting a true difference (increase) given a sample of size 15?

In RStudio

- `power.mean.t.onesample( )`

In ROIStat

- Go to Sample Size, Means
- Select One-sample Mean t Independent
- Toggle the power button 

Calculating Beta and Power for Variance

Calculating □ and Power for Variance

- In an improvement situation we have the following:
  - $H_0: \sigma_0 = 15$, $H_1: \sigma_0 \neq 15$
  - Normally distributed
  - n = 9
  - α of 0.05
- We obtained a sample whose mean was 37.5 and Std. Dev. was 17.5.

Calculating □ and Power for Variance

- What was the power to detect a **change** in the variability?

In RStudio

- `power.variance.onesample( )`

In ROIStat

- Go to Sample Size, Variances
- Select Select One-sample Variance
- Toggle the power button 

Calculating Beta and Power for Proportions

Calculating □ and Power for Proportions

H$_0$ H$_1$

- Historically, a process has had a defective rate of π = 0.1 (10%)
- The plant commissioned a team to reduce the defective rate. After several months of work, the team wanted to see if their work had resulted in a change in the defective rate.

Calculating □ and Power for Proportions

H_0 H_1

- They took a random sample of 100 items from the process under normal operating conditions and obtained a proportion defective, p, of 0.05.

- What was the power to detect this **change**? Assume a confidence level of 95%.

In RStudio

- `power.proportions.onesample.exact( )`

In ROIStat

- Go to Sample Size, Proportions
- Select One-sample Proportion - Exact
- Toggle the power button ☑ Power

Calculating Beta and Power for Rates

Calculating □ and Power for Rates

H_0 H_1

- You are on a team responsible improving the quality of potato chips that your company produces.

- Recent complaints have indicated that the number of undesirably dark (possibly burnt) chips in a bag is too high. (The historical average has been shown to be **λ = 3**.)

Calculating □ and Power for Rates

H_0 H_1

- Your team has identified several possible solutions, or actions that are expected to reduce the number of dark chips per bag.
- After acting on each of these possible solutions, a random sample of 70 bags of chips was taken throughout the day shift (representing 8 hours of production).

Calculating □ and Power for Rates

H_0 H_1

- The new average number of dark chips was found to be 2.4286.

- To make sure that did not make a Type I error and say that you have made an improvement when that is not the case, you chose an α of 0.01.

Calculating □ and Power for Rates

H_0 H_1

- What is the power associated with being able to detect this **decrease**?

In RStudio

- `power.count.poisson.onesample.exact( )`

In ROIStat

- Go to Sample Size, Rates
- Select One-sample Poisson - Exact
- Toggle the power button ☑ Power

Sample Size Calculations For Means

Sample Size Calculations

H_0 H_1

- For the industrial researcher, proper sample size is not an opinion

Sample Size Calculations

H_0 H_1

- Calculations exist to determine a proper sample size and these calculations are based upon the confidence and power the researcher desires in order to detect a change or difference of a given size or magnitude

Factors to Consider

H_0 H_1

- The effect size, Delta (Δ)
- The population variance (σ^2)
- The probability of committing a Type I error (α)
- The probability of committing a Type II error (□)
- The number of treatment levels or groups

Calculating Sample Size for Tests of Means

H_0 H_1

- One-sample tests of means, σ known
  - Assumptions:
  - Interval level or higher data
  - Normal distribution
  - One-sample test of means

Calculating Sample Size for Tests of Means

H_0 H_1

Formulas:

- Directional hypotheses

$$n = (z_{\alpha} + z_{\beta})^2 \frac{\sigma^2}{\Delta^2}$$

- Non-directional hypotheses

$$n = (z_{\alpha/2} + z_{\beta})^2 \frac{\sigma^2}{\Delta^2}$$

H_0 H_1

Example

- A Black Belt is studying the effect of a particular design modification.
- They want to know if it has effectively increased performance of a part by at least 7.5 units.
- The μ performance in the past has been 75, with a σ of 15. Assuming that an α of 0.05 and a □ of 0.20 are desired, how large a sample of parts must be drawn in order to test the hypothesis?

H_0 H_1

Example

$$n = (1.645 + 0.842)^2 \frac{15^2}{7.5^2} = 24.7 \text{ or } 25$$

- A Black Belt is studying the effect of a particular design modification.

- They want to know if it has effectively increased performance of a part by at least 7.5 units.

- The μ performance in the past has been 75, with a σ of 15. Assuming that an α of 0.05 and a □ of 0.20 are desired, how large a sample of parts must be drawn in order to test the hypothesis?

In RStudio

- `sample.size.mean.z.onesample( )`

In ROIStat

- Go to Sample Size, Means
- Select One-sample Mean z

Calculating Sample Size for Tests of Means

H_0 H_1

- In the case in which **σ** is unknown, it must be estimated

- This may be done by taking a small sample and estimating **σ**, by possibly using an upper end of a confidence interval

Calculating Sample Size for Tests of Means

H_0 H_1

- When **σ is unknown**, hypothesis tests for means will use the **t distribution**
- Unfortunately, the t distribution is based upon degrees of freedom, which is determined by sample size

Calculating Sample Size for Tests of Means

H_0 H_1

- Therefore, sample size must be solved iteratively, where the sample size is determined to be the smallest n that satisfies the following formula

Calculating Sample Size for Tests of Means

Formulas:

- Directional hypotheses

$$n \geq \left(t_{\alpha,(n-1)df} + t_{\beta,(n-1)df}\right)^2 \frac{\sigma^2}{\Delta^2}$$

- Non-directional hypotheses

$$n \geq \left(t_{\alpha/2,(n-1)df} + t_{\beta,(n-1)df}\right)^2 \frac{\sigma^2}{\Delta^2}$$

Calculating Sample Size for Tests of Means

- This sample size calculation will produce slightly larger values than the above equation where **s** was considered **known**

- For most reasonable selections of α and □, the sample size will be within about three units

Calculating Sample Size for Tests of Means

H_0 H_1

- One-sample tests of means, σ unknown
  - Assumptions:
  - Interval level or higher data
  - Normal distribution
  - One-sample test of means

In RStudio

- `sample.size.mean.t.onesample( )`

In ROIStat

- Go to Sample Size, Means
- Select One-sample Mean t Independent

Example

- If the requirements of a pull test are to be α = 0.05, β = 0.02, $\Delta\mu$ = 1 lbs, and σ = 2, what would the appropriate minimum sample size be for a non-directional test for means?

Sample Size Calculations For Variance

Calculating Sample Size for Tests of Variance

H_0 H_1

- Formula for non-directional hypotheses

$$\chi^2 = \frac{s^2(n-1)}{\sigma_0^2}$$

Calculating Sample Size for Tests of Variance

H_0 H_1

- For a non-directional test, we must consider two cases
    - One in which the variance increases
    - One in which the variance decreases

In RStudio

- `sample.size.variance.onesample( )`

In ROIStat

- Go to Sample Size, Variances
- Select One-sample Variance

Example

- If the requirements of a pull test are to be α = 0.05, β = 0.02, $\Delta\sigma$ = 1 lbs, and σ = 2, what would the appropriate minimum sample size be for a non-directional test for variances?
  - If the variance increases, σ = 3 (2 + 1)
  - If the variance decreases, σ = 1 (2 - 1)

Sample Size Calculations For Proportions

Calculating Sample Size for Tests of Proportions

H_0 H_1

- For a non-directional test, again, we must consider two cases
    - One in which the proportion increases
    - One in which the proportion decreases

In RStudio

- `sample.size.proportion.test.onesample.exact( )`

In ROIStat

- Go to Sample Size, Proportions
- Select One-sample Proportion - Exact

Example

- Assume that a company has been purchasing material from a particular supplier and the use of that material is associated with a current product defective rate π_0 of 0.03 (3%).

Example

- They have the opportunity to purchase the same material from a different supplier at a much lower cost.

- However, there are rumors that the defective rate that may result from the use of the new supplier's product may increase.

Example

- Since the company has the ability to conduct 100% inspection in an efficient and fast manner, they have conducted a cost/benefit analysis and determined that they will be ahead so long as the defective rate does not double (that is, it must be less than or equal to 0.06 or 6%).

Example

- This means π_0 is 0.03 and π_1 is 0.06.
- Since they have the opportunity to conduct a study involving a trial of the new supplier's material, they need to do a sample size calculation to determine how many units of product need to be run and tested to ascertain whether or not the defective rate had increased to 0.06 or more.

Example

- They want to hold **α** at 0.05 and □ at 0.05 (meaning having .95 or 95% Power) as well.