# LECTURE 3 – INFERENTIAL STATISTICS

Part C - Hypothesis Testing

**Book Chapter 9**

# Hypothesis Testing

- Hypothesis Testing is basically a **decision rule**, that enables you to *make a decision about a population parameter value* (rather than estimating it), based on sample evidence.
  - You can use Confidence Intervals for Hypothesis testing

- The test begins by considering two **hypotheses**. They are called the **null hypothesis** and the **alternative hypothesis**. These hypotheses contain opposing viewpoints.

- **$H_0$: The null hypothesis:** It is a statement about the population that either is believed to be true or is used to put forth an argument unless it can be shown to be incorrect beyond a reasonable doubt.

- **$H_a$: The alternative hypothesis:** It is a claim about the population that is contradictory to **$H_0$** and what we conclude when we reject **$H_0$**.

- Since the null and alternative hypotheses are contradictory, you must examine evidence to decide if you have enough evidence to reject the null hypothesis or not. The evidence is in the form of sample data.

- After you have determined which hypothesis the sample supports, you make a **decision.** There are two options for a decision. They are "**reject $H_0$**" if the sample information favors the alternative hypothesis or "**do not reject $H_0$**" or "**decline /fail to reject $H_0$**" if the sample information is insufficient to reject the null hypothesis.

# Hypothesis Testing

| $H_0$ | $H_a$ |
|---|---|
| equal (=) | not equal (≠) **or** greater than (>) **or** less than (<) |
| | less than (<) |
| | more than (>) |

- **$H_0$** *always has a symbol with an equal in it.*
  - For our purposes, we will always work with an = sign for the null hypothesis
- **$H_a$** *never has a symbol with an equal in it.*
  - The choice of symbol depends on the wording of the hypothesis test.
- We write this as:
  - **$H_0$:μ** = **$\mu_0$** where **$\mu_0$** is the hypothesized population mean
  - **$H_a$:μ** (( ≠ , >, or <) **$\mu_0$**

# Hypothesis Test – Probabilities of Correct & Incorrect Decisions

- Suppose you hypothesize a population mean $\mathbf{H_0{:}\mu = 10}$ against $\mathbf{H_a{:}\mu \neq 10}$
- Let us say that in fact the *unknown* population mean $\boldsymbol{\mu} = \mathbf{10}$.
  - <u>Incorrect Decision/Outcome:</u>
  - Suppose you decide on the basis of the hypothesis test conducted on the sample that *you will reject* $\mathbf{H_0{:}\mu = 10}$. Since you do not know the true population $\boldsymbol{\mu}$, there is a chance that you are *incorrectly rejecting* a true null hypothesis. This type of error is called a **Type I error**, and the probability of committing this error is called *level of significance* $\boldsymbol{\alpha}$.
  - <u>Correct Decision/Outcome:</u>
  - The probability that you are making a correct decision (Not rejecting $\boldsymbol{\mu} = \mathbf{10}$), therefore, is (1- $\boldsymbol{\alpha}$)
- Let us say that in fact the *unknown* population mean $\boldsymbol{\mu} \neq \mathbf{10}$.
  - <u>Incorrect Decision/Outcome:</u>
  - Suppose you decide on the basis of the hypothesis test conducted on the sample that you *will not reject* $\mathbf{H_0{:}\mu = 10}$. Since you do not know true population $\boldsymbol{\mu}$, there is a chance that you are *incorrectly not rejecting* a false null hypothesis. This type of error is called a **Type II error**, and the probability of committing this error is called $\boldsymbol{\beta}$.
  - <u>Correct Decision/Outcome:</u>
  - The probability that you are making a correct decision (Rejecting $\boldsymbol{\mu} = \mathbf{10}$), therefore, is (1- $\boldsymbol{\beta}$) and is called the *Power of the Test.*

# Hypothesis Test – Probabilities of Correct & Incorrect Decisions

| | **$H_0$ IS ACTUALLY** | **...** |
|---|---|---|
| **Decision** | True | False |
| **Do not reject $H_0$** | Correct Outcome | Type II error |
| **Reject $H_0$** | Type I Error | Correct Outcome |

- The four possible outcomes of a Hypothesis Test, as show in the table, are:
  1. **No error**: The decision is **not to reject $H_0$** when **$H_0$ is true.** The probability of this **correct outcome** = **1 - α**.
  2. **Type I error**: The decision is to **reject $H_0$** when **$H_0$ is true.** **α** = probability **P(Type I error).** This is also called the **Level of Significance**.
  3. **Type II error**: The decision is **not to reject $H_0$** when, in fact, **$H_0$ is false.** **β** = probability **P(Type II error)**.
  4. **No error**: The decision is to **reject $H_0$** when **$H_0$ is false.** The probability of this **correct outcome** is called the **Power of the Test** = **1 - β**.

# Hypothesis Test – Probabilities of Correct & Incorrect Decisions

- Things to Remember about Hypothesis Tests:
  - Even though the Table showed outcomes of the Hypothesis Test, we will never know if our decision is correct or not, *because we do not know the (null) hypothesized value of the population parameter.*
  - We can try to control the probabilities of **Type I** and **Type II** errors i.e., (**α** and **β** by keeping them as small as possible.
  - However, they are rarely zero and there is a direct tradeoff between them **<u>for a given sample size</u>**. Decreasing **α** will increase **β** and decrease the power of the test, *vice versa.*
  - Increasing the sample size can increase the Power of the Test.
  - **α** and **β** are decided upon *before the sample is collected* and the test is conducted, based on the <u>relative costs</u> of committing Type I error and Type II error.

# Deciding on which Error is more serious.

- **Example**: Suppose the null hypothesis, $\mathbf{H_0}$, is: Frank's rock climbing equipment is safe.
  - **Type I error**: Frank thinks that his rock climbing equipment may not be safe and so does not use it when, in fact, it really is safe.
  - **$\alpha$ = probability** that Frank thinks his rock climbing equipment may not be safe when, in fact, it really is safe.
  - **Type II error**: Frank thinks that his rock climbing equipment may be safe and uses it when, in fact, it is not safe.
  - **$\beta$ = probability** that Frank thinks his rock climbing equipment may be safe and uses it when, in fact, it is not safe.
- Notice that, in this case, <u>the error with the greater consequence is the Type II error</u>. (If Frank thinks his rock climbing equipment is safe, he will go ahead and use it.)
- So Frank should try to minimize $\boldsymbol{\beta}$ at the expense of an increased $\boldsymbol{\alpha}$.

# Performing Hypothesis Tests

## Before Collecting the Sample:

1. Decide on the population parameter value to be tested (null hypothesis value)
2. Decide on the nature of the alternative hypothesis. The test will therefore be (i) a two-sided test, (ii) lower-tailed test or (iii) upper-tailed test. Set up the Null and Alternate Hypotheses.
3. Decide on the level of significance **α**. (more sophisticated approaches may decide on what Power is needed )
4. Decide on the *statistic* to estimate the population parameter and identify its *sampling distribution* under the null hypothesis (i.e., **assuming null hypothesis is true**).
5. Convert the statistic to a *test statistic* and identify its sampling distribution, under the null hypothesis.
6. Calculate the *critical value* of the test statistic based on knowing the test statistic, its sampling distribution and **α**. Identify the associated *rejection region(s).*

## Collect the Sample

Collect the sample and calculate the test statistic. Calculate the **p-value**.

## Perform the Hypothesis Test

Compare the *magnitude* of the test statistic against the *magnitude* of the critical value.

**Reject the Null Hypothesis** if:

If the magnitude of the test-statistic > magnitude of the critical value;

Or

If **p-value** < **α**

Otherwise, **Fail to reject** the null hypothesis.

## Interpret the Results of Hypothesis Test

# Hypothesis Test of Unknown Population Mean $\mu$

| | Population $\sigma$ known | Population $\sigma$ unknown |
|---|---|---|
| Population Normal | $\bar{X} \sim \text{Normal}(\mu, \frac{\sigma}{\sqrt{n}})$. Use the standard normal z-distribution regardless of sample size. Case 1 | • $\bar{X} \sim \text{Normal}(\mu, \frac{S}{\sqrt{n}})$ if sample size >= 30. <br>• $\bar{X} \sim t_{n-1}(\mu, \frac{S}{\sqrt{n}})$ if sample size < 30. Case 2 |
| Population Not Known Normal | If sample size n > 30: <br>• $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ if $\sigma$ is known <br>• $\bar{X} \sim t_{n-1}(\mu, \frac{S}{\sqrt{n}})$ if $\sigma$ is not known <br>*If sample size << 30, we really have to assume the population distribution. To avoid this, collect a larger sample.* Case 3 & 3a | Obtain sampling distribution of $\bar{X}$ using Bootstrapping. (we will see later) |

1. The Null Hypothesis is $\mathbf{H_0{:}\mu = \mu_0}$.
2. Decide on the nature of the alternative hypothesis $\mathbf{H_a{:}\mu}$ (( $\neq$ , >, or <) $\boldsymbol{\mu_0}$
3. Decide on the level of significance ($\boldsymbol{\alpha}$)
4. $\bar{X}$ will be the statistic to estimate $\boldsymbol{\mu}$. Identify its sampling distribution of $\bar{X}$. These are the same rules as for confidence Intervals. Let us assume that it is $\mathbf{N(\mu_0, \frac{\sigma}{\sqrt{n}})}$. Notice that the sampling distribution of the estimator $\bar{X}$ is **centered on the value of the null hypothesis $\boldsymbol{\mu_0}$.**
5. Convert the statistic to a *test statistic*. If $\bar{X} \sim \mathbf{N(\mu_0, \frac{\sigma}{\sqrt{n}})}$, then we can convert this normal distribution to a standard normal distribution as $\mathbf{z} = (\bar{X} - \boldsymbol{\mu_0})/(\frac{\sigma}{\sqrt{n}})$. We will call it the **z-statistic**. $\bar{X}$ centered on $\boldsymbol{\mu_0}$ and correspondingly $\mathbf{z}$ is centered on 0.

# Performing the Hypothesis Test

6. Using $\alpha$ and the nature of the alternate hypothesis $H_a$:$\mu$ (( $\neq$ , >, or <) $\mu_0$ calculate the critical value ($Z_\alpha$ or $Z_{\alpha/2}$) and the **rejection region**.

7. We **reject** the null hypothesis, if the *test statistic* falls in the rejection region as evidenced by

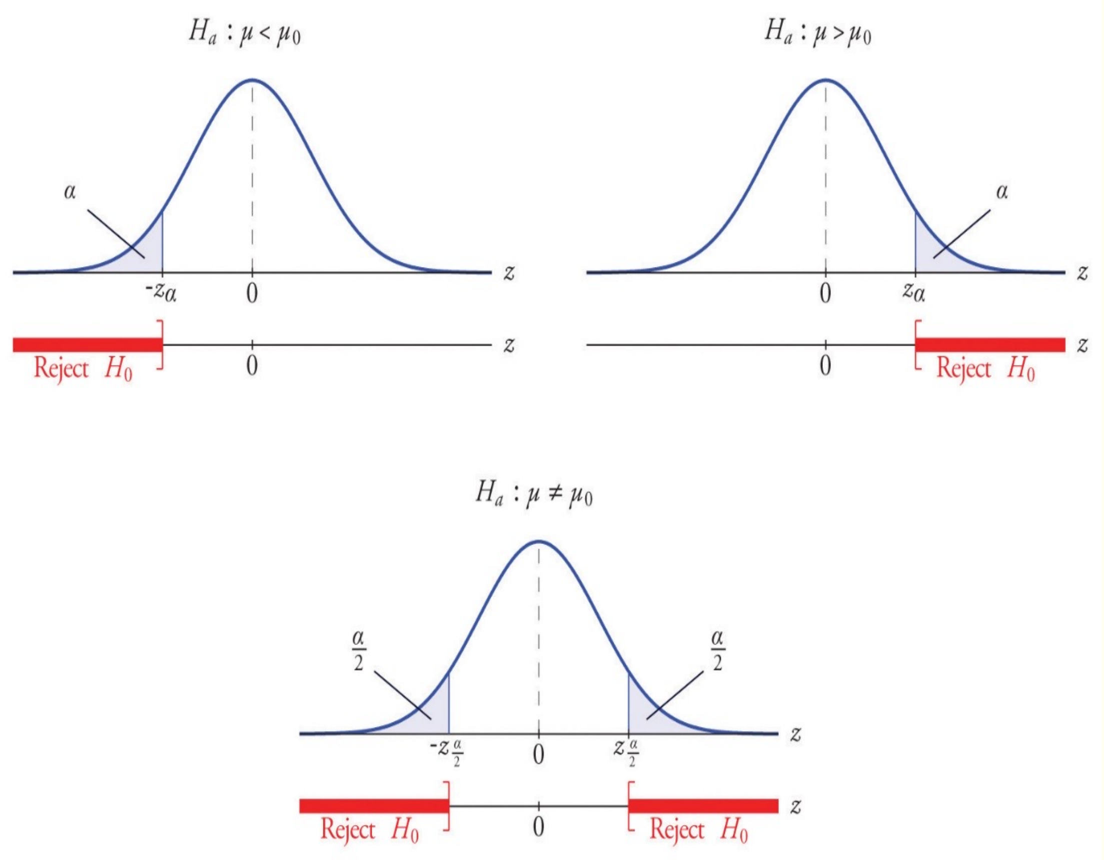
   *Test statistic < critical value at* $\alpha$ *for a lower-tailed test*

   *Test statistic > critical value at* $\alpha$ *for upper-tailed test*

   *Test statistic < critical value at* $\alpha/2$ **or** *Test statistic > critical value at* $\alpha/2$ for two-tailed test

8. If the test statistic (**z**-statistic) does not fall in the rejection region, we *Fail to Reject* $H_0$:$\mu = \mu_0$.

9. Alternatively, we **reject** the null hypothesis if the **p-value** < $\alpha$ for all three types of the test



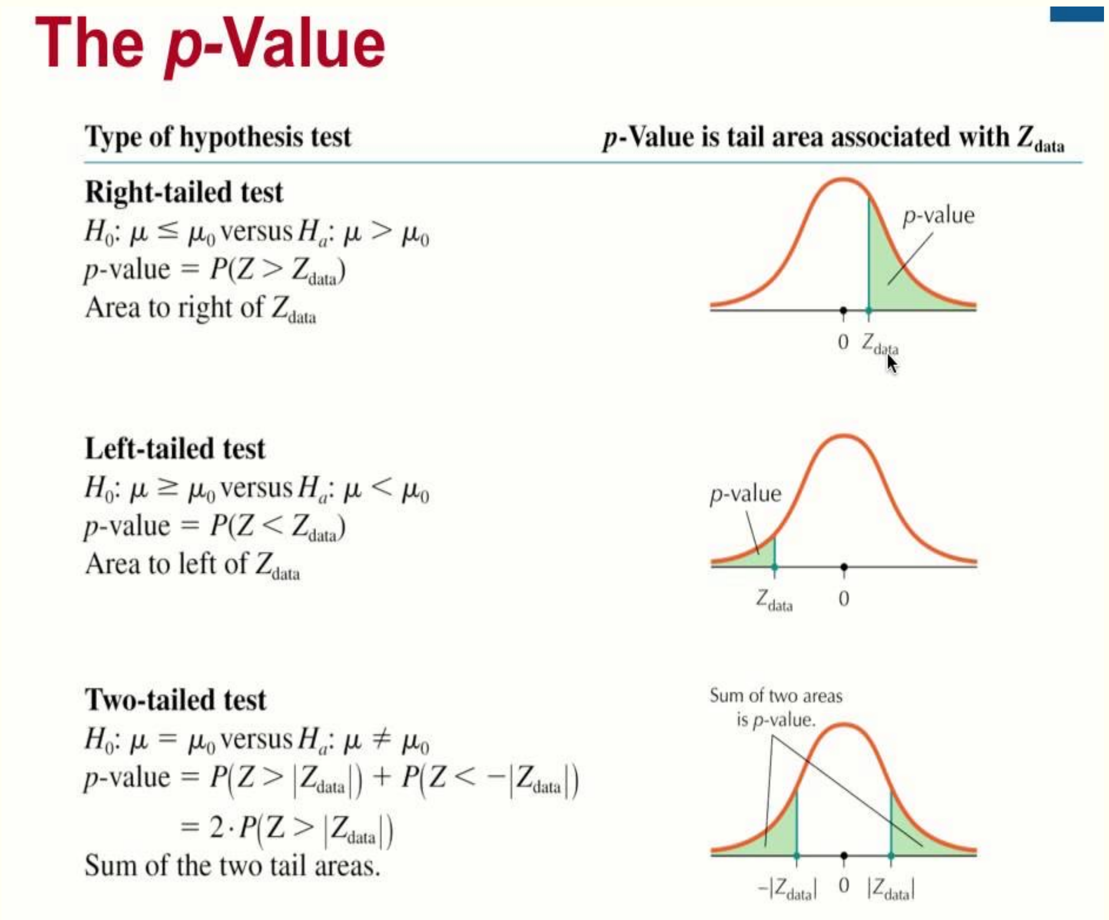

For $\alpha$ = **0.01**, $Z_\alpha$ = 2.33 and $Z_{\alpha/2}$ = 2.58

For $\alpha$ = **0.05**, $Z_\alpha$ = 1.645 and $Z_{\alpha/2}$ = 1.96.

For $\alpha$ = **0.10**, $Z_\alpha$ = 1.29 and $Z_{\alpha/2}$ = 1.645.

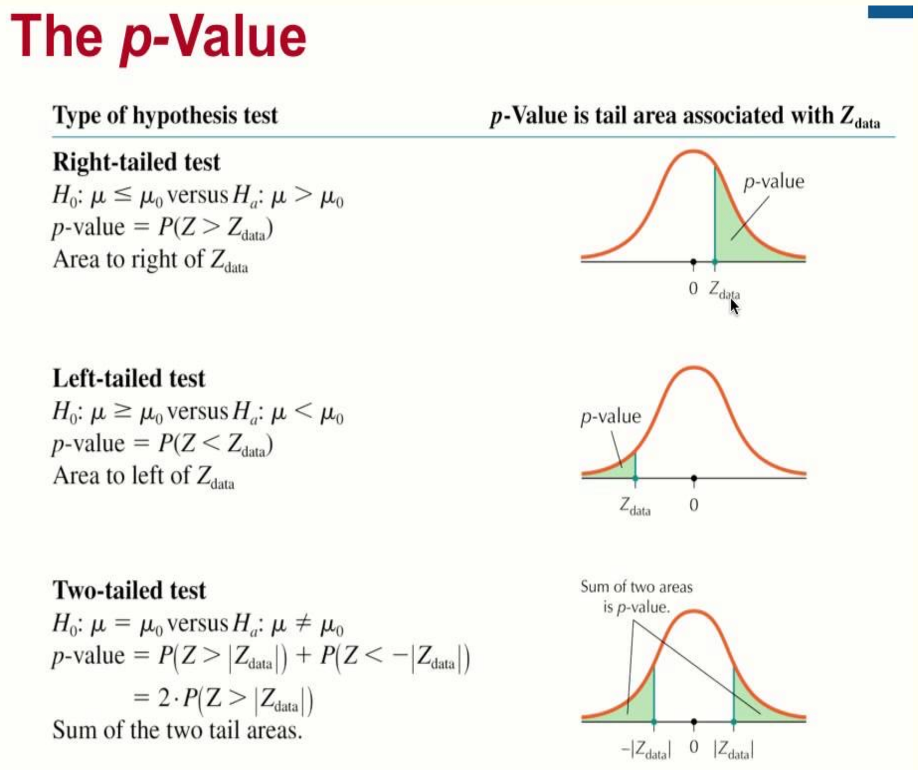
# The Test Statistic and its p-value

- Recall that Hypothesis Testing makes direct use of the Sampling Distribution of the statistic $\bar{X}$, **centered on the null hypothesized mean $\mu_0$**.
- Therefore, the distribution of the *test statistic* ($\bar{X}$ - $\mu_0$)/(Standard Error of Statistic $\bar{X}$ ) has a mean of 0 and a standard deviation of 1.
- It is usually a standard normal distribution (z) or the t-distribution (t), depending on the assumptions about the population.
- The probability of observing a value ***less than or equal to*** the test statistic is called the **p-value** of the test statistic.
- The p-value of a statistic is calculated **assuming** that the **null-hypothesis is true**.
- If **p-value < α** the test-statistic is in the rejection region and we **reject** the null-hypothesis.



# The Test Statistic and its p-value

- To calculate p-value:
  - Calculate the test-statistic based on ($\bar{X}$ - $\boldsymbol{\mu_0}$)/(Standard Error of Statistic $\bar{X}$). It will be a Z or a t value.
  - Check the type of test
    - Upper-tailed, find the probability of the test statistic *to the right* in the Z- or t- distribution
      i.e., as 1 – CDF or 1 – P(Z <= test-statistic)
    - Lower-tailed, find the probability of the test statistic *to the left* in the Z- or t- distribution
      i.e., as CDF or P(Z <= test-statistic)
    - Two-tailed, find the probability of the |*test statistic*| *to the right and multiply by 2.*
      i.e., 2(1 - P(Z > |*test-statistic*|))
    - |*test-statistic*| means magnitude only, ignore sign.
  - If the test statistic is a Z distribution, then use tables to find the closest p-value
  - If the test statistic is a t distribution, you may not be able to use the tables. You will have to get it from **R.**

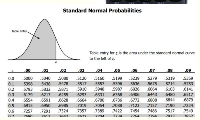


# Relationship between Hypothesis Tests & Confidence Intervals

- Hypothesis Testing uses many of the same concepts as Confidence Intervals
- In Confidence Intervals we are providing an interval estimate of the population parameter (say $\boldsymbol{\mu}$) using the sampling distribution of the sample statistic (say $\overline{X}$) and its standard error (say $\frac{\boldsymbol{\sigma}}{\sqrt{n}}$).
- In Hypothesis Testing we are testing the claim that the population parameter has a particular value (say $\mathbf{H_0{:}\mu = \mu_0}$) using the sampling distribution of the sample statistic (say $\overline{X}$) and its standard error (say $\frac{\boldsymbol{\sigma}}{\sqrt{n}}$) and converting $\overline{X}$ to a test statistic.
- In both, we are faced with the probability that we are incorrect in the estimate of $\boldsymbol{\mu}$ or concluding that $\boldsymbol{\mu = \mu_0}$. This is reflected by $\boldsymbol{\alpha}$.
- In Confidence Intervals
  - Probability ($\boldsymbol{\mu} - z_{\alpha/2} * \frac{\boldsymbol{\sigma}}{\sqrt{n}} \leq \overline{X} \leq \boldsymbol{\mu} + z_{\alpha/2} * \frac{\boldsymbol{\sigma}}{\sqrt{n}}$) = $(1 - \alpha)\%$
- In Hypothesis Testing
  - Probability $(\overline{X} - \boldsymbol{\mu_0})/(\frac{\boldsymbol{\sigma}}{\sqrt{n}})$ is used to obtain the p-value which is then compared with $\alpha$.
- We can use a Confidence Interval for a two-tailed hypothesis Test as follows:
  - If the hypothesized value $\boldsymbol{\mu_0}$ is not contained in the $(1 - \alpha)\%$ Confidence Interval we reject the null hypothesis $\mathbf{H_0{:}\mu = \mu_0}$ and conclude that $\mathbf{H_a{:}\mu \neq \mu_0}$ similar to a two-tailed test. Else, we fail to reject the null hypothesis.

# HYPOTHESIS TEST - PROBLEMS

Practice with calculator 

Practice with Tables 

# Hypothesis Testing – Book Example 9.14 – Page 514

- Jeffrey, as an eight-year old, established a **mean** time of **16.43** seconds for swimming the 25-yard freestyle, with a **standard deviation** of **0.8** seconds. His dad, Frank, thought that Jeffrey could swim the 25-yard freestyle *faster using goggles*. Frank bought Jeffrey a new pair of expensive goggles and timed Jeffrey for 15 25-yard freestyle swims.
For the 15 swims, Jeffrey's **mean time** was **16** seconds. Frank thought that the goggles helped Jeffrey to swim faster than the 16.43 seconds. Conduct a hypothesis test using a preset $\boldsymbol{\alpha} = 0.05$. Assume that the swim times for the 25-yard freestyle are **normal**. Also, construct the 95% confidence interval and calculate the **p-value** and compare it with $\boldsymbol{\alpha}$.

- **Solution:**

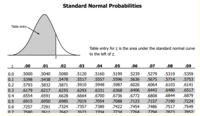
- The Null hypothesis is, $\mathbf{H_0}$: $\boldsymbol{\mu} = \boldsymbol{\mu_0} = 16.43$ seconds.

- The Alternative hypothesis, $\mathbf{H_a}$: $\boldsymbol{\mu} < 16.43$ seconds because Frank thought that the goggles helped Jeffrey to swim faster

- Therefore, we have a *lower-tailed test* at $\boldsymbol{\alpha} = 0.05$.

- We notice that the population standard deviation is known and the population is assumed normal. Therefore, we can assume that the sampling distribution of the estimator $(\bar{X})$ is Normal $(\boldsymbol{\mu_0}, \frac{\boldsymbol{\sigma}}{\sqrt{n}})$ and we can use the **z**-distribution.

Practice with calculator 

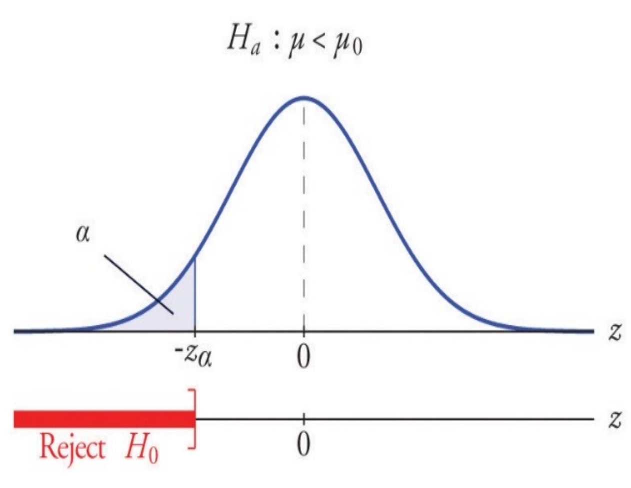
Practice with Tables 

# Hypothesis Testing – Book Example 9.14 – Page 514

- Calculate the test **z**-statistic **z** = $(\overline{X} - \mu_0)/(\frac{\sigma}{\sqrt{n}})$ = (16 – 16p.43 )/($\frac{\mathbf{0.8}}{\sqrt{15}}$) = -2.09173

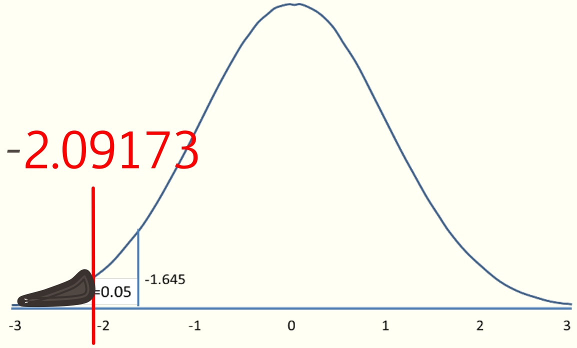
- At **α** = 0.05, the critical value = -1.645.

- Because test **z**-statistic -2.0917 is less than -1.645 (or the magnitude of test **z**-statistic is greater than magnitude of critical value), and falls in the critical (rejection) region, the null hypothesis **$H_0$: μ** = 16.43 seconds is **rejected**.
  - *At the 5% significance level, we conclude that Jeffrey swims faster using the new goggles. The sample data show there is sufficient evidence that Jeffrey's mean time to swim the 25-yard freestyle is less than 16.43 seconds.*




- The **p-value** = P(**Z** ≤ test **z**-statistic) = P(**Z** ≤ -2.0917) = 0.0182. So, since the **p-value= 0.0182** < **α** = 0.05 we **reject** the null hypothesis

- Using confidence intervals:
  The 95% CI for **μ** = ($\overline{X}$ - 1.96*$\frac{\sigma}{\sqrt{n}}$ ≤ **μ** ≤ $\overline{X}$ + 1.96*$\frac{\sigma}{\sqrt{n}}$) = (15.6, 16.40)
  Since the hypothesized **μ** = 16.43 is *outside the interval*, we **reject** the null hypothesis.

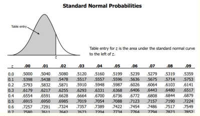



**p-value= 0.0182**

- The probability that we **reject** the null hypothesis incorrectly is **α** = 0.05.

- Suppose we set stricter standards and set **α** = 0.01, we *would not reject the null hypothesis* because p-value (0.0182) ≮ (not less than) α = 0.01

Practice with calculator 

Practice with Tables 


# Hypothesis Testing – Book problem 9.14 – Page 515

- The mean throwing distance of a football for a Marco, a high school freshman quarterback, is 40 yards, with a **standard deviation of 2 yards**. The team coach tells Marco to adjust his grip to get more distance. The coach records the distances for 20 throws. For the 20 throws, Marco's mean distance was 45 yards. The coach thought the different grip helped Marco throw farther than 40 yards. Conduct a hypothesis test using a preset $\alpha = 0.05$. Assume the throw distances for footballs are **normal**. Construct the 95% confidence interval for the mean throwing distance and calculate the **p-value**.

- **Solution**:

- The Null hypothesis is, $\mathbf{H_0}$: $\boldsymbol{\mu} = 40$ yards.

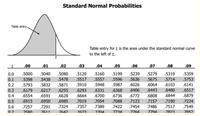
- The Alternative hypothesis, $\mathbf{H_a}$: $\boldsymbol{\mu} > 40$ yards because the coach thought the different grip helped Marco throw farther than 40 yards.

- Therefore, we have an *upper-tailed test* at $\boldsymbol{\alpha} = 0.05$.

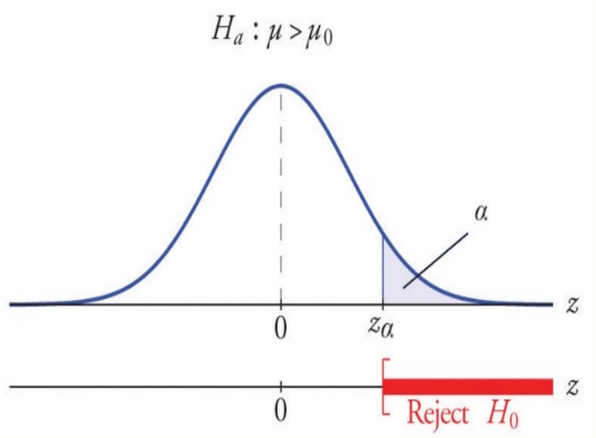
- The population standard deviation is known and the population is assumed normal. Therefore, we can assume that the sampling distribution of the estimator $(\overline{X})$ is Normal $(\boldsymbol{\mu}, {}^{\boldsymbol{\sigma}}/_{\sqrt{n}})$ and we can use the **z**-distribution.

Practice with calculator

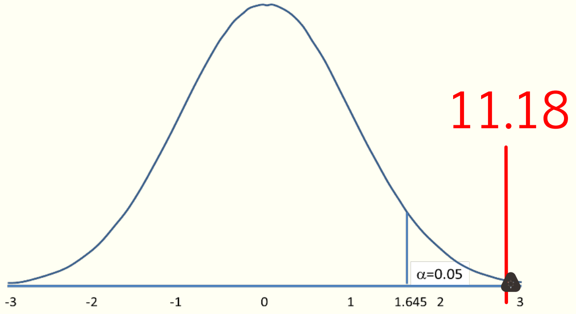
Practice with Tables



# Hypothesis Testing – Book problem 9.14 – Page 515




- Calculate the test **z**-statistic **z** = $(\bar{X} - \mu_0)/(\frac{\sigma}{\sqrt{n}})$ = $(45 - 40)/(\frac{2}{\sqrt{20}})$ = 11.18
- At $\alpha$ = 0.05, the critical value = 1.645 (there is 5% probability on the right tail of the standard normal)
- Because test **z**-statistic 11.18 is greater than 1.645, and falls in the critical (rejection) region, the null hypothesis **$H_0$**: **$\mu$** = 40 yards is **rejected**.
  - *At the 5% significance level we conclude that the sample provides evidence that a different grip helped Marco throw the ball farther than 40 yards.*
- The **p-value** = P(Z > test **z**-statistic) = P(**Z** > 11.18) = 1 – P(Z <= 11.18) = 0.000; since the **p-value= 0.000** < $\alpha$ = 0.05 we **reject** the null hypothesis
- Alternatively since he (1- $\alpha$ = 0.05)% = 95% CI for **$\mu$** = $(\bar{X} - 1.96 * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 * \frac{\sigma}{\sqrt{n}})$ = **(44.123, 45.877)** and the hypothesized **$\mu$** = 40 is *outside the interval*, we **reject** the null hypothesis.
- The probability that we **reject** the null hypothesis incorrectly is $\alpha$ = 0.05.
- Suppose we set stricter standards and set $\alpha$ = 0.01, we *would still reject the null hypothesis*, because **p-value (0.000)** < $\alpha$ = **0.01**.




**p-value= 0.000**

Practice with calculator

Practice with t-table

# Hypothesis Testing – Book Problem 9.46 – Page 541- Modified

- The cost of a daily newspaper varies from city to city. However, the variation among prices remains steady with a standard deviation of 20¢. A study was done to test the claim that the mean cost of a daily newspaper is $1.00. Twelve costs yield a mean cost of 95¢ with a standard deviation of 18¢. Do the data support the claim at the **1% level**? Construct the 99% confidence interval and calculate the **p-value**.

- The Null hypothesis is, $\mathbf{H_0}$: $\boldsymbol{\mu}$ = 100 cents (1 dollar, but the standard deviation is in cents).

- The Alternative hypothesis, $\mathbf{H_a}$: $\boldsymbol{\mu} \neq$ 100 cents because the study was done to test the claim that the mean cost of a daily newspaper is $1.00 and it is not claimed to be higher or lower.

- Therefore, we have an *two-tailed test* at $\boldsymbol{\alpha}$ = 0.01.

- The population standard deviation is known (20 cents), **but** the population is **not** given to be normal. We can use the Central Limit Theorem for a reasonable sample size; but we have only a sample size of 12.

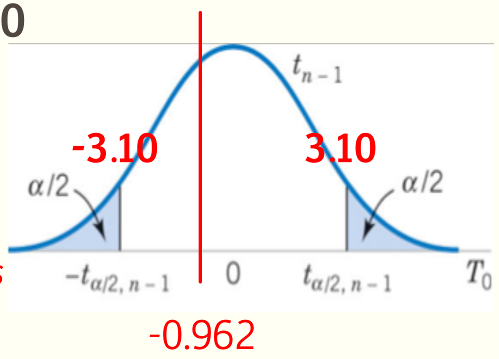
- Alternatively, let us assume that the population is normal. Then, we can assume that the sampling distribution of the estimator ($\bar{X}$) is a z-distribution. *Just to illustrate, I am going to assume that the population standard deviation is not known.* We will use the sample standard deviation of 18 cents instead. So, we use a t-distribution with 11 degrees of freedom from the sampling distribution of the sample mean.

Practice with calculator

Practice with t-table

# Hypothesis Testing – Book Problem 9.46 – Page 541-Modified

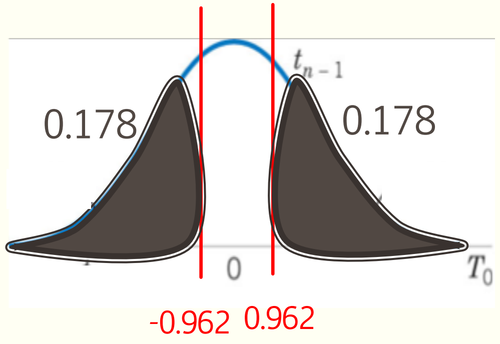
- Calculate the test **t**-statistic **t** = $(\bar{X} - \mu_0)/(\frac{s}{\sqrt{n}})$ = $(95 - 100)/(\frac{18}{\sqrt{12}})$ = -0.962
- At **α/2** = 0.005, using the t-table with 11 degrees of freedom, the critical values are **-3.10** and **+3.10** (there is 0.5% or 0.005 probability on each tail of the t-distribution)
- Because test **t**-statistic -0.962 is between (-3.10 and 3.10) it **does not fall** in the critical (rejection) regions, the null hypothesis $H_0$: **μ** = 100 cents is **not rejected**.
  - *At a significance level of 0.01, the sample does not provide evidence to reject the null hypothesis of 100 cents for the mean price of a newspaper.*




- The **p-value** for a two-tailed test is the **probability in both tails**
  = 2*(1 - P(**t** > |test **t**-statistic|))
  = 2*(1 - P(**t** > 0.92)) = 0.357

```
> 2*(1-pt(0.962,11))
[1] 0.3567202
```

- Since the **p-value= 0.357** > **α** = 0.01 we **fail to reject** the null hypothesis




- The 99% CI for **μ** = $(\bar{X} - 3.10^*\frac{s}{\sqrt{n}} \le \mu \le \bar{X} + 3.10^*\frac{s}{\sqrt{n}})$ = (78.9, 121.1) and the hypothesized **μ** = 100 is *inside the interval*, we **fail to reject** the null hypothesis.
- The probability that we **reject** the null hypothesis incorrectly is **α** = 0.01.
- Suppose we set looser standards and set **α** = 0.05, we *would still* fail to *reject the null hypothesis*, because **p-value (0.357)** > **α = 0.05**.

Practice with calculator

Practice with t-table

# Hypothesis Testing – Book Example 9.16 – Page 518

- Statistics students believe that the mean score on the first statistics test is 65. A statistics instructor thinks the mean score is higher than 65. He samples ten statistics students and obtains the scores 65; 65; 70; 67; 66; 63; 63; 68; 72; 71. He performs a hypothesis test using a 5% level of significance. The data are assumed to be from a **normal** distribution.

- **Solution**

- $\mathbf{H_0}$: $\mu = \mu_0 = 65$ $\mathbf{H_a}$: $\mu > 65$ (test is right-tailed); A 5% level of significance means that $\alpha = 0.05$.

- Data is **normal** but **population standard deviation is unknown**. The sample size is small (10) and therefore estimator ($\overline{X}$) is a **t-distribution** with 9 degrees of freedom.

- The sample mean ($\overline{X}$) and sample standard deviation (s) are calculated as 67 and 3.1972, respectively, from the data.

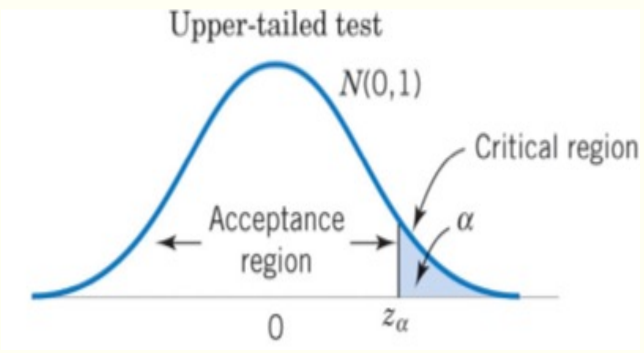
```
> stat_score <- c(65,65,70,67,66,63,63,68,72,71)
> mean(stat_score)
[1] 67
> sd(stat_score)
[1] 3.197221
```

Practice with calculator

Practice with t-table

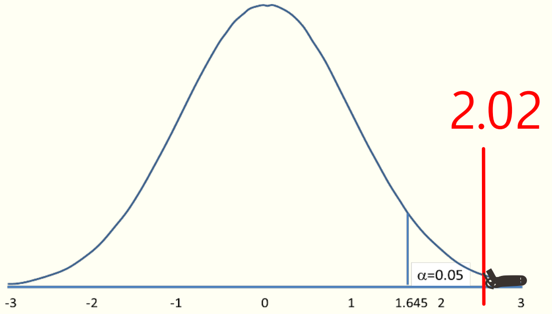
# Hypothesis Testing – Book Example 9.16 – Page 518

- Calculate the test **t**-statistic **t** = $(\bar{X} - \mu_0)/(\frac{s}{\sqrt{n}}) = (67 - 65)/(\frac{3.197}{\sqrt{10}}) = 2.02$
- At **α** = 0.05, the critical value from t-table with 9 degrees of freedom = 1.83
- Because test **t**-statistic 2.02 is > 1.83, it **falls** in the critical (rejection) region, the null hypothesis $\mathbf{H_0}$: **μ** = 65 is **rejected**. That is we **reject** the null hypothesis that 65 is the mean test score.
  - *At a significance level of 0.05, the sample provides evidence that the mean student score is greater than 65.*




- The **p-value** for the test is the **probability in the right tail** = 1 - P(**t** < test **t**-statistic) = 0.037

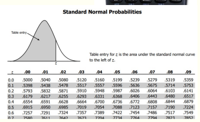
```
> print(1-pt(2.02,9))
[1] 0.037061
```




p-value= 0.037

- Since the **p-value= 0.037** < **α** = 0.05 we **reject** the null hypothesis

Practice with calculator 

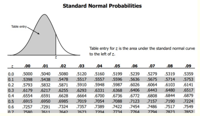
Practice with Tables 

# Hypothesis Testing – Book Example 9.17 – Page 519

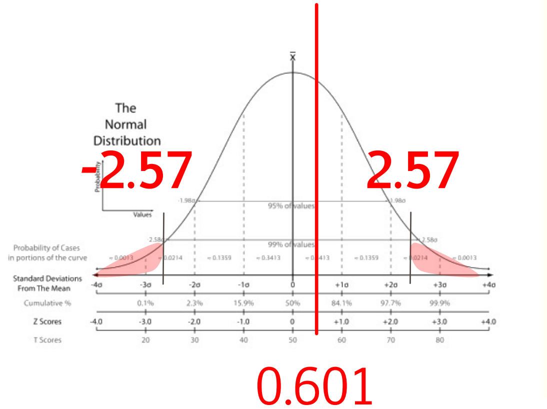
- Joon believes that 50% of first-time brides in the United States are younger than their grooms. She performs a hypothesis test to determine if the percentage is **the same or different from 50%**. Joon samples **100 first-time brides** and **53** reply that they are younger than their grooms. For the hypothesis test, she uses a 1% level of significance.
- **Solution:**
- The first thing we notice is that we are dealing with **population proportions**.
- Here **p**=0.5, n=100, p'=0.53.
- The Null hypothesis is, $\mathbf{H_0}$: $\mathbf{p} = \mathbf{p_0} = 0.5$
- The Alternative hypothesis, $\mathbf{H_a}$: $\mathbf{p} \neq 0.5$ because she performs a hypothesis test to determine if the percentage is **the same or different from 50%**. Therefore, we have a *two-tailed test* at $\boldsymbol{\alpha} = 0.01$.
- The sampling distribution p' ~ N($\mathbf{p_0}$, $\sqrt{\frac{p_0(1-p_0)}{n}}$) obtained as a normal approximation to the binomial distribution. **Note that (unlike for confidence intervals) we use the hypothesized $\mathbf{p_0}$ to calculate the standard error.**

Practice with calculator

Practice with Tables



# Hypothesis Testing – Book Example 9.17 – Page 519




- Calculate the test **z**-statistic **z** = $(p' - p_0) / \sqrt{\frac{p_0(1-p_0)}{n}}$

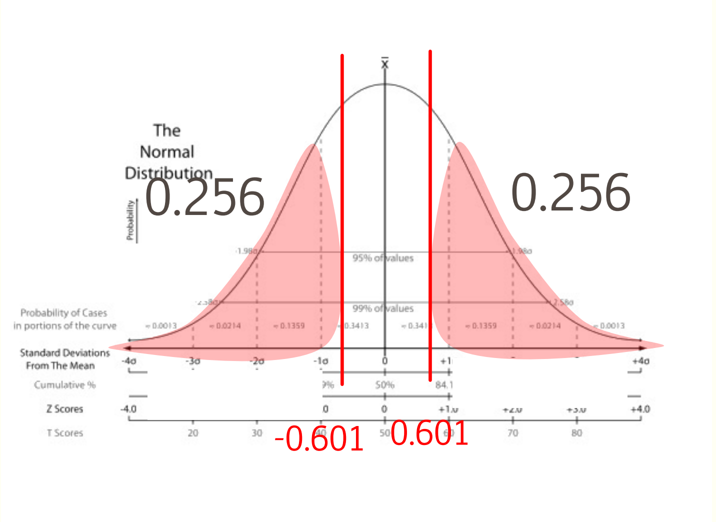
  $= (0.53 - 0.50)/(\sqrt{\frac{0.50(0.50)}{100}}) = 0.60$

- At $\alpha/2$ = 0.005, the critical values are -2.5783 and 2.5783.

- Because test **z**-statistic 0.601 is between (-2.5783 and 2.5783) it **does not fall** in the critical (rejection) regions, the null hypothesis $H_0$: **p** = 0.50 is **not rejected**.
  - *At a significance level of 0.01, the sample does not provide evidence to reject the null hypothesis that the ages of 50% of first-time brides are different than their grooms.*

- The **p-value** for a two-tailed test is the **probability in both tails** = 2* P(**Z** ≤ test **z**-statistic) = 2*0.258 = 0.516

- Alternatively, since the **p-value= 0.516** > $\alpha$ = 0.01 we **fail to reject** the null hypothesis



# POWER OF A HYPOTHESIS TEST

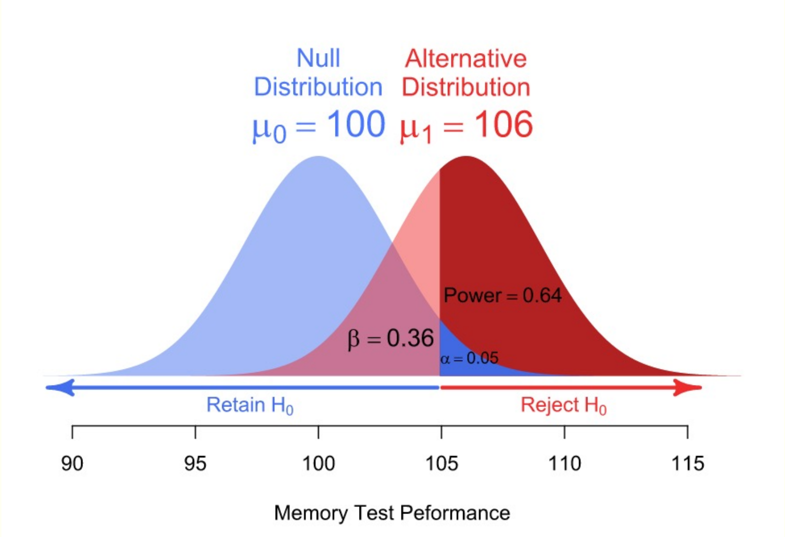
# Calculating **Power** of the test

- **α** (probability of a **Type I** error, or "significance level") is calculated assuming that the Null Hypothesis is true.

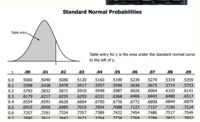
- **β** (probability of a **Type II** error) is calculated assuming that the Alternate Hypothesis is true, because it is the probability that we **do not reject** the null hypothesis, when the null hypothesis is false (i.e., the Alternate Hypothesis is true).

- The power of a test is (**1 - β**) represents the probability that we will *correctly reject a false null hypothesis.* It is very important since hypothesis tests with insufficient power will incorrectly accept false null hypothesis.

- However, note that we can never know the true **β** value because that requires centering the sampling distribution around a value *that is not the null hypothesis.* There are infinite possibilities for the alternate hypothesis generally. Hence, we also cannot calculate power because **power** = (**1-β**) unless we fix a value for the alternate hypothesis.

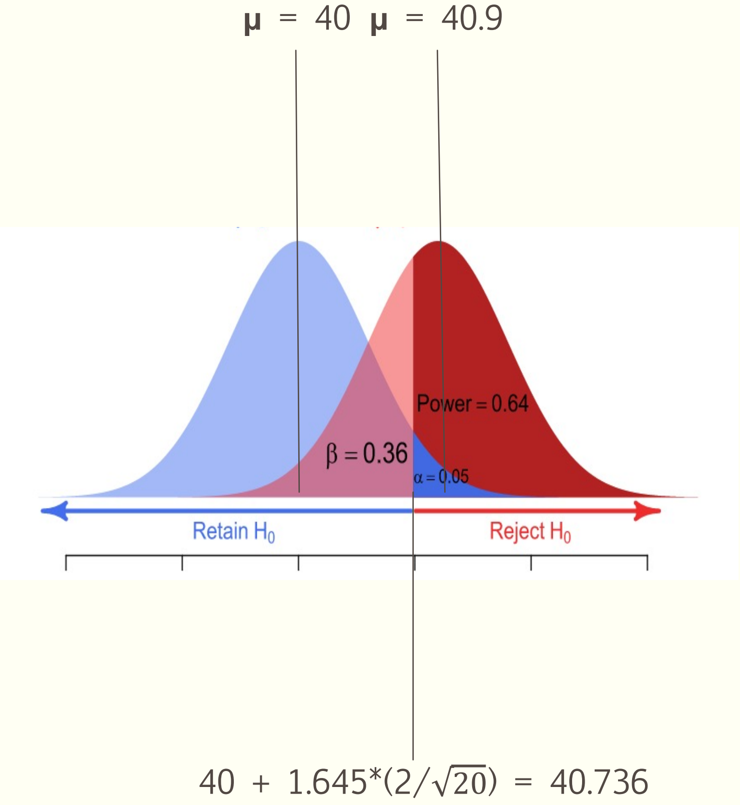


Practice with calculator

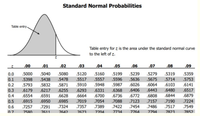

Practice with Tables


# Calculating β & Power for an *assumed* Alternative Hypothesis

- We go back to the Hypothesis Testing Example 2, where we had
  - The Null hypothesis is, $\mathbf{H_0}$: $\boldsymbol{\mu} = \boldsymbol{\mu_0}$ = 40 yards.
  - The Alternative hypothesis, $\mathbf{H_a}$: $\boldsymbol{\mu} > 40$.
  - An *upper-tailed test* at $\boldsymbol{\alpha} = 0.05$. *Critical value = 1.645.*
  - Test **z**-statistic $\mathbf{z} = (\bar{X} - \mu_0)/(\frac{\sigma}{\sqrt{n}}) = (45 - 40)/(\frac{2}{\sqrt{20}}) = 11.18$ with **p-value= 0.000**
  - **We rejected the null hypothesis**
- Suppose we assumed that the true population **μ** was 40.9 yards (That is the alternative hypothesis $\boldsymbol{\mu} > 40$ happened to be true).
- We now look for $\boldsymbol{\beta}$ = P(We **DO NOT Reject** | $\boldsymbol{\mu} = 40.9; \frac{2}{\sqrt{20}}$)
- Our critical value of $\mathbf{z} = 1.645$ under a mean of $\boldsymbol{\mu} = 40$ translates to an X- value of $40 + 1.645*(\frac{2}{\sqrt{20}}) = 40.736$
- $\boldsymbol{\beta} = P(X \leq 40.736 \mid \boldsymbol{\mu} = 40.9; \frac{2}{\sqrt{20}}) = 0.36$;
- Power = 1- $\boldsymbol{\beta}$ = 0.64
- Thus, **if** we knew or assumed what the true alternative hypothesis value was, we would be able to calculate Power. In practice, we assume different values and obtain power under each scenario. This is called ***Power Analysis***. The difference between $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_a$ is called ***effect size***.



Practice with calculator 

Practice with Tables 

# Calculating Power given an Effect Size

- Let us calculate the power for a test of $\mathbf{H_0{:}\mu} = 20$ for a *lower-tailed test*. Let the effect size we want to detect using our sample be 1. That is, we want to find the probability of correctly rejecting the null hypothesis if in fact $\mathbf{H_a{:}\mu} = 19$ is true .

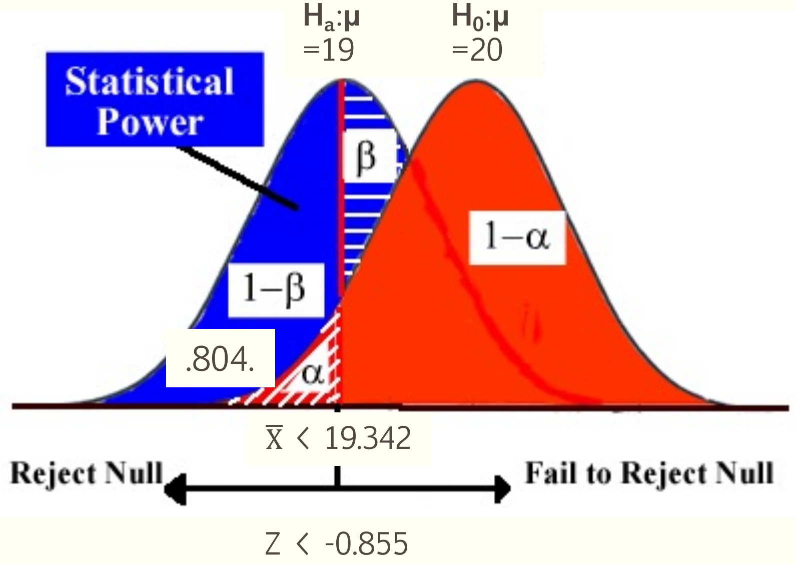
- Let the population be normal with a population standard deviation $\boldsymbol{\sigma} = 4$. Let our sample size be 100 and let $\boldsymbol{\alpha} = 0.05$.

- We know that $\bar{X} \sim$ Normal $(\boldsymbol{\mu}, \frac{\boldsymbol{\sigma}}{\sqrt{n}})$. So $\boldsymbol{\alpha} = 0.05$, *critical value = -1.645 (lower-tailed test)*

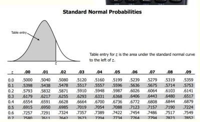
- Under the null hypothesis, $\bar{X} \sim$ Normal $(\mathbf{20}, \frac{\mathbf{4}}{\sqrt{100}}) \sim$ Normal $(\mathbf{20}, 0.4)$
  - So the Z-statistic is $(\bar{X} - \mathbf{20})/(0.4)$. If this is less than -1.645 we will correctly reject the null hypothesis.
  - That is, if $\bar{X} < 20 - 0.4 * 1.645$ i.e., $\bar{X} < 19.342$ we will reject the null hypothesis is our decision.

- The power then is the probability of finding $\bar{X} < 19.342$ under the alternative hypothesis: $\bar{X} \sim$ Normal $(\boldsymbol{19}, 0.4)$

- This probability is $P(Z < (19.342 - 19)/0.4) = P(Z < 0.855) = .804$. <u>This is the power of the test</u>.



Practice with calculator 

Practice with Tables 

# Calculating Sample Size, given Effect size and Power

- In the previous example, suppose we wanted a power of 0.9 given an effect size of 1 and $\alpha$ = 0.05. What should be the sample size?
- Under the null hypothesis
  - $\overline{X} \sim$ Normal $(\mathbf{20}, \frac{4}{\sqrt{n}}) \sim$ Normal $(\mathbf{20}, \frac{4}{\sqrt{n}})$
  - So the Z-statistic is $(\overline{X} - \mathbf{20})/(\frac{4}{\sqrt{n}})$. If this is less than -1.645 we will correctly reject the null hypothesis.
  - That is, if $\overline{X} < 20 - \frac{4}{\sqrt{n}} * 1.645$ we will reject the null hypothesis, if the null hypothesis is true.
- The power then is the probability of finding $\overline{X} < 20 - \frac{4}{\sqrt{n}} * 1.645$ under the alternative hypothesis: $\overline{X} \sim$ Normal $(\mathbf{19}, \frac{4}{\sqrt{n}})$
- This probability is $P(Z < (20 - \frac{4}{\sqrt{n}} * 1.645 - 19)/\frac{4}{\sqrt{n}}) = 0.9$ (Power).
- i.e., $(20 - \frac{4}{\sqrt{n}} * 1.645 - 19)/\frac{4}{\sqrt{n}}) < 1.3$
- The value of Z is approximately 1.3, so $(1 - \frac{4}{\sqrt{n}} * 1.645) < 1.3 * \frac{4}{\sqrt{n}}$.
- So, $(2.945 * \frac{4}{\sqrt{n}}) < 1$ so that $\sqrt{n} > 11.78$ so that $n > 139$.

# Power of a Hypothesis Test

- So, we cannot calculate Power without knowing an exact value for the Alternative Hypothesis (also called "effect size"). But, we do know the following:
  - The clearest way to increase the power of a test, everything else being constant, is to increase sample size.
  - Again, this happens through a reduction in standard error (SE) (i.e., a reduction in the standard deviation of the sampling distribution)
  - As the SE gets smaller, the test becomes more and more sensitive to detecting deviations from the hypothesized parameter

# BOOTSTRAPPING

# Resampling the sample - Bootstrapping

- What happens if you don't know the sampling distribution? Answer: **Bootstrapping**.
- Bootstrapping is a statistical resampling method developed by B. Efron in the late 1970s (e.g., Diaconis & Efron, 1983; Efron & Tibshirani, 1993)
- Basic idea: re-sample with replacement from the sample and construct an empirical "sampling distribution"; Obtain standard errors from this distribution for hypothesis testing.

# Book Example 9.16 – Page 518 – Using Bootstrapping

- Statistics students believe that the mean score on the first statistics test is 65. A statistics instructor thinks the mean score is higher than 65. He samples ten statistics students and obtains the scores 65; 65; 70; 67; 66; 63; 63; 68; 72; 71. *He wants to obtain a 95% confidence interval for the true mean score.* ~~The data are assumed to be from a **normal** distribution.~~

- The sample mean ($\bar{X}$) and sample standard deviation (s) are calculated as 67 and 3.1972, respectively, from the data.

- Suppose we **do not know** that the data is **normal**. We will not know the sampling distribution of $\bar{X}$ from theory, because we cannot use the Central Limit Theorem (sample size is small (10)) and we cannot use the t-distribution because the population is not known.

- We will try Bootstrapping to get the standard error of $\bar{X}$ as well as its bootstrap sampling distribution. We will also obtain the bootstrap confidence interval.

| | Population $\sigma$ known | Population $\sigma$ unknown |
|---|---|---|
| Population Normal | $\bar{X} \sim Normal(\mu, \frac{\sigma}{\sqrt{n}})$. Use the standard normal z-distribution regardless of sample size. Case 1 | • $\bar{X} \sim Normal(\mu, \frac{S}{\sqrt{n}})$ if sample size $\geq$ 30. • $\bar{X} \sim t_{n-1}(\mu, \frac{S}{\sqrt{n}})$ if sample size < 30. Case 2 |
| Population Not Known Normal | If sample size $n \geq 30$: • $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$ if $\sigma$ is known • $\bar{X} \sim t_{n-1}(\mu, \frac{S}{\sqrt{n}})$ if $\sigma$ is not known *If sample size ≪ 30, we really have to assume the population distribution. To avoid this, collect a larger sample.* Case 3 & 3a | Obtain sampling distribution of $\bar{X}$ using Bootstrapping. (we will see later) |

# Book Example 9.16 – Page 518 – Using Bootstrapping (**mean_bootstrap.R**)

- Original sample = 65; 65; 70; 67; 66; 63; 63; 68; 72; 71
- Example Bootstrapped samples:

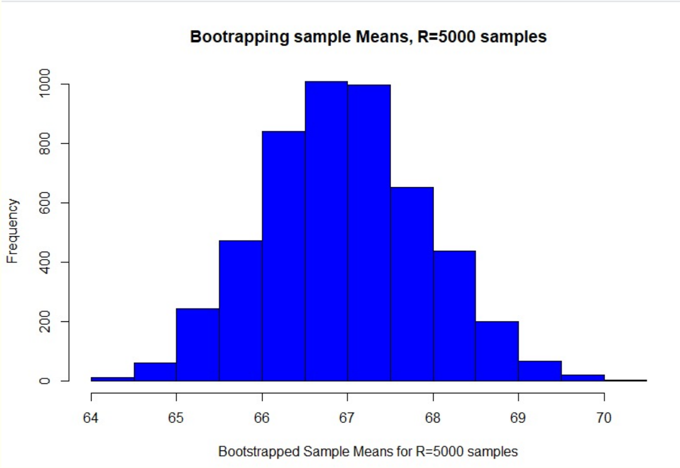
```
> # what bootstrap samples look like and their means
> #
> for (i in 1:5)  {
+     b_samp <- sample(stat_score, 10, replace=TRUE)
+    print(b_samp)
+    print(paste("mean of this bootstrapped sample is: ",mean(b_samp)))
+ }
 [1] 71 65 65 72 65 63 63 66 65 63
[1] "mean of this bootstrapped sample is:  65.8"
 [1] 65 71 70 65 70 72 65 72 72 65
[1] "mean of this bootstrapped sample is:  68.7"
 [1] 72 66 65 63 67 67 70 70 71 66
[1] "mean of this bootstrapped sample is:  67.7"
 [1] 66 68 71 71 63 63 63 65 66 68
[1] "mean of this bootstrapped sample is:  66.4"
 [1] 68 63 72 68 66 70 67 63 71 67
[1] "mean of this bootstrapped sample is:  67.5"
> 
```

- In *nonparametric bootstrapping*, cases from a raw data file are randomly selected with replacement to generate other data sets, usually with the same number of cases as the original
  - Because of sampling with replacement, the
    - same case can appear more than once in a generated data set
    - composition of cases will vary somewhat across the generated samples
- When nonparametric bootstrapping is repeated many times, it simulates random sampling from a population
- Standard errors are estimated in this method as the standard deviation in the empirical sampling distribution of the same estimator across all generated samples

# Book Example 9.16 – Page 518 – Using Bootstrapping (**mean_bootstrap.R**)

- We will obtain the sample means of 5000 bootstrap sample using the boot() function in R, draw the Histogram and get the 95% CI

```
#
library(boot)
stat_score <- c(65,65,70,67,66,63,63,68,72,71)
#
# We create a data frame of stat_score
#
df <- data.frame(stat_score)
print(df[1,])
#
# We have to create our own R function - mean_fun - boot() will use to re-sample n=10 observations,
# calculate the mean of the re_sampled observations, and retuen it back to bootstrap.
# boot() will do it R=k times, so you will have k boot-strapped means. In our case, k=5000 replicates.
#
mean_fun <- function(df, i)
  {
  d <- df[i, ]
  return(mean(d))
}
#
# The actual bootstrapping is now performed by the boot() function. The drop=FALSE retains each sample as a data frame.
#
bo <- boot(df[, "stat_score", drop = FALSE], statistic=meanfun, R=5000)
#
# bo($t) will contain R=5000 bootstrapped sample means
#
# We can do a histogram of bo$t to see the boot-strapped sampling distribution of sample means
#
hist(bo$t,
     main = "Bootrapping sample Means, R=5000 samples",
     xlab="Bootstrapped Sample Means for R=5000 samples",
     col="blue")
#
# We can print out the mean of all the boots-strapped sample means. This is the Expected value of X-bar
# Because this is bootstrapping, the expected value will not be exactly the population mean
# That is, it is biased
print(paste("Mean of Bootstrapped Sample Means",round(mean(bo$t),4)))
#
# The standard error is the standard deviation of the boot-strapped sampling distribution
#
print(paste("Standard Deviation of Bootstrapped Sample Means - Standard Error",round(sd(bo$t),4)))
#
# We can obtain a bot-strapped 95% confidence interval using boot.ci;
# bca means intervals are calculated using the adjusted bootstrap percentile (BCa) method.
#
boot.ci(bo, conf=0.95, type="bca")
```



# Book Example 9.16 – Page 518 – Using Bootstrapping (**mean_bootstrap.R**)

- Based on the bootstrap confidence interval, since 65 is NOT in the interval, the instructor can conclude that a null hypothesis of **H$_0$**: **$\mu$ = $\mu_0$** = 65 can be rejected,

```
> print(paste("Mean of Bootstrapped Sample Means",round(mean(bo$t),4)))
[1] "Mean of Bootstrapped Sample Means 66.9744"
> #
> # The standard error is the standard deviation of the boot-strapped sampling distribution
> #
> print(paste("Standard Deviation of Bootstrapped Sample Means - Standard Error",round(sd(bo$t),4)))
[1] "Standard Deviation of Bootstrapped Sample Means - Standard Error 0.9641"
> #
> # We can obtain a bot-strapped 95% confidence interval using boot.ci;
> # bca means intervals are calculated using the adjusted bootstrap percentile (BCa) method.
> #
> boot.ci(bo, conf=0.95, type="bca")
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 5000 bootstrap replicates

CALL :
boot.ci(boot.out = bo, conf = 0.95, type = "bca")

Intervals :
Level       BCa
95%   (65.2, 68.9 )
Calculations and Intervals on Original Scale
```

# Bootstrapping

- Bootstrapping is not a “magical” technique that can somehow compensate for
  - small or unrepresentative samples
  - distributions that are severely non-normal
  - the absence of independent samples for replication
- In fact, bootstrapping can potentially magnify the effects of unusual features in a data set