Chapter 5
Hypotheses Test of the Mean

Section 5.1
Logic of the Hypothesis Test

David W. Gerbing

The School of Business
Portland State University

- Logic of the Hypothesis Test
  - The Hypothesized Value
  - The Obtained t-value
  - The Rejection Region

5.1a
The Hypothesized Value

Two Forms of Statistical Inference

Confidence intervals and hypothesis tests

- The purpose of statistical inference of the mean is to estimate the unknown underlying population mean, μ, of which there are two forms
  - Confidence interval: An interval about the sample mean that specifies a range of plausible values of μ
  - Hypothesis test: Evaluate a specific value of μ, or range of values, specified by the analyst *before* the analysis begins
- **Hypothesis test** for a population mean: Evaluate the reasonableness of a presumed (hypothesized) value of μ by how close the sample mean, *m*, is to the proposed value
- Use the hypothesis test when there is particular interest regarding a single value or range of values of μ

Hypothesis Test Begins with the Hypothesized Value

Many analyses focus on a reference value of interest

- There are many applications in which an analysis of the mean focuses on a specific numerical value, such as
  - Salary: Do marketing executives in our favorite city make more or less than $100,000 salary, on average?
  - Sales: Following a new ad campaign, has average Gross Sales of 10 products for the subsequent quarter increased over the $856,000 baseline of sales over the last several years?
  - Surveys: For an item that assesses General Satisfaction of a product, measured on a 7-point scale from Strongly Disagree to Strongly Agree scored from 1 to 7, is the average response in the Agree region, that is, above a 4.0?

The Hypothesized Value as the Desired Value

Example of wanting to support the hypothesized value

- **Hypothesized value**: Population value (or values) assumed true for purposes of the analysis, a reference value of interest not necessarily assumed actually true in the real world
- In some situations the hypothesized value is the desired value
- Example in which the goal is to provide *support*
  - Is the machine that fills the cereal boxes set correctly?
  - The weights of the individual cereal boxes generally differ, so analyze their mean weight
  - The hypothesized population mean weight is the desired value, $\mu = 350$ g, the weight specified on each cereal box
  - Get a random sample of the weights of cereal box contents
  - Desired outcome: Obtain a sample mean, *m*, close to 350 g
  - Purpose of hypothesis test: Evaluate extent of "close to"

An Hypothesized Value that is Desired to be Rejected

Example of the hypothesized "for purposes of the test" only

- Unfortunately the phrase "hypothesis test" is a misnomer, a misleading terminology that can obscure understanding
- In practice, perhaps more often than not, the goal is to *refute* the so called "hypothesized value", such as a previously established baseline compared against a more recent result
- Example in which the goal is to *refute*
  - The hypothesized value is the previous average wait time before your customers were served, 39.4 sec
  - Implement a presumably more efficient service process
  - Now wish to detect an improvement, a smaller average population wait time than the hypothesized mean of 39.4
  - Collect a random sample of wait times with the new process
  - Desired outcome: Obtain m considerably below 39.4 sec
  - Purpose of hypothesis test: Evaluate "considerably below"

Impact of Sampling Error

Must consider the consequences of random variation

- The sample mean, m, varies across repeated random samples
- **Key Concept**: Because of sampling variability, even if a specified value of the population mean, μ, is true, a corresponding sample mean, m, generally does *not* equal μ
  - Even if the machine that fills the cereal boxes is set correctly, yielding a true process average of $\mu = 350$ g, the mean of a corresponding sample will likely not equal 350 g, $m \neq 350$ g
  - Even if the new service process did not change the average wait time of 39.4 sec, so that it is still true that $\mu = 39.4$, for (usually hypothetical) multiple samples half of all sample means will be below 39.4
- As always, sampling fluctuation obscures the underlying reality, the true population value, so an inference procedure such as hypothesis testing is required to infer the true underlying value

Three Basic Concepts: Notation

(Actual) Population Mean, μ

- Reality, typically unknown, but value of primary interest
- Ex: *Unknown* to the analyst, true mean is $\mu = 350.8$ g

Sample Mean, m

- What is observed from the collected data, an estimate of μ
- Ex: Sample mean calculated from the data is $m = 350.5$ g

Hypothesized Population Mean, μ_0

- Presumed value of μ, assumed true for purposes of the test
- Ex: Test the hypothesis that $\mu = 350$ g, so μ_0 is 350 g
- The hypothesis for a mean always focuses on a specific value of μ, a constant μ_0, such as 350 g or 39.4 secs or 52%

Null and Alternative Hypotheses

Hypotheses are always about unknown population values

- **Null hypothesis**: Value(s) of the population mean hypothesized assumed true for purposes of the test
- Following is the general form of the null hypothesis when only one specific hypothesized value, μ_0, a specific number such as 350 or 39.4, is presumed true for purposes of the test

$$H_0 : \mu = \mu_0$$

- **Alternative hypothesis**: Values of the population mean when the hypothesized value is not true, i.e., rejected

$$H_1 : \mu \neq \mu_0$$

- According to this alternative hypothesis, any value of μ larger than or smaller than the hypothesized value μ_0 invalidates the null hypothesis

5.1b
The Obtained t-value

Need for the Hypothesis Test

Must consider the consequences of random variation

- **Key Concept**: The hypothesis test provides evidence regarding if the sample mean, m, is sufficiently close or far away from the hypothesized value of μ, μ_0, to either render the hypothesized value reasonable or not
- For example, is
  - $m = 350.3$ close enough to 350 to render $\mu = 350$ reasonable?
  - $m = 36.8$ so much smaller than 39.4 to indicate a likely decrease from $\mu = 39.4$?
- Just knowing the separation between sample and hypothesized values of the mean is not enough to answer these questions, which can be formally assessed with the hypothesis test
- There needs to be some scale, some way to evaluate the size of the distance $m - \mu_0$

Given μ_0, What is an Unreasonable Sample Outcome?

Evaluate the separation between observed and hypothesized

- The issue is that the sample mean, m, randomly fluctuates from sample to sample, so the analyst must account for this variation in the assessment of any one m as it relates to μ_0
- If μ_0 is correct, the question to ask: What range does m typically vary around μ_0 over many, many samples?
- **Key Concept**: The logic of the hypothesis test assumes that the null hypothesized value, μ_0, is correct, and then assesses the size of the separation of m and μ_0 over many, many hypothetical repeated samples of data
- Assess the size of $m - \mu_0$, in terms of the standard error, which reflects the amount of variation of m over repeated samples
- **Test statistic**: Number of standard errors that separate the sample and hypothesized values

General Form of the Hypothesis Test

Standardize the distance between observed and hypothesized

- **Obtained t-value**: The test statistic for the evaluation of the hypothesized value of a mean is the standardized distance that separates m from μ_0

$$\text{obtained } t\text{-value}: \ t_m = \frac{m - \mu_0}{s_m}$$

- t_m: The estimated standard errors that separate the obtained sample mean, m, from the *hypothesized* mean, μ_0[1]
- To evaluate the size of the obtained $m - \mu_0$, compare the one calculated test statistic, t_m, against the hypothetical distribution of what would happen over many, many such t_m-values calculated over repeated sampling

[1] In the very unlikely event that the population standard error σ_m is known, the test statistic would be the corresponding z-value: $z_m = (m - \mu_0)/\sigma_m$

Distribution of t, from Mathematics, not Data

Evaluate a specific result against the t-distribution

- Need to obtain the probability of getting a specific m assuming the truth of the null hypothesis (regardless if it is actually thought or desired to be true)
- **t-distribution**: The mathematically derived distribution of the t-value over many, many hypothetical samples IF μ_0 is the true value of the population mean, μ
- **Degrees of freedom**: The specific t-distribution depends on the degrees of freedom, $df = n - 1$, where n is sample size
- A t-distribution is determined by mathematics, not data
- A t-distribution requires a set of assumptions necessary for its mathematical derivation
  - Each data value Y_i is independently sampled from a population of the same μ and σ
  - The sample mean, m, is normally distributed

5.1c
The Rejection Region

Assume the Null Hypothesis

Entire logic and all conclusions depend on assuming the null

- The version of the null hypothesis presented here is that the true population mean, μ, is the constant μ_0, which is always specified as a specific value such as 39.4 for a specific context
- **Key Concept**: IF the null hypothesis is true, THEN did the analysis provide expected or unexpected results?
- For hypothesis tests, *all* conclusions are conditional, given the truth of the reference value, μ_0:

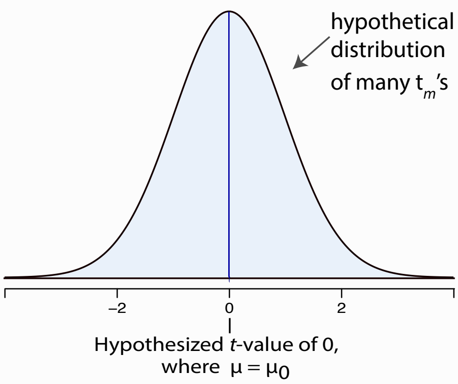
  IF $H_0 : \mu = \mu_0$ is true, THEN conclude ...
- If μ_0 is the true value of μ, the hypothetical distribution of
  - the sample mean, m, centers over μ_0
  - the corresponding transformation of m to the t-value, $t_m = (m - \mu_0)/s_m$, centers over 0

t-Distribution Assuming Null Hypothesis $\mu = \mu_0$

- For a population with $\mu = \mu_0$, *hypothetically* take very many samples, each of size n
- For *each* sample, calculate m and s_m, and then t_m:

$$t_m = \frac{m - \mu_0}{s_m}$$

- The result is an entire distribution of t-values
- When $\mu = \mu_0$, this *mathematically defined* distribution of t_m centers on $t = 0$



The Rejection Region

What happens when m is far from μ_0?

- **Key Concept**: Basis of the test is to compare the one t-statistic *calculated from the data*, t_m, to the entire set of outcomes given the mathematically defined distribution of t_m over many hypothetical samples
- **Rejection Region**: Tail areas of the t-distribution that represent unlikely events assuming a true value of $\mu = \mu_0$
- When t_m falls in the rejection region, m is far from the assumed true population mean $\mu = \mu_0$, so consider the assumption of a true μ_0 unreasonable and reject the null hypothesis

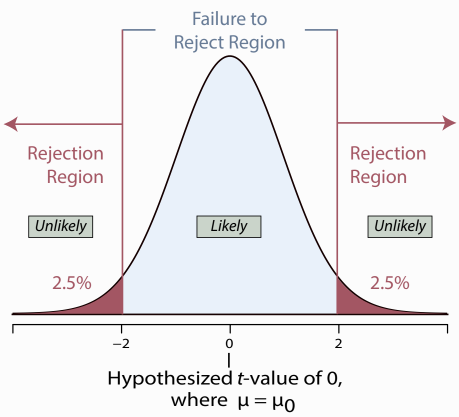
The $\alpha = 0.05$ Rejection Region

Define "unlikely" as a (conditional) probability of 5% or less

- **α (alpha)**: Specified before data analysis, the probability level that defines an event as unlikely
- Usually set to $\alpha = 0.05$ or 5%, which corresponds to a 95% range of sampling variation
- Usually willing to interpret either a positive or a negative deviation from μ_0, in which case define the rejection region with 2.5% in the upper tail and 2.5% in the lower tail
- **Cutoff** or critical value: Value of a distribution that cuts off a specified tail area, such as for specifying the rejection region
  - for $\alpha = .05$: $-t_{.025}$ and $t_{.025}$
  - in general: $-t_{\alpha/2}$ and $t_{\alpha/2}$
- The specific value of $t_{.025}$ for a given analysis depends on the specific t-distribution as indicated by $df = n - 1$, the degrees of freedom, but is usually approximately equal to 2

Ex: Two-Tailed $\alpha = 0.05$ Rejection Region, $df = 59$

- Suppose $n = 60$, then define the tails according to $\alpha = 0.05$ for $df = 59$ with $t_{.025} = 2.00$ and $-t_{.025} = -2.00$
- IF the hypothesized value is true, THEN a value of t_m in either tail of the t-distribution is *unlikely*
- IF $H_0 : \mu = \mu_0$ is true, and $t_m < -2.00$ or $t_m > 2.00$, THEN reject the value of μ_0 as *unlikely*



How far is m from μ_0?

Bases of test is the t-distribution

- Use the mathematically defined distribution of t to calculate the cutoff or critical values that define the rejection region, such as $-t_{.025}$ and $t_{.025}$
- Calculate the t-value obtained from the data, t_m, to see how many estimated standard errors the sample mean, m, is from μ_0

Large distance between actual and hypothesized values ...

- Yields a large value of t_m, in the rejection region
- Renders hypothesized value unreasonable

Small distance between actual and hypothesized values ...

- Yields a small value of t_m, *not* in the rejection region
- Demonstrates data consistent with hypothesized value

Index Subtract 2 from each listed value to get the Slide

- The End