BIOS:4120 – Introduction to Biostatistics
Unit 9: Hypothesis Testing

Knute D. Carter

Department of Biostatistics
The University of Iowa

October 7, 2025

Overview

- General Concepts
- Two-Sided Tests of Hypotheses
- One-Sided Tests of Hypotheses
- Types of Errors
- Power
- Sample Size Estimation

Hypothesis Testing

- Two major topic areas comprise statistical inference:
  - estimation,
  - hypothesis testing.
- A *Hypothesis Test* uses sample data to assess the plausibility of each of two competing hypotheses regarding an unknown parameter (or set of parameters).

Null and Alternative Hypotheses

- A *Statistical Hypothesis* is a statement or claim about an unknown parameter (or set of parameters).
- The *Null Hypothesis* generally represents what is assumed to be true before the experiment is conducted. This hypothesis is typically denoted as H_0.
- The *Alternative* or *Research Hypothesis* represents what the investigator is interested in establishing. This hypothesis is typically denoted as H_A.

Example: Null and Alternative Hypotheses

- Based on a publication of the National Center for Health Statistics, the mean serum cholesterol level for the general population of 20–74-year-old males is 211 mg/100 ml.

- Consider the subpopulation of 20–74-year-old males who are hypertensive and who smoke. Let μ denote the mean serum cholesterol level for this subpopulation.

- Assume μ is unknown.

Example: Null and Alternative Hypotheses

- Suppose we decide to collect sample data to test whether μ is comparable to 211 mg/100 ml, the mean serum cholesterol level for the general male population.
- In this setting, the hypotheses could be

$$H_0 : \mu = 211 \text{ mg/100 ml}$$

$$H_A : \mu \neq 211 \text{ mg/100 ml}$$

- The sample data, in particular $\bar{x}$, will be used to determine which of these two hypotheses is more plausible.

Null and Alternative Hypotheses

- Assume that the standard deviation of serum cholesterol levels for the population of hypertensive male smokers is known, and is given by $\sigma = 46$ mg/100 ml.
- Suppose we decide to proceed as follows:
  1. Collect serum cholesterol levels for a random sample of 36 hypertensive male smokers.
  2. Compute the sample mean $\bar{x}$.
  3. Reject H_0 in favor of H_A if $\bar{x} < 196$ or $\bar{x} > 226$. Otherwise, do not reject H_0, i.e., retain H_0.
- Is this a sensible decision rule? How was it devised?
- We will attempt to address the preceding questions in what follows.

Types of Errors in Hypothesis Testing

Test Result	True State of Nature	
	H_0 True	H_0 False
Retain H_0	Correct $(1-\alpha)$	
Reject H_0	Incorrect Type I Error (α)	

- A *Type I Error* occurs when the null hypothesis is true, yet based on the sample evidence, we reject the null hypothesis.
- The probability of committing a type I error is denoted by α.

Types of Errors in Hypothesis Testing

	True State of Nature	
Test Result	H_0 True	H_0 False
Retain H_0		Incorrect Type II Error (β)
Reject H_0		Correct ($1-\beta$)

- A *Type II Error* occurs when the null hypothesis is false, yet based on the sample evidence, we retain the null hypothesis.
- The probability of committing a type II error is denoted by β.

Types of Errors in Hypothesis Testing

Test Result	True State of Nature: H_0 True	True State of Nature: H_0 False
Retain H_0	Correct $(1-\alpha)$	Incorrect Type II Error (β)
Reject H_0	Incorrect Type I Error (α)	Correct $(1-\beta)$

- The probability $(1-\beta)$ is often called the *Power* of the test.
- The power reflects the probability of the test to establish the validity of a true alternative hypothesis.

Types of Errors in Hypothesis Testing

- Question: In the serum cholesterol level example, what is the probability of committing a type I error?
- We must find the following probability:

$$\alpha = P\left((\bar{X} < 196) \cup (\bar{X} > 226) \mid \mu = 211\right)$$

Types of Errors in Hypothesis Testing

- Question: In the serum cholesterol level example, assume that in reality $\mu = 219$ mg/100 ml, making the alternative hypothesis true. What is the probability of committing a type II error?

- We must find the following probability:

$$\beta = P\left(196 \leq \bar{X} \leq 226 \mid \mu = 219\right)$$

Types of Errors in Hypothesis Testing

- Note: Using a decision rule that results in a lower value of α will result in a higher value of β, and vice versa.

- To conclude the example, suppose that we draw our sample of 36 hypertensive male smokers, and obtain a sample mean serum cholesterol level of $\bar{x} = 228.7$ mg/100 ml.

- Based on our decision rule, we will reject H_0 in favor of H_A, and conclude that the mean serum cholesterol level for hypertensive male smokers is different than the mean level for the general male population (of 211 mg/100 ml).

Types of Errors in Hypothesis Testing

- This <u>does not mean</u> that H_A is guaranteed to be the true hypothesis.
- However, if H_0 is the true hypothesis, the probability of our test rejecting H_0 (as it did) is only 5% ($\alpha = 0.05$).
- Thus, the sample evidence supports H_A.

Types of Errors in Hypothesis Testing

- In hypothesis testing, if the sample evidence supports H_A, and we subsequently reject H_0 in favor of H_A, we claim that the results are *Statistically Significant*.
- If the sample evidence does not support H_A, and we subsequently fail to reject (retain) H_0, we claim that the results are *Not Statistically Significant*.
- In hypothesis testing, *Statistical Significance* is attained when the sample results render the null hypothesis implausible, and thereby justify rejecting the null hypothesis in favor of the alternative hypothesis.

Tests of Hypotheses

- In the general procedure for conducting a hypothesis test, the investigator chooses the value of α (before the sample data is collected).
- This value of α then leads to a decision rule for determining when H_0 should be rejected in favor of H_A, and when H_0 should not be rejected.
- The probability α is often called the *Level of Significance*.
- The smaller the value of α, the greater the 'burden of proof' required to reject H_0 in favor of H_A.
- The most common choices of α are 0.01, 0.05, and 0.10.

Two-Sided versus One-Sided Tests

- In the serum cholesterol level example, the hypotheses were

$$H_0 : \mu = 211 \text{ mg/100 ml}$$

$$H_A : \mu \neq 211 \text{ mg/100 ml}$$

- The baseline level of μ used in the formulation of the hypotheses (in this case, 211 mg/100 ml) is often denoted by μ_0.

Two-Sided versus One-Sided Tests

- Before the collection of the sample data, suppose we believe that the mean serum cholesterol level for males who are hypertensive and smokers exceeds the mean level for the general population of males.

- We may therefore wish to use an alternative hypothesis of the form

$$H_A : \mu > 211 \text{ mg/100 ml}$$

- Such an alternative hypothesis is sometimes called *Unidirectional*.

Two-Sided versus One-Sided Tests

- The null hypothesis can still be written as

$$H_0 : \mu = 211 \text{ mg/100 ml}$$

- However, some practitioners prefer formulating the null hypothesis as

$$H_0 : \mu \leq 211 \text{ mg/100 ml}$$

so that all possible values of μ are covered by H_0 and H_A.

Two-Sided versus One-Sided Tests

- Important Note: In class, we will always write the null hypothesis using a strict equality, since the theory of hypothesis testing is built upon such null hypotheses.

- Thus, for tests on a population mean μ, we will always write the null hypothesis as

$$H_0 : \mu = \mu_0$$

- For unidirectional alternative hypotheses, some authors write the null hypothesis as

$$H_0 : \mu \leq \mu_0$$

or

$$H_0 : \mu \geq \mu_0$$

Two-Sided versus One-Sided Tests

- Regardless of which convention you choose, remember that the equality is always represented under the null hypothesis, not the alternative hypothesis.

Two-Sided versus One-Sided Tests

- Hypotheses of the form

$$H_0 : \mu = \mu_0$$

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

form the basis for a *Two-Sided Test*.

Two-Sided versus One-Sided Tests

- Hypotheses of the form

$$H_0 : \mu = \mu_0$$

$$H_A : \mu > \mu_0$$

or the form

$$H_0 : \mu = \mu_0$$

$$H_A : \mu < \mu_0$$

form the basis for a *One-Sided Test*.

Steps in Hypothesis Testing (General)

1. Label and describe the parameter(s) of interest.
2. State both the null hypothesis H_0 and the alternative hypothesis H_A symbolically.
3. Select a value for α. The most common choices of α are 0.01, 0.05, and 0.10.
4. Specify the *Test Statistic* to be used. The test statistic is a number computed on the basis of the hypotheses and the sample data which is used in deciding whether to reject H_0.
5. Compute the numerical value of the test statistic.

Steps in Hypothesis Testing (General)

6. Compute the *p-value* for the test based on (i) the form of H_A, and (ii) the numerical value of the test statistic.

   The *p-value* (probability value) addresses the following question. Assuming H_0 is true, how likely would it be to observe a test statistic as extreme or more more extreme than the one we obtained in our study?

Steps in Hypothesis Testing (General)

- If the *p*-value is 'large,' the test statistic appears entirely plausible under the assumption that the null hypothesis is correct. We therefore do not have sufficient reason to question the validity of the null hypothesis.

- If the *p*-value is 'small,' the test statistic appears improbable under the assumption that the null hypothesis is correct. We therefore have sufficient cause to reject the null hypothesis in favor of the alternative hypothesis.

Steps in Hypothesis Testing (General)

7. Arrive at a conclusion by either (1) comparing the p-value to α, or (2) determining whether the test statistic falls into the *rejection region*.

- If the p-value $\leq \alpha$, we reject H_0 in favor of H_A.
  The alternative hypothesis has met the burden of proof.

- If the p-value $> \alpha$, we do not reject H_0.
  The alternative hypothesis has not met the burden of proof, so we continue to accept the 'status quo' hypothesis.

- The *Rejection Region* is a set of values for the test statistic that should lead to the rejection of H_0 in favor of H_A. This set of values is formulated based on (i) the form of H_A, and (ii) the choice of α.

Steps in Hypothesis Testing (General)

8. State the conclusion.
   - That is, whether or not H_0 should be rejected.
   - The conclusion should be stated in words related to the context of the problem.
   - It should also make reference to the p-value and to the level of significance α.

Steps for a z-test on a Population Mean

1. Label and describe the parameter of interest.
2. State the null hypothesis H_0 symbolically: $\mu = \mu_0$.
   State the alternative hypothesis H_A symbolically:
   $\mu \neq \mu_0$, $\mu < \mu_0$, $\mu > \mu_0$.
3. Select a value for α.

Steps for a z-test on a Population Mean

4. Specify the *Test Statistic* to be used.
   For the z-test, the test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

   Note: This test statistic can be used under either of the following settings.

- When the population is normal and σ is known.
- When n is 'large' ($n \geq 30$).
  (If n is 'large' and σ is unknown, s can be used in place of σ.)

Steps for a z-test on a Population Mean

5. Compute the numerical value of the test statistic.

6. Compute the p-value for the test from the normal tables.

H_A	p-value
$\mu \neq \mu_0$	$2P(Z > \|z\|)$
$\mu > \mu_0$	$P(Z > z)$
$\mu < \mu_0$	$P(Z < z)$

Steps for a z-test on a Population Mean

7. Arrive at a conclusion by either:
   (1) comparing the p-value to α; or
   (2) determining whether the test statistic falls into the rejection region.

H_A	Rejection Region
$\mu \neq \mu_0$	$z < -z_{\alpha/2}$ or $z > z_{\alpha/2}$
$\mu > \mu_0$	$z > z_\alpha$
$\mu < \mu_0$	$z < -z_\alpha$

8. State the conclusion: whether or not H_0 should be rejected.

Radon Example of a z-test on a Population Mean

- Radon is a colorless, odorless gas that is naturally released by rocks and soils and may concentrate in tightly closed houses.
- Because radon is slightly radioactive, there is a concern that it might pose a health hazard.
- Radon detectors are sold to homeowners worried about this risk, but household detectors may be inaccurate.
- Purdue University researchers placed 12 detectors in a chamber where they were exposed to 105 picocuries per liter (pCi/L) of radon over 3 days.

Radon Example of a z-test on a Population Mean

- The readings given by the 12 detectors are as listed below (Moore and McCabe, 2003):

91.9	97.8	111.4	122.3	105.4	95.0
103.8	99.6	96.6	119.3	104.8	101.7

- Assume that the readings follow a normal distribution.
- Also, assume (unrealistically) that you know the standard deviation of readings for all detectors of this type is $\sigma = 9$ pCi/L.
- At the 5% level, test whether there is statistically significant evidence that the mean detector reading differs from the true level of 105 pCi/L.

Radon Example of a z-test on a Population Mean

Lead Example of a z-test on a Population Mean

- Childhood lead poisoning has a significant impact on the health of children.
- Lead has adverse effects on nearly all of the organ systems in the body.
- A lead blood level of 10 μg/dl has been identified by the CDC as the threshold for lead poisoning in children.
- Since 1992, the Iowa Department of Public Health has recommended that all children under the age of six years be tested for lead poisoning.

Lead Example of a z-test on a Population Mean

- For children born from 1992–1996 in Scott county, 15.2% of those tested before the age of 6 had lead blood levels above 10 μg/dl.
- In a certain region of Scott county, suppose that 36 children are randomly sampled and tested for lead poisoning.
- For this sample, the mean lead blood level is 13.1 μg/dl with a standard deviation of 6.9 μg/dl.
- Test whether the true mean lead blood level for children in this region is above 10 μg/dl. Use $\alpha = 0.05$.

Lead Example of a z-test on a Population Mean

Equivalence of Confidence Intervals and Hypothesis Tests

I. Suppose a two-sided test of

$$H_0 : \mu = \mu_0$$

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

is conducted using α as the level of significance.

Suppose a $100(1 - \alpha)\%$ two-sided confidence interval is also constructed for μ.

Fact: The test will reject H_0 if and only if the confidence interval does not cover μ_0.

Equivalence of Confidence Intervals and Hypothesis Tests

II. Suppose a one-sided test of

$$H_0 : \mu = \mu_0$$

$$H_A : \mu < \mu_0$$

is conducted using α as the level of significance.

Suppose a $100(1-\alpha)\%$ one-sided confidence interval consisting of an upper confidence bound is also constructed for μ.

Fact: The test will reject H_0 if and only if the upper confidence bound for μ is less than μ_0.

Equivalence of Confidence Intervals and Hypothesis Tests

III. Suppose a one-sided test of

$$H_0 : \mu = \mu_0$$

$$H_A : \mu > \mu_0$$

is conducted using α as the level of significance.

Suppose a $100(1-\alpha)\%$ one-sided confidence interval consisting of a lower confidence bound is also constructed for μ.

Fact: The test will reject H_0 if and only if the lower confidence bound for μ is greater than μ_0.

Lead Example Equivalence of CIs and Hypothesis Tests

- In the lead poisoning example, use the sample data to construct a 95% lower confidence bound for the true mean lead blood level for children in the Scott county region. Does the lower bound exceed 10 μg/dl?

Steps for a t-Test on a Population Mean

1. Label and describe the parameter of interest.
2. State the null hypothesis H_0 symbolically: $\mu = \mu_0$.
   State the alternative hypothesis H_A symbolically:
   $\mu \neq \mu_0$, $\mu < \mu_0$, $\mu > \mu_0$.
3. Select a value for α.

Steps for a t-Test on a Population Mean

4. Specify the *Test Statistic* to be used.
   For the t-test, the test statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

   Here, t is based on $n - 1$ degrees of freedom (df).

   Note: This test statistic is generally used when n is below 30, σ is unknown, and the population distribution is normal (or at least approximately normal).

Steps for a t-Test on a Population Mean

5. Compute the numerical value of the test statistic.

6. Compute the p-value for the test using a statistical software package. Alternatively, find bounds for the p-value using the t tables.

H_A	p-value
$\mu \neq \mu_0$	$2P(T > \lvert t \rvert)$
$\mu > \mu_0$	$P(T > t)$
$\mu < \mu_0$	$P(T < t)$

Steps for a t-Test on a Population Mean

7. Arrive at a conclusion by either:
   (1) comparing the p-value to α; or
   (2) determining whether the test statistic falls into the rejection region.

H_A	Rejection Region
$\mu \neq \mu_0$	$t < -t_{\alpha/2}$ or $t > t_{\alpha/2}$
$\mu > \mu_0$	$t > t_{\alpha}$
$\mu < \mu_0$	$t < -t_{\alpha}$

8. State the conclusion: whether or not H_0 should be rejected.

Steps in Bounding a t-Test p-Value Using Table

Let t denote the observed value of the test statistic.

1. Consider $|t|$.

2. In the appropriate t table row, find two adjacent critical values t_1 and t_2 which bound $|t|$: $t_1 < |t| < t_2$.

3. In the top row of the table, find the upper tail areas p_1 and p_2 corresponding to t_1 and t_2.

4. For a one-sided test, the p-value is between p_1 and p_2; for a two-sided test, the p-value is between $2p_1$ and $2p_2$.

Example of Bounding a t-Test p-Value Using Table

In conducting a two-sided t-test based on a sample of size 5, suppose we obtain a test statistic of $t = -2.532$.
Find bounds for the p-value.

1. $|t| = +2.532$
2. From the row corresponding to $\text{df} = 4$, $|t| = 2.532$ is between $t_1 = 2.132$ and $t_2 = 2.776$.
3. From top row of table, $p_1 = 0.05$ and $p_2 = 0.025$.
4. Since we are conducting a two-sided test, the p-value is between $2p_2 = 0.05$ and $2p_1 = 0.10$.

Example for a t-Test on a Population Mean

- The group Led Zeppelin was one of the most successful bands of the rock era. Nearly all of their studio albums hit the #1 position on the *Billboard* charts, yet the band had very few hit singles. It has been claimed that Led Zeppelin failed to have much success on the singles charts because it tended to write songs that were too long for pop radio stations.
- The average Top 40 hit is about 3.5 minutes in length. Using a random sample of 16 Led Zeppelin songs, we will test the hypothesis that the mean length of all Led Zeppelin songs exceeds this Top 40 benchmark. We will conduct the test at the $\alpha = 0.05$ level.
- The lengths of the 16 songs (in minutes) are as follows:

2.43	4.97	5.55	2.67	2.72	2.10	4.92	4.82
5.40	3.67	4.33	8.55	10.43	4.70	6.80	5.85

Example for a t-Test on a Population Mean

Types of Errors and Power

- The probability of committing a type I error, the level of significance, is denoted by α.

$$\alpha = P(\text{Reject } H_0 | H_0 \text{ true})$$

- The probability of committing a type II error is denoted by β.

$$\beta = P(\text{Retain } H_0 | H_A \text{ true})$$

- The power of the test is given by $(1 - \beta)$.

$$(1 - \beta) = P(\text{Reject } H_0 | H_A \text{ true})$$

The power reflects the probability of the test to establish the validity of a true alternative hypothesis.

Power

- The power (or, equivalently, β) must be computed for a particular value of μ, say μ_A, represented under the alternative hypothesis.
- Different values of μ_A will correspond to different powers (different β's)

Power

Why are we concerned about power?

- If the power of a test is too low, then we have little chance of obtaining statistically significant results (i.e., of rejecting H_0), even if the alternative hypothesis is true.

- If the difference between μ_0 and μ_A is clinically meaningful, it is of practical importance to statistically detect this difference.

Power

- In most applications, if the difference between μ_0 and μ_A is deemed to be clinically meaningful, the power of the test to detect this difference should be at least 80%.
- The power of a test can be increased by:
  (i) increasing the sample size, or
  (ii) increasing α.

Evaluating the Power of a Test

- The power of a test is evaluated before the sample data is collected.

- To assess the power, we need to determine two important quantities: μ_A and σ.

  Note: Some authors use μ_1 in place of μ_A.

- The value μ_A is chosen so that the difference between μ_0 and μ_A is clinically meaningful.

- The absolute difference $|\mu_A - \mu_0|$ is called the *Effect Size*.

Evaluating the Power of a Test

As noted earlier, the value of σ may be based on

- s from a previous study or a 'pilot' study,
- an 'educated guess,'
- one-fourth the plausible range of the variable being measured:

$$\frac{1}{4}(\max - \min) \approx \sigma$$

We will discuss the evaluation of power only in the context of z-tests.

Power

The power computation is performed once the hypotheses and the level of significance have been determined. The computation involves three steps.

1. Determine μ_A and σ.

2. Using the rejection region for the test statistic, find the values of $\bar{x}$ that will lead to the rejection of H_0.

3. Assume that the mean of the sampled population is μ_A. Calculate the probability of obtaining a value for $\bar{X}$ that will lead to the rejection of H_0.

Lead Example: Power

- In the lead poisoning example, evaluate the power of the $\alpha = 0.05$, one-sided test under the assumption that $\mu_A = 14$ μg/dl.
- Recall that the hypotheses for the test were

$$H_0 : \mu = 10\, \mu\text{g/dl}$$

$$H_A : \mu > 10\, \mu\text{g/dl}$$

- Assume that $\sigma = 6.9$ μg/dl, and that the planned sample size for the study is $n = 36$.

Lead Example: Power

Serum Cholesterol Level Example: Power

- Consider the serum cholesterol level example.
- For the one-sided test, the hypotheses were

$$H_0 : \mu = 211 \text{ mg/100 ml}$$

$$H_A : \mu > 211 \text{ mg/100 ml}$$

- Evaluate the power of the one-sided test assuming that $\sigma = 46$ mg/100 ml, that $\mu_A = 230$ mg/100 ml, that $\alpha = 0.05$, and that the planned sample size for the study is $n = 36$.

Serum Cholesterol Level Example: Power

Sample Size Estimation

- **Problem**: Suppose we are planning a hypothesis test for a population mean μ using a level of significance α. Let μ_A denote a value of μ represented under the alternative hypothesis, such that the effect size $|\mu_A - \mu_0|$ represents a clinically important difference.

  What sample size would be required in order for the power of the test to be at least $(1 - \beta)$?

  We will address this question only in the context of *z*-tests.

  Note: Some authors use μ_1 in place of μ_A.

Sample Size Estimation

To complete the sample size calculation, the following quantities are required:

- α and β;
- for a two-sided test, the critical value $z_{\alpha/2}$;
  for a one-sided test, the critical value z_{α};
- the critical value z_{β};
- the effect size $|\mu_A - \mu_0|$; and
- the standard deviation σ.

Sample Size Estimation

Solution:

- For a two-sided test, take

$$n \geq \left[\frac{(z_{\alpha/2} + z_{\beta})\sigma}{|\mu_A - \mu_0|}\right]^2$$

- For a one-sided test, take

$$n \geq \left[\frac{(z_{\alpha} + z_{\beta})\sigma}{|\mu_A - \mu_0|}\right]^2$$

Lead Example: Sample Size Estimation

- In the lead poisoning example, consider again the one-sided test of the hypotheses

$$H_0 : \mu = 10\,\mu\text{g/dl}$$

$$H_A : \mu > 10\,\mu\text{g/dl}$$

- Suppose that $\sigma = 6.9$ μg/dl, and that the level of significance for the test is set at $\alpha = 0.05$.

a) Assuming that $\mu_A = 14$ μg/dl, the investigator is willing to risk only a 5% chance of failing to reject (i.e., of failing to detect the difference $|\mu_A - \mu_0|$). What sample size will be required?

Lead Example: Sample Size Estimation

Lead Example: Sample Size Estimation

b) Assuming again that $\mu_A = 14$ μg/dl, the investigator is willing to risk only a 1% chance of failing to reject. What sample size will be required?

Important Considerations in Hypothesis Testing

1. The conclusion in a hypothesis test depends critically on the choice of α.

   For example, in a hypothesis test on μ, suppose the *p*-value is 0.0446.

   If $\alpha = 0.05$, we reject H_0 (*p*-value $\leq \alpha$).

   If $\alpha = 0.01$, we retain H_0 (*p*-value $> \alpha$).

   In each of the preceding settings, the extent of evidence against H_0 is the same. However, the conclusions are different because the burden of proof required to reject H_0 is greater in the second setting than in the first.

Important Considerations in Hypothesis Testing

2. In a hypothesis test, if the sample results lead to the rejection of H_0, we say the results are *Statistically Significant*. If the discrepancy between the sample results and what is hypothesized under H_0 is clinically meaningful, we say the results are *Clinically Significant*.

   Statistical significance **does not** imply clinical significance.

Important Considerations in Hypothesis Testing

- To illustrate this point, suppose that a study geared towards American males is designed to determine whether an entirely vegetarian diet will increase life expectancy. The current life expectancy of American males is 74.1 years.

- In a prospective longitudinal study, assume that 3,240 randomly selected vegetarian seniors are monitored throughout the remainder of their lifetimes. In this sample, the mean lifespan was found to be $\bar{x} = 74.3$ years with a standard deviation of $s = 5.8$ years.

Important Considerations in Hypothesis Testing

- A test of

$$H_0 : \mu = 74.1 \text{ years}$$

$$H_A : \mu > 74.1 \text{ years}$$

  is conducted at the $\alpha = 0.05$ level. A p-value of 0.025 is obtained, leading to the rejection of H_0.

- Thus, the results are statistically significant.

- However, the sample of vegetarians only lived 0.2 years longer, on average, than the general population.

- Is this difference of clinical importance?

Important Considerations in Hypothesis Testing

3. Inferential procedures are only valid for random samples.

   Non-random samples, such as volunteer samples and convenience samples, are not scientific and should not be analyzed in a formal statistical context.