SWAYAM NPTEL COURSE ON

MINE AUTOMATION AND DATA ANALYTICS

By

Prof. Radhakanta Koner

Department of Mining Engineering

Indian Institute of Technology (Indian School of Mines) Dhanbad

Module 9 : Hypothesis Testing

Lecture 21A : Hypothesis Testing - I

CONCEPTS COVERED

1) What is Hypothesis Testing?
2) Size and Power of a Test.
3) Neyman - Pearson Paradigm of Hypothesis Testing
4) Types of hypothesis testing
5) Motivation example for hypothesis testing

"You can't prove a hypothesis; you can only improve or disprove it." – Christopher Monckton

What is Hypothesis Testing

- Motivating Example: Is a coin fair or unfair?
- A fair coin is said to have a probability of getting head P(H) = 0.5
- An unfair coin is said to have a probability of getting head P(H) = 0.6
- Let us suppose you have a coin that could be fair or unfair. You may toss the coin multiple times and observe the results. How would you test whether the coin is fair or unfair?

(i) Null Hypothesis (H_0):

The null hypothesis (H_0) is a statement about a population parameter or effect that is assumed to be true unless evidence suggests otherwise.

It represents the status quo or a baseline assumption.

Formally, the null hypothesis is denoted as H_0 and is typically expressed as equality.

(ii) Alternative Hypothesis (H_A):

The alternative hypothesis (H_A) is a statement that contradicts the null hypothesis.

It represents what the researcher is trying to provide evidence for.

What is Hypothesis Testing

- Motivating Example: Is a coin fair or unfair?
- A fair coin is said to have a probability of getting head P(H) = 0.5
- An unfair coin is said to have a probability of getting head P(H) = 0.6
- Let us suppose you have a coin that could be fair or unfair. You may toss the coin multiple times and observe the results. How would you test whether the coin is fair or unfair?

- <u>Hypothesis Testing:</u>
- Using samples, decide between a null hypothesis (H_0) and an alternative hypothesis (H_A)
- Fair Coin Example:

$$H_0 : P(H) = 0.5$$

$$H_A : P(H) = 0.6$$

- One of the most important statistical analysis methods with a wide range of applications.

What is Hypothesis Testing

- In summary, the null hypothesis represents the assumption to be tested, while the alternative hypothesis represents the researcher's claim or the possibility of an effect or difference.

- The goal of hypothesis testing is to gather evidence from sample data to decide whether to reject the null hypothesis in favor of the alternative hypothesis.

Accepting or Rejecting the Null Hypothesis

- **<u>Motivating Example: Is a coin fair or unfair?</u>**
- **Suppose we toss the coin 3 times.**
- **Possible outcomes are HHH, HHT, … , TTT**
- **For some outcomes, we will accept H_0 and others, we will reject H_0**
- **Let A be the set of all outcomes for which we accept H_0**
- **Every acceptance subset A corresponds to a test**

Size and Power of a Test

- Metric 1: Significance level (also called size) of a test, denoted α.
- Type I Error: Reject H_0 when H_0 is true
- Size of a test = α = P(Type I error) = P(Reject H_0 | H_0 is true)

- Metric 2: Power of a test, $1 - \beta$
- Type II error: Accept H_0 when H_A is true
- β = P(Type II error) = P(Accept H_0 | H_A is true)
- Power = $1 - \beta$ = P(Reject H_0 | H_A is true)

Computing the Size and Power for Unfair Coin Example

$H_0 : P(H) = 0.5$

$H_A : P(H) = 0.6$

- Toss 3 times = { HHH , HHT, HTH, THH, THT , HTT , TTH , TTT}
- If acceptance subset A = ∅
- Always reject H_0
- $\alpha = 1$, $\beta = 0$
- If acceptance subset A = { HHH, HHT, HTH, THH, THT, HTT, TTH, TTT}
- Always accept H_0
- $\alpha = 0$, $\beta = 1$
- If acceptance subset A = { HHT, HTH, THH, THT, HTT, TTH }
- $\alpha = P(A^C \mid P(H) = 0.5) = 2/8 = 0.25$
- $\beta = P(A \mid P(H) = 0.6) = 3(0.4)^2(0.6) + 3(0.4)(0.6)^2 = 0.72$
- The value α, called the level of significance of the test, is usually set in advance, with commonly chosen values being α = .1, .05, .005.

Neyman-Pearson Paradigm of Hypothesis Testing

X1,X2,X3,…,Xn ~ iid X

- **H_0: Null Hypothesis on the distribution of X, H_A: Alternative Hypothesis**
- **Test: Defined by an acceptance set A**
- **If samples fall in A, accept H_0; otherwise, reject H_0**

- **<u>Two Errors:</u>**
- **Type I Error: Reject H_0 when H_0 is true**
- **Type II error: Accept H_0 when H_A is true**

- **<u>Two Metrics</u>**
- **Size of a test = α = P(Type I error) = P(Reject H_0 | H_0 is true)**
- **Power = 1 – β = P(Reject H_0 | H_A is true)**

Types of Hypothesis Testing

Simple Hypothesis:

- A hypothesis that completely specifies the distribution of the samples is called a simple hypothesis.
- Example:
  1) Coin Toss ; P(Heads) = 0.5 , P(Heads) = 0.8
  3) Normal (μ, 3) samples ; $\mu = 1$, $\mu = -1$ etc.,
- Simple null vs simple alternative

Types of Hypothesis Testing

Composite Hypothesis

- A hypothesis that does not specify the distribution of the samples is called a **Composite hypothesis**.

Example 1: Coin Toss ;

- Null: P(Heads) = 0.5 (coin is fair), simple
- Alternative: P(Heads) ≠ 0.5 (coin is unfair), composite

Example 2: Normal (μ, 3) samples ;

- Null: μ = 0 (some effect is not present), simple
- Alternative: μ > 1 (the effect is present), composite

- **Simple / Composite Null vs Composite Alternative**

Types of Hypothesis Testing
Standard Tests: One Sample

$X1, X2, X3, \ldots, Xn \sim$ iid X

$E(X) = \mu$; $Var(X) = \sigma^2$

- **Testing for mean,**

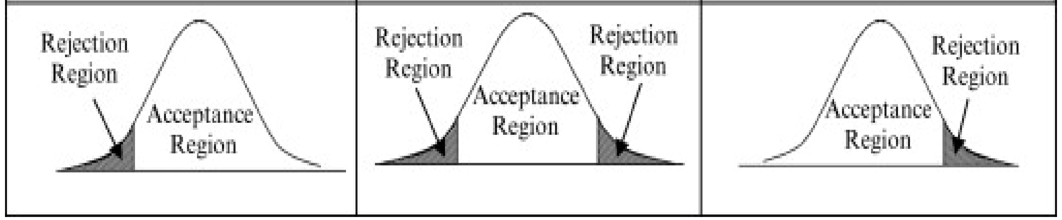
Null $H_0 : \mu = c$

Alternative :

- **Right tail test, $H_A: \mu > c$**
- **Left tail test, $H_A: \mu < c$**
- **Two tail test, $H_A: \mu \neq c$**

Two Cases: known or unknown variance

One-Tailed Test (Left Tail)	Two-Tailed Test	One-Tailed Test (Right Tail)
$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X < \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X \neq \mu_0$	$H_0 : \mu_X = \mu_0$ $H_1 : \mu_X > \mu_0$

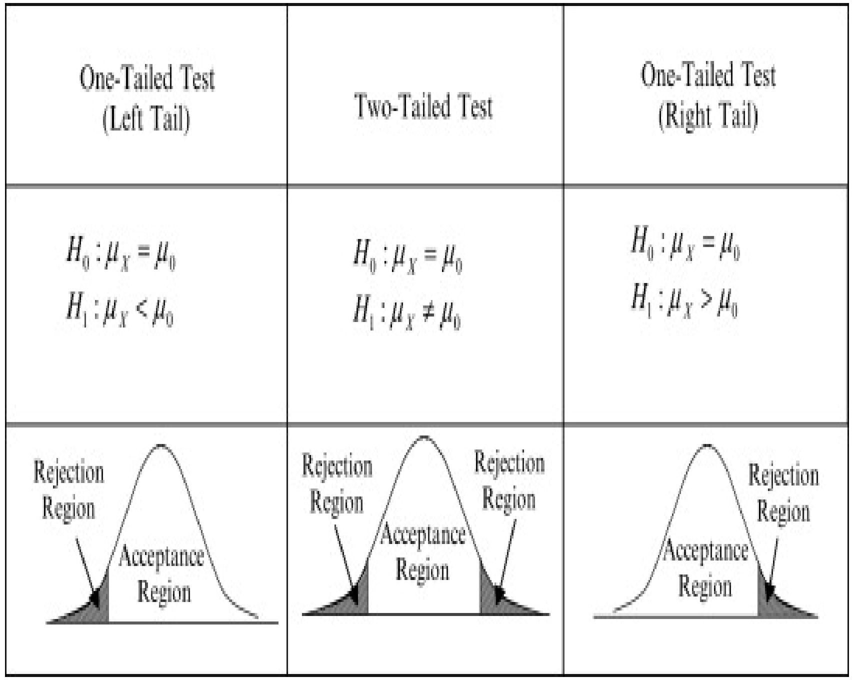


Z-score values for Rejection Regions

99% Confidence level (i.e alpha = 0.01):
Left-tailed test: z = -2.33
Two-tailed test: z = +- 2.55
(the critical z-values are +2.55 and -2.55)
Right-tailed test: z = +-2.33

95% Confidence level (i.e alpha = 0.05):
Left-tailed test: z = -1.65
Two-tailed test: z = +-1.96
(the critical z-values are -1.96 and 1.96)
Right-tailed test: z = +1.65

90% Confidence level (i.e alpha = 0.1):
Left-tailed test: z = -1.2
Two-tailed test: z = +-1.65
(the critical z-values are -1.65 and 1.65)
Right-tailed test: z = +1.2



Standard Tests: One Sample

$$X1, X2, X3, \ldots, Xn \sim \text{iid } X$$

$$E(X) = \mu \; ; \; Var(X) = \sigma^2$$

- Testing for variance,

Null $H_0 : \sigma = c$

Alternative :

- Right tail test, $H_A: \sigma > c$
- Left tail test, $H_A: \sigma < c$
- Two-tail test, $H_A: \sigma \neq c$

Standard Tests: Two Sample

X1,X2,X3,…,Xn ~ iid X

Y1,Y2,Y3,…,Yn ~ iid Y

E(X) = μ1 ; Var(X) = $\sigma 1^2$

E(X) = μ2 ; Var(X) = $\sigma 2^2$

- **Testing to compare means**

Null H_0 : μ1 = μ2

Alternative: H_A : μ1 ≠ μ2

- **Testing to compare variances**

Null H_0 : σ1 = σ2

Alternative: H_A : σ1 ≠ σ2

Goodness of fit testing

Samples: X1,X2,X3,…,Xn

- **Problem: Do the samples follow a certain distribution?**

- **Examples :**
- **Integer Samples Xi ∈ { 0,1,2, … }. Is the distribution Poisson?**
- **Continuous Samples Xi ∈ (- ∞ , ∞) . Is the distribution normal?**

Observations

- In all examples, the questions seem to be reasonably posed in a statistical hypothesis testing framework.
- In most cases, the null and/or alternative are composite
- In all cases, the confidence of the testing is very important.

How do you quantify confidence?

- With the help of an alpha value concept (or)
- A notion called P - value is used to quantify confidence.
- The P-value is known as the probability value.
- It is defined as the probability of getting a result that is either the same or more extreme than the actual observations.
- The P-value is known as the level of marginal significance within the hypothesis testing that represents the probability of occurrence of the given event.
- The P-value is used as an alternative to the rejection point to provide the least significance at which the null hypothesis would be rejected.
- If the P-value is small, then there is stronger evidence in favor of the alternative hypothesis.

Covering concepts (through an example)

Suppose that a construction firm has just purchased a large supply of cables that have been guaranteed to have an average breaking strength of at least 7,000 psi.

To verify this claim, the firm has decided to take a random sample of 10 of these cables to determine their breaking strengths. They will then use the result of this experiment to ascertain whether or not they accept the cable manufacturer's hypothesis that the population mean is at least 7,000 pounds per square inch.

A statistical hypothesis is usually a statement about a set of parameters of a population distribution. It is called a hypothesis because it is not known whether or not it is true.

A primary problem is to develop a procedure for determining whether or not the values of a random sample from this population are consistent with the hypothesis

For instance,
consider a particular normally distributed population having an unknown mean value θ and known variance 1.

The statement "θ is less than 1" is a statistical hypothesis that we could try to test by observing a random sample from this population.

If the random sample is deemed to be consistent with the hypothesis under consideration, we say that the hypothesis has been "accepted"; otherwise, we say that it has been "rejected"

Important Note: In accepting a given hypothesis, we are not actually claiming that it is true but rather we are saying that the resulting data appear to be consistent with it.

For instance,
in the case of a normal $(\theta, 1)$ population,

If a resulting sample of size 10 has an average value of 1.25, then although such a result cannot be regarded as being evidence in favor of the hypothesis "$\theta < 1$," it is not inconsistent with this hypothesis, which would thus be accepted.

On the other hand, if the sample of size 10 has an average value of 3, then even though a sample value that large is possible when $\theta < 1$, it is so unlikely that it seems inconsistent with this hypothesis, which would thus be rejected.

Significance Levels

Consider a population having distribution F_θ, where θ is unknown, and suppose we want to test a specific hypothesis about θ.

We shall denote this hypothesis by H_0 and call it the null hypothesis. For example, if F_θ is a normal distribution function with mean θ and variance equal to 1, then two possible null hypotheses about θ are

$$(a)\ H_0 : \theta = 1$$
$$(b)\ H_0 : \theta \leq 1$$

Thus, the first of these hypotheses states that the population is normal with mean 1 and variance 1, whereas the second states that it is normal with variance 1 and a mean less than or equal to 1.

Note: the null hypothesis in (a), when true, completely specifies the population distribution, whereas the null hypothesis in (b) does not. A hypothesis that, when true, completely specifies the population distribution is called a ***simple*** **hypothesis**; one that does not is called a ***composite*** **hypothesis**.

Suppose now that in order to test a specific null hypothesis H_0, a population sample of size n — say X1, . . . , X_n — is to be observed. Based on these n values, we must decide whether or not to accept H_0.

A test for H_0 can be specified by defining a region C in n-dimensional space with the provison that the hypothesis is to be rejected if the random sample X1, . . . , Xn turns out to lie in C and accepted otherwise. The region C is called the critical region. In other words, the statistical test determined by the critical region C is the one that

$$\text{accepts } H_0 \quad \text{if} \quad (X_1, X_2, \ldots, X_n) \notin C$$

$$\text{and} \quad \text{rejects } H_0 \quad \text{if} \quad (X1, X_2, \ldots, Xn) \in C$$

For instance, a common test (will cover in next lecture) of the hypothesis that θ, the mean of a normal population with variance 1, is equal to 1 has a critical region given by

$$C = \{(X_1, \ldots, X_n) : |\bar{X} - 1| > 1.96 / \sqrt{n}\}$$

Thus, this test calls for rejection of the null hypothesis that $\theta = 1$ when the sample average differs from 1 by more than 1.96 divided by the square root of the sample size.

Important Note:

When developing a procedure for testing a given null hypothesis H_0 that, in any test, two different types of errors can result.

1. The first of these, called a type I error, is said to result if the test incorrectly calls for rejecting H_0 when it is indeed correct.
2. The second, called a type II error, results if the test calls for accepting H_0 when it is false.

The objective of a statistical test of H_0 is not to explicitly determine whether or not H_0 is true but rather to determine if its validity is consistent with the resultant data.

Hence, with this objective, it seems reasonable that H_0 should only be rejected if the resultant data are very unlikely when H_0 is true.

The classical way of accomplishing this is

(i) First, specify a value α and

(ii) then require the test to have the property that whenever H_0 is true its probability of being rejected is never greater than α.

The value α, called the level of significance of the test, is usually set in advance, with commonly chosen values being α = .1, .05, .005.

In other words,

the classical approach to testing H_0 is to fix a significance level α and then require that the test have the property that the probability of a type I error occurring can never be greater than α.

REFERENCES

- Introduction to Probability and Statistics for Engineers and Scientists, Sixth Edition, Sheldon M. Ross
- Statistical Methods Combined Edition (Volume I& II), N G Das

CONCLUSION

- Defined the Hypothesis Testing with examples.
- Discussed how to compute the Size and Power of a Test.
- Discussed the Neyman-Pearson Paradigm of Hypothesis Testing.
- Types of hypothesis testing
  a. Standard test – one sample
  b. Standard tests – two sample
  c. Goodness of fit testing
- Motivation example for hypothesis testing with significance level (alpha).


THANK YOU