MATH F113
Probability and Statistics

BITS Pilani
Pilani Campus

Dr. Shivi Agarwal
Department of Mathematics

8.3 Hypothesis testing

- We want to apply the results of previous chapter to test claims about population parameters, like $\mu \geq 10$, or $p = 0.1$ for binomial population with $n=100$, etc.
- Claims will be tested based on information available from samples.

Null and Alternate hypotheses

- If a certain claim is made about population parameter then it is compared against another hypothesis, exclusive from the first, which is accepted when original hypothesis is rejected.

Hypothesis testing:

Always in any statistical study, we have 2 hypotheses:

The null hypothesis(H_0): a preconceived notion regarding the parameter value. The hypothesis to be contested.

Alternate hypothesis(H_1): The claimed or proposed value. The hypothesis which we wish to establish.

Guidelines while deciding H_0 and H_1:

H_0 denotes the specific numerical value that could be the actual value. This is termed as '*null value*' and denoted by (θ_0). Always includes the '*statement of equality*'.

Whatever *we claim* is taken as *alternate hypothesis.*

We hope that evidence will lead us to reject H_0 and thereby accept H_1

We always estimate a parameter called '*test statistic*'.

We make a *decision by observing the value of the test statistic* whose probability distribution is known under the assumption that the null value is true value of the parameter estimated θ_0.

If the test statistic assumes a value that is rarely seen when $\theta=\theta_0$, then we reject the null hypothesis and accept the alternate hypothesis.

Choice of hypotheses

- To test the claim of the drug company that its drug cures on average 25 cases, customer would choose as alternate hypothesis ‘it cures less than 25 cases’ as this was his intent in challenging the null hypothesis.
- The border value is included in the null hypothesis and is called the null value.

Type of errors

- We can have one of the 4 situations.

	H_0 accepted	H_0 rejected
H_0 true	No error	Type I error
H_0 false	Type II error	No error

- **Type-I error: Rejecting null hypothesis when it is actually true; Prob(type-I error) = α**
- **Type-II error: Failed to reject null hypothesis when it is false; Prob(type-II error) = β**
- **Power of a test (1-β): Probability of rejecting null hypothesis when it is false**

Possible end results:

Methods must be designed so that the probabilities of *making either type of error is reasonably small.*

The set of values of the test statistic that leads us to reject the null hypothesis is termed as '*Critical Region*'.

Definition 8.3.1

A type I error is an error made when the null hypothesis is rejected, in spite of it being true. *The probability of committing a type I error is called the 'level of significance' of the test* and is denoted by 'α'.

Probability of Type I error =

$$P[H_0 \text{ is rejected} | H_0 \text{ is true}].$$

Definition 8.3.2

Consider a test of hypothesis. A type II error is an error that is made when the null hypothesis is not rejected when, in fact, the research theory is true. The probability of committing a type II error is denoted by β

Definition 8.3.3

Consider a test of hypothesis. *The probability that the null hypothesis will be rejected when, in fact, the research theory is true* is called the power of the test.

Note: We will either fail to reject to the null hypothesis with probability β or we reject the null hypothesis with probability power

$$\beta + \text{power} = 1$$

Note: Our objective is always to keep α and β small and the power of the test to be high. This is usually achieved by choosing a appropriate sample size.

If the sample size is fixed, when Type I error is small, Type II error will be large and vice versa. So one may have to weigh which error should be controlled while designing the criterion.

Constructing null and alternative hypotheses

- **One-tailed and two-tailed test:**

$H_0: \theta \geq \theta_0$	$H_0: \theta \leq \theta_0$	$H_0: \theta = \theta_0$
$H_1: \theta < \theta_0$	$H_1: \theta > \theta_0$	$H_1: \theta \neq \theta_0$
(Left Tailed)	(Right Tailed)	(Two Tailed)

Example

From long experience of coca-cola company, it is known that yield is normally distributed with mean of 500 units and standard deviation 96 units. For a modified process, yield is 435 units for a sample of size 50. At 5% significance level, does the modified process decreased the yield?

$$H_0 : \mu \geq 500$$

$$H_1 : \mu < 500 \text{ (Left Tailed)}$$

Example

A department store manager determines that a new billing system will be cost effective only if the mean monthly account is more than Rs. 170 . A random sample of 400 monthly account is drawn, for which the sample mean is Rs.178 . It is known that the accounts are approximately normally distributed with s.d. of Rs. 65 . At 5% level of significance, can we conclude that the new system will be cost-effective?

$$H_0 : \mu \leq 170$$

$$H_1 : \mu > 170 \text{ (Right Tailed)}$$

Example

A drug is given to 10 patients, and the increments in their blood pressure were recorded as 3, 6, 2, 4, 4, 1, 6, 0, 0, 2. Is it reasonable to believe that the drug has no effect on change of the mean blood pressure? Test at 5% significance level, assuming that the population is normal with variance 1.

$$H_0 : \mu = 0$$

$$H_1 : \mu \neq 0 \text{ (Two Tailed)}$$

Exercise 1

To test the null hypothesis that population mean $\mu=4$ against alternate hypothesis $\mu=5$, a test is designed based on a random sample of size 49 which rejects null hypothesis if the observed sample mean $\bar{x} > 4.3$. Find the probabilities of Type I and Type II errors if the population variance is 9.

Solution

$\bar{X} \sim N(\mu, \sigma/\sqrt{n})$, $\sigma = 3$.

Probability of Type I error is probability of rejecting null hypothesis when it is assumed true, i.e. $\mu=4$.

Thus

$\alpha = P(\bar{X} > 4.3)$

$= P((\bar{X}-4)/(3/7)) > (4.3-4)/(3/7))$

$= P(Z > 0.7)$

$= 1 - F(0.7)$

$= 1 - 0.7257$ (by standard normal table).

$= 0.2743$

β = probability of accepting null hypothesis when it is false = Probability of accepting null hypothesis when alternate hypothesis is true.

Now $X \sim N(\mu, \sigma/\sqrt{n})$, $\sigma = 3$. random sample of size 49.

Thus probability of type II error is

$P(Z \leq (4.3 - 5)/(3/7)) = F(-1.63) = 0.0516.$

Q 8.3.26 : Suppose we want to test

H_0 : $p \geq .7$

H_1 : $p < .7$

based on a sample size 10.

(a) Find the critical region for an α =.05 level test.

(b) If when the data are gathered, $x = 5$, will H_0 be rejected? What type error might you be making?

Testing of Hypothesis / Test of significance

- On the basis of a sample, $x_1, x_2, \ldots\ldots, x_n$ of size n drawn from a population, we are going to reject or accept a null hypothesis H_0. The method adopted for this is known as ***testing of Hypothesis.***

- Since it is not possible to make a decision with perfect certainty, we have to say how much confidence we can place on such a decision, so these tests are often called ***test of significance.***

8.4 Significance testing

In this method, we first set H_0 and H_1.

But *we do not preset* α and specify a rigid critical region.

Instead, we *first evaluate the test statistic*.

Then we *find the probability of observing a value of the test statistic* at least as extreme as the value noted under the assumption that $\theta=\theta_0$.

8.4 Significance testing

This probability is referred by many names like critical level, descriptive level of significance or "P value" of the test. Book notation is "P value".

We reject *H_0 if we find P to be very small*.

If we have a industry set *maximum acceptable level of risk* α, then we can compare the P value to the preset value α.

If $P \leq \alpha$, *then reject the null hypothesis* at the stated level of significance. This method is widely accepted today.

Sometimes in hypothesis testing there could be an error in case you are on the margin of critical region. This can be avoided using significance testing.

Test types:

Right-tailed test: *P value is to the right side* of the test statistic.

Left-tailed test: *P value is to the left side* of the test statistic.

Two-tailed test: If the distribution of the test statistic is *symmetric*, it is logical to double the one-tailed P value. If it is *asymmetric*, then P value is approx. double the one-tailed P value.

Q33: It is thought that more than 15 % of the furnaces used to produce steel in the U.S. are still open-hearth furnaces. To verify this contention, a random sample of 40 furnaces is selected and examined.

(a) Set the appropriate null alternative hypotheses required to support the statement contention.

Q33: (b) When the data are gathered, it is found that 9 of the 40 furnaces inspected are open-hearth furnaces. Use the normal approximation to the binomial distribution to find the P value for the test. Do you think that H_0 should be rejected? Explain. To what type of error are you now subject?

Q. 42

Lasers are now used to detect structural movement in bridges and large buildings. These lasers must be accurate. In laboratory testing of one such laser, measurements of error made by the device are taken. The data obtained are used to test $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$ A sample of 25 measurements yields sample mean of 0.03 mm over 100 m and $s = 0.1$. Find the P value for the two tailed test. Do you think H_0 will be rejected?

8.5 Hypothesis and significance tests on the mean:

Normally in hypothesis testing for the mean, we had assumed that we known the value of σ. This might not be true most of the times.

A hypothesis test on mean can typically take one of the following three types:

Right-tailed test: $H_0: \mu \leq \mu_0$
$H_1: \mu > \mu_0$

Left-tailed test: $H_0: \mu \geq \mu_0$
$H_1: \mu < \mu_0$

Two-tailed test: $H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$

Hypothesis Testing

1. Set up the appropriate null and alternate hypothesis.
2. Critical point for preset value of α. (level of significance)
3. Define the criterion for critical region.
4. Calculations
5. Conclusions

In one-sided tests, the *inequality in the alternate hypothesis points towards the critical region*.

Hence in a right-tailed test, the natural region leading towards rejection of the null hypothesis is the right-tailed region of the distribution of the test statistic.

Also, we know that if X is normal,

$$(\overline{X} - \mu_0)/(\sigma/\sqrt{n})$$

follows a standard normal-distribution. Tests based on this statistic are termed as ‘Z’ tests.

Popln dist	Popln s.d. σ	Sample size n	Test Statistic	Dist of statistic
Normal	known	any	$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	Normal (by linear expression)
Normal	Not known	small	$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$	T-distribution (**estimator** S of s.d.)
Any	Known	At least 25	$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$	Approx normal (By central LT)
Any	unknown but **estimate** s known	At least 25	$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$	Approx normal (CLT and approx of s.d. by estimate s)

Summary:

$$\text{Test Statistic: } Z = \frac{\overline{X} - \mu_0}{\dfrac{\sigma}{\sqrt{n}}}$$

Null Hypothesis H_0	Alternative Hypothesis H_1	Reject null hypothesis if
$\mu \geq \mu_0$	$\mu < \mu_0$	$Z < -z_\alpha$
$\mu \leq \mu_0$	$\mu > \mu_0$	$Z > z_\alpha$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert Z \rvert > z_{\alpha/2}$

Summary (Small Sample Formula)

$$\text{Test Statistic:} \quad t = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Null Hypothesis H_0	Alternative Hypothesis H_1	Reject null hypothesis if
$\mu \geq \mu_0$	$\mu < \mu_0$	$t < -t_{n-1,\alpha}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$t > t_{n-1,\alpha}$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert t \rvert > t_{n-1,\alpha/2}$

Example

According to the norms established for a mechanical aptitude test, persons who are above 18 years old should average 73.2 with a standard deviation of 8.6. If 45 randomly selected persons of that age averaged 76.7, test the null hypothesis $\mu = 73.2$ against the alternative hypothesis $\mu > 73.2$ at the 0.01 level of significance.

(Assume data is normal):

Solution: Here Null hypothesis is $H_0 : \mu = 73.2 \; (= \mu_0)$

Alternative hypothesis is $H_1 : \mu > 73.2$

Level of significance is $\alpha = 0.01$

Q 36. A new 8-bit microcomputer chip has been developed that can be reprogrammed without removal from the microcomputer. It is claimed that a byte of memory can be programmed on average in less than 14 seconds. Let X, the time required to reprogram a byte of memory has normal distribution.

(a) Set up the appropriate null and alternate hypothesis needed to verify this claim.

(b) What is the critical point for an $\alpha = 0.05$ level test based on a sample of size 15.

Q 36. (c) These data are obtained on X, the time required to reprogram a byte of memory

11.6	14.7	12.9	13.3	13.2
13.1	14.2	15.1	12.5	15.3
13.3	13.4	13.0	13.8	12.3

Test the null hypothesis. Can H_0 be rejected at the $\alpha = 0.05$ level? To what type error are you now subject?

What if for two-sided test? H_1: $\mu \neq 14$

Q 40. The Elbe River is important in the ecology of central Europe, as it drains much of the region. Due to increased industrialization, it is feared that the mineral content in the soil is being depleted. This will be reflected in an increase in the level of certain minerals in the water of Elbe. A study of the river conducted in 1982 indicated that the mean silicon level was 4.6 mg/l.

(a) Set up the appropriate null and alternate hypothesis needed to gain evidence to support the contention that the mean silicon concentration in the river has increased. Assume that the silicon concentration in the river is normally distributed.

innovate achieve lead

Q 40. (b) A sample of size 20 yields sample mean = 5.2 with s = 1.4. Find the P value for the test. Do you think that H_0 should be rejected?

Eg: As per health ministry advisory, the maximum acceptable level of fluoride in drinking water is on an average 8 mg/l. Due to increased industrialization and other factors, it is feared that the fluoride content in the underground water has reached to an alarming level making it unsafe for drinking. In order to support the contention that the fluoride content in the underground water has reached to an alarming level, the following values of fluoride content (in units of mg/l) are observed in different samples of underground water (Assume data is normal):

10.7	9.2	8.4	12.5	10.9	7.4	9.2	8.7
7.9	9.1	11.6	9.3	8.4	7.7	8.5	

(a) Choosing a suitable test statistic, make a sketch of the critical region at 0.005 level of significance, and use the same to draw your conclusion from the test.

(b) Using P value of the test, decide whether the contention can be supported at 0.001 level of significance.

Example

- The standard specification requires percentage of carbon content of a certain variety of steel to be at most 0.05. For 12 samples of this steel, the percentage of carbon content were found to have an average 0.0483 and standard deviation 0.00117. Do these data reasonably confirm to the standard specification? (Assume that the population of percentages of carbon content is normal and perform the test at 1% significant level). What if specification required percentage exactly 0.05? At 11 degrees of freedom, $P(T < 2.718) = 0.99$, $P(T < 3.106) = 0.995$.