Testing of hypothesis

Basis definitions

- **Population:** The group of individual under study is called Population.
- **Sample:** A finite subset of statistical individuals in a population is called Sample.
- **Sample size:** The number of individuals in a sample is called sample size.
- **Simple Random Sampling:** A random sample is one in which each unit of population has an equal chance (say p) of being included in it and this probability is independent of the previous drawing.
- **Parameter:** The statistical constant of population is called parameter e.g. Mean μ, Variance σ^2 etc.

Basic definitions

- **Statistics:** Statistical measures computed from the sample observations e.g. sample mean $\bar{x}$, sample variance s^2 etc.
- **Estimator:** Any function of the random sample$x_1, x_2, \ldots, x_n$that are being observed, say $T_n(x_1, x_2, \ldots, x_n)$ is called a statistic, it is a random variable and it is used to estimate an unknown parameter θ of the distribution, it is called an estimator. A particular value of the estimator say $T_n(x_1, x_2, \ldots, x_n)$ is called an estimate of θ.

Basic definitions and terminology

- **Standard error:** The standard deviation of the sampling distribution of statistic is called standard error.
- **Utility of standard error:** If the discrepancy between the observed and the expected value of a statistic is greater than z_α times the standard error, the null hypothesis is rejected at α level of significance.

Test of significance

- A very important aspect of the sampling theory is the study of the *test of significance,* which enable us to decide on the basis of sample results

(i) The deviation between the observed sample statistic and the hypothetical parameter value

Or

(ii) The deviation between two independent sample statistics

is significant or might be attributed to chance or fluctuations of sampling.

Null and Alternate hypothesis

- **Null Hypothesis:** According to *Prof. R.A. Fisher,* null hypothesis is the hypothesis which is tested for the possible rejection under the assumption that it is true. It is usually denoted by H_0.

- **Alternate hypothesis:** Any hypothesis which is complimentary to null hypothesis is called an alternate hypothesis, usually denoted by H_1.

- **For example:** if we want to test the null hypothesis that the population has specified mean μ_0 (say), i.e. $H_0: \mu = \mu_0$

Then the alternate hypothesis could be

(i) $H_1: \mu \neq \mu_0$ *(two tailed alternative)*

(ii) $H_1: \mu > \mu_0$ *(one tail right tail)* *(iii)* $H_1: \mu < \mu_0$ *(one tail left tail)*

Error in sampling

- **Type I error:** Reject H_0, when it is true.
- **Type II error:** Accept H_0, when H_1 is true.

If we write $P[Reject\ H_0|H_0] = \alpha$

And $P[Accept\ H_0|H_1] = \beta$

Then α and β are called the sizes of type I and type II error.

The power of a test is the probability of rejecting H_0 given that a specific alternative is true.

The power of a test can be computed as $\mathbf{1} - \boldsymbol{\beta}$.

Critical region and level of significance

- A region corresponding to the statistic (t) in the sample space S which amounts to the rejection of H_0 is termed as critical region or region of rejection. If ω is the critical region and t is the value of the statistic based on the random sample of size n, then

$$P(t \in \omega | H_0) = \alpha, P(t \in \overline{\omega} | H_1) = \beta$$

Where $\overline{\omega}$, the complementary set of ω, is called the acceptance region.

We have $\omega \cup \overline{\omega} = S$ and $\omega \cap \overline{\omega} = \phi$.

The probability α that a random value of the statistic t belongs to the critical region is known as the level of significance. (level of significance is the size of type I error)

One tailed and two tailed test

- In any test, the critical region is represented by the portion of the area under the probability curve of the sampling distribution of the test statistic.
- **One tailed test:** A test of any statistical hypothesis where the alternative hypothesis is one tailed is called one tailed test.
- **For example :** Null Hypothesis: $H_0: \mu = \mu_0$ against the alternative hypothesis $H_1: \mu > \mu_0$(one tailed Right tailed)

or $H_1: \mu < \mu_0$ (one tailed left tailed)

In right tailed test, the critical region lies entirely in the right tail of the sampling distribution of $\bar{x}$, while for left tailed, the critical region is entirely in the left tail of the distribution.

One tailed and two tailed test

- **Two tailed test:** A test of the statistical hypothesis where the alternative hypothesis is two tailed such as $H_0: \mu = \mu_0$ against the alternative hypothesis $H_1: \mu \neq \mu_0$, is known as two tailed test and in such case the critical region is given by the portion of the area lying in both tails of the probability curve of the test statistic.

Critical values or significant values

- The value of the test statistic which separates the critical region and the acceptance region is called the critical value or significant value. It depends upon:

(i) The level of significance used and

(ii) The alternative hypothesis, whether is two tailed or one tailed.

Critical Value α	Level of significance		
	1%	5%	10%
Two tailed test	$\lvert z_\alpha \rvert = 2.58$	$\lvert z_\alpha \rvert = 1.96$	$\lvert z_\alpha \rvert = 1.645$
Right tailed test	$z_\alpha = 2.33$	$z_\alpha = 1.645$	$z_\alpha = 1.28$
Left tailed test	$z_\alpha = -2.33$	$z_\alpha = -1.645$	$z_\alpha = -1.28$

Procedure for testing of Hypothesis

- Set up Null and alternative hypothesis
- Choose the appropriate level of significance
- Compute test statistic: $z = \frac{\left(t - E(t)\right)}{S.E.(t)}$, under H_0
- Conclusion: we compare the computed value of z with the significant value at the given level of significance.

Problems

Q: A sample of 900 members has a mean 3.4. cms. And standard deviation s.d. 2.61 cms. Is the sample from a large population of mean 3.25 cms. and s.d. 2.61.cms. If the population is normal and its mean is unknown, find the 95% and 98% fiducial limits of true mean.

Q: An insurance agent has claimed that the average age of policy holders who insure through him is less than the average for all agents, which is 30.5 years. A random sample of 100 policy holders who had insured through him gave the following age distribution. Use that data to test his claim at 5% level of significance.

Age last birthday	No. of persons			
16-20	12			
21-25	22			
26-30	20			
31-35	30			
36-40	16			

Q: The mean muscular endurance score of a random sample of 60 subjects was found to be 145 with s.d. of 40. Construct a 95% confidence interval for the true mean. Assume the sample size to be large enough for normal approximation. What size of sample is required to estimate the mean within 5 of the true mean with 95% confidence.

Test of significance for difference of means

Q: The means of two single large samples of 1000 and 2000 members are 67.5 inches and 68 inches respectively. Can thee sample be regarded as drawn from the same population of standard deviation 2.5 inches?

Q: The average hourly wage of a sample of 150 workers in a plant A was Rs. 2.56 with a standard deviation of Rs. 1.08. The average hourly wage of a sample of 200 workers in plant B was Rs. 2.87 with a standard deviation of Rs. 1.28. Can an applicant safely assume that th hourly wages paid by plant B are higher than those paid by plant A?

Q: In a certain factory there are two independent processes manufacturing the same item. The average weight in a sample of 250 items produced from one process is found to be 120 ozs. With a standard deviation of 12 ozs. While the corresponding figures in a sample of 400 items from the other process are 124 and 14. Obtain the standard error of difference between the sample means. Is this difference significant? Also find the 99% confidence limits for the difference in the average weights of items produced by the two processes respectively.

Small sample test (t-test)

- Suppose we want to test:

(i) If a random sample $x_i, i = 1,2,\ldots,n$ of size n has been drawn from a normal population with a specified mean (say) μ_0.

(ii) If the sample mean differ significantly from the hypothetical value μ_0 of the population mean.

The statistic: $t = \frac{\bar{x}-\mu_0}{S/\sqrt{n}}$, where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ follow Student's t distribution with $(n-1)$ degree of freedom.

Assumptions for student’s t test

- The parent population from which the sample is drawn is normal.
- The sample observations are independent i.e. sample is random.
- The population standard deviation is unknown.

Q: The mean weekly sales of soap bars in a departmental stores was 146.3 bars per store. After an advertising campaign the mean weekly sales in 22 stores for a typical week increased to 153.7 and showed a standard deviation of 17.2. Was the advertising campaign successful?

Q: A random sample of 10 boys had the following I.Q.'s: 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these data support the assumption of a population mean I.Q. of 100? Find a reasonable range in which most of mean I.Q. values of samples of 10 boys lie.

T-test for difference of means

14·2·10. t-Test for Difference of Means. Suppose we want to test if two independent samples x_i ($i = 1, 2, \ldots, n_1$) and y_j, ($j = 1, 2, \ldots, n_2$) of sizes n_1 and n_2 have been drawn from two normal populations with means μ_X and μ_Y respectively.

Under the null hypothesis (H_0) *that the samples have been drawn from the normal populations with means μ_X and μ_Y and under the assumption that the population variance are equal, i.e.,* $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ (say), the statistic

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \ldots(14{\cdot}7)$$

where $\bar{x} = \frac{1}{n_1}\sum_{i=1}^{n_1} x_i, \qquad \bar{y} = \frac{1}{n_2}\sum_{j=1}^{n_2} y_j$

and $S^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_i (x_i - \bar{x})^2 + \sum_j (y_j - \bar{y})^2\right]$ $\qquad \ldots[14{\cdot}7(a)]$

Q:Samples of two electric light bulbs were tested for length of life and following data were obtained:

	Type I	**Type II**
Sample size	$n_1 = 8$	$n_2 = 7$
Sample means	$\bar{x}_1 = 1{,}234$ hrs.	$\bar{x}_2 = 1{,}036$ hrs.
Sample s.d.	$s_1 = 36$ hrs.	$s_2 = 40$ hrs.

Is the difference in the means sufficient to warrant that type I is superior to type II regarding the length of life?

Q: The height of six randomly chosen sailors are (inches): 63,65,68,69,71 and 72. Those of 10 randomly chosen soldiers are 61,62,65,66,69,69,70,71,72 and 73. Discuss the light that these data throw on the suggestion that sailor are on the average taller than soldiers.

	X	X-68	(X-68)^2
	63	-5	25
	65	-3	9
	68	0	0
	69	1	1
	71	3	9
	72	4	16
sum	408	0	60
mean	68		
	s.d.	3.162278	
	varaince	10	

	Y	Z=Y-68	Z^2=(Y-68)^2
	61	-7	49
	62	-6	36
	65	-3	9
	66	-2	4
	69	1	1
	69	1	1
	70	2	4
	71	3	9
	72	4	16
	73	5	25
sum	678	-2	154
mean	67.8		
	variance	15.36	

Q: To test the claim that the resistance of electric wire can be reduced by at least 0.05 ohm by alloying, 25 values obtained for each alloyed wire and standard wire produced the following results:

	Mean	**Standard deviation**
Alloyed wire	0.083 ohm	0.003 ohm
Standard wire	0.136 ohm	0.002 ohm

Test at 5% level whether or not the claim is substantiated?

Paired t-test

- **Paired t-test For Difference of Means**. Let us now consider the case when (the sample sizes are equal. i.e. $n_1 = n_2 = n$ (say).

(i) the two samples are not independent but the sample observations are paired together. i.e. the pair of observations (x_i, y_i). $(i = 1,2 \dots n)$ corresponds to the same (i^{th})sample unit. The problem is to test if the sample means differ significantly or not.

Q: A certain stimulus administered to each of the 12 patients resulted in the following increase of blood pressure: 5,2,8,-1,3,0,-2,1,5,0,4 and 6. Can it be concluded that the stimulus will, in general be accompanied by an increase in blood pressure?

	d	d^2		
	5	25		
	2	4		
	8	64		
	-1	1		
	3	9		
	0	0	variance	8.743056
	-2	4		
	1	1		
	5	25		
	0	0		
	4	16		
	6	36		
sum	31	185		
mean d	2.583333			

Q: In a certain experiment to compare two types of animal foods A and B, the following results of increase in weights were observed in animals.

		1	2	3	4	5	6	7	8	Total
Increase in weight in lb	Food A	49	53	51	52	47	50	52	53	407
	Food B	52	55	52	53	50	54	54	53	423

Examine the case when the same set of eight animals were used in both the foods.

Can we conclude that Food B is better than Food A

	x	y	d=x-y	d^2
	49	52	-3	9
	53	55	-2	4
	51	52	-1	1
	52	53	-1	1
	47	50	-3	9
	50	54	-4	16
	52	54	-2	4
	53	53	0	0
sum	407	423	-16	44

F-test for the equality of two population

- Suppose that we want to test (i) whether the independent samples x_i $(i = 1,2,3, \ldots ,\ n)$ and y_j $(j = 1,2,3, \ldots m)$ have been drawn from the normal population with the same variance σ^2 or (ii) whether the two independent estimates of the population variance are homogeneous or not.
- Test statistic F is given by $F = \frac{S_X^2}{S_Y^2}$, where $S_X^2 > S_Y^2$

Q: Pumpkins were grown under two experimental conditions. Two random samples of 11 and 9 pumpkins show the sample standard deviation of their weights are 0.8 and 0.5 respectively. Assuming that the weight distribution are normal, test the hypothesis that the true variances are equal against that they are not, at 10% level. [given $P\left(F_{10,8} \geq 3.35\right) = 0.05$ and $P\left(F_{8,10} \geq 3.07\right) = 0.05$]

Q: Two random samples gave the following results:

Sample	Size	Sample mean	Sum of square of deviations about mean
1	10	15	90
2	12	14	108

Test whether the samples come from the same normal population at 5% level of significance.

Chi- square (χ^2) test

- **Chi-square Test of Goodness of Fit.** A very powerful test for testing the significance of the discrepancy between theory, and experiment was given by Prof. Karl Pearson in 1900 and is known as "Chi·square test of goodness of fit."
- If O_i and E_i $(i = 1,2,3, \ldots, k)$, be the set of observed and expected frequencies, then

$$\chi^2 = \sum_{i=0}^{k} \frac{[O_i - E_i]^2}{E_i}, \sum_{i=i}^{k} O_i = \sum_{i=i}^{k} E_i$$

Follows chi- square distribution with $(k-1)$ degree of freedom.

Condition for the validity of χ^2test

- The sample observations should be independent.
- Constraints on the cell frequencies, if any, should be linear, e.g.$\sum_{i=i}^{k} O_i = \sum_{i=i}^{k} E_i$
- N, The total frequency should be reasonably large greater than 50.
- No theoretical cell frequency should be less than 5. If any theoretical cell frequency is less than 5, then it is pooled with the proceeding or succeeding frequency so that the pooled frequency is more than 5 and finally adjust the degree of freedom.

Q: The demand for a particular spare part in a factory was found to vary from day to day. In a sample study the following information was obtained:

Days	Mon	Tues	Wed	Thurs	Fri	Sat
No. of parts	1124	1125	1110	1120	1126	1115

Test the hypothesis that the number of parts demanded does not depend on the day of week. Given the value of chi-square significance at 5,6,7 d.f. are respectively 11.07, 12.59, 14.07 at 5% level of significance.

Days	Observed Oi	Expected Ei	(Oi-Ei)^2	(Oi-Ei)^2/Ei
Mon	1124	1120	16	0.01428571
Tue	1125	1120	25	0.02232143
Wed	1110	1120	100	0.08928571
Thu	1120	1120	0	0
Fri	1126	1120	36	0.03214286
Sat	1115	1120	25	0.02232143
sum	**6720**	**6720**	**202**	**0.18035714**

Days	Observed Oi	Expected Ei	(Oi-Ei)^2	(Oi-Ei)^2/Ei
Mon	1124	1120	16	0.01428571
Tue	1125	1120	25	0.02232143
Wed	1110	1120	100	0.08928571
Thu	1120	1120	0	0
Fri	1126	1120	36	0.03214286
Sat	1115	1120	25	0.02232143
sum	**6720**	**6720**	**202**	**0.18035714**

Q: When the first proof of 392 pages of a book of 1200 pages were read, the distribution of printing mistakes were found to as follow:

No. of mistakes in a page x	0	1	2	3	4	5	6
No. of pages f	275	72	30	7	5	2	1

Fit a Poisson distribution to the above data and test the goodness of fit.