Sampling Theory

A large collection of individuals (or) attributes (or) numerical data can be understood as population or universe.

A finite subset of the universe is called sample. The number of individuals in a sample is called a sample size.

If the sample size (n) is less than (or) equal to 30 the sample is said to be small, otherwise it is a large sample.

The process of selecting a sample from the population is called as sampling.

The selection of an individual (or) item from the population in such a way that each has the same chance of being selected is called as random sampling.

In sampling where a member of the population may be selected more than once is called as sampling with replacement.

If a member cannot be chosen more than once is called sampling without replacement.

Sampling Distribution :-

Suppose that we have different sample of size n drawn from a population. For each and every sample of size n we can compute quantities like, mean, standard deviation etc, & these will not be the same. If we group these characteristics according to their frequencies, the frequency distributions so generated are called sampling distribution. The sample distribution of Large samples is assumed to be a normal distribution.

Standard Error :-

The standard ~~dev~~ deviation of a sampling distribution is also called the standard error (S.E).

The reciprocal of the standard error is called precision.

Test of Hypothesis:

The hypothesis formulated for the purpose of its rejection under the assumption that it is true is called the Null Hypothesis. Any hypothesis which is complimentary to the null hypothesis is called Alternative Hypothesis.

Significance level

The probability level below which ~~the~~ leads to the rejection of the hypothesis is known as significance level. This probability is conventionally fixed at 0.05 (or) 0.01 being 5% (or) 1%. These are called significance levels.

Test of significance

The process ~~of~~ which helps us to decide about the acceptance of rejection of the hypothesis is called the test of significance.

Let us suppose that we have a normal population with mean μ & S.D σ. If $\bar{x}$ is the sample mean of a random sample of size n the quantity Z defined by

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

is called the Standard Normal variate (SNV).

Test of significance of large samples:

Test of significance of proportion :-

~~not~~ In probability distribution, we have noticed that the normal distribution is the limiting form of binomial distribution when n is large & neither ~~poor~~ p nor q is small.

According to Binomial distribution, mean is np & S.D is $\sqrt{npq}$

Let us consider the proportion of success:

(i) mean proportion of success $= \frac{np}{n} = p$ (ii) S.D proportion success $= \frac{\sqrt{npq}}{n} = \sqrt{pq/n}$

Let x be the observed no of successes in a sample size n, & $\mu = np$ be the expected no of successes. Then the associated ~~normal~~ standard normal variate Z is defined as $Z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{npq}}$, If $|Z| > 2.58$ we conclude that diff is highly significant & rejected the hypothesis

APPENDIX - C

χ^2_α - Critical Values of the Chi-squared Distribution

	α									
ν	0.30	0.25	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.001
1	1.074	1.323	1.642	2.706	3.841	5.024	5.412	6.635	7.879	10.827
2	2.408	2.773	3.219	4.605	5.991	7.378	7.824	9.210	10.597	13.815
3	3.665	4.108	4.642	6.251	7.815	9.348	9.837	11.345	12.838	16.268
4	4.878	5.385	5.989	7.779	9.488	11.143	11.668	13.277	14.860	18.465
5	6.064	6.626	7.289	9.236	11.07	12.832	13.388	15.086	16.750	20.517
6	7.231	7.841	8.558	10.645	12.592	14.449	15.033	16.812	18.548	22.457
7	8.383	9.037	9.803	12.017	14.067	16.013	16.622	18.475	20.278	24.322
8	9.524	10.219	11.03	13.362	15.507	17.535	18.168	20.090	21.955	26.125
9	10.656	11.389	12.242	14.684	16.919	19.023	19.679	21.666	23.589	27.877
10	11.781	12.549	13.442	15.987	18.307	20.483	21.161	23.209	25.188	29.588
11	12.899	13.701	14.631	17.275	19.675	21.920	22.618	24.725	26.757	31.264
12	14.011	14.845	15.812	18.549	21.026	23.337	24.054	26.217	28.300	32.909
13	15.119	15.984	16.985	19.812	22.362	24.736	25.472	27.688	29.819	34.528
14	16.222	17.117	18.151	21.064	23.685	26.119	26.873	29.141	31.319	36.123
15	17.322	18.245	19.311	22.307	24.996	27.488	28.259	30.578	32.801	37.697
16	18.418	19.369	20.465	23.542	26.296	28.845	29.633	32.000	34.267	39.252
17	19.511	20.489	21.615	24.769	27.587	30.191	30.995	33.409	35.718	40.790
18	20.601	21.605	22.760	25.989	28.869	31.526	32.346	34.805	37.156	42.312
19	21.689	22.718	23.900	27.204	30.144	32.852	33.687	36.191	38.582	43.820
20	22.775	23.828	25.038	28.412	31.410	34.170	35.020	37.566	39.997	45.315
21	23.858	24.935	26.171	29.615	32.671	35.479	36.343	38.932	41.401	46.797
22	24.939	26.039	27.301	30.813	33.924	36.781	37.659	40.289	42.796	48.268
23	26.018	27.141	28.429	32.007	35.172	38.076	38.968	41.638	44.181	49.728
24	27.096	28.241	29.553	33.196	36.415	39.364	40.270	42.980	45.558	51.179
25	28.172	29.339	30.675	34.382	37.652	40.646	41.566	44.314	46.928	52.620
26	29.246	30.434	31.795	35.563	38.885	41.923	42.856	45.642	48.290	54.052
27	30.319	31.528	32.912	36.741	40.113	43.194	44.140	46.963	49.645	55.476
28	31.391	32.620	34.027	37.916	41.337	44.461	45.419	48.278	50.993	56.893
29	32.461	33.711	35.139	39.087	42.557	45.722	46.693	49.588	52.336	58.302
30	33.530	34.800	36.250	40.256	43.773	46.979	47.962	50.892	53.672	59.703

Chi-square distribution :-

Chi-square distribution provides a measure of correspondence b/w the theoritical frequencies and observed frequencies.

If O_i is observed frequency &
E_i is expected (or) estimated frequency.

Then chi-square is denoted by χ^2 is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$, degrees of freedom $= n-1$.
i.e, dof(ν) $= n-1$. | ν = 'nu'

& $\sum f$ = total frequency (or) sum of observed frequency.

Note :

If the calculated value of chisquare (χ) is less than the table value at specified level of significance then the hypothesis is accepted, otherwise rejected.

DOF (ν)	5% level (0.05)	1% level (0.01)
1	3.841	6.634
2	5.991	9.210
3	7.815	11.345
4	9.488	13.287
5	11.07	15.086
6	12.592	16.812
7	14.067	18.475

Note :

1) Formula to find E_i = (total frequency) × (probability of x_i).

2) In the case of Binomial distribution dof $= n-1$.

3) In the case of Poisson distribution dof $= n-2$.

4) Binomial Distribution, $P(x) = {}^{n}C_{x} p^{x} q^{n-x}$

5) Poisson Distribution, $P(x) = \frac{m^{x} e^{-m}}{x!}$

Problems

1) If a coin is tossed 200 times, 118 times heads & 82 times tail were observed. Test the hypothesis that the coin is fair at 5% & 1% level of significance.

Sol:-

x	Head (x_1)	Tail (x_2)
f (or) O_i	118 (O_1)	82 (O_2)

Here observed frequencies are

O_1 (or) $f_1 = 118$ $\quad$ O_2 (or) $f_2 = 82$

$\therefore \Sigma f$ = total frequency $= 118 + 82 = 200$

$E_i = (\Sigma f)\,[P(x)]$

$P(x_1)$ is probability of getting head.

ie $P(x_1) = \frac{1}{2}$

$P(x_2)$ is probability of getting tail

$P(x_2) = 1/2$

$$E_1 = 200 \times P(x_1) = 200\left(\frac{1}{2}\right) = 100$$

$$E_2 = 200 \times P(x_2) = 200\left(1/2\right) = 100$$

$$\therefore \chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

$$= \frac{(118-100)^2}{100} + \frac{(82-100)^2}{100}$$

$$\boxed{\chi^2 = 6.48}$$

Degree of freedom $(\nu) = n - 1$
$= 2 - 1$
$= 1$

Calculated value is 6.48, but table value at 5% is 3.84

ie $6.48 \not< 3.84$ $\therefore$ hypothesis is rejected at 5% level.

But table value at 1% is 6.64.

$\therefore 6.48 < 6.64$

$\therefore$ The hypothesis is accepted at 1% level.

2) A die is thrown 60 times and the frequency distribution for the number appearing on the face x is given by following table

x	1	2	3	4	5	6
Observed frequency (f)	15	6	4	7	11	17
	O_1	O_2	O_3	O_4	O_5	O_6

Test the hypothesis that the die is unbiased, given that
$\chi^2_{0.05}(5) = 11.07$, $\chi^2_{0.01}(5) = 15.09$

Here 5 represents DOF (ν) = n-1 = 6-1 = 5

Soln:- $\Sigma f = 15+6+4+7+11+17$
$= 60$

$E_1 = (\Sigma f)(P(x_1)) = 60 \times \frac{1}{6} = 10$

$E_2 = (\Sigma f)(P(x_2)) = 60 \times \frac{1}{6} = 10$

llly $E_3 = E_4 = E_5 = E_6 = 60 \times \frac{1}{6} = 10$

$P(x_1)$ is probability of getting 1 when total outcome is 6 ie $\frac{1}{6}$.
llly $P(x_2) = \frac{1}{6}$ .... $P(x_6) = \frac{1}{6}$
(because given is single die)

$$\therefore \chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(15-10)^2}{10} + \frac{(6-10)^2}{10} + \frac{(4-10)^2}{10} + \frac{(7-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(17-10)^2}{10}$$

$$\boxed{\chi^2 = 13.6.}$$

Degree of freedom (ν) = n-1
= 6-1
= 5

$\therefore 13.6 \nless 11.07$, $\therefore$ Hypothesis is rejected at 5% level.
$13.6 < 15.09$, $\therefore$ Hypothesis is accepted at 1% level.

3) A set of five coins are tossed 320 times and the result are shown in the table. Test the hypothesis that the data follows a binomial distribution associated with a coin.

No. of heads	0	1	2	3	4	5
frequency	6	27	72	112	71	32

Sol:- $\Sigma f = 6 + 27 + 72 + 112 + 71 + 32$

$\Sigma f = 320$

& $E_i = (\Sigma f) \times P(x_i)$

Here we have to find E_i by using Binomial distribution, ie $P(x) = {}^nC_x p^x q^{n-x}$

$\therefore \quad P(x) = {}^5C_x (1/2)^x (1/2)^{5-x}$

p = Probability of getting head = $\frac{1}{2}$
$q = 1-p = 1-1/2 = 1/2$
n = no. of coins = 5.

$\therefore$ ~~$E_1 (x=0) = {}^5C_0 (1/2)^0$~~

$E_1 (x=0) = 320 \times {}^5C_0 (1/2)^0 (1/2)^{5-0}$

$= 320 \times 0.03125$

$= 10$

$E_2 (x=1) = 320 \times {}^5C_1 (1/2)^1 (1/2)^{5-1} = 320 \times {}^5C_1 (1/2) (1/2)^4$

$= 320 \times 0.15625$

$= 50$

$E_3 (x=2) = 320 \times {}^5C_2 (1/2)^2 (1/2)^3$

$= 320 \times$

$= 100$

$E_4 (x=3) = 320 \times {}^5C_3 (1/2)^3 (1/2)^2$

$= 100$

$E_5 (x=4) = 320 \times {}^5C_4 (1/2)^4 (1/2)^1$

$= 50$

$E_6 (x=5) = 320 \times {}^5C_5 (1/2)^5 (1/2)^0$

$= 10$

$$\therefore \chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2} + \frac{(O_3-E_3)^2}{E_3} + \frac{(O_4-E_4)^2}{E_4} + \frac{(O_5-E_5)^2}{E_5} + \frac{(O_6-E_6)^2}{E_6}$$

$$= \frac{(6-10)^2}{10} + \frac{(27-50)^2}{50} + \frac{(72-100)^2}{100} + \frac{(112-100)^2}{100} + \frac{(71-50)^2}{50} + \frac{(32-10)^2}{10}$$

$$\boxed{\chi^2 = 78.68}$$

Dof (ν) $= n-1$

$= 6-1$

$\nu = 5$

at 5% level :- $78.68 \nless 11.07$, $\therefore$ hypothesis rejected at 5% level.

at 1% level : $78.68 \nless 15.09$, $\therefore$ hypothesis rejected at 1% level.

$\therefore$ Hypothesis is not accepted at both 5% & 1% levels.

4) A dice is thrown 264 times and the number appearing on the face (x) is given in the following table. Calculate the value of χ^2.

x	1	2	3	4	5	6
$f_{(O_i)}$	40	32	28	58	54	60

Soln. $\Sigma f = 40+32+28+58+54+60$

$= 264$

$\therefore E_1 = (\Sigma f)(P(x_1))$

$= 264 \times \frac{1}{6} = 44$

$\therefore E_2 = E_3 = E_4 = E_5 = E_6 = 44.$

$$\therefore \chi^2 = \sum_{i=1}^{n} \frac{(O_i-E_i)^2}{E_i} = \frac{(40-44)^2}{44} + \frac{(32-44)^2}{44} + \frac{(28-44)^2}{44} + \frac{(58-44)^2}{44} + \frac{(54-44)^2}{44} + \frac{(60-44)^2}{44}$$

$$\chi^2 = \frac{968}{44} = 22$$

Scanned with CamScanner

*5) 4 coins are tossed 100 times and the following results were obtained. Fit a binomial distribution for the data & test the goodness fit. ($\chi^2_{0.05} = 9.49$ for 4 d.f).

No. of heads	0	1	2	3	4
Frequency	5	29	36	25	5

Solⁿ:- $\Sigma f = 5+29+36+25+5$
$= 100$

$E_i = (\Sigma f)(P(x_i))$

where $P(x_i) = {}^nC_x p^x q^{n-x}$ (binomial distribution).

where $n = 4$, $p = 1/2$, $q = 1/2$.

$\therefore E(x=0) = (100)\, {}^4C_0 (1/2)^0 (1/2)^{4-0}$
$= 7$

$E(x=1) = (100)\, {}^4C_1 (1/2)^1 (1/2)^{4-1}$
$= 26$

$E(x=2) = 100\, {}^4C_2 (1/2)^2 (1/2)^2$
$= 37.$

$E(x=3) = (100) \times {}^4C_3 (1/2)^3 (1/2)^1$
$= 24$

$E(x=4) = 100 \times {}^4C_4 (1/2)^4 (1/2)^0$
$= 6.$

$\therefore$

O_i	5	29	36	25	5
E_i	7	26	37	24	6

$$\therefore \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(5-7)^2}{7} + \frac{(29-26)^2}{26} + \frac{(36-37)^2}{37} + \frac{(25-24)^2}{24} + \frac{(5-6)^2}{6}$$

$$= \frac{4}{7} + \frac{9}{26} + \frac{1}{37} + \frac{1}{24} + \frac{1}{6}$$

$\boxed{\chi^2 = 1.15}$

at 5% level $\chi^2 = 9.49$ for 4 d.f

$\therefore 1.15 < 9.49$

Thus the hypothesis that the fitness is good can be accepted.

6) A survey of 240 families with 3 children each, the distribution shown in the following table. The data consists with the hypothesis is the data the male & female works are equally probable ?. Use Chisquare at 5% & 1% level of significance.

No. of children (x)	3B 0G	2B 1G	1B 2G	0B 3G
No. of families (f)	37	101	84	18

Sol^n: $\Sigma f = 37 + 101 + 84 + 18$
$= 240.$

$p(x) = {}^nC_x \, p^x q^{n-x}$,

$n = 3$, $x =$ no. of boys, $p = 1/2$, $q = 1/2$.

p -> probability of having boys is 1/2
$q = 1 - p$
$= 1/2$

$E_i = (\Sigma f)(P(x_i))$

$E_1(x=3) = (240) \times {}^3C_3 (1/2)^3 (1/2)^0$
$= 30.$

$E_2(x=2) = (240) \times {}^3C_2 (1/2)^2 (1/2)^1$
$= 90$

$E_3(x=1) = (240) \times {}^3C_1 (1/2)^1 (1/2)^2$
$= 90$

$E_4(x=0) = (240) \times {}^3C_0 (1/2)^0 (1/2)^1$
$= 30$

$$\therefore \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \frac{(O_4 - E_4)^2}{E_4}$$

$$= \frac{(37 - 30)^2}{30} + \frac{(101 - 90)^2}{90} + \frac{(84 - 90)^2}{90} + \frac{(18 - 30)^2}{30}$$

$\chi^2 = 8.177$

& DOF $(\nu) = n - 1 = 4 - 1 = 3.$

at 5% level: $8.17 \nless 7.82$
$\therefore$ Hypothesis rejected at 5% level.

at 1% level: $8.17 < 11.34$
$\therefore$ Hypothesis accepted at 1% level.

7) The following table gives the no. of road accidents that occured in a city during various days of a week. Test the hypothesis that the accidents are uniformly distributed over all the days of a week.

days	Sun	Mon	Tue	wed	Thur	Fri	Sat
No. of accidents	14	16	8	12	11	9	14

Sol: $\Sigma f = 14+16+8+12+11+9+14 = 84.$

$E_i = (\Sigma f)[P(x_i)]$

$E_1 = (84)\, P(x_1)$

$= 84 \times \frac{1}{7} = 12$

$\therefore E_2 = E_3 = E_4 = E_5 = E_6 = E_7 = 12.$

$P(x_1)$ = probability of getting sunday in a week ie $1/7$.

IIIly $P(x_2) = P(x_3) = P(x_4) = P(x_5) = P(x_6) = P(x_7) = 1/7$

$$\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(14-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(8-12)^2}{12} + \frac{(12-12)^2}{12} + \frac{(11-12)^2}{12} + \frac{(9-12)^2}{12} + \frac{(14-12)^2}{12}$$

$$\boxed{\chi^2 = 4.166}$$

DOF $(\nu) = n-1$

$= 7-1$

$= 6$

for DOF 6, 5% & 1% significant level values are 12.59 & 16.81.

$\therefore$ at 5% level :- $4.166 < 12.59$

$\therefore$ Hypothesis accepted.

at 1% level :- $4.166 < 16.81$

$\therefore$ Hypothesis is accepted.

Hence the hypothesis is accepted at both 5% & 1% level.

8) A sample analysis of examination results of 500 students was made. It was found that 220 students had failed, 170 had secured third class, 90 had secured second class & 20 had secured first class. Do these figures support the general examination results which is in the ratio 4:3:2:1 for the respective categories. ($\chi^2_{0.05} = 7.81$, for 3 df).

Sol:

	x	fail	3^{rd} class	2^{nd} class	1^{st} class
	f	220	170	90	20
		O_1	O_2	O_3	O_4

$\Sigma f = 220 + 170 + 90 + 20 = 500$, DOF $= n-1 = 4-1 = 3$.

The expected result in general is given in term of ratio, ie $4:3:2:1$.

when $E_i = (\Sigma f)[P(x_i)]$

∴ the expected frequencies in the respective category are

$E_1 = 500 \times \frac{4}{10} = 200$

$E_2 = 500 \times \frac{3}{10} = 150$

$E_3 = 500 \times \frac{2}{10} = 100$

$E_3 = 500 \times \frac{1}{10} = 50$

$$\therefore \chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(220-200)^2}{200} + \frac{(170-150)^2}{150} + \frac{(90-100)^2}{100} + \frac{(20-50)^2}{50}$$

$$= \frac{400}{200} + \frac{400}{150} + \frac{100}{100} + \frac{900}{50}$$

$$\chi^2 = 23.67 \not< \chi^2_{0.05} = 7.81$$

∴ The hypothesis is rejected.

9) Fit a poisson distribution to the following data of test the ~~hypo~~ goodness of fit at 5% level of significance. ($\chi^2_{0.05}(3) = 7.815$)

x	0	1	2	3	4
f	122	60	15	2	1
	O_1	O_2	O_3	O_4	O_5

Soln: $\Sigma f = 122 + 60 + 15 + 2 + 1 = 200$

$E_i = (\Sigma f) \cdot P(x)$. Here to find $P(x)$ we have to use poisson's distribution.

ie $P(x) = \dfrac{m^x e^{-m}}{x!}$

$m = \mu = \dfrac{\Sigma fx}{\Sigma f} = \dfrac{0 + 60 + 30 + 6 + 4}{200} = \dfrac{100}{200} = \dfrac{1}{2} = 0.5$

$\boxed{m = 0.5}$

$\therefore E_1 (x=0) = 200 \times \dfrac{(0.5)^0 e^{-0.5}}{0!} = 121.3$

$E_2 (x=1) = 200 \times \dfrac{(0.5)^1 e^{-0.5}}{1!} = 60.64$

$E_3 (x=2) = 200 \times \dfrac{(0.5)^2 e^{-0.5}}{2!} = 15.16$

$E_4 (x=3) = 200 \times \dfrac{(0.5)^3 e^{-0.5}}{3!} = 2.52.$

$E_5 (x=4) = 200 \times \dfrac{(0.5)^4 e^{-0.5}}{4!} = 0.3159.$

$$\chi^2 = \Sigma \frac{(O_i - E_i)^2}{E_i} = \frac{(122 - 121.3)^2}{121.3} + \frac{(60 - 60.64)^2}{60.64} + \frac{(15 - 15.16)^2}{15.16} + \frac{(2 - 2.52)^2}{2.52} + \frac{(1 - 0.3159)^2}{0.3159}$$

$\boxed{\chi^2 = 1.60}$

DOF $(\nu) = n - 2 = 5 - 2 = 3$

at 5% level :- $\chi^2 = 1.60 < 7.815$

$\therefore$ the hypothesis is accepted //