Chi-Square Variate

If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

and $Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2$ is a chi-square variate with

1 d.f. (degree of freedom).

In general, if X_i $(i = 1, 2 - n)$ are n independent normal variates with μ_i mean, and variance σ_i^2,

then $\chi^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2$ is a chi-square variate with n d.f.

P.d.f of χ^2-variate ; when $\chi^2 \sim \chi^2_{(n)} \rightarrow$ d.f.

$$f(x) = \frac{1}{2^{n/2}\,\Gamma n/2}\, e^{-\frac{x}{2}} \cdot x^{(n/2)-1} \quad ; \; 0 \le x < \infty$$

$\rightarrow$ Gamma distribution

Exer: If $X \sim \chi^2_{(n)}$, then $X/2 \sim \gamma(n/2)$.

[Note: If $X \sim \gamma(n)$, then $f(x) = \frac{1}{\Gamma n} e^{-y} \cdot x^{n-1}$.
γ ↓ Gamma]

Put $Y = X/2$.

$$g(y) = f(x) \left| \frac{dx}{dy} \right| = \frac{1}{2^{n/2}\,\Gamma n/2}\, e^{-y} (2y)^{(\frac{n}{2}-1)} \cdot 2$$

$$= \frac{1}{\Gamma n/2}\, e^{-y} \cdot y^{(n/2-1)} \quad ; \; 0 \le y < \infty.$$

Hence $X/2 \sim \gamma(n/2)$

Additive property:

The sum of independent chi-square variates is also a χ^2-variates i.e If $X_i \sim \chi^2_{(n_i)}$, then $\sum_{1}^{k} X_i \sim \chi^2_{(\sum_{i=1}^{k} n_i)}$.

(Proof can be thought using MGF of χ^2 variate i.e $(1-2t)^{-n_i/2}$)

Application: (i) Goodness of fit: (ii) Estimate population variance if $(\sigma^2 = \sigma_0^2)$.

If O_i is a set of observed frequencies and E_i is the corresponding set of expected frequencies.

then

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

[Condition $\Sigma O = \Sigma E$) (for large value n)

Eg! The following figures show the distribution of digits in number chosen at random from a telephone directory.

Digit:	0	1	2	3	4	5	6	7	8	9	Total.
Freq:	1026	1107	997	966	1075	933	1107	972	964	853	10,000

Test whether the digits may be taken to occur equally frequently in the directory

Sol: Here: H_0: digits occur equally f...

i.e. each digit 0, 1, 2, ... 9 has expected freq: $\frac{10000}{10} = 1000$.

Calculations for χ^2

Digit	OF (Observed)	EF (Expected)	$(O-E)^2$	$(O-E)^2/E$
0	1026	1000	676	0.676
—	—	—	—	—
9	853	1000	21609	21.609

Sum =

$\therefore \chi^2 = \sum \frac{(O-E)^2}{E} = 58.54$

Degree of freedom $= 10 - 1 = 9$. [10 frequencies subject to condition $\sum O = \sum E$.]

Tabulated value of $\chi^2_{0.05}$ for 9 d.f $= 16.919$.

χ^2 (Calculated) is must greater than tabulated value.

$\Rightarrow$ we reject the Null hypothesis.

Eg: The following table gives the number of aircraft accidents that occur during the various days of the week. Find whether the accidents are uniformly distributed over week.

Sun	M	Tues	Wed	Thurs	Fri	Sat
14	16	8	12	11	9	14

Under H_0 : $E_i = 12 \ \forall i$.

$$\therefore \chi^2 = \frac{(14-12)^2}{12} + \frac{(16-12)^2}{12} + \dots \frac{(14-12)^2}{12}$$

$$= 4.17.$$

$d.f = 7-1 = 6.$

Tabulated value : = 12.59 (will be given in the question)

Calculated χ^2 < tabulated χ^2

H_0 is accepted.

Eg: A survey of 320 families with 5 children each revealed the distribution:

No of boys	5	4	3	2	1	0
" " girls	0	1	2	3	4	5
" " Families	14	56	110	88	40	12

Is male and female births are equally probable?

Sol: H_0: equally probable

Let p be the prob of male birth

Then $p = 1/2$.

$p(r)$ = prob of 'r' male in a family of 5 Children

$$= {}^5C_r\, p^r q^{5-r} = \binom{5}{r}\left(\tfrac{1}{2}\right)^5$$

$$f(r) = N \cdot p(r) = 320 \times \binom{5}{r}\left(\tfrac{1}{2}\right)^5$$

$$= 10 \times \binom{5}{r}.$$

12

6.

ie $f(0) = 10$ $\quad f(3) = 100$

$f(1) = 50$ $\quad f(4) = 50$

$f(2) = 100$ $\quad f(5) = 10.$

OF	EF	$(O-E)^2$	$(O-E)^2/E$
14	10	16	1.6
56	50	36	0.72
110	100	100	1.00
88	100	144	1.44
40	50	100	2.0
12	10	4	0.40
			7.16.

$\chi^2 = 7.16.$ (Calculated)

tabulated $\chi^2_{0.05}$ for $6-1=5$ (d.f) is 11.07

Ho is accepted.

Exer: Fit a Poisson distribution to the following data and test goodness of fit.

X	0	1	2	3	4	5	6
f:	275	72	30	7	5	2	1.

Sol: $\bar{x} = \frac{\sum f_i x_i}{N} = 0.482$ Under H_0. (choose $\lambda = 0.4$

$\therefore f(r) = N p(r) = 392 \times \frac{e^{-0.482} (0.482)^r}{\lfloor r}$.

OF	275	72	30	7	5	2	1
EF	242.1	116.7	28.1	4.5	0.5	0.1	0.

$\overbrace{7 \quad 5 \quad 2 \quad 1}^{15}$

$\underbrace{4.5 \quad 0.5 \quad 0.1 \quad 0.}_{5.1}$

$\chi^2 = \sum \frac{(O-E)^2}{E} = 40.937.$

grouping

$d.f = 7-1-1-3 = 2.$

$\lambda = \bar{x}$ (choosen)

$\sum O = \sum E$

$\chi^2_{0.05}$ for 2 d.f $= 5.99.$

Hence Poisson distribution is not a good fit to the given data.

x ——— x

Q: If is believed that the precision of an instrument is no more than 0.16. write down the null and alternative hypothesis for testing this belief. Carry out the test at 1% Level given 11 measurements of the same subject on the instruments 12
2.5, 2.3, 2.4, 2.3, 2.5, 2.7, 2.5, 2.6, 2.6, 2.7, 2.5. 6.

Sol:

$H_0: \sigma^2 = 0.16.$

$H_1: \sigma^2 > 0.16.$

From given data: $\bar{x} = \frac{27.6}{11} = 2.51$

$\Sigma(x-\bar{x})^2 = 0.1891$

The test statistic is:

$$\chi^2 = \frac{ns^2}{\sigma^2} = \frac{\Sigma(x-\bar{x})^2}{\sigma^2} = \frac{0.1891}{0.16}$$

$$= 1.182$$

$\chi^2_{1\%}$ with 10 d.f $= 23.2$

Ho may be accepted.

$H_0: \sigma^2 = \sigma_0^2$

the statistic:

$$\chi^2 = \frac{\Sigma(x_i-\bar{x})^2}{\sigma_0^2} = \frac{1}{\sigma_0^2}\left(\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}\right)$$

$$= \frac{ns^2}{\sigma_0^2}.$$

follow Chi-square distribution with (n-1) d.f.

Ho rejected 10.5%.

Test of Significance for Difference of Means.

Let $\bar{x}_1$ be the mean of a random sample of size n_1 from a population with mean μ_1 and variance σ_1^2 and let $\bar{x}_2$ be the mean of an independent random sample of size n_2 from another population with mean μ_2 and variance σ_2^2. 12 6.

If sample sizes are large:

$$\bar{x}_1 \sim N(\mu_1, \sigma_1^2/n_1) \text{ and}$$
$$\bar{x}_2 \sim N(\mu_2, \sigma_2^2/n_2)$$

$\bar{x}_1 - \bar{x}_2$ is also a normal variate. The corresponding

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - E(\bar{x}_1 - \bar{x}_2)}{S.D \text{ of } (\bar{x}_1 - \bar{x}_2)} \sim N(0,1)$$

Under the hypothesis; H_0: No significant difference b/w ~~sample~~ mean

i.e $\mu_1 = \mu_2$.

$$\therefore \quad E(\bar{x}_1 - \bar{x}_2) = E(\bar{x}_1) - E(\bar{x}_2) = \mu_1 - \mu_2 = 0.$$
$$V(\bar{x}_1 - \bar{x}_2) = V(\bar{x}_1) + V(\bar{x}_2) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}.$$

$\therefore$ The test statistic becomes:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1).$$

Remark: If $\sigma_1^2 = \sigma_2^2 = \sigma^2$ $\therefore$ $Z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{1/n_1 + 1/n_2}}$

H_0 rejected 10.59.

Eg: The means of two single large samples 1000 and 2000 members are 67.5 inches and 68.0 inches resp. Can the samples be regarded as drawn from the same population of s.d 2.5 inches?. (5% Level of significance).

Sol: $n_1 = 1000$ $n_2 = 2000$

$\bar{x}_1 = 67.5$ $\bar{x}_2 = 68.0$.

H_0: $\mu_1 = \mu_2$ and $\sigma = 2.5$

H_1: $\mu_1 \neq \mu_2$ (two tailed)

Test statistic:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma^2 (1/n_1 + 1/n_2)}} = \frac{67.5 - 68}{2.5\sqrt{1/1000 + 1/2000}} = -5.1$$

Since $|Z| > 3$ ∴ difference is highly significant.

H_0 is rejected.

i.e. Samples are not from a same population

Exer: 150 workers in plant A → $2.56 = \bar{x}_1$; $s_1 = 1.08$
200 " " " B → $\bar{x}_2 = 2.87$; $s_2 = 1.28$

Can you conclude wages of plant B worker are higher?

$n_1 = 150$, $n_2 = 200$.

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 < \mu_2$ (single tailed)

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = -2.46$$

at 5%

$|Z| \leq 1.645$

H_0 rejected

⇒ B has higher salary.