19-04-2022

Tuesday, April 19, 2022 10:20 AM

23-04-2021

Friday, April 23, 2021 11:47 AM

Unit -5 (Hypothesis Testing)

* Procedure of Z-test:

1. Set up null hypothesis H_0.
2. Set up alternative hypothesis H_1 or H_A.
3. Level of significance α.
4. Calculation of test-statistic $Z = \frac{t - E(t)}{S.E(t)}$
5. Comparison of Z and Z_α.

Z_α

$Z \to Z_\alpha$

* Formula for Standard error:-

$$S.E(t) = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{pq}{n}}$$

$$E(t) = P$$

p – prob. of accepting null hyp.

q – prob. of failure

n – sample size.

Pb:- In a sample of 1000 people in Maharashtra, 540 people like rice and the rest like wheat. Can we assume that rice & wheat eaters are equally popular in this state at 1% level of significance.

$\alpha \to 0$

$\alpha = 1\%$

$\mu \neq \mu_0$

$\mu > \mu_0$

$\mu < \mu_0$

Sol^n:

H_0 : Rice & wheat eaters are equally popular.

H_1 : Rice & wheat eaters are not equally popular.

$\boxed{\text{2-tailed}} \to Z_\alpha$

p-statistic

Level of significance, $\alpha = 1\% = 0.01$.

i.e., $Z_\alpha = 2.58$

i.e., $Z_\alpha = 2.58$

$P = \mathcal{E}(\bar{z})$

$Z = ?$

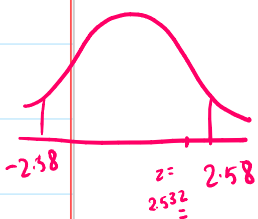
$\mathcal{E}(p) = P = \frac{1}{2}$, $p = \frac{540}{1000} = 0.54$, $q = 1 - 0.54 = 0.46$.

$n = 1000$.

Now, $Z = \frac{\bar{z} - \mathcal{E}(\bar{z})}{S.\mathcal{E}(\bar{z})} = \frac{p - P}{S.\mathcal{E}(p)}$, $S.\mathcal{E} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.54 \times 0.46}{1000}}$

$$= \frac{0.54 - 0.50}{\sqrt{\frac{0.54 \times 0.46}{1000}}}$$

$$= 2.532$$



$Z = 2.532$ & $Z_\alpha = 2.58$, difference is less than 1%.

Since the difference betn Z and Z_α is not significant, so may accept null hypothesis.

i.e., we can assume that rice & wheat eaters are equally popular in Maharashtra.

→ Normal distribution, $n > 30$. $\alpha = 5\%$, $\alpha = 1\%$.

1. Z-test
2. F-test
3. t-test

Chi-square test → Non-parametric

parametric test

Non-parametric test.

§ chi-square distribution:

2-D plane
3-D space
3

$var(X)$
$E(X^2) - \{E(X)\}^2$

$$X, \quad Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X-\mu}{\sigma}.$$

The square of a standard normal variate is known as a chi-square variate with 1 degree of freedom.

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

$$Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2 \text{ is a chi-square variate with 1 d.f.}$$

In general, if X_i ($i = 1$ to n), are n independent normal variates with mean μ_i & variance σ_i^2, then –

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

§ Conditions for the Validity of χ^2-test :- χ^2-test is an approximate test for large values of n. For the validity of χ^2-test of 'goodness of fit' between theory & experiment, the foll^n conditions must be satisfied:

1. The sample observation should be independent.
2. Constraints on the cell frequencies, if any, should be linear.
3. N, the total frequency should be reasonably large, i.e $N > 50$.
4. No theoretical frequency should be less than 5.

× Goodness of fit :- A very powerful test for testing the significance of the discrepancy between theory & experiment was given by Prof. Karl Pearson in the yr 1900 and is known as " chi-square test of goodness of fit".

It enables us to find if the deviation of the experiment from theory is just by chance or is it really due to inadequacy of the theory to fit the obs. data.

If f_i ($i = 1$ to n) is a set of observed (experimental) frequencies and e_i ($i = 1$ to n) is the corresponding set of expected (hypothetical) frequencies, then Karl Pearson's chi-square, given by—

$$\rightarrow \chi^2 = \sum_{i=1}^{n} \frac{(f_i - e_i)^2}{e_i}, \quad \left(\sum_{i=1}^{n} f_i = \sum_{i=1}^{n} e_i \right)$$

follows chi-square distribution $(n-1)$ degrees of frequencies.

$d.f. \longrightarrow$ d.f

$\rightarrow \chi^2$ ( ) |

$$> \chi^2_\alpha(n-1) \downarrow$$

$\alpha\% \longrightarrow$ a.T

* **Decision rule:-** H_0 is accepted if $\chi^2 \leq \chi^2_\alpha(n-1)$

and H_0 is rejected if $\chi^2 > \chi^2_\alpha(n-1)$,

where χ^2 is the calculated value of chi-square &

6
1%
$\chi^2_\alpha(n-1)$ is the tabulated value of chi-square for $(n-1)$ d.f and level of significance α.

$\chi^2_\alpha(n-1)$

Pb :- The demand for a particular spare part in a factory was found to vary from day-to-day. In a sample study, the foll^n info. was obtained:

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

$n-1$

d.f $= 6-1$
$= 5$

Test the hypothesis that the no. of parts demanded does not depend on the day of the week.

(Given: the values of χ^2 significance at 5, 6, 7 d.f. are resp. 11.07, 12.59, 14.07 at the 5% level of significance)

$\chi^2_{5\%}(5) = 11.07$

$\chi^2_{5\%}(6) = 12.59$

$\chi^2_{5\%}(7) = 14.07$

Sol^n: H_0: no. of parts demanded does not depend on the day of the week.

H_1: no. of parts demanded depend on the day of the week.

H_1 : no. of parts demanded depend on the day of the week.

$\alpha = 5\%$

Days	Frequencies		$(f_i - e_i)^2$	$\frac{(f_i - e_i)^2}{e_i}$
	Obs. (f_i)	Exp. (e_i)		
Mon	1124	1120	16	0.014
Tues	1125	1120	25	0.022
Wed	1110	1120	100	0.089
Thurs	1120	1120	0	0
Fri	1126	1120	36	0.032
Sat	1115	1120	25	0.022
Total	6720	6720		0.179

Under the null hypothesis, the exp. frequencies of the spare part demanded on each day of the week

$= \frac{1}{6}(1124 + 1125 + 1110 + 1120 + 1126 + 1115)$

$= 1120$

$\chi^2 = \sum_{i=1}^{6} \frac{(f_i - e_i)^2}{e_i}$

$\Sigma f_i = \Sigma e_i$

$\therefore \chi^2 = \sum_{i=1}^{6} \frac{(f_i - e_i)^2}{e_i} = 0.179$

$\chi^2 < \chi^2_{5\%}(5)$

The tabulated value of $\chi^2_{0.05}(5)$ is 11.07.

Since calculated value is less than the tabulated value, the diffⁿ is not significant and hence we can accept the null hypothesis.

That is, no. of parts demanded does not depend on the day of the week.

Pb A survey of 800 families with four children each revealed the foll^n distribution :-

No. of boys	: 0	1	2	3	4
No. of girls	: 4	3	2	1	0
No. of families	: 32	178	290	236	64

5
$5-1$
$=4$

Is this result consistent with the hypothesis that male & female births are equally probable? $(\chi^2_{5\%}(4) = 9.488)$

Sol^n: H_0 : male & female births are equally probable.

H_1 : " " " " " not equally probable.

Under null hypothesis,

$$p = \text{Probability of male birth} = \frac{1}{2} = q$$

$$p(r) = \text{prob. of } r \text{ male births in a family of } 4 = {}^4C_r p^r q^{4-r}$$

$$= {}^4C_r \left(\tfrac{1}{2}\right)^r \left(\tfrac{1}{2}\right)^{4-r}$$

$$= {}^4C_r \left(\tfrac{1}{2}\right)^4$$

The frequency of r male births is -

$$f(r) = N \cdot p(r) = 800 \times {}^4C_r \left(\tfrac{1}{2}\right)^4 = 50\, {}^4C_r.$$

$f(0) = 50 \cdot {}^4C_0 = 50;$

$f(1) = 50 \cdot {}^4C_1 = 200,$

$f(2) = 50 \cdot {}^4C_2 = 300,$

$f(3) = 50 \cdot {}^4C_3 = 200$

$f(4) = 50 \cdot {}^4C_4 = 50.$

Male births

	ob. (f_i)	exp (e_i)	$(f_i-e_i)^2$	$\frac{(f_i-e_i)^2}{e_i}$
0	32	50	324	6.48
1	178	200	844	2.42
2	290	300	100	0.33
3	236	200	1296	6.48
4	64	50	196	3.92
				19.63

$$\chi^2 = \sum_{i=1}^{5} \frac{(f_i - e_i)^2}{e_i} = 19.63, \qquad \chi^2_{0.05}(4) = 9.488$$

Since $\chi^2 > \chi^2_{0.05}(4)$, so the difference is significant & H_0 is rejected

Thus, male & female births are not equally probable.