

Azmayen Fayek Sabil

ID: 190042122

Assignment (Q12-04)

==

Chi-square test

Example 4.9

We want to test

H_0 : Accidents are uniformly distributed over the week

H_1 : Accidents are not uniformly distributed over the week

We know,

$$\chi^2 = \sum \frac{(O-E)^2}{E} \quad \text{with d.f.} = n-1$$

here, O = observed value

E = expected value

As,

E and \bar{x} (arithmetic mean) are same,

$$E = \frac{\text{total number of accidents}}{\text{total days}} = \frac{14 + 16 + 8 + 20 + 11 + 9 + 14}{7} = \frac{92}{7} = 13.14$$

Now,

Day	O	E	$(O-E)^2$	$(O-E)^2/E$
Sun	14	13.14	0.7396	—
Mon	16	13.14	8.1796	—
Tue	8	13.14	26.4196	—
Wed	20	13.14	47.0596	—
Thu	11	13.14	4.5796	—
Fri	9	13.14	17.8996	—
Sat	14	13.14	0.7396	—

So, we get, $\sum \frac{(O-E)^2}{E} = 7.98$ and $d.f. = n-1 = 7-1 = 6$

The calculated value of χ^2 is $\chi^2_{\text{cal}} = 7.98$ and the table value of χ^2 for d.f. 5 at 5% level of significance is $\chi^2_{\text{tab}} = 12.59$

since $\chi^2_{\text{cal}} < \chi^2_{\text{tab}}$ so the null hypothesis is accepted.

Example 52

We want to test;

H_0 : The populations are homogeneous with respect to ...

We know,

$$\chi^2 = \sum \frac{(O-E)^2}{E}, \text{ with d.f.} = (r-1)(c-1) \text{ and margin total} = 3$$

here,

O = observed frequency

$$E = \frac{\text{row total} \times \text{col total}}{\text{total number of observation}}$$

Age group	Drama	Talkshows	Cinema	Total
Under 30	120	30	50	200
30-40	10	75	15	100
45 and above	10	30	60	100
TOTAL	140	135	125	400

O	E	$(O-E)^2$	$(O-E)^2/E$
120	70	2500	-
10	35	625	-
10	35	-	-
30	67.5	-	-
75	33.75	-	-
30	33.75	-	-
50	62.5	-	-
15	31.25	-	-
60	31.25	-	-

$$\chi^2 = \sum \frac{(O-E)^2}{E} = 180.5 \quad \text{d.f.} = (3-1)(3-1) = 4$$

The calculated value of χ^2 is $\chi^2_{\text{cal}} = 180.5$ and the table value of χ^2 for d.f. 4 at 5% is $\chi^2_{\text{tab}} = 9.4877$

$\chi^2_{\text{cal}} > \chi^2_{\text{tab}}$. So the hypothesis is rejected.

$$ge = 81 + 35 + 33.75 + 62.5 = 212.25$$

$$gt = 10 + 10 + 10 + 10 = 40$$

ge is not gt. So the hypothesis is rejected.

Decision Rule: $gt > ge$

Ex. Example 19.6 :

1. $H_0: M = 98.5$

2. $H_1: M \neq 98.5$

3. $\alpha = 0.05$

4. Reject the null if $T \leq T_{0.05}$

5. Subtracting 98.5 from each value and ranking the differences without regard to their sign,

Observation X	$\frac{X-98.5}{\sqrt{1.8}(1.8)}$	$ X-98.5 $	G	Rank	$\frac{S(X-98.5)}{G} = 55$
97.5	-1	1	0.21	4	
95.2	-3.3	3.3	0.21	12	To solve later
97.3	-1.2	1.2	0.21	6	
96.0	-2.5	2.5	0.21	10	
:	:	:			
97.7	-1	1			
97.1	-1	1			
97.7	-1	1			
97.7	-1	1			
94.9	-3.6	3.6		13	

$$\text{So that } T^- = 4 + 12 + 6 + \dots + 13 = 95$$

$$T^+ = 8 + 2 = 10 \text{ and } T = 10$$

From the table 10, we find that $T_{0.05} = 21$ for $n=19$

$T_{\text{obs}} < T_{0.05}$. Null rejected //

Example 19.7-8

3 stream sample

i) $H_0: M_1 = M_2$

$281 - 20 = 61$

ii) $H_1: M_1 \neq M_2$

$SPL = 33.4 \text{ dB}$

iii) $\alpha = 0.05$

$200 - 2$

iv) reject null hyp if $T \leq T_{0.05}$

$S = \text{relative loudness}$

v)

x	y	$D = (x - y)$	$ D $	rank	$R(+)$	$R(-)$
16	4	12	12	7.5	✓	
12	18	-6	6	5		✓
22	10	12	12	9.5	✓	3.0 (not used)
16	14	2	2	1.5	✓	
1	1	1	1	1		
22	12	10	10	7.5	✓	

$T^- = 8, T^+ = 47$

$T = \min(8, 47) = 8$

critical value

$T_{0.05} = 8$

$$\frac{SPL - 33.4}{S} = 5$$

$T \leq T_{0.05} \rightarrow$ so null must be rejected

$281 - 20 = 61$

at 10pm

Ex) Hypo example 2

Ex) Hypo example 2

Given that,

$$H_0: \mu = 138$$

$$H_1: \mu = 142$$

$$\alpha = 0.05$$

$$\text{standard deviation} = 2$$

Now,

z_0	$\bar{x} - \mu_0$	standard deviation	100	$(y - x) \cdot 3$	P	α
	$\frac{\bar{x} - \mu_0}{\text{standard deviation}}$	8.0	81	21	7	0.1
		8.1	80	20	6	0.1
		8.2	79	19	5	0.1
		8.3	78	18	4	0.1
		8.4	77	17	3	0.1
		8.5	76	16	2	0.1
		8.6	75	15	1	0.1
		8.7	74	14	0.1	0.1
		8.8	73	13	0.01	0.1
		8.9	72	12	0.001	0.1
		9.0	71	11		
		9.1	70	10		
		9.2	69	9		
		9.3	68	8		
		9.4	67	7		
		9.5	66	6		
		9.6	65	5		
		9.7	64	4		
		9.8	63	3		
		9.9	62	2		
		10.0	61	1		
		10.1	60	0		
		10.2	59			
		10.3	58			
		10.4	57			
		10.5	56			
		10.6	55			
		10.7	54			
		10.8	53			
		10.9	52			
		11.0	51			
		11.1	50			
		11.2	49			
		11.3	48			
		11.4	47			
		11.5	46			
		11.6	45			
		11.7	44			
		11.8	43			
		11.9	42			
		12.0	41			
		12.1	40			
		12.2	39			
		12.3	38			
		12.4	37			
		12.5	36			
		12.6	35			
		12.7	34			
		12.8	33			
		12.9	32			
		13.0	31			
		13.1	30			
		13.2	29			
		13.3	28			
		13.4	27			
		13.5	26			
		13.6	25			
		13.7	24			
		13.8	23			
		13.9	22			
		14.0	21			
		14.1	20			
		14.2	19			
		14.3	18			
		14.4	17			
		14.5	16			
		14.6	15			
		14.7	14			
		14.8	13			
		14.9	12			
		15.0	11			
		15.1	10			
		15.2	9			
		15.3	8			
		15.4	7			
		15.5	6			
		15.6	5			
		15.7	4			
		15.8	3			
		15.9	2			
		16.0	1			
		16.1	0			

Critical value for 0.05 significance level is 1.645 (greater than)

Now,

$$\bar{x} = \mu_0 + z_{0.05} \cdot \frac{6}{\sqrt{n}}$$
$$= 138 + 1.645 \cdot 2$$

$$= 141.30$$

The corresponding z value;

$$z = \frac{141.30 - 142}{2}$$

$$= -0.35$$

Accepted H_0 .

Decision rule: $\bar{x} > z_{0.05} \cdot \frac{6}{\sqrt{n}} \geq T$

Ex(Cov problem):

Given that, $f(x, y) = \begin{cases} \frac{6}{5}(x^2 + 2xy) & ; 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & ; \text{otherwise} \end{cases}$

The marginal density function of x is,

$$\begin{aligned} g(x) &= \frac{6}{5} [x^2 y + x y^2] \Big|_0^1 \\ &= \frac{6}{5} (x^2 + x) \\ &= \frac{6}{5} x(x+1) \end{aligned}$$

$$h(y) = \frac{6}{5} \left[\frac{y^3}{3} + x y^2 \right] \Big|_0^1$$

$$= \frac{6}{5} \left[\frac{1}{3} + y \right]$$

$$= \frac{2}{5} (1 + 3y)$$

$$\mu_x = \int x \cdot \frac{6}{5} x(x+1) dx$$

$$= \frac{6}{5} \int (x^3 + x^2) dx = \frac{6}{5} \left[\frac{x^4}{4} + \frac{x^3}{3} \right] \Big|_0^1$$

$$= \frac{6}{5} \left(\frac{1}{4} + \frac{1}{3} \right) = \frac{3+4}{12} = \frac{7}{10}$$

$$\mu_y = \int y \cdot \frac{2}{5} (1 + 3y) dy = \frac{3}{5}$$

$$E[XY] = \frac{6}{5} \int \int xy(n^3y + 2n^2y^2) dy dx$$

$$= \frac{6}{5} \int \int (n^3y + 2n^2y^2) dy dx = \frac{5}{12}$$

$$\text{cov}(X, Y) = \frac{5}{12} - (7 \times 10 \times 3 \times 5)$$

$$= -1500 \text{ (nm)}$$

$$\int [e^{3n} + e^{3n}]^{\frac{3}{2}} = 0.58$$

$$(n+2n)^{\frac{3}{2}} =$$

$$(1+n)n^{\frac{3}{2}} =$$

$$H_0: \mu = 3.65$$

$$G = 0.16 \quad \int [e^{3n} + e^{3n}]^{\frac{3}{2}} = 0.58$$

$$H_1: \mu \neq 3.65$$

$$n = 36$$

$$[e^3 + \frac{1}{e}]^{\frac{3}{2}} =$$

$$\bar{x} = 3.45$$

$$\alpha = 0.05 \rightarrow \text{Z} < -1.96$$

$$Z > 1.96$$

$$Z = \frac{\bar{x} - \mu_0}{G\sqrt{n}} = \frac{3.45 - 3.65}{0.16/\sqrt{36}} = \frac{-20}{-7.5} = 2.67$$

$$H_0 \rightarrow \text{X} = \int [e^3 + \frac{1}{e}]^{\frac{3}{2}} = 0.58 (e^3 + e^{-3})^{\frac{3}{2}} =$$

$$\frac{d}{dt} = \frac{D+8}{21} = \left(\frac{1}{7} + \frac{1}{3}\right)^{\frac{3}{2}} =$$

$$\frac{d}{dt} = 0.5 (e^3 + e^{-3})^{\frac{3}{2}} \cdot 8 = 0.5$$

$$n=40, \bar{x}=2.8, \sigma=0.35, \alpha=0.01, z_{\alpha/2} = 2.57, z > 2.57$$

$$H_0: \mu = 2.4$$

$$H_1: \mu \neq 2.4$$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = 7.22$$

$$\mu_{\bar{x}} = \frac{\bar{x}}{2} \pm z \cdot \frac{\sigma}{\sqrt{n}}$$

$$= 2.8 \pm 2.57 \times \frac{0.15}{\sqrt{40}}$$

$$= (2.26, 3.4)$$

n	y	ny	n^2
10	5	50	100
12	6	72	144
13	7	91	169
12	9	108	144
18	13	234	324
70	90	600	1026

$$\frac{600 - \frac{70 \times 90}{5}}{1026 - \frac{70^2}{5}}$$

0.87

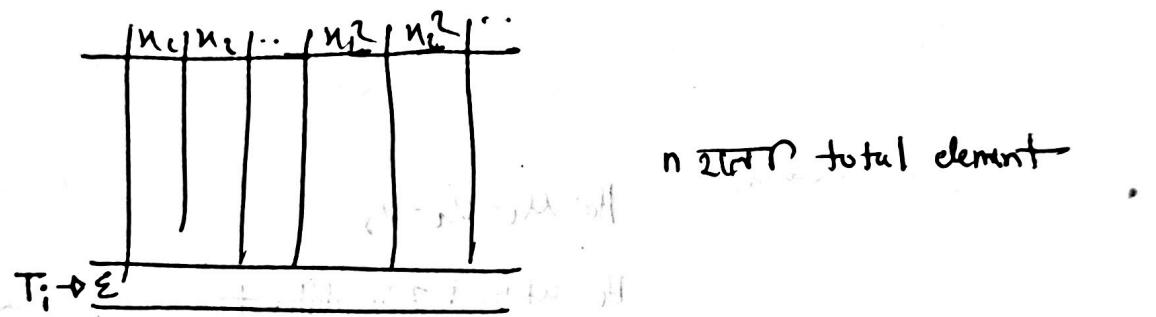
Formulas

ANOVA

(i) hypothesis, H_0, H_1 तथा, & तरीका

(ii) group तरीके n_1, n_2, \dots, n_k लागतहरा and $n_1^2, n_2^2, \dots, n_k^2$ तरीका

with 'E' of each column



(iii) $T = \sum T_i$

$$CT = \frac{T^2}{n}$$

(iv) $SST = \sum T_i^2 - CT$

$$SSB = \left(\sum \frac{T_i^2}{n_i} \right) - CT$$

$$SSW = SST - SSB$$

(v) degree of freedom तरीका, mean squares

$$df_{SSB} = k-1$$

$$MSB = SSB/k-1 = \checkmark$$

$$df_{SSW} = n-k$$

$$MSW = SSW/n-k = \checkmark$$

$$df_{SST} = n-1$$

$$MSB/MSW = F\text{-ratio}$$

compute F-ratio with
F-table

if F-ratio accepted then rejected

✓

Non parametric test ✓

$$\# Z = \frac{S - np}{\sqrt{np(1-p)}} = \frac{S - 0.5n}{0.5\sqrt{n}} \quad \text{here } p=0.5$$

① math section added

n = total number of ...

S = number of plus signs

wilcoxon signed rank test;

x	$x - \text{mean}$	$ x - \text{mean} $	rank
-			
-			
-			

rank: 1, 2, 4, 6, 8, 9
 ① ② ③ ④ ⑤ ⑥
 ⑦ ⑧ 9.5 25
 9, 10, 11, 11

$$6+7 = 13 = 13/2 = 6.5$$

$$9+10 = 19 = 9.5$$

then $T^+ = \sum \text{rank of all positives}$

$T^- = \sum \text{rank of all negatives}$

table P2176 $T_{\text{table}} = 4$ then compare

(T with T_{table})

$$\} T = T^+ \cancel{+ T^-}$$

reject it

$H_1: M \neq M_0 \quad T < T_\alpha$

$H_1: M > M_0 \quad T^- < T_{2\alpha}$

$H_1: M < M_0 \quad T^+ < T_{2\alpha}$

Spearman rank correlation coefficient:

$$\pi_s = 1 - \frac{6 \sum d^2}{n(n^2-1)} \quad d = R_1 - R_2$$

||

Chi-square test ✓

Goodness of fit

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

O = observed value

E = expected value = mean

df = n - 1

test of homogeneity

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

$$df = (r-1)(c-1)$$

O = observed value

E = $\frac{\text{row total} \times \text{col total}}{\text{total no. of observations}}$

Ex hypo

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} \quad \bar{x} = \mu_0 + z \cdot \sigma_{\bar{x}}$$

or

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} \quad \bar{x} = \mu_0 \pm z \frac{\sigma}{\sqrt{n}} \quad (\text{two-tailed})$$

if rejected,
then $\mu_0 = \bar{x} \pm z \frac{\sigma}{\sqrt{n}}$

~~$\mu_0 = 820$~~

~~$\mu_0 =$~~

~~820 ± 80~~

~~$\alpha = 0.05$~~

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \times$$

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

Correlation - correlation

$$\text{cov}(X, Y) = E[XY] - \mu_X \mu_Y$$

Correlation coefficient $\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{V(X)V(Y)}}$

~~Var(x) = E(x^2)~~

$$\mu_X = E[X] = n \int g(x)$$

$$\mu_Y = E[Y] = n \int h(y)$$

$$\text{cov}_n = \frac{\sum_{i=1}^n x_i y_i}{n}$$

$$V[X] = E[X^2] - (E[X])^2$$

$$\sqrt{\left(\sum_{i=1}^n x_i^2 - \frac{(\sum x_i)^2}{n}\right) \cdot \left(\sum_{i=1}^n y_i^2 - \frac{(\sum y_i)^2}{n}\right)}$$

probable error, $P.E(n) = \frac{1 - \pi/2}{\sqrt{n}} \times 0.6745$

\rightarrow ~~approx n error~~
error calculate

$$6 \times P.E(n) < n \rightarrow$$
 good value

Rank correlation;

$$R = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

$$d = (R_1 - R_2)$$

\rightarrow individual rank
~~rank corr 2 rank~~

Regression analysis

$$y = a + bx$$

$$a = \bar{y} - b\bar{x} \quad b = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$\hat{y} - \bar{y} = b(x - \bar{x})$$

$$r = \sqrt{b \times d} \rightarrow b \rightarrow n, d \rightarrow y \text{ regression coefficient}$$

connection coefficient

$$\text{coefficient of determination, } r^2 = \frac{\sum (y - \bar{y})^2}{\sum (y - \hat{y})^2}$$

ANOVA

$$SST = SS_B + SSW$$

$$SS_B, \text{ dof} = k-1 \quad \textcircled{2}$$

$$SS_W, \text{ dof} = n-k \quad \textcircled{3}$$

$$SST, \text{ dof} = n-1 \quad \textcircled{1}$$

$$\text{mean square of } SS_B = \frac{SS_B}{k-1} = MS_B$$

$$\text{ " " " } SSW = \frac{SS_W}{n-k} = MS_W$$

$$F\text{-ratio} = \frac{MS_B}{MS_W}$$