Run Test

Run test is a technique used to determine randomness of sample observations.

It is based on number of run exhibited in the order of the sample observations.

By the term 'run' we mean consecutive occurrences of any characteristic in the observations.

The hypotheses for the run test are usually of the form:

$$H_0: sample\ observations\ are\ random$$

$$H_1: sample\ observations\ are\ not\ random$$

For test procedure, let a sequence of random observations contains n_1 unit of one kind, say, of $X_1$and n_2 number of another kind, say, X_2.

In this context the hypotheses can also be stated as:

$$H_0: probability\ distribution\ of\ X_1\ and\ X_2\ are\ not\ identical$$

$$H_0: probability\ distribution\ of\ X_1\ and\ X_2\ are\ identical$$

Further let U be total number of runs of both kinds observed in the sample.

For $n_1\ and\ n_2 > 20$, it is observed that the number of runs (U) is normally distributed with mean and variance given by

$$mean = E(U) = \frac{2n_1n_2}{n_1 + n_2} + 1$$

Or,

$$E(U) = \frac{2n_1n_2 + n}{n}$$

and

$$V(U) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}$$

Or,

$$V(U) = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n - 1)}$$

So, test statistic is

$$Z_0 = \frac{U - E(U)}{\sqrt{V(U)}}$$

Finally, the null hypothesis is rejected by usual Z-test method.

Solved Problem

#.1

Given observations- (*L stands for Local number plate and O stands for number plates of other states*)

L L O L L L L O O L L L L O L O O L L L L O L O O L L L L L O L L L O L O L L L L O O L

O O O O L L L L O L O O L L L O

Test whether the arrangement of L's and O's are random.

Solution-

Here,

$$H_0: the\ arrangement\ of\ L'sand\ O's\ are\ random.$$

$$H_1: the\ arrangement\ of\ L'sand\ O's\ are\ not\ random.$$

$$\#\ of\ L's\ (n_1) = 38 \qquad \#\ of\ O's\ (n_2) = 22 \qquad total\ (n) = 38 + 22 = 60$$

The number of runs $(U) = 28$ (*It is obtained by number consecutive occurrences of L's and O's*)

Calculation of U is shown below-

L L (1) O (2) L L L L (3) O O (4) L L L L (5) O (6) L (7) O O (8) L L L L (9) O (10) L (11) O O (12) L L L L L (13) O (14) L L L (15) O (16) L (17) O (18) L L L L (19) O O (20) L (21) O O O O (22) L L L L (23) O (24) L (25) O O (26) L L L (27) O (28)

$$E(U) = \frac{2n_1 n_2 + n}{n} = \frac{2 \times 38 \times 22 + 60}{60} = 28.87$$

$$V(U) = \frac{2n_1 n_2(2n_1 n_2 - n)}{n^2(n-1)} = \frac{2 \times 38 \times 22(2 \times 38 \times 22 - 60)}{60^2(60-1)} = \frac{2695264}{212400} = 12.69$$

Now,

$$Z_0 = \frac{U - E(U)}{\sqrt{V(U)}} = \frac{28 - 28.87}{\sqrt{12.69}} = -0.244$$

Here, it is not true that $|Z_0| \geq Z_{0.05}(= 1.64)$, so null hypothesis is not rejected.

#.2

The number of defective pieces observed in 24 shifts are:

15	11	17	14	16	12	19	17	21	15	17	19	21
14	22	16	19	12	16	14	18	17	24	13.		

Test for randomness of observations.

Solution-

Here,

$$H_0: Given\ observations\ are\ random.$$

$$H_1: Given\ observations\ are\ not\ random.$$

Given observations are categorized into two groups according as whether they lie above or below median.

For calculation of median, we create stem-and-leaf plot-

```
1 | 1223444
1 | 5566677778999
2 | 1124
```

Here, $n = 24$.

$$M_d = \left(\frac{n+1}{2}\right)^{th} value = \left(\frac{24+1}{2}\right)^{th} value = 12.5^{th} value$$

$$= 12^{th} value + 0.5 \times \left(13^{th} value - 12^{th} value\right)$$

$$= 16 + 0.5 \times (17 - 16) = 16.5$$

Writing 'a' for values above median and 'b' for values below median, given observations are written as

b	b	a	b	b	b	a	a	a	b	a	a	a
b	a	b	a	b	b	b	a	a	a	b		

Count of 'a' $(n_1) = 12$

Count of 'b' $(n_2) = 12$

Total $(n) = 12 + 12 = 24$

The number of runs of a and b $(U) = 13$

$$E(U) = \frac{2n_1n_2 + n}{n} = \frac{2 \times 12 \times 12 + 24}{24} = 13$$

$$V(U) = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)} = \frac{2 \times 12 \times 12(2 \times 12 \times 12 - 24)}{24^2(24-1)} = 5.74$$

Now,

$$Z_0 = \frac{U - E(U)}{\sqrt{V(U)}} = \frac{13 - 13}{\sqrt{5.74}} = 0$$

Here, it is not true that $|Z_0| \geq Z_{0.05} (= 1.64)$, so null hypothesis is not rejected.

Contingency Table Test

(Also called Chi-square Test of Independence)

It is a statistical test procedure used to test whether two random variables are independent or not.

Let X and Y be two random variables, denoting two characteristics of some population. Chi-square contingency table test is used to observe whether there is any association (or relationship) between X and Y.

The null and alternative hypotheses of the test are of the form:

$$H_0: The\ two\ variables\ are\ independent\ (or, there\ is\ no\ relationship\ between\ two\ variables)$$

$$H_1: The\ two\ variables\ are\ not\ independent\ (Or, there\ is\ association\ between\ the\ two\ variables)$$

For test procedure, a contingency table of different categories of X and different categories of Y is constructed.

A contingency table is a two-way classified frequency table of two jointly distributed random variables. An example of contingency table is shown below-

Sales of different categories of cars of difference brands last year-

Category	**Brands**				
	Hyundai	**Toyota**	**Nissan**	**Maruti**	**Ford**
Excellent	**35**	**24**	**20**	**14**	**34**
Superior	**65**	**23**	**61**	**72**	**23**
Medium	**72**	**65**	**75**	**28**	**33**
Economic	**34**	**34**	**34**	**34**	**29**

Let there be 'r' different categories of variable in different rows and 'c' different categories of variables in different columns of the contingency table.

If $x_1, x_2, \dots \dots \dots \dots, x_r$ are different categories of X and $y_1, y_2, \dots \dots \dots . y_c$ are different categories of Y. Further let $\boldsymbol{O_{ij}}$ denote, observed frequency corresponding to category $\boldsymbol{x_i}$ of X, where i = 1, 2, ……. , r and category $\boldsymbol{y_j}$ of Y, where j = 1, 2, ……. , c, then contingency table looks as follows:

X	Y					
	y_1	y_2	……….	y_j	………	y_c
x_1	O_{11}	O_{12}	……….	O_{1j}``	……….	O_{1c}
x_2	O_{21}	O_{22}	……….	O_{2j}	……….	O_{2c}
…….	…….	…….	……..	…….	……..	…….
…….	…….	…….	……..	…….	……..	…….
x_i	O_{i1}	O_{i2}	……..	O_{ij}	……..	O_{ic}
…….	…….	…….	…….	…….	…….	…….
…….	…….	…….	…….	…….	…….	…….
x_r	O_{r1}	O_{r2}	…….	O_{rj}	…….	O_{rc}

For calculation of test statistic, first of all column totals as well as row totals and grand total are obtained.

Let O_{io} denotes sum of all frequencies in i^{th} row.

O_{oj} denotes sum of all frequencies in j^{th} column.

$O_{..}$ denotes grand sum of all frequencies in entire table.

Now probability that X takes value x_i and Y takes value y_j simultaneously is

$$P(X = x_i \cap Y = y_j)$$

If the null hypothesis $H_0: X \text{ and } Y \text{ are independent}$ is true, then

$$P(X = x_i \cap Y = y_j) = P(X = x_i).P(Y = y_j)$$

But we have,

$$P(X = x_i) = \frac{O_{io}}{O_{oo}}$$

$$P(Y = y_j) = \frac{O_{oj}}{O_{oo}}$$

So,

$$P(X = x_i \cap Y = y_j) = \frac{O_{io}}{O_{oo}} . \frac{O_{oj}}{O_{oo}}$$

Let, $O_{oo} = N$ and $P(X = x_i \cap Y = y_j) = p_{ij}$, then

$$p_{ij} = \frac{O_{io}}{N} . \frac{O_{oj}}{N}$$

If p_{ij} is multiplied by N, then in fact, we get expected frequency corresponding to i^{th} value of X and j^{th} value of Y if the null hypothesis that X and Y are independent is true and it is denoted as E_{ij}.

Thus,

$$Expected\ frequency\ (E_{ij}) = p_{ij} \times N = \frac{O_{io}}{N} . \frac{O_{oj}}{N} \times N = \frac{O_{io} \times O_{oj}}{N}$$

Thus,

$$E_{ij} = \frac{i^{th}\ row\ total \times j^{th}\ column\ total}{grand\ total}$$

Now the values of expected frequencies E_{ij} for entire observations for rows from 1 to r and for columns from 1 to c are calculated.

Then the value of following test statistic is calculated-

$$\sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_i - E_i)^2}{E_i}$$

Above statistic has chi-square distribution with (r-1)(c-1) degrees of freedom, i.e.,

$$\sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{(r-1)(c-1)}$$

Let,

$$\chi_0^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_i - E_i)^2}{E_i}$$

Finally H_0 is rejected at α level of significance, if

$$\chi_0^2 \geq \chi^2_{\alpha,(r-1)(c-1)}$$

Solved Problems

#.1

The distribution of person's blood group according to gender are recorded for 400 patients and is summarized as follows-

Gender	Blood Group			
	O	A	B	AB
Male	100	40	45	10
Female	110	35	55	5

Is there any association between blood group and gender. Test at 5% level of significance?

Solution-

Here:

$$H_0: there\ is\ no\ association\ between\ blood\ group\ and\ gender$$

$$H_1: there\ is\ association\ between\ blood\ group\ and\ gender$$

Calculation of totals-

Gender	Blood Group				Total
	O	A	B	AB	
Male	100	40	45	10	195
Female	110	35	55	5	205
Total	210	75	100	15	400

Here $r = 2, c = 4, row1\ total = 195, row2\ total = 205, column1\ total = 210, column2\ total = 75, column3\ total = 100, column4\ total = 15, N = 400$

Calculation of expected frequencies using formula stated below-

$$E_{ij} = \frac{i^{th} row\ total \times j^{th} column\ total}{grand\ total}$$

Gender	Blood Group				Total
	O	A	B	AB	
Male	$\frac{195 \times 210}{400} = 102.4$	$\frac{195 \times 75}{400} = 36.6$	$\frac{195 \times 100}{400} = 48.8$	$\frac{195 \times 15}{400} = 7.3$	195
Female	$\frac{205 \times 210}{400} = 107.6$	$\frac{205 \times 75}{400} = 38.4$	$\frac{205 \times 100}{400} = 51.2$	$\frac{205 \times 15}{400} = 7.7$	205
Total	210	75	100	15	400

Calculation of test statistic-

	Observed Freq. (O_{ij})	Expected Freq. (E_{ij})	$(O_{ij}-E_{ij})^2$	$(O_{ij}-E_{ij})^2/E_{ij}$
Male,O	100	102.4		0.056
Male,A	40	36.6		0.316
Male,B	45	48.8		0.296
Male,AB	10	7.3		0.999
Female,O	110	107.6		0.054
Female,A	35	38.4		0.301
Female,B	55	51.2		0.282
Female,AB	5	7.7		0.947
Total				3.251

The test statistic is

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_i - E_i)^2}{E_i} = 3.251$$

Critical value is

$$\chi^2_{\alpha,(r-1)(c-1)} = \chi^2_{0.05,(2-1)(4-1)} = 7.832$$

Since it is not true that $\chi_0^2 \geq 7.832$, so null hypothesis is not rejected.

#.2

A random sample of 200 married men, all retired was classified according to level of education and number of children with following result-

Level of Education	Number of Children		
	0-1	2-3	Over 3
Elementary	14	37	32
Secondary	19	42	17
College	12	17	10

Test the hypothesis at 0.05 level of significance that the number of children is independent of level of education.

Solution-

Here:

$$H_0: number\ of\ children\ is\ independent\ of\ level\ of\ education$$

$$H_1: number\ of\ children\ is\ not\ independent\ of\ level\ of\ education$$

Calculation of totals-

Level of Education	Number of Children			Total
	0-1	2-3	Over 3	
Elementary	14	37	32	83
Secondary	19	42	17	78
College	12	17	10	39
Total	45	96	59	200

Calculation expected frequencies-

Level of Education	Number of Children			Total
	0-1	2-3	Over 3	
Elementary	83*45/200=18.675	83*96/200=39.84	83*59/200=24.48	83
Secondary	78*45/200=17.55	78*96/200=37.44	78*59/200=23.01	78
College	39*45/200=8.78	39*96/200=18.72	39*59/200=11.5	39
Total	45	96	59	200

Calculation of test statistic-

	Observed Freq. (O_{ij})	Expected Freq. (E_{ij})	$(O_{ij}-E_{ij})^2$	$(O_{ij}-E_{ij})^2/E_{ij}$
Elem,0-1	14	18.675		1.17
Elem,2-3	37	39.84		0.202
Elem,Over3	32	24.48		2.307
Sec,0-1	19	17.55		0.120
Sec,2-3	42	37.44		0.555
Sec,Over3	17	23.01		1.57
Col,0-1	12	8.775		1.185
Col,2-3	17	18.72		0.158
Col,Over3	10	11.5		0.197
Total				7.464

The test statistic is

$$\chi_0^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_i - E_i)^2}{E_i} = 7.464$$

Critical value is

$$\chi^2_{\alpha,(r-1)(c-1)} = \chi^2_{0.05,(3-1)(3-1)} = 9.488$$

Since it is not true that $\chi_0^2 \geq 9.488$, so null hypothesis is not rejected.

Goodness-of-Fit Test

It is a statistical test procedure used to test whether population from which samples are drawn follow a specified distribution and it is also used to validate a hypothesis.

The null and alternative hypotheses of the test are usually of the form

$H_0: sample\ drawn\ are\ from\ specified\ population\ (with\ known, or\ partially\ known\ or\ unknown\ parameters)$

Or,

$$H_0: sample\ drawn\ validates\ specified\ hypothesis$$

$$H_1: sample\ drawn\ are\ not\ from\ specified\ population$$

Or,

$$H_1: sample\ drawn\ does\ not\ validate\ specified\ hypothesis$$

For the test procedure sample observations are categorized into 'k' number of classes and observed frequency in each class is noted (it is denoted as O_i, meaning observed frequency of i^{th} class)

Then expected frequency of each class is computed as E_i according to specification in null hypothesis.

To test whether sample observations drawn are from specified population, expected frequencies are calculated from the parameters of the population, if known, or if they are partially known or unknown, then sample statistics are used as estimates of parameters.

Then following statistic is considered-

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

which has chi-square distribution with k – p – 1 degrees of freedom, i.e,

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

where 'p' is number of unknown parameters which are estimated from sample observations, and 'k' is number of classes into which sample observations are categorized.

If the null hypothesis is true, let

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Finally the null hypothesis is rejected at α level of significance, if

$$\chi_0^2 \geq \chi_{\alpha,k-p-1}^2$$

Solved Problems

#.1

A die is rolled 60 times and following outcomes were observed:

Side	1	2	3	4	5	6
# of times observed	8	9	13	7	15	8

Is the die fair? Test at 5% level of significance.

Solution-

Here the null and alternative hypothesis are:

$$\boldsymbol{H_0}: \ \boldsymbol{the\ die\ is\ fair}.$$

$$\boldsymbol{H_1}: \ \boldsymbol{the\ die\ is\ not\ fair}.$$

If $\boldsymbol{H_0}$ is true, then probability of getting face x is

$$\boldsymbol{p(x)} = \frac{\mathbf{1}}{\mathbf{6}}$$

Working Table-

Side	Obs. Freq. (O_i)	$p(x)$	Exp. Freq. $(E_i) = N.p(x)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
1	8	1/6	10	4	0.4
2	9	1/6	10	1	0.1
3	13	1/6	10	9	0.9
4	7	1/6	10	9	0.9
5	15	1/6	10	25	2.5
6	8	1/6	10	4	0.4
Total	N=60				5.2

Now test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = 5.2$$

The critical value is $\chi_{\alpha,k-p-1}^2$.

Here, α = 0.05, k = 6, p = 0 (since no parameter of population distribution is estimated from sample observations)

So, $\chi^2_{\alpha,k-p-1} = \chi^2_{0.05,6-0-1} = 11.07$

Since it is not true that $\chi_0^2 \geq 11.07$, so null hypothesis is not rejected.

#.2

In 50 random samples of a manufacturing product the number of samples containing defective items is noted below :

# of defective items (x)	0	1	2	3	4	5
Frequency (f)	4	13	17	12	3	1

At 5% level of significance, can we assert that samples are drawn from a binomially distributed population with p = 0.30?

Solution-

Here the null and alternative hypothesis are:

$$\boldsymbol{H_0 :\ samples\ drawn\ are\ from\ binomially\ distributed\ population\ with\ p = 0.30}$$

$$\boldsymbol{H_1 :\ samples\ drawn\ are\ not\ from\ binomially\ distributed\ population\ with\ p = 0.30}$$

If $\boldsymbol{H_0}$ is true, then $X \sim B(n,p)$ where n = 5 and probability of getting x number of defectives is

$$\boldsymbol{p(x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{5}{x} 0.30^x (1-0.30)^{5-x}}$$

Working Table-

# of defectives	Obs. Freq. (O_i)	$p(x)$	Exp. Freq. $(E_i) = N.p(x)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
0	4	0.1681	8.405		2.309
1	13	0.36.2	18.01		1.394
2	17	0.3087	15.435		0.1587
3	12	0.1323	6.615		4.384
4	3	0.0284	1.42		1.758
5	1	0.0024	0.12		6.453
Total	N=50				16.457

Now test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = 16.457$$

The critical value is $\chi^2_{\alpha,k-p-1}$.

Here, α = 0.05, k = 6, p = 0 (since no parameter of population distribution is estimated from sample observations)

So, $\chi^2_{\alpha,k-p-1} = \chi^2_{0.05,6-0-1} = 11.07$

Since $\chi^2_0 \geq 11.07$, so null hypothesis is rejected.

#.3

4 coins are tossed 150 times and following number of heads (X) are obtained. Test whether binomial distribution is a good fit to following data :

x	0	1	2	3	4
f	28	62	46	10	4

Solution-

Here the null and alternative hypothesis are:

$$\boldsymbol{H_0 : \ binomial\ distribution\ is\ a\ good\ fit\ to\ given\ data}$$

$$\boldsymbol{H_1 : \ binomial\ distribution\ is\ not\ a\ good\ fit\ to\ given\ data}$$

If $\boldsymbol{H_0}$ is true, i.e., if $X \sim B(n,p)$ where n = 4, then probability of getting value x is

$$\boldsymbol{p(x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{4}{x} p^x (1-p)^{4-x}}$$

Since the value of parameter p is not known, so estimate it from given data as follows-

We know that for $B(n,p)$ distribution,

$$mean = np \quad \dots\dots\dots\dots \ (i)$$

Now,

$$mean(\bar{x}) = \frac{1}{N}\sum fx = \frac{0\times 28 + 1\times 62 + 2\times 46 + 3\times 10 + 4\times 4}{28+62+46+10+4} = \frac{200}{150} = \frac{4}{3}$$

From (i)

$$\frac{4}{3} = 4 \times p$$

So,

$$p = \frac{1}{3}$$

So, fitted binomial distribution is $X \sim B(4,\ 1/3)$

Working table for goodness-of-fit test-

x	Obs. Freq. (O_i)	$p(x)$	Exp. Freq. $(E_i) = N.p(x)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
0	28	0.1975	29.63		
1	62	0.3951	59.26		
2	46	0.2963	44.44		
3	10	0.0988	14.81		
4	4	0.0123	1.85		
Total	N=150				0.899

Now test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = 0.899$$

The critical value is $\chi^2_{\alpha,k-p-1}$.

Here, α = 0.05, k = 5, p = 1 (since one parameter of population distribution is estimated from sample observations)

So, $\chi^2_{\alpha,k-p-1} = \chi^2_{0.05,5-1-1} = 7.815$

Since it is not true that $\chi_0^2 \geq 7.815$, so null hypothesis is not rejected.

#.4

The number of accidents per week at a certain junction was checked for 50 weeks and result is shown below:

X (# of accidents)	0	1	2	3 or more
Frequency	22	18	10	0

Assuming that the observations are independent test the hypothesis that no. of accidents follow Poisson distribution.

Solution-

Here the null and alternative hypothesis are:

$$\boldsymbol{H_0 : \ number\ of\ accidents\ follow\ Poisson\ distribution}$$

$$\boldsymbol{H_1 : \ number\ of\ accidents\ do\ not\ follow\ Poisson\ distribution}$$

If $\boldsymbol{H_0}$ is true, then $X \sim P(\lambda)$, where is to be estimated from given observations and also probability of getting x number of accidents is

$$\boldsymbol{p(x) = \frac{e^{-\lambda}\lambda^x}{x!}}$$

Since the value of parameter λ is not known, so estimate it from given data as follows-

We know that for $Poisson(\lambda)$ distribution,

$$mean = \lambda \quad \dots\dots\dots\dots \ (i)$$

Now,

$$mean(\bar{x}) = \frac{1}{N}\sum fx = \frac{0\times 22 + 1\times 18 + 2\times 10 + 3\times 0}{22+18+10+0} = \frac{38}{50} = 0.76$$

From (i)

$$0.76 = \lambda$$

Or,

$$\lambda = 0.76$$

Hence,

$$p(x) = \frac{\boldsymbol{e^{-0.76}(0.76)^x}}{\boldsymbol{x}!}$$

Working Table-

x	Obs. Freq. (O_i)	$p(x)$	Exp. Freq. $(E_i) = N.p(x)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
0	22	0.4676	23.38		0.0818
1	18	0.3554	17.77		0.0029
2	10	0.1350	6.75		1.5611
3	0	0.0342	1.71		1.7107
Total	N=50				3.3567

Now test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = 3.3567$$

The critical value is $\chi^2_{\alpha,k-p-1}$.

Here, α = 0.05, k = 4, p = 1 (since one parameter of population distribution is estimated from sample observations)

So, $\chi^2_{\alpha,k-p-1} = \chi^2_{0.05,4-1-1} = 5.991$

Since it is not true that $\chi_0^2 \geq 5.991$, so null hypothesis is not rejected.

#.5

Genetic theory states that children having one parent of blood group A and other of B will always be one of the three blood groups A, AB, B with proportions of 1 : 2 : 1. A report states that out of 300 children having one parent of blood group A and other of B, 30% were found to be of blood group A, 45% of blood group AB and rest of blood group B. Does the report validate the genetic theory?

Solution:

Here the null and alternative hypothesis are:

$$\boldsymbol{H_0} : \ \boldsymbol{report\ validates\ genetic\ theroy}$$

$$\boldsymbol{H_1} : \ \boldsymbol{report\ does\ not\ validate\ genetic\ theroy}$$

If $\boldsymbol{H_0}$ is true then probability of a children having blood group 'A', 'AB' and 'B' should be (since given proportion is 1 : 2 : 1)

$$\frac{1}{4}, \frac{2}{4}, \frac{1}{4}$$

Working Table-

x	Obs. Freq. (O_i)	$p(x)$	Exp. Freq. $(E_i) = N.p(x)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
A	30% of 300 = 90	¼	75	225	3
AB	45% of 300= 135	2/4	150	225	1.5
B	25% of 300 = 75	¼	75	0	0
Total	N=300				4.5

Now test statistic is

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = 4.5$$

The critical value is $\chi^2_{\alpha,k-p-1}$.

Here, α = 0.05, k = 3, p =0 (since no parameter of population distribution is estimated from sample observations)

So, $\chi^2_{\alpha,k-p-1} = \chi^2_{0.05,3-0-1} = 5.991$

Since it is not true that $\chi_0^2 \geq 5.991$, so null hypothesis is not rejected.