$$= \frac{12}{30(30+1)}\left[\frac{57^2}{10} + \frac{153^2}{10} + \frac{255^2}{10}\right] - 3(30+1)$$

$$= 0.0129 \times 9168.3 - 93$$

$$\therefore \quad W = 25.2711$$

The χ^2_α value at $\alpha = 0.05$ level of significance with degrees of freedom $k - 1 = 3 - 1 = 2$ is 5.991.

Conclusion

Since $W > 5.991$, we reject H_0 and conclude that there is a significant differences in mean among the three samples.

4.5 One Sample Run Test

The run test is used to determine the randomness with which the sample items have been selected. This test can also be used to detect departures in randomness of a sequence of quantitative measurements over time, caused by trends or periodicities.

"A run is a subsequence of one or more identical symbols representing a common property of the data (or) A run is a sequence of identical elements that are preceded and followed by different elements or no element at all."

For example, suppose that 12 people have been selected to constitute a committee, let us denote the male by M and female by F. Arrange the people according to the sex say

$$\underline{MM} \quad \underline{FFF} \quad \underline{M} \quad \underline{FF} \quad \underline{MMMM}$$

$$1 \quad 2 \quad 3 \quad 4 \quad 5$$

Such a grouping are called runs. Here there are total of 5 runs. It seems that some relationship exists between randomness and the number of runs.

4.5.1 Working rule

First combine the observations from both samples and arrange them in ascending order. Now assign the letter A to each observation corresponding to first sample and letter B to second sample. We get a sequence consisting of the symbols A and B. In case of tie, break the tie in such a way that to get maximum number of runs. Now, we set up

Null hypothesis

H_0 : Observations are generated randomly

Alternative hypothesis

H_1 : Observations are not randomly generated (two-tailed test)

Test statistic

Let R = Number of runs and n_1 and n_2 be the number of items in the first and second samples respectively.

When n_1 and n_2 are large ($n_1, n_2 \geq 8$), the number of runs R can be closely approximated by the normal distribution with

$$Z = \frac{R - E(R)}{\sqrt{V(R)}} \sim N(0,1)$$

where $$E(R) = \mu = \frac{2n_1n_2}{n_1+n_2} + 1$$

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

$$\therefore \quad Z = \frac{R-\mu}{\sigma} \sim N(0,1)$$

If $|Z| \leq 1.96$ accept H_0 at 5% level, otherwise H_0 is rejected.

Solved Problem 4.27

In an industrial production line items are inspected periodically for detectives. The following is a sequence of defective items (D) and non-defective items (N) produced by these production line.

$DD\ NNN\ D\ NN\ DD\ NNNNN\ DDD\ NN\ D\ NNNN\ D\ N\ D$

Test whether the defectives are occurring at random or not at 5% level of significance.

Solution

Null hypothesis

H_0 : Defectives occurring at random

Alternative hypothesis

H_1 : Defectives not occurring at random.

Level of significance $\alpha = 5\%$

Test statistic

$$\underline{DD}\ \underline{NNN}\ \underline{D}\ \underline{NN}\ \underline{DD}\ \underline{NNNNN}\ \underline{DDD}\ \underline{NN}\ \underline{D}\ \underline{NNNN}\ \underline{D}\ \underline{N}\ \underline{D}$$

1 2 3 4 5 6 7 8 9 10 11 12 13

$\therefore$ R = number of runs = 13

Here $n_1 = 11, n_2 = 17$

$$\therefore \quad Z = \frac{R-\mu}{\sigma} \sim N(0,1)$$

$$Z = \frac{R - \dfrac{2n_1n_2}{n_1+n_2} + 1}{\sqrt{\dfrac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}}} \sim N(0,1)$$

$$= \frac{13 - \dfrac{2 \times 11 \times 17}{11+17} + 1}{\sqrt{\dfrac{2 \times 11 \times 17(2 \times 11 \times 17 - 11 - 17)}{(11+17)^2(11+17-1)}}}$$

$$Z = \frac{13 - 14.36}{\sqrt{6.11}} = 0.55$$

The value of Z_α at 5% level of significance is 1.96.

Conclusion

Since $|Z| < 1.96$, we accept our null hypothesis and conclude that the defectives are occurring at random.

Solved Problem 4.28

The following are the prices in Rs.1 kg of a commodity from 2 random samples of shops from 2 cities *A* & *B*.

City A	:	2.73	3.82	4.35	3.23	4.74	3.65	3.8	4.15	
		2.76	2.85	3.25	3.45	3.85	4.45	4.95	3.95	4.72
City B	:	3.75	5.37	4.78	3.69	4.75	4.85	6.0	4.8	4.9
		3.84	3.9	4.8	5.23	6.1	3.6	3.83		

Apply the run test to examine whether the distribution of prices of commodity in the two cities is the same.

Solution

Null hypothesis

H_0 : The distribution of prices of commodity in the 2 cities is same.

Alternative hypothesis

H_1 : The distribution of prices of commodity in the 2 cities is not same.

Level of significance $\alpha = 5\%$

Test statistic

Let us combine the observations from both cities and arrange them in ascending order. Now assign the letter A to city A and letter B to city B as follows

1	2.73 2.76 2.85 3.23 3.25 3.45	A A A A A A
2	3.6	B
3	3.65	A
4	3.69 3.75	B B
5	3.8 3.82	A A
6	3.83 3.84	B B
7	3.85	A
8	3.9	B
9	3.95 4.15 4.35 4.45 4.72 4.74	A A A A A A
10	4.75 4.78 4.8 4.8 4.85 4.9	B B B B B B
11	4.95	A
12	5.23 5.27 6.0 6.1	B B B B

$\therefore$ R = number of runs = 12

Here $n_1 = 17$ and $n_2 = 16$

where $$\mu = \frac{2n_1n_2}{n_1+n_2} + 1 = 17.485$$

and $$\sigma = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2.(n_1+n_2-1)}} = \sqrt{7.977} = 2.824$$

$$\therefore \quad |Z| = \frac{12 - 17.485}{2.824} = \frac{5.485}{2.824} = 1.94$$

The value of Z_α at 5% level of significance is 1.96.

Conclusion

Since $|Z| < 1.96$, we accept H_0 and conclude that the distribution of prices of commodity in the 2 cities is the same.