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 proceded 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 F. arrange the people according to the sex say

$$\underset{1}{MM} \quad \underset{2}{FFF} \quad \underset{3}{M} \quad \underset{4}{FF} \quad \underset{5}{MMMM}$$

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.

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.

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 sample respectively.

Where n_1 and n_2 are large and $(n_1, n_2 \geq 8)$, 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 $V(R) = \sigma^2 = \frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_{1+}n_2)^2(n_{1+}n_2-1)}$

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

If $|Z| \le Z_\alpha$ are accept H_0 and reject H_0 when $|Z| > Z_\alpha$, where Z_α is the table value of Z for the level of significance α.

1. In an industrial production line items are inspected periodically for defectives. 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: $H_0: Defectives\ occuring\ at\ random$

Alternative Hypothesis: $H_1: Defectives\ not\ occuring\ at\ random$

Level of significance: $\alpha = 5\%$

Test Statistic:

DD	NNN	D	NN	DD	$NNNNN$	DDD	NN	D	$NNNN$	D	N	D
1	2	3	4	5	6	7	8	9	10	11	12	13

R =Number of runs=13, $n_1 = 11$ and $n_2 = 17$

$$Z = \frac{R-\mu}{\sigma}$$

$$Z = \frac{R - \frac{2n_1n_2}{n_1+n_2} + 1}{\sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_{1+}n_2)^2(n_{1+}n_2-1)}}}$$

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

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

At $\alpha = 0.05,\ \ Z_\alpha = 1.96$

$|Z| \le Z_\alpha$ We accept H_0.

2. The following are the prices in Rs. 1 kg of a commodity from 2 random samples of shop from 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 two cities is same.

Alternative Hypothesis: H_1: the distribution of prices of commodity in the two cities is not same.

Level of significance: $\alpha = 5\%$

Test Statistic:

Arrange the observations in ascending order

		1				2	3
2.73	2.76	2.85	3.23	3.25	3.45	3.6	3.65
A	A	A	A	A	A	B	A

4		5		6		7	8
3.69	3.75	3.8	3.82	3.83	3.84	3.85	3.9
B	B	A	A	B	B	A	B

		9						10			
3.95	4.15	4.35	4.45	4.72	4.74	4.75	4.78	4.8	4.8	4.85	4.9
A	A	A	A	A	A	B	B	B	B	B	B

11		12		
4.95	5.23	5.27	6	6.1
A	B	B	B	B

R =Number of runs=12, $n_1 = 17$ and $n_2 = 16$

$$Z = \frac{R-\mu}{\sigma}$$

$$Z = \frac{R - \frac{2n_1n_2}{n_1+n_2} + 1}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_{1+}n_2)^2(n_{1+}n_2 - 1)}}}$$

$$= \frac{12 - \frac{2*17*16}{17+16} + 1}{\sqrt{\frac{2*17*16(2*17*16-17-16)}{(17+16)^2(17+16-1)}}}$$

$$= \frac{12 - 17.485}{\sqrt{7.977}} = 1.94$$

At $\alpha = 0.05,\ \ Z_\alpha = 1.96$

$|Z| \leq Z_\alpha$ We accept H_0.

3. The production manager of a large undertaking randomly paid 10 visits to the work site in a month. The number of workers who reported late for duty were found to be 2,4,5,1,6,3,2,1,7 and 8 respectively. Use the run test for randomness at $\alpha = 0.05$ to check the claim of the production superintendent that on an average not more than 3 workers report late for duty.

Solution:

Here, A= the average above 3

B= the average below 3

Given, $\begin{matrix} 2 & 4 & 5 & 1 & 6 & 3 & 2 & 1 & 7 & 8 \\ B & A & A & B & A & - & B & B & A & A \end{matrix}$

The above sequence can be written as

B	AA	B	A	BB	AA
1	2	3	4	5	6

R =Number of runs=6, $n_1 = 5$ and $n_2 = 4$

$$Z = \frac{R-\mu}{\sigma}$$

$$Z = \frac{R - \frac{2n_1n_2}{n_1+n_2} + 1}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_{1+}n_2)^2(n_{1+}n_2 - 1)}}}$$

$$= \frac{6 - \frac{2*5*4}{5+4} + 1}{\sqrt{\frac{2*5*4(2*5*4-5-4)}{(5+4)^2(5+4-1)}}}$$

$$= \frac{6-5.444}{\sqrt{1.914}} = 0.4016$$

At $\alpha = 0.05,\ \ Z_\alpha = 1.96$

$|Z| \leq Z_\alpha$ We accept H_0.

HOME WORK

1. A technician is asked to analyze the results of 22 items made in preparation run. Each item has been measured and compared to engineering specifications. The order of acceptance $'a'$ and rejections $'r'$ is $aarrrarraaaaarrarrraara$

   Determine whether it is a random sample or not. Use $\alpha = 0.05$

2. In 30 tosses of a coin, the following sequence of head (H) and tails (T) is obtained $HTTHTHHHTHHTTHTHTHHTHTTHTHHTHT$

(a)Determine the number of runs.

(b)Test at 0.10 level of significance, whether the sequence is random.

3. After a television debate between two political candidates, a telephone line is open to viewers wishing to express their opinions on which the democratic (D) or the republican (R) candidate won the debate. The following sequence represents 19 opinions of viewers in the order in which they telephoned. Using a run test and a significance level of 5% does the sequence indicate a non random order? $RRDDRDDRRRRDDRDRDDD$

Kologorov- Smirnov Test

The K-S test is, another measure of the goodness of fit of a frequency distribution, as was the chi-square test. The advantages of this test are

1. It is a more powerful test

2. It is easier to use, since it does not require that the data be grouped in way.

The test statistics for K.S test is the maximum absolute deviation of expected relative frequency F_e and the observed relative frequency F_o. It is denoted by D_n.

$$D_n = max|F_e - F_o|$$

1. Apply the K-S test to check that the observed frequencies match with the expected frequencies which are obtained from Normal distribution. (Given at n=5, $D_{0.01}$= 0.510).

Test Score	51-60	61-70	71-80	81-90	91-100
Observed Frequency	30	100	440	500	130
Expected Frequency	40	170	500	390	100

Solution:

Null Hypothesis: H_0: the distribution follows a normal distribution.
Alternative Hypothesis:
H_1: the distribution does not follows a normal distribution.
Level of Significance : $\alpha = 0.10$

Observed Frequency	Observed cumulative frequency	Observed relative frequency (F_o)	Expected frequency	Expected cumulative frequency	Expected relative frequency (F_e)	$D = \|F_e - F_o\|$
30	30	$\frac{30}{1200} = 0.025$	40	40	$\frac{40}{1200} = 0.033$	0.008
100	130	$\frac{130}{1200} = 0.108$	170	210	$\frac{210}{1200} = 0.175$	0.067
440	570	$\frac{570}{1200} = 0.475$	500	710	$\frac{710}{1200} = 0.592$	0.117

500	1070	$\frac{1070}{1200} = 0.891$	390	1100	$\frac{1100}{1200} = 0.920$	0.029
130	1200	$\frac{1200}{1200} = 1$	100	1200	$\frac{1200}{1200} = 1$	0

Test Statistic: $D_n = |F_e - F_o|$
K.S Statistic

$$D_n = max|F_e - F_o| = 0.117$$

Conclusion

The tabulated value of $D_n for\ n = 5$ and $\alpha = 0.10$is 0.510
Calculated value$<$ Table value
We accept null Hypothesis
i.e.) the distribution follows a normal distribution

2. Kevin morgan, national sales manager of an electronics firm, has collected the following salary statistics on his field sales force earnings. He has both observed frequencies and expected frequencies if the distribution of salaries is normal. At the 0.05 level of significance, can Kevin conclude that the distribution of sales force earnings is normal?

Earnings in thousands	25-30	31-36	37-42	43-48	49-54	55-60	61-66
Observed Frequency	9	22	25	30	21	12	6
Expected Frequency	6	17	32	35	18	13	4

Solution:

Null Hypothesis: H_0: the distribution of sales force earnings is normal.
Alternative Hypothesis:
H_1: the distribution of sales force earnings is not normal.
Level of Significance : $\alpha = 0.05$

Observed Frequency	Observed cumulative frequency	Observed relative frequency (F_o)	Expected frequency	Expected cumulative frequency	Expected relative frequency (F_e)	$D = \|F_e - F_o\|$
9	9	$\frac{9}{125} = 0.072$	6	6	$\frac{6}{125} = 0.048$	0.024

22	31	$\frac{31}{125} = 0.248$	17	23	$\frac{23}{125} = 0.184$	0.064
25	56	$\frac{56}{125} = 0.448$	32	55	$\frac{55}{125} = 0.440$	0.008
30	86	$\frac{86}{125} = 0.688$	35	90	$\frac{90}{125} = 0.720$	0.032
21	107	$\frac{107}{125} = 0.856$	18	108	$\frac{108}{125} = 0.864$	0.008
12	119	$\frac{119}{125} = 0.952$	13	121	$\frac{121}{125} = 0.968$	0.076
6	125	$\frac{125}{125} = 1$	4	125	$\frac{125}{125} = 1$	0

Test Statistic: $D_n = |F_e - F_o|$
K.S Statistic

$$D_n = max|F_e - F_o| = 0.064$$

Conclusion

The tabulated value of $D_n for\ n = 7$ and $\alpha = 0.05$is 0.486

Calculated value< Table value

We accept null Hypothesis

i.e.) the distribution of sales force earnings is normal.

Sign Test

Sign test is conducted under the following circumstances.

1. When there are pair of observations on two things being compared
2. For any given pair, each of the two observations is made under similar conditions
3. No assumptions are made regarding the parent population.

Sign test for paired data

Let $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$ are a sample of paired observations on two random variables x and y. Let $d_i = x_i - y_i; i = 1,2, \ldots n$. Measurements are such that the deviations d_i can be expressed in terms of positives or negative signs and also $d_i's$ are independent. We take the two hypothesis as:

Null Hypothesis: $H_0: p = 0.5$

Alternative Hypothesis: $H_1: p \neq 0.5$

the standard error of the proportion p is given by $\sigma_p = \sqrt{\frac{pq}{n}}$

The two limits of acceptance region at 5% level of significance are

$$p + 1.96\sigma_p \ \ and\ p - 1.96\sigma_p$$

It is important to note that if both np and nq are not greater than 5, then we must use the binomial distribution instead of normal distribution test.

Working Rule

1. By omitting the zero difference, find the Number of positive deviation in $d_i = x_i - y_i$. Let it be k
   When $n \leq 30$
2. Find $p' = P(u \leq k) = \left(\frac{1}{2}\right)^n \sum_{x=0}^{k} \binom{n}{x}$ if $k$is the number of positive deviations

$p' = P(u \geq k) = \left(\frac{1}{2}\right)^n \sum_{x=k}^{n} \binom{n}{x}$ if $k$is the number of negative deviations

3. If $p' \leq 0.05$, reject the null hypothesis at 5% level and accept H_0 if $p' > 0.05$

   When $n > 30$

   Find $Z = \frac{u - \frac{n}{2}}{\sqrt{\frac{n}{4}}} = \frac{u - np}{\sqrt{npq}}$

   Where $E(u) = Mean = np = n * \frac{1}{2} = \frac{n}{2}$ and

   $V(u) = npq = n * \frac{1}{2} * \frac{1}{2} = \frac{n}{4}$

   u the number of negative deviations.

n= number of given items.

If $|Z| \leq 1.96$ we accept H_0 at 5% level of significance, otherwise we reject H_0

If $|Z| \leq 2.58$ we accept H_0 at 1% level of significance, otherwise we reject H_0

1. The following data show the employee's rates of defective work before and after a change in the wage incentive plan. Compare the following two sets of data to see whether the charge lowered the defective units produced. Using the sign test with $\alpha = 0.01$

Before :	8	7	6	9	7	10	8	6	5	8	10	8
After :	6	5	8	6	9	8	10	7	5	6	9	8

Solution

Null Hypothesis: $H_0: p = 0.5$

Alternative Hypothesis: $H_1: p \neq 0.5$

Level of significance: $\alpha = 0.01$

Test statistics

$$- \quad - \quad + \quad - \quad + \quad - \quad + \quad + \quad 0 \quad - \quad -0$$

$$n = 4 + 6 = 10$$

$$k \longrightarrow no. of\ negative\ deviation$$

$$k = 6$$

$$p' = P(u \geq k) = \left(\frac{1}{2}\right)^n \sum_{x=k}^{n} \binom{n}{x} \qquad (Since, np < 5)$$

$$= \left(\frac{1}{2}\right)^{10} \sum_{x=6}^{10} \binom{10}{x}$$

$$= \left(\frac{1}{2}\right)^{10} \left[\binom{10}{6} + \binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right]$$

$$= 0.000976[210 + 120 + 45 + 10 + 1]$$

$$= 0.000976 * 386$$

$$p' = 0.3767$$

Conclusion:

$$p' > 0.05,$$

$$Accept\ H_0$$

2. An automotive engineer is investigating 2 different types of metering devices for an electronic fuel injection system to determine whether they differ in their fuel mileage performance. The system is installed on 12 different cars and a test is run with each metering device on each car. The observed fuel mileage performance data are given in the following table. Use the sign test to determine whether the median fuel mileage performance is the same for both devices using 5% level of significance

Car	:	1	2	3	4	5	6
Device I	:	17.6	19.4	19.5	17.1	15.3	15.9
Device II	:	16.8	20	18.2	16.4	16	15.4

Car	:	7	8	9	10	11	12
Device I	:	16.3	18.4	17.3	19.1	17.8	18.2
Device II	:	16.5	18	16.4	20.1	16.7	17.9

Solution

Null Hypothesis: $H_0 : p = 0.5$

Alternative Hypothesis: $H_1 : p \neq 0.5$

Level of significance: $\alpha = 5\%$

Test statistics

$$- \quad + \quad - \quad - \quad + \quad - \quad + \quad - \quad - \quad + \quad - \quad -$$

$$n = 4 + 8 = 12$$

$$k \longrightarrow no. of\ negative\ deviation$$

$$k = 8$$

$$p' = P(u \geq k) = \left(\frac{1}{2}\right)^n \sum_{x=k}^{n} \binom{n}{x} \qquad (Since, np < 5)$$

$$= \left(\frac{1}{2}\right)^{12} \sum_{x=8}^{12} \binom{12}{x}$$

$$= \left(\frac{1}{2}\right)^{12} \left[\binom{12}{8} + \binom{12}{9} + \binom{12}{10} + \binom{12}{11} + \binom{12}{12}\right]$$

$$= 0.000244[495 + 220 + 66 + 12 + 1]$$

$$= 0.000244 * 794$$

$$p' = 0.1937$$

Conclusion:

$p' > 0.05, Accept\ H_0$.

One Sample sign test

1. The following data represent the number of hours that a rechargeable hedge trimmer operates before a recharge is required. 1.5, 2.2, 0.9, 1.3, 2.0, 1.6, 1.8, 1.5, 2.0, 1.2 and 1.7. Use the sign test to test the hypothesis of the 0.05 level of significance that this particular trimmer operates with a median of 1.8 hours before requiring a recharge.

Solution

Null Hypothesis: $H_0: \mu = 1.8$

Alternative Hypothesis: $H_1: \mu > 1.8$. (one-tailed test)

Level of significance: $\alpha = 0.05$

Test Statistic

Given data is

1.5, 2.2, 0.9, 1.3, 2.0, 1.6, 1.8, 1.5, 2.0, 1.2 and 1.7

$$- \quad + \quad - \quad - \quad + \quad - \quad 0 \quad - \quad + \quad - \quad -$$

$$n = 7 + 3 = 10$$

$$k \longrightarrow \ no. of\ negative\ sign$$
$$k = 7,$$
$$u = 3\ , no\ of\ positive\ sign$$

$$p' = P(u \geq k) = \left(\frac{1}{2}\right)^n \sum_{x=k}^{n} \binom{n}{x}$$

$$= P(u \geq 3) = \left(\frac{1}{2}\right)^{10} \sum_{x=3}^{10} \binom{10}{x}$$

$$= \left(\frac{1}{2}\right)^{10} \left[\binom{10}{3} + \binom{10}{4} + \binom{10}{5} + \binom{10}{6} + \binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right]$$

$$= 0.000976[120 + 210 + 252 + 210 + 120 + 45 + 10 + 1]$$
$$= 0.000976 * 968$$
$$p' = 0.94476$$

Conclusion:

$p' > 0.05, Accept\ H_0$.

2. The following data in tons, are the amounts of sulphur oxides emitted by a large industrial plant in 40 days.

24	15	20	29	19	18	22	25	27	9
17	20	17	6	24	14	15	23	24	26
19	23	28	19	16	22	24	17	20	13
19	10	23	18	31	13	20	17	24	14

Use the sign test to test the null hypothesis $\mu = 21.5$ against the altera hypothesis $\mu > 21.5$ at the 0.01 level of significance.

Solution

Null Hypothesis: $H_0: \mu = 21.5$

Alternative Hypothesis: $H_1: \mu > 21.5$. (one-tailed test)

Level of significance: $\alpha = 0.01$

Test Statistic

Given data is

$$+ \quad - \quad - \quad + \quad - \quad - \quad + \quad + \quad + \quad -$$
$$- \quad - \quad - \quad - \quad + \quad - \quad - \quad + \quad + \quad +$$
$$- \quad + \quad + \quad - \quad - \quad + \quad + \quad - \quad - \quad -$$
$$- \quad - \quad + \quad - \quad + \quad - \quad - \quad - \quad + \quad -$$

$$n = 16 + 24 = 40,$$

$$k \longrightarrow no. of\ negative\ sign$$
$$k = 24,$$
$$u = 16\ , no\ of\ positive\ sign$$
$$p = \frac{1}{2}, \qquad q = \frac{1}{2}, \quad (p + q = 1)$$
$$Z = \frac{u - np}{\sqrt{npq}}$$
$$Z = \frac{16 - 40 * 0.5}{\sqrt{40 * 0.5 * 0.5}} = \frac{16 - 20}{\sqrt{10}} = -1.26$$

$$|z| = 1.26$$

Conclusion

Table value $Z_{\propto}\ at\ \alpha = 0.01$is 2.33

$$1.26 < 2.33,$$
$$Accept\ H_0$$

Home Work

1. The following are the measurements of breaking strength of a certain kind
   of 2 inch cotton ribbon in pounds.

163	164	160	189	161	171	158	151	169	162
163	139	172	165	148	166	172	163	187	173