RUN TEST FOR RANDOMNESS

Course Instructor: Santosh Chhatkuli

Function of the test

One of the most important assumption of all types of statistical tests is that the sample must be as representative of the population as possible. It is because the sample results are used in decision regarding the population from which the samples are drawn. This requires the randomness of the data in the sample. The randomness of data can be tested by simple run test.

A run test can be used to test

- Serial randomness of nominal-scale categories
- Serial randomness of numerical data

What is run?

A run test uses the number of runs in a sequence of sample data to test for randomness in the order of data.

A run is defined as a sequence of identical symbols. It is a sequence of like elements, bounded on both sides by either unlike elements or no elements.

Example 1

In an industrial production line, items are inspected periodically for defective itmes. The following is a sequence of defective items, D, and non-defective items, N, produced by this production line.

DDNNNDNNDDNNNNNDDDNNDNNNNDND

Find the number of runs in the sequence.

Solution:

Here n = number of observations in the sequence = 28

The given sequence is

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

Here, the number of runs (r) =13

n_1 = Number of elements of the first category i.e D in the sequence = 11

n_2 = Number of element of the second category i.e. N in the sequence = 17

Example 2

The life-time of 19 successively produced storage batteries are as follows:

145	152	148	155	176	134	184	132	145	162
165	185	174	198	179	194	201	169	182	

Find the number of runs.

Solution:

To find the number of runs in the sequence of numerical data, we first find the median or any cut-off point and find the no. of runs above or below the median.

We use following categorization rule:

Use – or A sign if data falls below the median ($X_i < m_d$)

Use + or B sign if data falls above the median ($X_i > m_d$)

Omit those data which are tied to the median

Data Array:

132	134	145	145	148	152	155	162	165	169
174	176	179	182	184	185	194	198	201	

Obs. No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Life-time	145	152	148	155	176	134	184	132	145	162	165	185	174	198	179	194	201	169	182
Sign	-	-	-	-	+	-	+	-	-	-	-	+	+	+	+	+	+	omit	+

Hence the number of runs = 6

Test Assumptions

- The sample data are arranged according to some ordering scheme such as the order in which the sample values were obtained.
- Each data value can be categorized into one of two separate categories

Test hypothesis

We first set null and alternative hypothesis

Case I (Two tailed test)

H_0: the order of sample data is random (the sequence is random)

H_1: the order of sample data is not random

Case II (Left tailed test)

H_0: the order of sample data is random

H_1: too few runs (there is specific grouping or trend)

Case III (Right tailed test)

H_0: the order of sample data is random

H_1: too many runs (there is periodicity in the data)

Test statistic: small sample case

Small sample case ($n_1 < 20$, $n_2 < 20$)

We first find the n_1 and n_2 , where $n_1 =$ number of elements of first type and $n_2 =$ number of element of second type. If both n_1 and n_2 are smaller than 20 or one of them is smaller than 20, we consider it as small sample case. In this case we use exact distribution or r.

Test statistic = Number of runs in the sequence

Test statistic: large sample case

Large sample case (n1 ≥ 20 and n2 ≥ 20)

When n_1 and n_2 increase in size, the sampling distribution of r approaches to a normal probability distribution with

$$\text{mean } (\mu_r) = \frac{2\, n_1 n_2}{n} + 1$$

and standard deviation (σ_r) = $\sqrt{\dfrac{(2\, n_1 n_2)(2\, n_1 n_2 - n)}{n^2(n-1)}}$

Then Z transformation of r is given by,

$$Z_r = \frac{r - \mu_r}{\sigma_r}$$

The statistic Z follows standard normal distribution with mean 0 and standard deviation of 1.

Test statistic = Z score

Decision rule: small sample case

The total number of runs in the sequence indicates the sample random or not. If very large no. of runs occur, systematic short-period cyclical fluctuations seem to be influencing the scores. If very few runs occur, then it would indicate a pattern or a trend (clustering of scores). We use following decision rule.

Hypothesis	Decision Rule
Case I	Accept H_0 if $r_L < r < r_U$ otherwise reject H_0
Case II	Accept H_0 if $r > r_L$ otherwise reject H_0
Case III	Accept H_0 if $r < r_U$ otherwise reject H_0

Decision rule: large sample case

For large sample case we use Z approximation of r use following decision rule.

Hypothesis	Decision Rule
Case I	Accept H_0 if $-Z_{\alpha/2} < Z_r < +Z_{\alpha/2}$ otherwise reject H_0
Case II	Accept H_0 if $Z_r > -Z_\alpha$
Case II	Accept H_0 if $Zr < +Z_\alpha$

Critical value or Z

Significance level (α)	Two-tailed test	Left tailed test	Right tailed test
5 %	1.96	- 1.65	+ 1.65
1 %	2.58	- 2.33	+ 2.33

Numericals (Small sample case)

- At a small soda factory, the amount of soda put into each 12-ounce bottle by the bottling machine varies slightly for each filling. The plant manager suspects that the machine has a random pattern of overfilling and under-filling the bottles. The following are the results of filling 18 bottles, where O denotes 12 ounces or more of soda in a bottle and U denotes less than 12 ounces of soda.

  U U U O O O O U U O O
  O U O U U U U

  Using the runs test at the 5% significance level, can you conclude that there is a nonrandom pattern of overfilling and under-filling such bottles?

- The students in a statistics class were asked if they could be a good random number generator. Each student was asked to write down a single digit from 0 through 9. The data were collected starting at the front left of the class, moving row by row, to the back right of the class. The sequence of digits were as follows:

  7 4 3 6 9 5 4 4 4 3 6 3 3 7 7 7 6 3 6 7 6 9 6 7 3 7 7 3 4 6

  Do these data show a randomness about the median value of 4.5 at 5 % level of significance?

Numerical (Large sample case)

- The following arrangement of men (M) and women (W) lined up to purchase tickets for a rock concert.

  MWMWMMMWMWMMMWWMMMMWWMWMMMWMMMWWWMWMMMWMWMMMMW
  WMMWW

  Test for randomness of the arrangement at $\alpha = 0.05$ level of significance