Wald–Wolfowitz Runs test

Dr.E.S.Jeevanand,
Associate Professor
Union Christian College. Aluva

Wald–Wolfowitz Runs test

- The runs test is a shortened version of the full name: the Wald–Wolfowitz runs test, so named after mathematicians Abraham Wald and Jacob Wolfowitz.
- A runs test is a statistical procedure that examines whether a sequence of data is occurring randomly from a specific distribution. That is run test is used for examining whether or not a set of observations constitutes a random sample from an infinite population. Test for random ness is of major importance because the assumption of randomness underlies statistical inference.
- The run test for randomness is carried out in a random model in which the observations vary around a constant mean. The observation in the random model in which the run test is carried out has a constant variance, and the observations are also probabilistically independent.

Understanding a Run

- A run is defined as a series of increasing values or a series of decreasing values. The number of increasing, or decreasing, values is the length of the run. In short a run and the length of the run can be defined as
- ***Run*** *– sequence of similar events, items, or symbols that is followed by an event, item, or symbol that is mutually exclusive from the first event, item, or symbol*
- ***Length*** *– number of events, items, or symbols in a run*
- For example, a series of 20 coin tosses might produce the following sequence of heads (H) and tails (T).
- H H T T H T H H H H T H H T T T T T H H
- The number of runs for this series is nine. There are 11 heads and 9 tails in the sequence.

One Sample Wald–Wolfowitz Runs test

- In this case the **Null hypothesis and Alternative hypothesis: are**
- H_0: Sample value come from a random Sequence

  (or the given Sequence is random)
- H_1: Sample value come from a non-random Sequence

  (or the given Sequence is non- random)
- **Test Statistics is the number of runs r.**
- For finding the number of runs, the observations are listed in their order of occurrence. Each observation is denoted by '+' sign if it is more than the previous observation by a '-' sign if it is less than the previous observation. If the observation is same as previous observation put '0'. The total number of runs up (+) and down (-) is counted.

One Sample Wald–Wolfowitz Runs test

- **Critical value** for test is obtained from the table for **Wald–Wolfowitz runs test one sample for** a given value of n and desired level of significance α. Let this value be r_{cl} (lower value) and r_{cu} (upper value).
- Decision criteria is

  Accept H_0 if $r_{cl} \leq r \leq r_{cu}$. **Other wise reject H_0.**
- For large Sample: When sample size is greater than 25, the critical values r_{cl} and r_{cu} is obtained using normal approximation as r_{cl} = μ-1.96σ and r_{cu} = μ+1.96σ at 5% level of significant. Where

$$\mu = \frac{2n-1}{3} \text{ and } \sigma = \sqrt{\frac{16n-29}{90}}.$$

Example-I

- Data on value of imports of selected agricultural production inputs by a country in million rupees during recent 12 years is given below. Is the sequence random?
- 5.2, 5.5, 3.8,2.5, 8.3, 2.1, 1.7, 10.0,10.0, 6.9, 7.5, 10.6.
- The hypothesis are
- H_0: the Sequence is random
- H_1: the Sequence is non random

5.2	
5.5	+
3.8	-
2.5	-
8.3	+
2.1	-
1.7	-
10	+
10	0
6.9	-
7.5	+
10.6	+

Example I

- Here n=11, the number of runs r=7, Critical value for α = 0.05 from the table are r_{cl} (lower value) = 4 and r_{cu} (upper value) = 10.
- Here $r_{cl} \leq r \leq r_{cu}$. i.e. observed r lies between 4 and 10. So H_0 is accepted and conclude that the sequence is random

Example II

- Seasonal rainfall at a district for 25 years is given below test whether the sequence is random
- 25.34, 49.35, 39.62, 42.90, 57.66, 24.89, 50.63, 52.63, 38.47, 43.25, 50.83, 22.06, 22.04, 24.31, 45.13, 42.83, 46.94, 57.50, 30.70, 48.37, 38.45, 44.00, 50.03, 30, 54.2, 56

Example II Large Sample Case

Table to calculate runs			
25.34		24.31	+
49.35	+	45.13	+
39.62	-	42.83	-
42.90	+	46.94	+
57.66	+	57.50	+
24.89	-	30.70	-
50.63	+	48.37	+
52.63	+	38.45	-
38.47	-	44.00	+
43.25	+	50.03	+
50.83	+	30.00	-
22.06	-	54.20	+
22.04	-	56.00	+

Here n= 26, r=17

$\mu = \frac{2n-1}{3} = \frac{52-1}{3} = 17$

and $\sigma = \sqrt{\frac{16n-29}{90}} = \sqrt{\frac{416-29}{90}} = 2.074$

So r_{cl} = μ-1.96σ =17-1.96×2.074

=17-4.065 =12.935

and

r_{cu} = μ+1.96σ =17+ 1.96×2.074

=17+4.065 =21.065

Here $r_{cl} \leq r \leq r_{cu}$. i.e. observed r lies between 12.935 and 21.065. So H_0 is accepted and conclude that the sequence is random.

Wald–Wolfowitz Two sample Runs test

- Wald-Wolfowitz Two-sample Run test is used to examine whether two random samples came from population having same distribution. This test can detect differences in average or spread or any other important aspect between the two population.
- **Hypothesis**
- H_0: Sample value came from the population having same distribution
- H_1: Sample value came from the population having different distribution

Wald–Wolfowitz Two sample Runs test

- **Test Statistics is the number of runs r.**
- Let ***r*** denote the number of runs. To obtain ***r***, list the $n_1 + n_2$ observations from two samples in ascending order of magnitude. Denote observations from one sample as x's and other by y's. Count the number of runs. In case x and y observation are same value, place the observation x(y) first if run of x(y) observation is counting.
- **Critical value** for test is obtained from the table for **Wald–Wolfowitz runs test two sample** for a given value of n_1 and n_2 desired level of significance α. Let this value be r_c. Note that this a one sided hypothesis.

Wald–Wolfowitz Two sample Runs test

- Decision criteria is

Accept H_0 if $r > r_c$. **Or reject H_0 if** $r > r_c$.

- For large Sample: When Sample size is greater than 25, the critical values r_c is obtained using normal approximation as r_c = μ-1.96σ at 5% level of significant. Where

$$\mu = \mathbf{1} + \frac{\mathbf{2}\boldsymbol{n_1 n_2}}{\boldsymbol{n_1 + n_2}}$$

and

$$\sigma = \sqrt{\frac{\mathbf{2}\boldsymbol{n_1 n_2}(\mathbf{2}\boldsymbol{n_1 n_2 - n_1 - n_2})}{(\boldsymbol{n_1 + n_2})^{\mathbf{2}}(\boldsymbol{n_1 + n_2} - \mathbf{1})}}.$$

Example-III

- To determine if a new hybrid seeding procedures a bushier flowering plant, following data was collected. Examine whether the data indicate that the new hybrid produces larger shrubs than the current verity.
- The hypothesis are
- H_0: x and y population are same
- H_1: there is some difference in growth of x and y shrubs

Shrub's Growth (in inches)	
Hybrid	Current Variety
x	y
31.80	35.35
32.80	27.60
39.20	21.30
36.00	24.80
30.00	36.70
34.50	30.00
37.40	

Example-III

Table to calculate runs	
21.30	y
24.80	y
27.60	y
30.00	y
30.00	x
31.80	x
32.80	x
34.50	x
35.35	y
36.00	x
36.70	y
37.40	x
39.20	x

Test Statistics, the number of runs

r = 6.

For n_1 = 7 and n_2 = 6 and α=0.05 value of r_c = 3.

Since $r > r_c$ we accept H_0 and conclude that x and y have identical distribution.

Example IV

- Traffic conjunction on roads and highways costs industry loss of billions of dollars annually as workers to get to and from work. In order to over coms this company decided to introduce flexitime, which involve workers to determine their own schedules (provided they work a full shift). A 32 workers are selected for testing the effectiveness of the new pattern. Their travel time for normal shift of 8 A.M (x) is recorded on a Wednesday and on the next Wednesday they are allowed to opt the flexi time (y) and the travel time is recorded. These results are listed in the following table. Examine whether the data indicate that the the travel time under the flextime program are different from times to arrive at work at 8 A.M.?
- The hypothesis are
- H_0: x and y population are same
- H_1: there is some difference in travel time of x and y.

Example IV

x	y	x	y	x	y
34	31	13	12	19	18
35	31	69	71	48	51
43	44	18	13	29	33
46	44	53	55	24	21
16	15	18	19	51	50
26	28	41	38	40	38
68	63	25	23	26	22
38	39	17	14	20	19
61	63	26	21	19	21
52	54	44	40	42	38
68	65	30	33		

Example IV
Table for calculating runs

12	y	19	y	24	x
13	x	19	y	25	x
13	y	19	x	26	x
14	y	19	x	26	x
15	y	20	x	26	x
16	x	21	y	28	y
17	x	21	y	29	x
18	y	21	y	30	x
18	x	22	y	31	y
18	x	23	y	31	y
				33	y
				33	y

Example IV
Table for calculating runs

34	x	44	y	61	x
35	x	44	y	63	y
38	x	44	x	63	y
38	y	46	x	65	y
38	y	48	x	68	x
38	y	50	y	68	x
39	y	51	x	69	x
40	y	51	y	71	y
40	x	52	x		
41	x	53	x		
42	x	54	y		
43	x	55	y		

Example IV

Test Statistics, the number of runs r = 27.

$n_1 = n_2 = 32$,

$$\mu = \mathbf{1} + \frac{\mathbf{2n_1n_2}}{\boldsymbol{n_1+n_2}} = 1 + \frac{2\times32\times32}{32+32} = 33$$

and

$$\sigma = \sqrt{\frac{\mathbf{2n_1n_2(2n_1n_2-n_1-n_2)}}{\mathbf{(n_1+n_2)^2(n_1+n_2-1)}}} = \sqrt{\frac{\mathbf{2\times32\times32\ (2\times32\times32-32-32)}}{\mathbf{(32+32)^2(32+32-1)}}} = 15.746$$

- So r_c = μ-1.96σ = 33 – 1.96×15.746 =1.138
- Since $r > r_c$ we accept H_0 and conclude that x and y have identical distribution.

The End

You can wait for next presentation for mann-Whitney and Kruskal-Wallis Test!

Contact details esjmaths@gmail.com

9847915101, 8089684892