PABNA UNIVERSITY OF SCIENCE AND TECHNOLOGY

Department of Information and Communication Engineering (ICE)

Assignment on : Run Test and Rank Sum Test

Course Code: STAT-2201

Course Title: Engineering Statistics

Submitted by:	Submitted to:
Sazia Afrin Labonna	**Dr. Md. Sarwar Hosain**
Roll: 210621 Session: 2020-21 2nd Year 2nd Semester Department of ICE, PUST	Associate Professor Department of Information and Communication Engineering(ICE), Pabna University of Science and Technology, Pabna.

Run Test and Rank Sum Test

The Runs Test For Randomness:

A nonparametric test for randomness is provided by the theory of runs.To understand what a run is,consider a sequence made up of two symbols, a and b, such as

a a|b b b|a|b b|a a a a a|b b b|a a a a|……………(1)

In tossing a coin,for example, a could represent "heads" and b could represent "tails". Or in sampling the bolts produced by a machine a could represent "defective" and b could represent "nondefective".

A run is defined as a set of identical symbols contained between two different symbols or no symbol. Proceeding from left to right in sequence (1),the first run ,indicated by a vertical bar, consists of two a's; similarly,the second run consists of three b's,the third run consists of one a,etc. there are seven run in all.

It seems clear that some relationship exists between randomness and the number of runs. Thus for the sequence

a|b|a|b|a|b|a|b|a|b|a|b|……………….(2)

there is a cyclic pattern, in which we go from a to b, back to a again,etc,which we could hardly believe to be random. In such case we have too many runs.

On the other hand , for the sequence

a a a a a a|b b b b|a a a a a|b b b|……………….(3)

there seems to be a trend pattern, in which the a's and b's are grouped together. In such case there are too few runs and we would not consider the sequence to be random.

Thus a sequence would be considered nonrandom if there are either too many or too few runs, and random otherwise. To quantity this idea, suppose that we form all possible sequences consisting of N_1 a's and N_2 b's, for a total of N symbols in all($N_1 + N_2 = N$). The collection of all these sequence provides us with a sampling distribution: Each sequence has an associated number of runs, denoted by V. in this way we are led to the sampling distribution of the statistic V. It can be shown that this sampling distribution has a mean and variance given, respectively, by the formulas

$$\mu_v = \frac{2N_1N_2}{N_1+N_2} + 1 \qquad \sigma_v^2 = \frac{2N_1N_2(2N_1N_2-N_1-N_2)}{(N_1+N_2)^2(N_1+N_2-1)} \text{.........................(4)}$$

By using formulas (4), we can test the hypothesis of randomness at appropriate levels of significance. It turns out that if both N_1 and N_2 are at least equal to 8, then the sampling distribution of V is very nearly a normal distribution. Thus

$$z = \frac{V-u_v}{\sigma_v} \text{............................(5)}$$

is normally distributed with mean 0 and variance 1, and thus Appendix II can be used.

Rank Sum Test(THE MANN-WHITNEY U TEST)

The Mann–Whitney U Test is a non-parametric test used to determine whether two independent samples come from the same distribution**.** It is often seen as the non-parametric alternative to the two-sample t-test**,** which assumes normality.

Step 1: Combine all simple values in an array from the smallest to the largest and assign ranks to all these values. If two or more sample values are identical, the sample values are each assigned a rank equal to the mean of the ranks that would otherwise be assigned.

Step 2: Find the sum of the ranks for each of the samples. Denote these sums by R_1 and R_2 where N_1 and N_2 are the respective sample sizes. For convenience, choose N_1 as the smaller size if they are unequal,so that $N_1 <= N_2$. A significant difference between the rank sums R_1 and R_2 implies a significant difference between the samples.

Step 3: To test the difference between the rank sums, use the statistics

$$U=N_1N_2+\frac{N(N_1+1)}{2}-R_1 \text{.............................}(1)$$

The sampling distribution of U is symmetrical and has a mean and variance given, respectively, by the formulas

$$\mu_u=\frac{N_1N_2}{2} \qquad \sigma^2{}_u=\frac{N_1N_2(N_{1+}N_2+1)}{12} \text{..................}(2)$$

If N_1 and N_2 are both at least equal to 8, it turns out that the distribution of U is nearly normal, so that

$$Z=\frac{U-u_U}{\sigma_U} \text{.....................}(3)$$

Is normally distributed with mean 0 and variance 1.

A value corresponding to sample (1) is given by the statistic

$$U=N_1N_2+\frac{N(N_1+1)}{2}-R_2 \text{.......................}(4)$$

and has the same sampling distribution as statistic (1), with the mean and variance of formulas (2). Statistic (4) is related to (1), for U_1 and U_2 are the values corresponding to statistics (1) and (4),respectively, then we have the result

$$U_1+U_2=N_1N_2 \text{.......................}(5)$$

We also have

$$R_1+R_2=\frac{N(N+1)}{2} \text{....................}(6)$$

Where $N= N_1 +N_2$.