(E) 1

## The run's test

- We look at a test for testing randomness of observed data on the basis of the order of which they were obtained.

Definition: Run - A run is a succession of identical letters (or other kinds of symbols) which is preceded and followed by different letters or no letters at all.

- Consider the following arrangement of effective (d) and non-effective (n) pieces produced in the given order by a certain machine.

$$\underbrace{n\,n\,n\,n\,n}_{1}\ \underbrace{d\,d\,d\,d}_{2}\ \underbrace{n\,n\,n\,n\,n\,n\,n\,n\,n\,n}_{3}\ \underbrace{d\,d}_{4}\ \underbrace{n\,n}_{5}\ \underbrace{d\,d\,d\,d}_{6}\ \underbrace{n}_{7}\ \underbrace{d\,d}_{8}\ \underbrace{n\,n}_{9}$$

hence 9 runs.

- The total number of runs appearing in an arrangement of this kind is often a good indicator of a possible lack of randomness
- If there are too few runs, we might suspect a definite grouping or clustering or perhaps a trend.
- If there are too many runs, we might suspect some sort of repeated alternating pattern
- In our illustration, there seems to be some sort of <u>clustering</u> (few runs)

- If we have $n_1$ letters of some kind and $n_2$ letters of another kind, there are:

$\binom{n_1+n_2}{n_1}$ possible arrangements of these letters

- All arrangements are regarded to be equally likely.

Example

1) If we have 2 a's and b's, then the no of possible arrangements of these letters is

$$\binom{2+2}{2} = \binom{4}{2} = \frac{4!}{2!\,2!} = \frac{4\times3\times2\times1}{2\times2\times1\times1} = 6$$

These arrangements are:

| | | | | |
|---|---|---|---|---|
| aabb | abab | abba | baba | bbaa |
| 2r | 4r | 3r | 4r | 2r |

baab
3r

2) We have five letters e's which are divided into 3 runs using vertical bars to represent the 5 letters into 3 runs. We find that there are 6 possibilities i.e.

e|e|eee    e|ee|ee    ee|ee|ee

eee|e|e    ee|e|ee    e|eee|e

- Let there be $n_1$ letters of one kind and $n_2$ letters of another kind.
- Let $u$ be the number of runs formed by these $n_1 + n_2$ letters
- If $u = 2k$, where $k$ is a positive integer then we have $k$ runs of $n_1$ letters and $k$ runs of the $n_2$ letters.
- Number of ways of which $n_1$ letters can form $k$ run is $\binom{n_1 - 1}{k-1}$ and number of ways in which $n_2$ letters can for $k$ runs is $\binom{n_2 - 1}{k-1}$

$$p\left(u = 2k \text{ runs}\right) = F(u) = \frac{2\binom{n_1-1}{k-1}\binom{n_2-1}{k-1}}{\binom{n_1+n_2}{n_1}}$$

if $u = 2k+1$

$$F(u) = p(u_1 = k+1,\ u_2 = k) + p(u_1 = k,\ u_2 = k+1)$$

$$= \binom{n_1-1}{k}\binom{n_2-1}{k-1} + \binom{n_1-1}{k-1}\binom{n_2-1}{k}$$

$$= \frac{\binom{n_1-1}{k}\binom{n_2-1}{k-1} + \binom{n_1-1}{k-1}\binom{n_2-1}{k}}{\binom{n_1+n_2}{n_1}}$$

when $n_1$ and $n_2$ are small the test of randomness is based on $u$ (the number of runs) are ussally performed with the use of special tables. We reject the null hypothesis of randomness at $\alpha$ level of significance if $u < u_{\alpha/2}$ or

$u \geq u_{\alpha/2}$

where $u_{\alpha/2}$ is the largest value for which $P(u \leq u_{\alpha/2})$ does not exceed $\alpha/2$ and $u_{\alpha/2}$ is the smallest value for which $P(u \geq u_{\alpha/2})$ does not exceed $\alpha/2$

<u>Example</u>

Checking on the palm tree that were planted many years ago along a country road a county official obtained the following arrangements of health (H) and diseased (D) trees i.e.

H H H H D D D H H H H H H H

D D H H D D D D

Test at $\alpha = 0.05$ whether this arrangement can be regarded to be random

Solution

We wish to test

$H_0$: Arrangement is random

$H_1$: Arrangement is not random

No of H's, $n_1 = 13$ $\qquad u_{\alpha/2} = u_{0.05/2} = 6$

No of D's, $n_2 = 9$

We reject $H_0$ if $u \le 6$ or $u \ge 7$

From the data $u = 6$, since $6 \le 6$ we reject $H_0$ and conclude that the arrangement of healthy and diseased tree is not random.

- If $n_1$ and $n_2$ are both greater or equal to 10, it is considered reasonable to assume that the distribution of $u$ can be closely approximated by the normal.
- To perform the test on the basis of this assumption we need the following results.

<u>Theorem</u>: Under the null hypothesis of randomness the mean and variance of $u$ are

$$E[u] = \frac{2n_1 n_2}{n_1 + n_2} + 1$$

$$Var(u) = \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)}$$

<u>Example</u>

The following series of boys and girls were selected haphazardly in a core education school

G BB GG B G BBB G G G G B G BG BB GB G G G
B B G B GG B G BG BBG

Test for randomness at $\alpha = 0.01$

Solution

We want to test $H_0$: Arrangement is random

$H_1$: Arrangement is not random

(6)

If $u$ is the number of runs, than

$$E[u] = \frac{2n_1n_2}{n_1+n_2} - 1$$

$$var(u) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{n_1}$$

We use the test statistic

$$Z = \frac{(u \pm 1/2) - E[u]}{\sqrt{var(u)}}$$

We reject $H_0$ if $Z \le -Z_{0.005} = -2.575$ $= \alpha/2$

or if $Z \geq Z_{0.005} = 2.575$

No of girls $G = n_1 = 20$
No of boys $B = n_2 = 18$

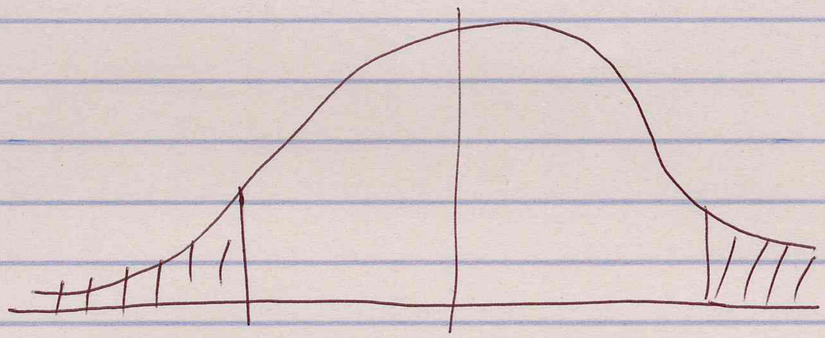
$u = 25$ (runs)

$$E[u] = \frac{2(20)(18)}{20+18} + 1 = 19.947$$

$$Var(u) = \frac{2(20)(18)[2(20)(18) - 20 - 18]}{(20+18)^2(20+18-1)}$$

$$= 9.190$$

$$Z = \frac{25 - 1/2 - 1.947}{\sqrt{9.190}}$$




Since

$$-2.57 < u < 2.57$$

$$-2.57 < 1.50 < 2.57$$

~~we accept~~ $H_0$ ~~and~~
We do not reject $H_0$ and say the arrangement is random (1.50 is in the acceptance region)

Exercise

1) The following are the amounts of money in dollars spend by 16 persons at a certain amusement bar

10.15 9.85 13.75 8.63 11.09 15.63

6.65 9.27 8.80 11.45 10.29 9.51 13.80

10.0 7.48 9.11

(a) Use the sign-test at $\alpha = 0.05$ to test the null hypothesis that on average a person spends 9 dollars at the bar.

(b) Rework the preceding problem using the sign-rank test based on the normal approximation to the distribution of the test statistic