Math 362: Mathematical Statistics II

Le Chen
le.chen@emory.edu

Emory University
Atlanta, GA

Last updated on April 24, 2021

2021 Spring

Chapter 14. Nonparametric Statistics

Plan

Whether the sample are random at all?

E.g. Whether the number of successful strikes are random? $\alpha = 0.05$.

Whether the sample are random at all?

E.g. Whether the number of successful strikes are random? $\alpha = 0.05$.

Whether the sample are random at all?

E.g. Whether the number of successful strikes are random? $\alpha = 0.05$.

Year	Number of Strikes	% Successful, y_i
1881	451	61
1882	454	53
1883	478	58
1884	443	51
1885	645	52
1886	1432	34
1887	1436	45
1888	906	52
1889	1075	46
1890	1833	52
1891	1717	37
1892	1298	39
1893	1305	50
1894	1349	38
1895	1215	55
1896	1026	59
1897	1078	57
1898	1056	64
1899	1797	73
1900	1779	46
1901	2924	48
1902	3161	47
1903	3494	40
1904	2307	35
1905	2077	40

Sol. Compute the run-up and run-down:

$1 \to$
$2 \to$
$3 \to$
$4 \to$
$5 \to$
$6 \to$

$7 \to$
$8 \to$
$9 \to$
$10 \to$

$11 \to$
$12 \to$

$13 \to$
$14 \to$

$15 \to$
$16 \to$
$17 \to$

$18 \to$

Sol. Compute the run-up and run-down:

Year	Number of Strikes	% Successful, y_i		$\operatorname{sgn}(y_i - y_{i-1})$
1881	451	61	1 $\longrightarrow$	$-$
1882	454	53	2 $\longrightarrow$	$+$
1883	478	58	3 $\longrightarrow$	$-$
1884	443	51	4 $\longrightarrow$	$+$
1885	645	52	5 $\longrightarrow$	$-$
1886	1432	34	6 $\longrightarrow$	$+$
1887	1436	45		$+$
1888	906	52	7 $\longrightarrow$	$-$
1889	1075	46	8 $\longrightarrow$	$+$
1890	1833	52	9 $\longrightarrow$	$-$
1891	1717	37	10 $\longrightarrow$	$+$
1892	1298	39		$+$
1893	1305	50	11 $\longrightarrow$	$-$
1894	1349	38	12 $\longrightarrow$	$+$
1895	1215	55		$+$
1896	1026	59	13 $\longrightarrow$	$-$
1897	1078	57	14 $\longrightarrow$	$+$
1898	1056	64		$+$
1899	1797	73	15 $\longrightarrow$	$-$
1900	1779	46	16 $\longrightarrow$	$+$
1901	2924	48	17 $\longrightarrow$	$-$
1902	3161	47		$-$
1903	3494	40		$-$
1904	2307	35	18 $\longrightarrow$	$+$
1905	2077	40		

$\Big\}\ w = 18$

Theorem Let W be the number of runs up and down in a sequence of $n \geq 2$ observations.

If the sequence is random, then

$$\mathbb{E}(W) = \frac{2n-1}{3} \quad \text{and} \quad \mathrm{Var}(W) = \frac{16n-29}{90}.$$

Moreover, when n is large, namely, $n \geq 20$, then

$$\frac{W - \mathbb{E}(W)}{\sqrt{\mathrm{Var}(W)}} = \frac{W - [2n-1]/3}{\sqrt{[16n-29]/90}} \quad \overset{approx}{\sim} \quad N(0,1).$$

Theorem Let W be the number of runs up and down in a sequence of $n \geq 2$ observations.

If the sequence is random, then

$$\mathbb{E}(W) = \frac{2n-1}{3} \quad \text{and} \quad \mathrm{Var}(W) = \frac{16n-29}{90}.$$

Moreover, when n is large, namely, $n \geq 20$, then

$$\frac{W - \mathbb{E}(W)}{\sqrt{\mathrm{Var}(W)}} = \frac{W - [2n-1]/3}{\sqrt{[16n-29]/90}} \quad \overset{approx}{\sim} \quad N(0,1).$$

Theorem Let W be the number of runs up and down in a sequence of $n \geq 2$ observations.

If the sequence is random, then

$$\mathbb{E}(W) = \frac{2n-1}{3} \quad \text{and} \quad \mathrm{Var}(W) = \frac{16n-29}{90}.$$

Moreover, when n is large, namely, $n \geq 20$, then

$$\frac{W - \mathbb{E}(W)}{\sqrt{\mathrm{Var}(W)}} = \frac{W - [2n-1]/3}{\sqrt{[16n-29]/90}} \quad \overset{approx}{\sim} \quad N(0,1).$$

Sol. (Continued) $n = 25,\ w = 18$

$$\mathbb{E}(W) = \frac{2 \times 25 - 1}{3} = 16.3$$

and

$$\mathrm{Var}(W) = \frac{16 \times 25 - 29}{90} = 4.12.$$

Hence, the z-score is

$$z = \frac{18 - 16.3}{\sqrt{4.12}} = 0.84.$$

The critical region is

$$C = \{z : |z| \geq z_{\alpha/2} = z_{0.025} = 1.96\}$$

The p-value is

$$2 \times \mathbb{P}(Z > 0.84) = 0.4009084$$

Conclusion: Fail to reject. ■

Sol. (Continued) $n = 25$, $w = 18$

$$\mathbb{E}(W) = \frac{2 \times 25 - 1}{3} = 16.3$$

and

$$\text{Var}(W) = \frac{16 \times 25 - 29}{90} = 4.12.$$

Hence, the z-score is

$$z = \frac{18 - 16.3}{\sqrt{4.12}} = 0.84.$$

The critical region is

$$C = \{z : |z| \geq z_{\alpha/2} = z_{0.025} = 1.96\}$$

The p-value is

$$2 \times \mathbb{P}(Z > 0.84) = 0.4009084$$

Conclusion: Fail to reject. ■

Sol. (Continued) $n = 25$, $w = 18$

$$\mathbb{E}(W) = \frac{2 \times 25 - 1}{3} = 16.3$$

and

$$\text{Var}(W) = \frac{16 \times 25 - 29}{90} = 4.12.$$

Hence, the z-score is

$$z = \frac{18 - 16.3}{\sqrt{4.12}} = 0.84.$$

The critical region is

$$C = \left\{ z : |z| \geq z_{\alpha/2} = z_{0.025} = 1.96 \right\}$$

The p-value is

$$2 \times \mathbb{P}(Z > 0.84) = 0.4009084$$

Conclusion: Fail to reject. ∎

Sol. (Continued) $n = 25$, $w = 18$

$$\mathbb{E}(W) = \frac{2 \times 25 - 1}{3} = 16.3$$

and

$$\mathrm{Var}(W) = \frac{16 \times 25 - 29}{90} = 4.12.$$

Hence, the z-score is

$$z = \frac{18 - 16.3}{\sqrt{4.12}} = 0.84.$$

The critical region is

$$C = \{z : |z| \geq z_{\alpha/2} = z_{0.025} = 1.96\}$$

The p-value is

$$2 \times \mathbb{P}(Z > 0.84) = 0.4009084$$

Conclusion: Fail to reject. ■

Sol. (Continued) $n = 25$, $w = 18$

$$\mathbb{E}(W) = \frac{2 \times 25 - 1}{3} = 16.3$$

and

$$\text{Var}(W) = \frac{16 \times 25 - 29}{90} = 4.12.$$

Hence, the z-score is

$$z = \frac{18 - 16.3}{\sqrt{4.12}} = 0.84.$$

The critical region is

$$C = \{z : |z| \geq z_{\alpha/2} = z_{0.025} = 1.96\}$$

The p-value is

$$2 \times \mathbb{P}(Z > 0.84) = 0.4009084$$

Conclusion: Fail to reject. ■

R code:

```
library("snpar")
y <- c(
0,1,0,1,0,1,1,0,1,0,1,1,0,1,1,0,1,1,0,1,0,0,0,1

)
runs.test(y, exact = FALSE)
runs.test(y, exact = TRUE)
```

Output:

```
> runs.test(y, exact = FALSE)

        Approximate runs rest

data:  y
Runs = 18, p-value = 0.03256
alternative hypothesis: two.sided

> runs.test(y, exact = TRUE)

        Exact runs test

data:  y
Runs = 18, p-value = 0.01624
alternative hypothesis: two.sided
```

Remark The procedure that we learnt is an approximation. There is a big discrepancy for the above two p-values: one that we obtained through formula and one that is obtained by the r function.

Thanks for learning statistics with me through the semester !