ORIGIN ≡ 1

Kruskal-Wallis Test

The Kruskal-Wallis is a non-parametric analog to the One-Way ANOVA F-Test of means. It is useful when the k samples appear not to come from underlying Normal Distributions, or when variance in the different samples are of greatly different magnitudes (non-homogeneous). As with other rank-based tests, it does not have as much power as the fully parametric tests, but nevertheless enjoys wide use. Note that when the number of samples k=2, this test is identical to the Mann-Whitney Test.

Data Structure:

k groups with not necessarily the same numbers of observations and different means.

Let index i,j indicate the ith column (treatment class) and jth row (object).

One-Way ANOVA					
	Treatment Classes:				
Objects (Replicates)	**#1**	**#2**	**#3**	**…**	**#k**
1					
2					
3					
…					
n	n1	n2	n3		nk
means:	**X1bar**	**X2bar**	**X3bar**	**…**	**Xkbar**

Assumptions:

- Observations in each class(block) are a random sample.
- Observations in each block are independent of observations in other class.
- Underlying distribution of observations in each cell are continuous.
- Measurement scale is at least ordinal.

Zar Example 10.10

Hypotheses:

H_0: Δ= 0   < No population differences

H_1: Δ ≠ 0   < **Two Sided Test**

	number	loc	rank	RankSum
4	8.2	1	10	
3	9.6	1	12	
5	10.2	1	13	
2	12.1	1	14	
1	14	1	15	64
7	5.1	2	2	
8	5.5	2	4	
10	6.3	2	6	
9	6.6	2	7	
6	8.4	2	11	30
14	4.1	3	1	
15	5.4	3	3	
13	5.8	3	5	
11	6.9	3	8	
12	7.3	3	9	26

Criterion for Normal Approximation:

IF $n_i \geq 5$ THEN Normal Approximation Applies

OTHERWISE use Special Tables e.g. Rosner Table 15 p. 844

Normal Approximation:

Rank Data and Sum:

- Pool the data over all treatment classes - Total sample size $N = \Sigma n_i$
- Assign Data to Ranks. In the case of ties, t observations in a rank are assigned the appropriate average rank.
- Compute the Rank Sum (R_i) for each treatment class i.

$$N := 15 \qquad k := 3$$
$$R_1 := 64 \qquad n_1 := 5$$
$$R_2 := 30 \qquad n_2 := 5$$
$$R_3 := 26 \qquad n_3 := 5$$

Test Statistic:

$$i := 1 .. 3$$

$$H_S := \frac{12}{N \cdot (N+1)} \cdot \sum_i \frac{(R_i)^2}{n_i} - 3 \cdot (N+1) \qquad \text{< where } R_i \text{ are the Rank sums}$$
$$\text{for each treatment class i}$$

$$H_S = 8.72$$

IF no ties, THEN: $H := H_S$ **< no correction factor in this case...**

OTHERWISE:

$$m := 0$$

correction factor > $\Omega := 1 - \dfrac{\displaystyle\sum_{j=1}^{m} \left[(t_j)^3 - t_j\right]}{N^3 - N}$ **< t represent the number of observations that are tied in groups 1 to g**

$$H_X := \frac{H_S}{\Omega} \qquad \text{< Corrected Test Statistic}$$

Sampling Distribution of the Test Statistic H:

If Assumptions hold, H_0 is true, and assuming Normal Approximation then $H \sim \chi^2_{(k-1)}$

Critical Value of the Test:

< Probability of Type I error must be explicitly set

$$\alpha := 0.05$$
$$C := qchisq(1 - \alpha, k - 1) \qquad C = 5.9915 \quad \text{< Normal approximation} \qquad \text{< Note: df = (k-1)}$$

$$C := 5.780 \qquad \text{< Critical Value from Zar Appendix B-13}$$

Decision Rule:

IF H > C THEN REJECT H_0 OTHERWISE ACCEPT H_0

$$H = 8.72$$

Probability Value:

$$P := (1 - pchisq(H, k - 1)) \qquad P = 0.0128 \qquad \text{< Normal approximation}$$

Zar Example 10.11

Hypotheses:

H_0: $\Delta = 0$ **< No population differences**

H_1: $\Delta \neq 0$ **< Two Sided Test**

pH	pond	rank	RankSum	n
7.68	1	1		
7.69	1	2		
7.7	1	3.5		
7.7	1	3.5		
7.72	1	8		
7.73	1	10		
7.73	1	10		
7.76	1	17	55	8
7.71	2	6		
7.73	2	10		
7.74	2	13.5		
7.74	2	13.5		
7.78	2	20		
7.78	2	20		
7.8	2	23.5		
7.81	2	26	132.5	8
7.74	3	13.5		
7.75	3	16		
7.77	3	18		
7.78	3	20		
7.8	3	23.5		
7.81	3	26		
7.84	3	28	145	7
7.71	4	6		
7.71	4	6		
7.74	4	13.5		
7.79	4	22		
7.81	4	26		
7.85	4	29		
7.87	4	30		
7.91	4	31	163.5	8

Criterion for Normal Approximation:

- **IF** $n_i \geq 5$ **THEN Normal Approximation Applies**

    OTHERWISE use Special Tables e.g. Rosner Table 15 p. 844

Normal Approximation:

Rank Data and Sum:

- Pool the data over all treatment classes - Total sample size $N = \Sigma n_i$
- Assign Data to Ranks. In the case of ties, t observations in a rank are assigned the appropriate average rank.
- Compute the Rank Sum (R_i) for each treatment class i.

$$k := 4$$

$$R := \begin{pmatrix} 55 \\ 132.5 \\ 145 \\ 163.5 \end{pmatrix} \qquad n := \begin{pmatrix} 8 \\ 8 \\ 7 \\ 8 \end{pmatrix} \qquad N := \sum n \qquad N = 31$$

Test Statistic:

$$i := 1 .. k$$

$$H_S := \frac{12}{N \cdot (N + 1)} \cdot \sum_i \frac{\left(R_i\right)^2}{n_i} - 3 \cdot (N + 1) \qquad \text{< where } R_i \text{ are the Rank sums}$$
$$\text{for each treatment class i}$$

$$H_S = 11.8761$$

IF no ties, THEN: $H_X := H_S$ **< no correction factor...**

OTHERWISE:

$$m := 7$$

$$t := \begin{pmatrix} 2 \\ 3 \\ 3 \\ 4 \\ 3 \\ 2 \\ 3 \end{pmatrix} \qquad \begin{array}{l} \text{< t represent the number of} \\ \text{tied observations in each of} \\ \text{the m groups of tied ranks} \end{array}$$

correction factor > $\Omega := 1 - \dfrac{\displaystyle\sum_{j=1}^{m} \left[\left(t_j\right)^3 - t_j\right]}{N^3 - N}$ $\Omega = 0.9944$

$$H := \frac{H_S}{\Omega} \qquad H = 11.9435 \qquad \text{< Corrected Test Statistic in this case}$$

Distribution of the Test Statistic H:

If H_0 is true and assuming Normal Approximation then $H \sim \chi^2_{(k-1)}$

Critical Value of the Test:

< Probability of Type I error must be explicitly set

$$\alpha := 0.05$$

$$C := \text{qchisq}(1 - \alpha, k - 1) \qquad C = 7.8147 \qquad \text{< Normal approximation} \qquad \text{< Note: df = (k-1)}$$

Decision Rule:

IF H > C THEN REJECT H_0 OTHERWISE ACCEPT H_0

Probability Value:

$$P := (1 - \text{pchisq}(H, k - 1)) \qquad P = 0.0076 \qquad \textbf{< Normal approximation}$$

Prototype in R:

```
#KRUSKAL-WALLIS TEST
#ZAR EXAMPLE 10.10
ZAR=read.table("c:/DATA/Biostatistics/ZarEX10.10R.txt")
ZAR
attach(ZAR)
kruskal.test(number,loc)
```

Kruskal-Wallis rank sum test

data: number and loc
Kruskal-Wallis chi-squared = 8.72, df = 2, p-value = 0.01278

```
#ZAR EXAMPLE 10.11
ZAR=read.table("c:/DATA/Biostatistics/ZarEX10.11R.txt")
ZAR
attach(ZAR)
kruskal.test(pH,pond)
```

Kruskal-Wallis rank sum test

data: pH and pond
Kruskal-Wallis chi-squared = 11.9435, df = 3, p-value = 0.007579