

克拉斯卡－瓦立斯檢定 Kruskal-Wallis Test

意義

小樣本統計中，替代 ANOVA 的無母數統計方法。

公式

$$K = (N - 1) \frac{\sum_{i=1}^g n_i (\bar{r}_i - \bar{R})^2}{\sum_{i=1}^g \sum_{j=1}^{n_i} (r_{ij} - \bar{R})^2}$$

N=樣本數

n_i =第 i 組樣本數

r_{ij} =第 i 組的第 j 個樣本等級

\bar{r}_i =第 i 組的樣本等級平均

\bar{R} =全樣本等級平均

範例

表為近十年台灣近海、沿岸與養殖烏魚的漁獲量，以克拉斯卡－瓦立斯檢定三種漁獲區域的漁獲量有無顯著差異。

year	offshore	coastal	aquaculture
2011	221	256	2422
2012	756	572	1664
2013	514	144	2637
2014	526	225	2728
2015	993	296	1931
2016	1511	287	1511
2017	1207	222	2050
2018	965	167	2390
2019	1026	585	1985
2020	791	221	1326

由題目得知樣本數

N=30

$n_{offshore} = 10$ 、 $n_{coastal} = 10$ 、 $n_{aquaculture} = 10$

$N - 1 = 30 - 1 = 29$ -----①

30 個樣本等級排序如下：

offshore		coastal		aquaculture	
n	等級	n	等級	n	等級
221	3.5	144	1	1326	20
514	10	167	2	1511	21.5
526	11	221	3.5	1664	23
756	14	222	5	1931	24
791	15	225	6	1985	25
965	16	256	7	2050	26
993	17	287	8	2390	27
1026	18	296	9	2422	28
1207	19	572	12	2637	29
1511	21.5	585	13	2728	30

$$\bar{r}_{offshore} = \frac{3.5 + 10 + 11 + 14 + 15 + 16 + 17 + 18 + 19 + 21.5}{10} = 14.5$$

$$\bar{r}_{coastal} = \frac{1 + 2 + 3.5 + 5 + 6 + 7 + 8 + 9 + 12 + 13}{10} = 6.65$$

$$\bar{r}_{aquaculture} = \frac{20 + 21.5 + 23 + 24 + 25 + 26 + 27 + 28 + 29 + 30}{10} = 25.35$$

$$\bar{R} = \frac{10 \times 14.5 + 10 \times 6.65 + 10 \times 25.35}{30} = 15.5$$

$$n_{offshore}(\bar{r}_{offshore} - \bar{R})^2 = 10 \times (14.5 - 15.5)^2 = 10$$

$$n_{coastal}(\bar{r}_{coastal} - \bar{R})^2 = 10 \times (6.65 - 15.5)^2 = 783.225$$

$$n_{aquaculture}(\bar{r}_{aquaculture} - \bar{R})^2 = 10 \times (25.35 - 15.5)^2 = 970.225$$

$$\sum_{i=1}^g n_i(\bar{r}_i - \bar{R})^2 = 10 + 783.225 + 970.225 = 1763.45 \text{-----} \textcircled{2}$$

$\sum_{i=1}^g \sum_{j=1}^{n_i} (r_{ij} - \bar{R})^2$ 依下表計算

n	等級	$(r_{ij} - \bar{R})^2$	n	等級	$(r_{ij} - \bar{R})^2$	n	等級	$(r_{ij} - \bar{R})^2$
221	3.5	144	144	1	210.25	1326	20	20.25
514	10	30.25	167	2	182.25	1511	21.5	36
526	11	20.25	221	3.5	144	1664	23	56.25
756	14	2.25	222	5	110.25	1931	24	72.25
791	15	0.25	225	6	90.25	1985	25	90.25
965	16	0.25	256	7	72.25	2050	26	110.25
993	17	2.25	287	8	56.25	2390	27	132.25
1026	18	6.25	296	9	42.25	2422	28	156.25
1207	19	12.25	572	12	12.25	2637	29	182.25
1511	21.5	36	585	13	6.25	2728	30	210.25

$$\sum_{i=1}^g \sum_{j=1}^{n_i} (r_{ij} - \bar{R})^2 = 254 + 926.25 + 1066.25 = 2246.5 \text{-----} \textcircled{3}$$

$\therefore K = 29 \times \frac{1763.45}{2246.5} = 22.764$ ，p 值達顯著，三種區域的漁獲量有顯著差異。