Kruskal-Wallis Test

Here our dataset is of (711S) Art sector which is small in size. The data gathered from the survey which is divided into 9 different Segments namely DIVERSITY IN POLICIES AND PROCEDURES, INCLUSIVITY BY POLICIES AND PRACTICES, DIVERSITY BY RECRUITMENT PROCESS, DIVERSITY IN POLICIES AND PROCEDURES, INCLUSIVITY BY POLICIES AND PRACTICES, WORKPLACE INCLUSION AND DIVERSITY BY RECRUITMENT PROCESS, ORGANIZATION'S CULTURE, INCLUSION BY DIFFERENT NEEDS and WORKPLACE INCLUSION AND DIVERSITY BY CAREER DEVELOPMENT AND EMPLOYEE RETENTION coded as Q1 to Q9 in the dataset.

The dataset has responses from 6 employees with various job position. Here in the second column the 2 in front of Q1 indicates the median of all the Q1 for all employees which is DIVERSITY IN POLICIES AND PROCEDURES. Likewise, for all different segments with respect to the job positions that make 62 rows.

The dataset looks like:

Segment	Medians
Q1	2
Q1	4
Q1	3
Q1	3
Q1	4
Q1	1
Q1	0
Q1	1
Q2	4
Q2	3
Q2	5
Q2	4

- The outcome of Kruskal-Wallis tells us if there is difference among segments but doesn't tell us which segments are different. In order to determine which segments, differ we need to use Dunn's test as a post-hoc analysis of Kruskal-Wallis test.
- Assumptions:
    1. One-way data.
    2. Dependent variable is ordinal, interval or ratio.
    3. Independent variable is a factor with two or more levels (Two or more segments).
    4. Observations between segments are independent.

5. In order to be a test of medians, the distribution of values for each group need to be of similar spread.

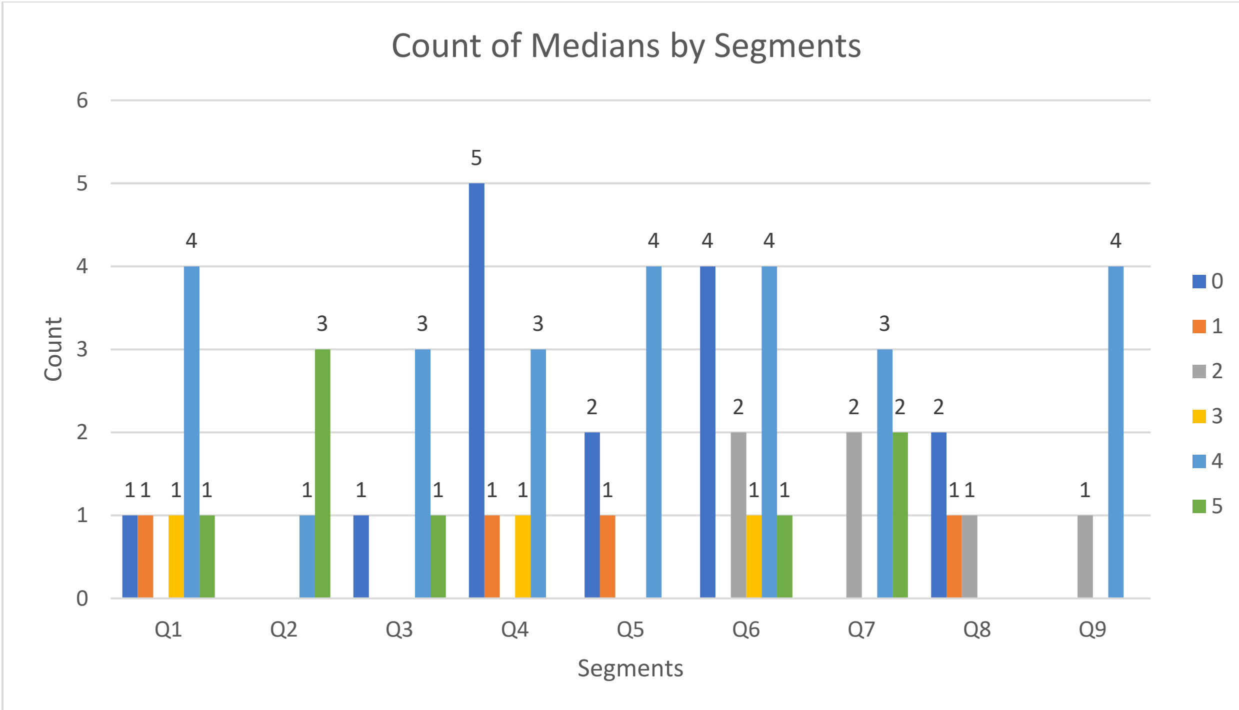
- Hypothesis:
- H0: The segments sampled from populations are with identical distributions.
- H1: The segments are sampled from populations with different distributions.

- Interpretation
- Significant results can be reported as “There was a significant difference in values among segments”.
- Post-hoc analysis (“There was a significant difference in values between segments from (Q1-Q9)).

```
> #Reading the Excel file.
> #Harsh3 <- read_excel(file.choose())
> #View(Harsh3)
> #head(Harsh3)
> #Converting the Segments into factor for analysis.
> Harsh3$Segment<-ordered(as.factor(Harsh3$Segment))
> bartlett.test( Medians~ Segment, data = Harsh3)

	Bartlett test of homogeneity of variances

data:  Medians by Segment
Bartlett's K-squared = 9.5098, df = 8, p-value = 0.3011

> 
```


Here one of the assumptions for Bartlett test is the p-value needs to be greater than 0.05. For our test the p-value is 0.3011 which is greater than 0.05.

- The below bar plot shows the count of median ranks for different segments.



- Test

```
> #Kruskal-Wallis test
> kruskal.test( Medians~ Segment, data = Harsh3)

        Kruskal-Wallis rank sum test

data:  Medians by Segment
Kruskal-Wallis chi-squared = 19.615, df = 8, p-value = 0.01189
```

We reject the Null Hypothesis by stating that p value is less than 0.1 alpha level and conclude that the segments have different distributions from the population.

- **Effect size**

Here We found out that the p-value for Kruskal-Wallis test is significant. So, after

running the Effect size Statistics for the Kruskal–Wallis test provides us the degree to which one group has data with higher ranks than another group. They are related to the probability that a value from one group will be greater than a value from another group.

An effect size of 1 indicates that the measurements for each group are entirely greater or entirely less than some other group, and an effect size of 0 indicates that there is no effect; that is, that the segments are absolutely stochastically equal.

Interpretations for Vargha and Delaney's *A* and Cliff's *delta* come from Vargha and Delaney (2000).

Maximum Vargha and Delaney's *A* or Cliff's *delta*

Here, the *multiVDA* function is used to calculate Vargha and Delaney's *A* (VDA), Cliff's *delta* (CD), and *r* between all pairs of segments. The function identifies the comparison with the most extreme VDA statistic **with 1 for [(Q2-Q8)].** That is, it identifies the most **disparate** segments.


```
$comparison
Comparison
 "Q2 - Q8"

$statistic
VDA
  1

$statistic.m
VDA.m
    1
```

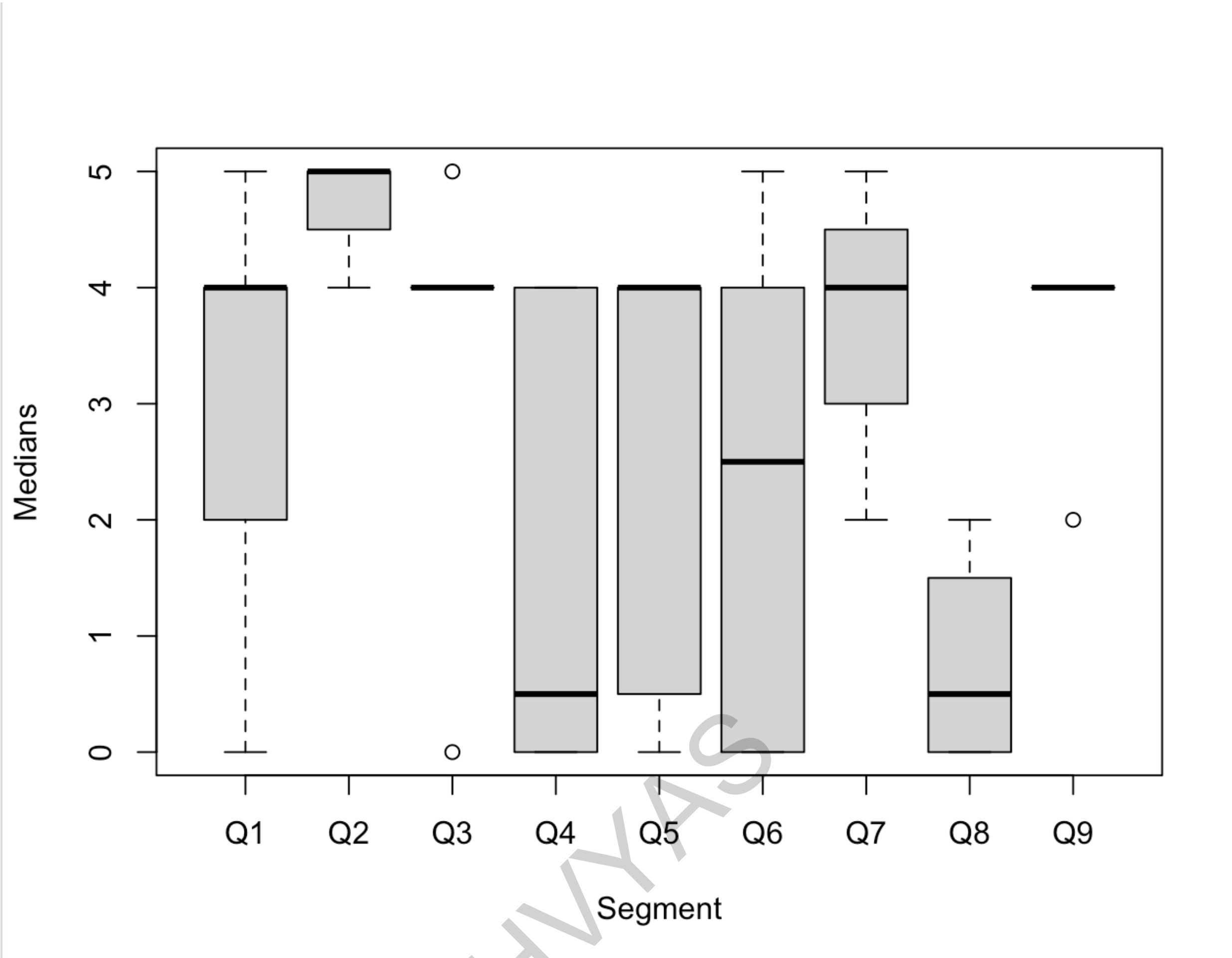
```
> library(rcompanion)
> library(coin)
> multiVDA(x = Harsh3$Medians,
+          g = Harsh3$Segment)
$pairs
   Comparison   VDA     CD      rg VDA.m  CD.m   rg.m
1     Q1 - Q2 0.141 -0.718 -0.7190 0.859 0.718 0.7190
2     Q1 - Q3 0.425 -0.150 -0.1500 0.575 0.150 0.1500
3     Q1 - Q4 0.731  0.462  0.4620 0.731 0.462 0.4620
4     Q1 - Q5 0.598  0.196  0.1960 0.598 0.196 0.1960
5     Q1 - Q6 0.625  0.250  0.2500 0.625 0.250 0.2500
6     Q1 - Q7 0.393 -0.214 -0.2140 0.607 0.214 0.2140
7     Q1 - Q8 0.859  0.718  0.7190 0.859 0.718 0.7190
8     Q1 - Q9 0.450 -0.100 -0.1000 0.550 0.100 0.1000
9     Q2 - Q3 0.800  0.600  0.6000 0.800 0.600 0.6000
10    Q2 - Q4 0.962  0.924  0.9250 0.962 0.924 0.9250
11    Q2 - Q5 0.929  0.858  0.8570 0.929 0.858 0.8570
12    Q2 - Q6 0.906  0.812  0.8120 0.906 0.812 0.8120
13    Q2 - Q7 0.768  0.536  0.5360 0.768 0.536 0.5360
14    Q2 - Q8 1.000  1.000  1.0000 1.000 1.000 1.0000
15    Q2 - Q9 0.900  0.800  0.8000 0.900 0.800 0.8000
16    Q3 - Q4 0.760  0.520  0.5200 0.760 0.520 0.5200
17    Q3 - Q5 0.657  0.314  0.3140 0.657 0.314 0.3140
18    Q3 - Q6 0.675  0.350  0.3500 0.675 0.350 0.3500
19    Q3 - Q7 0.471 -0.058 -0.0571 0.529 0.058 0.0571
20    Q3 - Q8 0.850  0.700  0.7000 0.850 0.700 0.7000
21    Q3 - Q9 0.560  0.120  0.1200 0.560 0.120 0.1200
22    Q4 - Q5 0.364 -0.272 -0.2710 0.636 0.272 0.2710
23    Q4 - Q6 0.396 -0.208 -0.2080 0.604 0.208 0.2080
24    Q4 - Q7 0.179 -0.642 -0.6430 0.821 0.642 0.6430
25    Q4 - Q8 0.588  0.176  0.1750 0.588 0.176 0.1750
26    Q4 - Q9 0.200 -0.600 -0.6000 0.800 0.600 0.6000
27    Q5 - Q6 0.524  0.048  0.0476 0.524 0.048 0.0476
28    Q5 - Q7 0.286 -0.428 -0.4290 0.714 0.428 0.4290
29    Q5 - Q8 0.732  0.464  0.4640 0.732 0.464 0.4640
30    Q5 - Q9 0.343 -0.314 -0.3140 0.657 0.314 0.3140
31    Q6 - Q7 0.286 -0.428 -0.4290 0.714 0.428 0.4290
32    Q6 - Q8 0.729  0.458  0.4580 0.729 0.458 0.4580
33    Q6 - Q9 0.317 -0.366 -0.3670 0.683 0.366 0.3670
34    Q7 - Q8 0.964  0.928  0.9290 0.964 0.928 0.9290
35    Q7 - Q9 0.571  0.142  0.1430 0.571 0.142 0.1430
36    Q8 - Q9 0.025 -0.950 -0.9500 0.975 0.950 0.9500
```

Post-hoc test: Dunn test for multiple comparisons of segments

Here the Kruskal–Wallis test is significant, a post-hoc analysis can be performed to determine which segments differ from each other group. The most popular post-hoc test for the Kruskal–Wallis test is the Dunn test. Because the post-hoc test will produce multiple *p*-values, adjustments to the *p*-values can be made to avoid inflating the possibility of making a type-I error. There are a variety of methods for controlling the familywise error rate or for controlling the false discovery rate.

When there are many *p*-values to evaluate, it is useful to condense a table of *p*-values to a compact letter display format. In the output, segments are separated by letters. **Segments sharing the same letter are not significantly different.** Compact letter displays are a clear and succinct way to present results of multiple comparisons.

```
boxplot(Harsh3$Medians~Harsh3$Segment,xlab = "Segment",ylab = "Medians")
```


Medians
0
1
2
3
4
5
Q1
Q2
Q3
Q4
Q5
Q6
Q7
Q8
Q9
Segment

```
> ##Dunn's Test
> DT=dunnTest(Medians ~ Segment, data = Harsh3,method="bh")
Warning messages:
1: In if (tmp$Eclass != "factor") { :
  the condition has length > 1 and only the first element will be used
2: In if (tmp$Eclass == "numeric") STOP("RHS variable must be a factor.") :
  the condition has length > 1 and only the first element will be used
3: Segment was coerced to a factor.
> DT
Dunn (1964) Kruskal-Wallis multiple comparison
  p-values adjusted with the Benjamini-Hochberg method.

   Comparison           Z       P.unadj       P.adj
1     Q1 - Q2 -1.86321279 0.0624323345 0.22475640
2     Q1 - Q3 -0.36458835 0.7154187105 0.83080883
3     Q2 - Q3  1.39103227 0.1642156421 0.36948519
4     Q1 - Q4  1.68224606 0.0925211368 0.25621238
5     Q2 - Q4  3.27740597 0.0010476562 0.01885781
6     Q3 - Q4  1.83634341 0.0663068836 0.21700435
7     Q1 - Q5  0.68805460 0.4914183998 0.61003663
8     Q2 - Q5  2.38851697 0.0169165271 0.12179900
9     Q3 - Q5  0.96312700 0.3354837935 0.52510507
10    Q4 - Q5 -0.89661619 0.3699237564 0.51220212
11    Q1 - Q6  0.92347905 0.3557576105 0.53363642
12    Q2 - Q6  2.70630989 0.0068035524 0.08164263
13    Q3 - Q6  1.18235392 0.2370653006 0.40639766
14    Q4 - Q6 -0.87919960 0.3792930634 0.50572408
15    Q5 - Q6  0.13752568 0.8906152987 0.91606145
16    Q1 - Q7 -0.70811450 0.4788741564 0.61569534
17    Q2 - Q7  1.23566828 0.2165818930 0.38984741
18    Q3 - Q7 -0.27092380 0.7864496432 0.83271139
19    Q4 - Q7 -2.36288702 0.0181331965 0.10879918
20    Q5 - Q7 -1.35183492 0.1764281318 0.37361251
21    Q6 - Q7 -1.65685696 0.0975483981 0.25083874
22    Q1 - Q8  1.95815357 0.0502119958 0.20084798
23    Q2 - Q8  3.30940034 0.0009349604 0.03365858
24    Q3 - Q8  2.09738199 0.0359597744 0.18493598
25    Q4 - Q8  0.67808399 0.4977184386 0.59726213
26    Q5 - Q8  1.34499014 0.1786284194 0.33845385
27    Q6 - Q8  1.34686120 0.1780249392 0.35604988
28    Q7 - Q8  2.49783883 0.0124952990 0.11245769
29    Q1 - Q9 -0.31359697 0.7538271576 0.82235690
30    Q2 - Q9  1.43436661 0.1514676720 0.36352241
31    Q3 - Q9  0.04596301 0.9633397349 0.96333973
32    Q4 - Q9 -1.78326990 0.0745423788 0.22362714
33    Q5 - Q9 -0.91348128 0.3609894755 0.51982484
```

```
33    Q5 - Q9 -0.91348128 0.3609894755 0.51982484
34    Q6 - Q9 -1.12774173 0.2594290103 0.42452020
35    Q7 - Q9  0.32056952 0.7485366413 0.84210372
36    Q8 - Q9 -2.05404765 0.0399710803 0.17986986
```

Adjusted p-values of different pair of Segments which are < 0.1 are statistically significant.

(alpha=0.1).

Refer to Boxplot for matching the segments by similar alphabets.

Post-hoc analysis:

After Performing Dunn's test, we found out the test result for differences between segments. (Refer to Boxplot) to see whether the median of one group is greater or lesser than the corresponding group.
Here this are the segments having significant p values that are(Q2-Q4), (Q2-Q5), (Q2-Q6), (Q4-Q7), (Q2-Q8),and (Q7-Q8).

Q1) DIVERSITY IN POLICIES AND PROCEDURES,
Q2) INCLUSIVITY BY POLICIES AND PRACTICES,
Q3) DIVERSITY BY RECRUITMENT PROCESS
Q4) DIVERSITY IN POLICIES AND PROCEDURES,
Q5) INCLUSIVITY BY POLICIES AND PRACTICES,
Q6) WORKPLACE INCLUSION AND DIVERSITY BY RECRUITMENT PROCESS,
Q7) ORGANIZATION'S CULTURE,
Q8) INCLUSION BY DIFFERENT NEEDS
Q9) WORKPLACE INCLUSION AND DIVERSITY BY CAREER DEVELOPMENT AND EMPLOYEE RETENTION