

Basic Course on R: Distribution-Free ANOVA

Elizabeth Ribble*

28 Oct - 1 Nov 2019

Contents

1	Kruskal-Wallis Test	2
2	Friedman's Test	5
3	Summary of Procedures	8

*emcclel3@msudenver.edu

1 Kruskal-Wallis Test

The single-factor ANOVA model for comparing population or treatment means assumed that for all groups, random samples were drawn from normal populations each having the same variance. This normality assumption is required for a valid F test, but the next procedure for testing the equality of the centers of the distributions only requires that the populations have the same continuous distribution.

The null hypothesis is that all of the group centers are the same and the alternative is that at least one of the group centers is different. The Kruskal-Wallis test examines the validity of these hypotheses by working on ranks of the data without assuming the data come from a specific distribution.

The procedure starts by ranking all the data together on the assumption of the null, that if the centers of the groups are equal, the ranks from all the groups will be intermingled. If the null is false, then some samples will consist mostly of observations having small ranks in the combined sample, whereas others will consist mostly of observations having large ranks.

The Kruskal-Wallis test statistic is a measure of the extent to which the sums of the ranks within each group deviate from their common expected value, and the null is rejected if the computed value of the statistic indicates too great a discrepancy between observed and expected rank averages.

Example. The accompanying observations on axial stiffness index resulted from a study of metal-plate connected trusses in which five different plate lengths (4 in., 6 in., 8 in., 10 in., and 12 in.) were used.

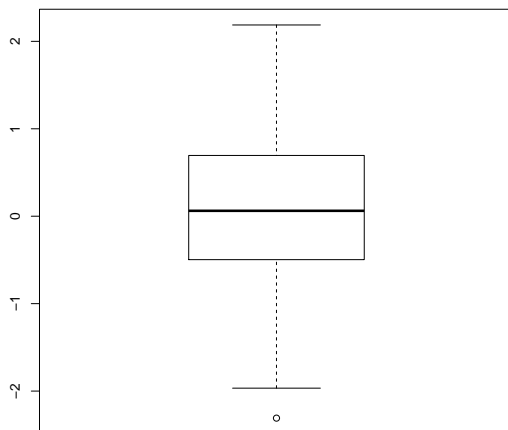
	<i>i</i> = 1 (4"):	309.2	309.7	311.0	316.8	326.5	349.8	409.5		
	<i>i</i> = 2 (6"):	331.0	347.2	348.9	361.0	381.7	402.1	404.5		
	<i>i</i> = 3 (8"):	351.0	357.1	366.2	367.3	382.0	392.4	409.9		
	<i>i</i> = 4 (10"):	346.7	362.6	384.2	410.6	433.1	452.9	461.4		
	<i>i</i> = 5 (12"):	407.4	410.7	419.9	441.2	441.8	465.8	473.4		
									r_i	\bar{r}_i
<i>Ranks</i>	<i>i</i> = 1:	1	2	3	4	5	10	24	49	7.00
	<i>i</i> = 2:	6	8	9	13	17	21	22	96	13.71
	<i>i</i> = 3:	11	12	15	16	18	20	25	117	16.71
	<i>i</i> = 4:	7	14	19	26	29	32	33	160	22.86
	<i>i</i> = 5:	23	27	28	30	31	34	35	208	29.71

Here we read in the data and note how a boxplot of normally distributed data looks:

```
axial <- read.csv("stiffness.csv", header=T)
attach(axial)

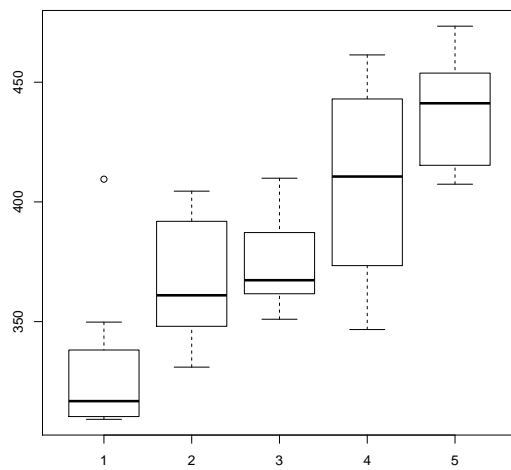
## The following objects are masked from axial (pos = 10):
##
## lengths, stiffness
## The following objects are masked from axial (pos = 20):
##
## lengths, stiffness
## The following objects are masked from axial (pos = 50):
##
## lengths, stiffness

set.seed(123)
boxplot(rnorm(100))
```



If we look at the boxplot of the stiffness values for each length, we see that the means are probably different and that the boxplots don't look very normal:

```
boxplot(stiffness~lengths)
```



The output of the Kruskal-Wallis test confirms that the means are indeed different. The p -value is quite small, indicating at least two of the group centers are different:

```
kruskal.test(stiffness, lengths)

##
##  Kruskal-Wallis rank sum test
##
## data:  stiffness and lengths
## Kruskal-Wallis chi-squared = 20.122, df = 4, p-value = 0.0004724
```

2 Friedman's Test

In many cases we have two factors of interest instead of just one, as in the two-way ANOVA we've already looked at, but for one of the factors we know that the observations within groups will be more homogeneous (less variable) than between the groups with respect to the response values.

In this case, the levels of the factors are called **blocks** and we treat it differently from the other factor, since the presence or absence of a statistically significant result may be due to this extraneous variation rather than to the presence or absence of factor effects. Here we will not consider any interactions between the two factors, and we will not test the blocks because we already know they should have differing means.

The purpose of utilizing blocks is to filter out the variation due to the blocks. For example, a single subject can receive all treatments and the blocking is done to control for variability between subjects (and each subject acts as its own control). These are called “repeated-measures” designs: the units within a block are different instances of a treatment application. Blocks can also be taken as different time periods, locations, or observers.

Let's take a look at an example using data on tennis rackets and the speed of the ball coming of the racket. In one experiment four different string tensions were used, and balls projected from a machine were hit by 18 different players. The rebound speed (km/h) was determined for each tension-player combination. Consider the following data from a similar experiment with 6 players.

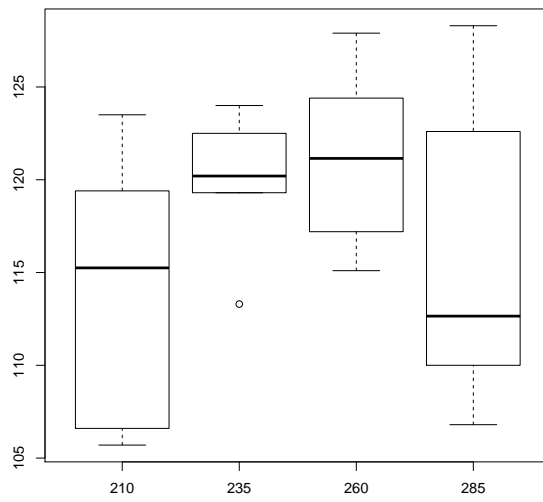
```
tennis <- read.csv("tennis.csv", header=T)
head(tennis)

##      tension player speed
## 1         210      1 105.7
## 2         210      2 116.6
## 3         210      3 106.6
## 4         210      4 113.9
## 5         210      5 119.4
## 6         210      6 123.5

attach(tennis)

## The following objects are masked from tennis (pos = 10):
##
##      player, speed, tension
## The following objects are masked from tennis (pos = 20):
##
##      player, speed, tension
## The following objects are masked from tennis (pos = 50):
##
##      player, speed, tension

boxplot(speed~tension)
```



So we see there are possibly varying mean speeds between the different tensions (and the speeds are not necessarily normal) and we would like to filter out the person-to-person variability.

Friedman's test ranks each observation within the blocks (here, people) and then the average rank is computed for the factor of interest (in this case, racket tension). When the null hypothesis that all of the tension median speeds are the same, the average ranks should be close to one another, but when the null is false, they will be far from one another. Friedman's test statistic measure the discrepancy between the expected value of each rank average under the null and the observed rank averages.

In R, the function is `friedman.test()` with arguments `y` (response), `groups` (factor of interest), and `blocks` (blocking factor):

```
friedman.test(speed, tension, player)

##
##  Friedman rank sum test
##
## data:  speed, tension and player
## Friedman chi-squared = 7.8, df = 3, p-value = 0.05033
```

At a 10% significance level, we are able to say the data provide sufficient evidence that there is a difference in mean speed between the rackets because our p -value is less than 0.10. Note what happens if we ignore the blocks and use Kruskal-Wallis:

```
kruskal.test(speed, tension)

##
##  Kruskal-Wallis rank sum test
##
## data:  speed and tension
## Kruskal-Wallis chi-squared = 3.9684, df = 3, p-value = 0.2649
```

The p -value is much larger without using blocks.

3 Summary of Procedures

Here is a table that summarizes when to use each of the procedures we've covered:

	Test one mean	Compare 2 means	Compare at least 2 means (one-way)	Compare at least 2 means (two-way)
Population(s) assumed normally distributed	one-sample t -test	two-sample t -test	ANOVA (F -test)	ANOVA (F -test)
Distribution-free	Signed rank test	Wilcoxon rank sum test	Kruskal-Wallis rank sum test	Friedman rank sum test