

Homework 9

Stat 135: Concepts of Statistics
UC Berkeley, Fall 2022

Due: **November 10th, 11:59PM**

- Read sections 11.4.8, 12.1 and 12.2 of the textbook. Go over the slides on ANOVA and multiple testing.
 - Your homework must be submitted to Gradescope **as a single PDF file**.
 - To obtain full credit, please **write your answers clearly and show your reasoning**.
-

Problem 1

One researcher collected data to see whether there exists difference in the energy use of three gas ranges across seven different days. He was in the process of performing the ANOVA analysis when his laptop crashed and all his data were erased. Help him restore the lost information and complete the following ANOVA table for him (fill in every "?" cell in the table, not just the p-value). Show your work and the R code you used.

Source	df	SS	MS	F	p-value
Between	?	64.2	?	?	?
Within	?	?	2.39		
Total	20	?	?		

$df(\text{between}) = \text{number of groups} - 1 = 3 - 1 = 2$

$df(\text{within}) = df(\text{total}) - df(\text{between}) = 20 - 2 = 18$

$SS(\text{within}) = df(\text{within}) * MS(\text{within}) = 18 * 2.39 = 43.02$

$SS(\text{total}) = SS(\text{within}) + SS(\text{between}) = 64.2 + 43.02 = 107.22$

$MS(\text{total}) = SS(\text{total}) / df(\text{total}) = 107.22 / 20 = 5.361$
 $MS(\text{between}) = SS(\text{between}) / df(\text{between}) = 64.2 / 2 = 32.10$

$F = MS(\text{between}) / MS(\text{within}) = 32.10 / 2.39 = 13.43$
 $p\text{-value} = \text{fdist}(13.43, 2, 18) = 0.0003$

Completed Chart:

Source	df	SS	MS	F	p-value
Between	2	64.2	32.10	13.43	0.0003
Within	18	43.02	2.39		
Total	20	107.22	5.361		

Problem 2

A researcher did an experiment where they randomly placed each of 21 men on a low, medium, or high intensity exercise regimen. The weight loss of each participant was recorded in the following table and can be found in the file `weight_loss.csv`.

Low	Medium	High
6.89	8.54	10.39
7.87	8.32	14.22
10.26	8.76	19.91
5.70	9.30	12.78
6.82	10.45	16.33
9.23	8.94	11.21
11.20	10.37	14.55

Denote the population average weight loss from the low, medium, and high exercise regimens respectively by μ_1, μ_2, μ_3 .

- (a) Calculate the sums of squares and fill out an ANOVA table to test $H_0 : \mu_1 = \mu_2 = \mu_3$ against the alternative that at least one of μ_1, μ_2, μ_3 is not equal to the others. Explicitly show how to compute each term in the ANOVA table, and check your answers with the `aov()` function. Provide a conclusion for your test.

```

```{r}
low = c(6.89, 7.87, 10.26, 5.7, 6.82, 9.23, 11.2)
mid = c(8.54, 8.32, 8.76, 9.3, 10.45, 8.94, 10.37)
high = c(10.39, 14.22, 19.91, 12.78, 16.33, 11.21, 14.55)
total = c(6.89, 7.87, 10.26, 5.7, 6.82, 9.23,
11.2, 8.54, 8.32, 8.76, 9.3, 10.45, 8.94, 10.37, 10.39, 14.22, 19.91, 12.78, 16.33, 11.21, 14.55)

```

```{r}
low_mean = c(6.89, 7.87, 10.26, 5.7, 6.82, 9.23, 11.2)-mean(low)
mid_mean = c(8.54, 8.32, 8.76, 9.3, 10.45, 8.94, 10.37)-mean(mid)
high_mean = c(10.39, 14.22, 19.91, 12.78, 16.33, 11.21, 14.55) -mean(high)
SSw= sum(low_mean^2+ mid_mean^2+high_mean^2)
SSb=((mean_low-mean_total)^2+(mean_mid-mean_total)^2+(mean_high-mean_total)^2)*7
MSb=SSb/2
MSw=SSw/18
f=MSb/MSw
```

```

#calculate sum of squares

| | sum | avg | variance | count |
|--------|-------|----------|-----------|-------|
| low | 57.97 | 8.281429 | 4.039581 | 7 |
| medium | 64.68 | 9.24 | 0.7335667 | 7 |
| high | 99.39 | 14.19857 | 10.45608 | 7 |

Anova

| sources of variability | SS | df | MS | F | p-value |
|------------------------|----------|-----------|-------------------------------|---------------------------------------|---------|
| between | 141.2109 | 2 | $141.2109/2$
$= 70.60534$ | MS_B / MS_W | |
| within | 91.37539 | $21-3=18$ | $91.37539/18$
$= 5.076409$ | $= 70.60534/5.076409$
$= 13.90852$ | |
| total | | | | | |

Checking with aov() function:

```

{r}
val <- c(6.89, 7.87, 10.26, 5.7, 6.82, 9.23,
11.2, 8.54, 8.32, 8.76, 9.3, 10.45, 8.94, 10.37, 10.39, 14.22, 19.91, 12.78, 16.33, 11.21, 14.55)
type <- c('low', 'low', 'low', 'low', 'low', 'low', 'low', 'mid', 'mid', 'mid', 'mid', 'mid', 'mid', 'mid',
'high', 'high', 'high', 'high', 'high', 'high', 'high', 'high')
aov.model <- aov(val ~ type)
summary(aov.model)

```

```

      type      Df Sum Sq Mean Sq F value    Pr(>F)
Residuals  18   91.38     5.08
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

The **conclusion** of the test is that the p-value is statistically significant to reject the null that there is no difference in the mean of low, medium, and high intensity exercise regimen.

(b) What is the problem if you were to do multiple pairwise t-tests instead of ANOVA?

Solution: Every time when a t-test is performed, there is a chance that the test makes a Type I error. By running two t-test, this increases the chance of making a error to 10%. Additionally, the new error rate of multiple t-tests is not necessarily equal to multiplying the chance of type I error, which is typically 5%, by the number of tests. Instead, an ANOVA test controls the Type I error to remain at the certain

level. Therefore, any statistically significant result you get from an ANOVA test is not the result of running many rounds of tests.

- (c) If a Bonferroni simultaneous hypothesis test uses a significance level of α/K where K is the number of tests used, adapt this idea to compute Bonferroni simultaneous 95% confidence intervals for μ_1, μ_2, μ_3 .

Solution:

Critical Value is $t_{0.025/3, df=21-3=18} = t_{0.008333, df=18} = 2.6393$.

The margin of error for each of the three intervals are (critical value)*(sqrt(MS within groups)/(number of samples in each group))

$$= 2.6393 * (\sqrt{5.0076409/7}) = 2.2476.$$

Then, the corresponding three confidence intervals are:

$$(a) \mu_{low} : (\bar{X}_{low} \pm 2.2476) = (8.281429 \pm 2.2476) = (6.0338, 10.5290)$$

$$(b) \mu_{mid} : (\bar{X}_{mid} \pm 2.2476) = (9.24 \pm 2.2476) = (6.9924, 11.4876)$$

$$(c) \mu_{high} : (\bar{X}_{high} \pm 2.2476) = (14.19857 \pm 2.2476) = (11.95097, 16.44617)$$

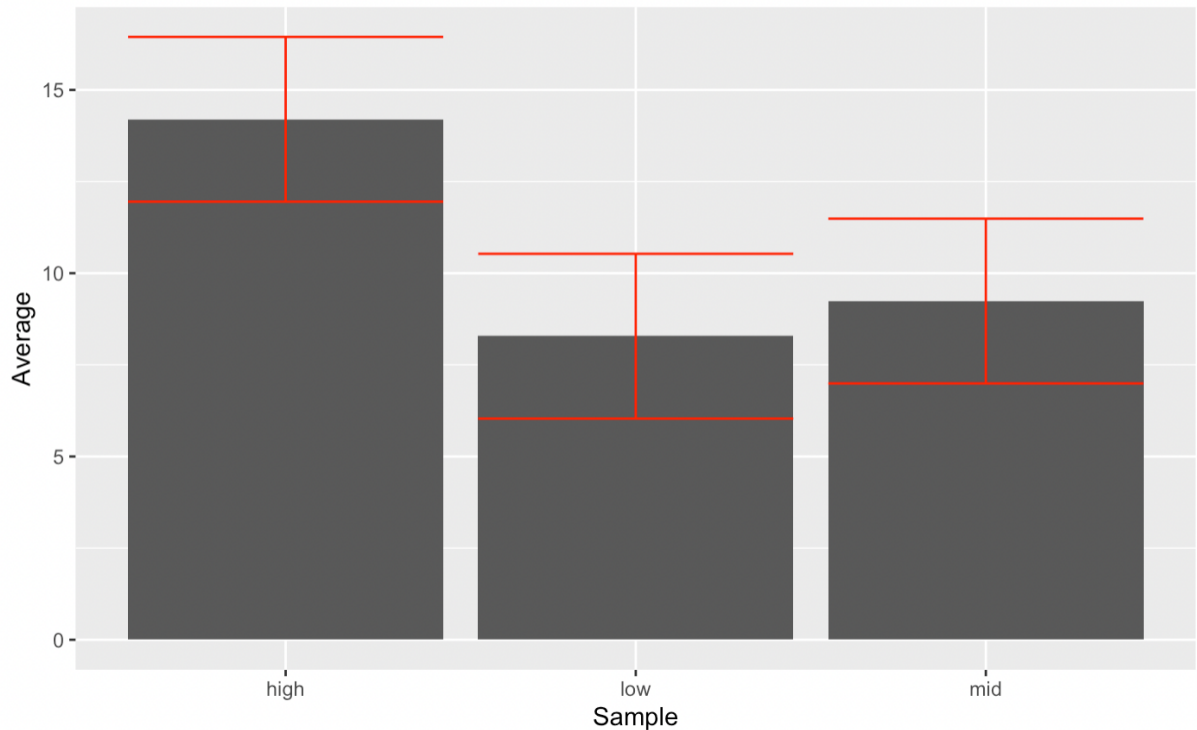
- (d) Create a visualization that compares your confidence intervals from part (c) using `geom_errorbar()`. What would you expect to see if the null hypothesis from part (a) were true? Does your visualization agree or disagree with your conclusion from part (a)?

Solution: If the null hypothesis from part (a) were true, then there should be an overlap between all three confidence intervals found in (c).

```

library(ggplot2)
df <- data.frame(
  Sample=c("low", "mid", "high"),
  Average=c(8.281429, 9.24, 14.19857),
  SE = c(2.2476, 2.2476, 2.2476) )
ggplot(data = na.omit(df)) + #don't bother plotting the NA
  geom_bar(stat = "identity", aes(x = Sample, y = Average)) +
  geom_errorbar(
    aes(x=Sample,
        ymin = Average - SE,
        ymax = Average + SE),
    color = "red"
  )

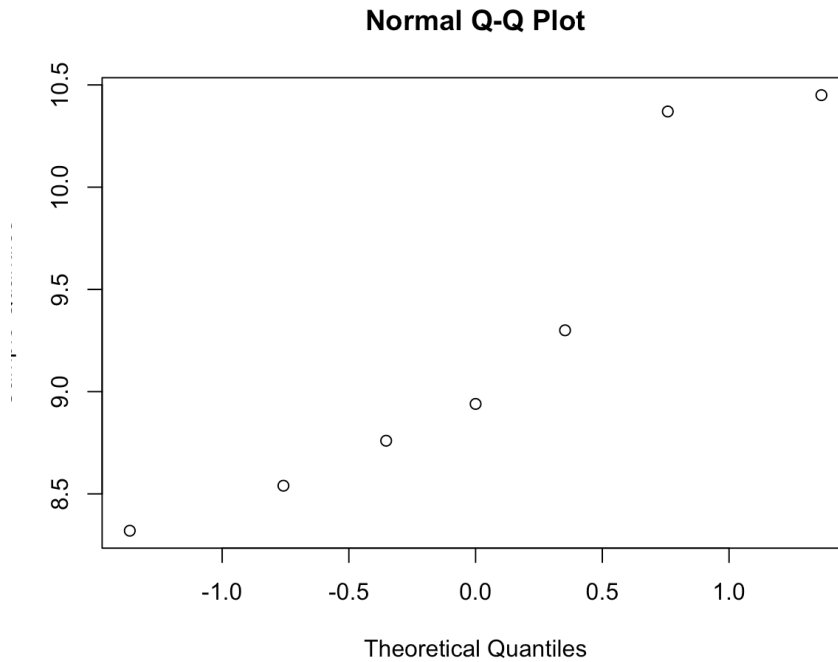
```



Low and Medium's error bars overlap while high's error bar do not seem to overlap with low's. Therefore, the mean of high is different from low and medium's mean.

- (e) Investigate the validity of the ANOVA normal assumption for this data. Do you think the normal assumption is reasonable? Regardless of your answer, perform an appropriate non-parametric test. Do you reach the same conclusion?

Solution: Assuming the normality of the data is inappropriate because the sample size for each low, medium, high population is small. Furthermore, when we plot the qqplot of each low, medium, and high, we can see that the data do not follow the linear line. For example, the image attached below is the qqplot of medium. We can see that the data do not follow a linear trend line, which shows that the normality assumption cannot be made.



Now, I perform a **non-parametric** test.

| Low | Medium | High |
|-----|--------|------|
| 3 | 6 | 13 |
| 4 | 5 | 18 |
| 11 | 7 | 21 |
| 1 | 10 | 17 |
| 2 | 14 | 20 |
| 9 | 8 | 16 |
| 15 | 12 | 19 |

- mean Low rank: 6.428571
- mean Medium rank: 8.857143
- mean High rank: 17.71429

$$\text{K-statistics:} = \frac{12}{21(22)}(7(6.428571^2 + 8.857143^2 + 17.71429^2)) - 66 = 78.83118 - 66 = 12.83118$$

We can now compute the p-value and decide whether to reject or not with the correct Type I error:

p-value = $P(K \geq 12.83118) = 0.0016359$, where $K \sim \chi^2_2$. Since p-value is statistically significant, we reject the null that each sample comes from populations with the same distribution. Both **parametric and nonparametric** reach the same conclusion that reject the null hypothesis.

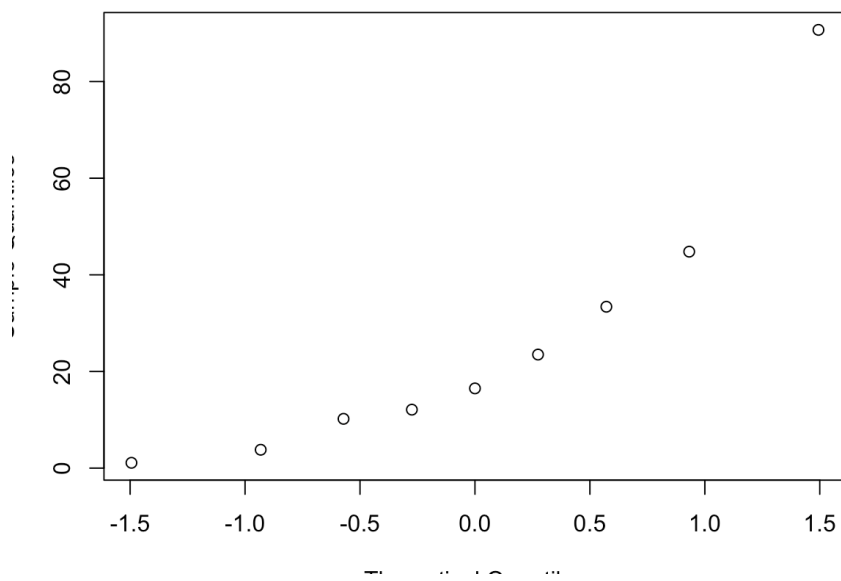
Problem 3

Samples of three different types of phones are tested for how many drops they can experience before breaking. Test whether this is a significance difference among the types, and if there is, determine which types are significantly different. Use both a parametric and a non-parametric technique. Explain which test is more appropriate. (You need to enter the data into R by yourself)

| Type A | Type B | Type C |
|--------|--------|--------|
| 1.1 | 10.3 | 13.4 |
| 3.8 | 18.3 | 20.8 |
| 10.2 | 25.2 | 22.3 |
| 12.1 | 43.9 | 31.5 |
| 16.5 | 46.2 | 44.2 |
| 23.5 | 60.3 | |
| 33.4 | | |
| 44.8 | | |
| 90.7 | | |

Which Test is Appropriate: Non-parametric vs Parametric Non-parametric test is appropriate because the sample size of Type C is very small and assuming normality is not appropriate. Furthermore, since we are not given the underlying distribution of these samples, it is not appropriate to use parametric test. I have plotted the QQplot for each group, but the QQplot of small data set is not accurate. Firstly, here is the qqnorm plot of Type A.

Normal Q-Q Plot



We can see that the data do not follow a linear trend, suggesting that the normality cannot be assumed about the data. Since at least one of the data do not follow normality, it is inappropriate to use parametric test. Type B and Type C's dataset are so small that plotting them on a qqplot would not give too much insight into the normality of the data.

Parametric test

For a-b pair:

```

{r}
a = c(1.1, 3.8, 10.2, 12.1, 16.5, 23.5, 33.4, 44.8, 90.7)
b = c(10.3,18.3,25.2,43.9,46.2,60.3)
c = c(13.4,20.8,22.3,31.5,44.2)
type <- c('a','a','a','a','a','a','a','a','a','b','b','b','b','b','b')
val_wob <- c(1.1, 3.8, 10.2, 12.1, 16.5, 23.5, 33.4, 44.8, 90.7,10.3,18.3,25.2,43.9,46.2,60.3)
aov.model <- aov(val_wob ~ type)
summary(aov.model)

```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|--------|
| type | 1 | 219 | 219.0 | 0.353 | 0.563 |
| Residuals | 13 | 8070 | 620.8 | | |

Since $p = 0.563$, p-value is not statistically significant, therefore we **fail to reject the null** that a and b have the same population mean.

For a-c pair:

```

{r}
a = c(1.1, 3.8, 10.2, 12.1, 16.5, 23.5, 33.4, 44.8, 90.7)
b = c(10.3,18.3,25.2,43.9,46.2,60.3)
c = c(13.4,20.8,22.3,31.5,44.2)
type <- c('a','a','a','a','a','a','a','a','a','c','c','c','c','c','c')
val_wob <- c(1.1, 3.8, 10.2, 12.1, 16.5, 23.5, 33.4, 44.8, 90.7,13.4,20.8,22.3,31.5,44.2)
aov.model <- aov(val_wob ~ type)
summary(aov.model)

```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|--------|
| type | 1 | 0 | 0.1 | 0 | 0.988 |
| Residuals | 12 | 6806 | 567.2 | | |

Since $p = 0.988$, p-value is not statistically significant, therefore we **fail to reject the null** that a and c have the same population mean.

For b-c pair:

```

{r}
a = c(1.1, 3.8, 10.2, 12.1, 16.5, 23.5, 33.4, 44.8, 90.7)
b = c(10.3,18.3,25.2,43.9,46.2,60.3)
c = c(13.4,20.8,22.3,31.5,44.2)
type <- c('b','b','b','b','b','b','c','c','c','c','c','c')
val_wob <- c(10.3,18.3,25.2,43.9,46.2,60.3,13.4,20.8,22.3,31.5,44.2)
aov.model <- aov(val_wob ~ type)
summary(aov.model)

```

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|-----------|----|--------|---------|---------|--------|
| type | 1 | 157.3 | 157.2 | 0.594 | 0.461 |
| Residuals | 9 | 2384.2 | 264.9 | | |

Since $p = 0.461$, p-value is not statistically significant, therefore we **fail to reject the null** that b and c have the same population mean.

For a-b-c pair:

```

> ```{r}
a = c(1.1, 3.8, 10.2, 12.1, 16.5, 23.5, 33.4, 44.8, 90.7)
b = c(10.3,18.3,25.2,43.9,46.2,60.3)
c = c(13.4,20.8,22.3,31.5,44.2)
type <- c('a','a','a','a','a','a','a','a','a','b','b','b','b','b',
val <- c(1.1, 3.8, 10.2, 12.1, 16.5, 23.5, 33.4, 44.8,
90.7,10.3,18.3,25.2,43.9,46.2,60.3,13.4,20.8,22.3,31.5,44.2)
aov.model <- aov(val ~ type)
summary(aov.model)
```

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
type	2	251	125.4	0.247	0.784
Residuals	17	8630	507.7		

Since  $p = 0.784$ , p-value is not statistically significant, therefore we **fail to reject the null** that a, b, and c have the same population mean.

## Non-Parametric test

For a-b pair:

```

3 > ```{r}
4 a = c(1.1, 3.8, 10.2, 12.1, 16.5, 23.5, 33.4, 44.8, 90.7)
5 b = c(10.3,18.3,25.2,43.9,46.2,60.3,NaN,NaN,NaN)
6 c = c(13.4,20.8,22.3,31.5,44.2,NaN,NaN,NaN,NaN)
7 #label <- c(rep("a",9), rep("b",9), rep("c",9))
8 #val <- c(a,b,c)
9 #df = data.frame(type=label, values =val)
0 label <- c(rep("a",9), rep("b",9))
1 val <- c(a,b)
2 df = data.frame(type=label, values =val)
3 kruskal.test(values ~type, data=df)
4 > ```

```

```

Kruskal-Wallis rank sum test

data: values by type
Kruskal-Wallis chi-squared = 1.3889, df = 1, p-value = 0.2386

```

Since  $p = 0.2386$ , p-value is not statistically significant, therefore we **fail to reject the null** that a and b have the same population mean.

For a-c pair:

```

```{r}
a = c(1.1, 3.8, 10.2, 12.1, 16.5, 23.5, 33.4, 44.8, 90.7)
b = c(10.3,18.3,25.2,43.9,46.2,60.3,NaN,NaN,NaN)
c = c(13.4,20.8,22.3,31.5,44.2,NaN,NaN,NaN,NaN)
#label <- c(rep("a",9), rep("b",9), rep("c",9))
#val <- c(a,b,c)
#df = data.frame(type=label, values =val)
label <- c(rep("a",9), rep("c",9))
val <- c(a,c)
df = data.frame(type=label, values =val)
kruskal.test(values ~type, data=df)
```

Kruskal-Wallis rank sum test

data: values by type
Kruskal-Wallis chi-squared = 0.36, df = 1, p-value = 0.5485

```

Since  $p = 0.5485$ , p-value is not statistically significant, therefore we **fail to reject the null** that a and c have the same population mean.

**For b-c pair:**

```

```{r}
a = c(1.1, 3.8, 10.2, 12.1, 16.5, 23.5, 33.4, 44.8, 90.7)
b = c(10.3,18.3,25.2,43.9,46.2,60.3,NaN,NaN,NaN)
c = c(13.4,20.8,22.3,31.5,44.2,NaN,NaN,NaN,NaN)
#label <- c(rep("a",9), rep("b",9), rep("c",9))
#val <- c(a,b,c)
#df = data.frame(type=label, values =val)
label <- c(rep("b",9), rep("c",9))
val <- c(b,c)
df = data.frame(type=label, values =val)
kruskal.test(values ~type, data=df)
```

Kruskal-Wallis rank sum test

data: values by type
Kruskal-Wallis chi-squared = 0.3, df = 1, p-value = 0.5839

```

Since  $p = 0.5839$ , p-value is not statistically significant, therefore we **fail to reject the null** that b and c have the same population mean.

**For a-b-c pair:**

```

3 }
4 a = c(1.1, 3.8, 10.2, 12.1, 16.5, 23.5, 33.4, 44.8, 90.7)
5 b = c(10.3, 18.3, 25.2, 43.9, 46.2, 60.3, NaN, NaN, NaN)
6 c = c(13.4, 20.8, 22.3, 31.5, 44.2, NaN, NaN, NaN, NaN)
7 label <- c(rep("a", 9), rep("b", 9), rep("c", 9))
8 val <- c(a, b, c)
9 df = data.frame(type=label, values=val)
10 kruskal.test(values ~ type, data=df)
11

```

```

 Kruskal-Wallis rank sum test

```

```

data: values by type
Kruskal-Wallis chi-squared = 1.4851, df = 2, p-value = 0.4759

```

Since  $p = 0.4759$ , p-value is not statistically significant, therefore we **fail to reject the null** that a, b, and c have the same population mean.

**Conclusion:** Both parametric and nonparametric test conclude that there are significant differences between each type.

## Problem 4

Show that for comparing two groups with the Kruskal-Wallis test is equivalent to the Mann-Whitney test.

## MANN whitney test $\leftrightarrow$ Krushal Wallis test

### i) small sample

#### Mann-Whitney:

- ① Sample 1: sample for which ranks are smaller  
sample 2: sample with larger rank
- ② count observations in sample 2 that are smaller than sample 1's observations. This count is  $U$ .

### ii) Large sample

| Mann-Whitney                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            | Kruskal - |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <ol style="list-style-type: none"> <li>① rank all observation, regardless of the group they are in</li> <li>② Add the ranks in the smaller group. Sum of all ranks should equal to <math>\frac{N(N+1)}{2}</math> where <math>N</math> is the total number of observations.</li> <li>③ <math>U</math>, the # of observations in sample 2 that are smaller than observation in sample 1, is <math>U = mn + \frac{m(m+1)}{2} - R_1</math> where <math>R_1</math> is the sum of ranks in sample 1.</li> </ol> <p>Then, <math>H</math> is <math>\frac{12}{N(N+1)}</math></p> |           |