Non - Parametric Methods

Correlation, Regression, t-tests, ANOVA etc. are all parametric methods because they involve estimation of a parameter with a statistic and the statistic is assumed to follow a theoretical distribution (e.g., normal distribution).
Non-parametric methods do not require a theoretical distribution.
Almost all non-parametric tests involve ranking the raw data to analyze the data as Ranks can reduce the effects of extreme values.

When to use non-parametric tests:

1) When our area of study is better represented by the median.
2) When data is not normally distributed. To know if our data is normally distributed or not, we will check skewness and kurtosis of the data. For normally distributed data, value of skewness and kurtosis will be 1.
3) When data has outliers that cannot be removed.
4) When data is not normally distributed, and sample size is small

We need large sample set for non-normally distributed to perform parametric tests. For example:

Parametric analysis	Sample size requirements for non-normal data
1-sample t test	Greater than 20
2- sample t test	Each group should have more than 15 observations
One way ANOVA	For 2-9 groups, each group should have more than 15 observations. For 10-12 groups, each group should have more than 20 observations.

So, in case if we have to use non-parametric tests, there are some alternatives for parametric tests.
1) Mann-Whitney Rank Sum test – alternative to independent samples t tests.
2) Wilcoxon matched-pairs signed ranks test – alternative to dependent samples t test.

3) Kruskal- Wallis test – alternative to one way ANOVA
4) Friedman rank sum test – alternative to one way repeated measures ANOVA

The Wilcoxon-Mann-Whitney Rank-Sum Test (Wilcoxon W or Mann-Whitney U test)

Use this test to compare differences between two independent groups when dependent variables are either ordinal or continuous. This method is extension of the parametric t test by replacing observations using their ranks. The idea is that if there is no group difference, then the ranks should be random in both groups.

Step1: Rank the scores ignoring the groups
For example- We have two groups and their scores.

Group1	Group2
12	13
17	19
14	20

Their rank table will be:

Rank1	Rank2	
1	2	
4	5	
3	6	
8	13	SUM

Step2: Identify all possible combinations
We will take all possible combinations of ranks for the three scores in group with smaller sum(group1) assuming H0 is true. Here, there is no group difference, so the ranks are random.

Table 20.1 Possible Rankings and Their Sum from First of Two Groups of Three Scores Each

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]	[,11]	[,12]
[1,]	1	1	1	1	1	1	1	1	1	1	2	2
[2,]	2	2	2	2	3	3	3	4	4	5	3	3
[3,]	3	4	5	6	4	5	6	5	6	6	4	5
Sum	6	7	8	9	8	9	10	10	11	12	9	10

	[,13]	[,14]	[,15]	[,16]	[,17]	[,18]	[,19]	[,20]
[1,]	2	2	2	2	3	3	3	4
[2,]	3	4	4	5	4	4	5	5
[3,]	6	5	6	6	5	6	6	6
Sum	11	11	12	13	12	13	14	15

figure (1)

Here, there are 20 possible combinations.

Step 3: Calculate a proportion

Let smaller sum from the rank table be T, so T=8
Among all these possible combinations, four combinations have the sum of the ranks<=T.
So, if the data were distributed randomly(i.e. no group effect), then 20% of the samples would have a T less than or equal to the one that we obtained.
The proportion times 2(2*20% = 40%) serves as an estimation of the p value for a two tailed test.
Apparently, 40% is higher than alpha, thus we cannot reject the null hypothesis and say that there is no group difference.

```
y <- c(12,17,14,13,19,20)
group <- c(1,1,1,2,2,2)

wilcox.test(y ~group, paired = FALSE, exact = TRUE)
```

```
>
> y <- c(12,17,14,13,19,20)
> group <- c(1,1,1,2,2,2)
> wilcox.test(y ~group, paired = FALSE, exact = TRUE)

        Wilcoxon rank sum exact test

data:  y by group
W = 2, p-value = 0.4
alternative hypothesis: true location shift is not equal to 0

>
```

Here, W = T – n1(n1+1)/2, n1 is the number of the observations in the smaller group.

Wilcoxon Matched-Pairs Signed Ranks Test

This is an alternative to correlated samples t test.
Suppose we have a group of children and we calculate the number of statements(inferences) in the stories written by them when they were young and when they were older.

Younger Children	Older Children	Difference	Rank
12	18	−6	−3
4	16	−12	−6
8	24	−16	−7
10	6	4	2
2	13	−11	−5
22	25	−3	−1
20	28	−8	−4
T+ = 2	T − = 26		

Figure (2)

R code:

```r
#Wilcoxon Matched-Pairs Signed Ranks Test

inference <- c(12,4,8,10,2,22,20,18,16,24,6,13,25,28)
Age <- rep(c(1,2), each=7)
wilcox.test(inference ~ Age, alternative="less",
            paired= TRUE, exact=TRUE)
```

```
>
>
> inference <- c(12,4,8,10,2,22,20,18,16,24,6,13,25,28)
> Age <- rep(c(1,2), each=7)
> wilcox.test(inference ~ Age, alternative="less",
+             paired= TRUE, exact=TRUE)

        Wilcoxon signed rank exact test

data:  inference by Age
V = 2, p-value = 0.02344
alternative hypothesis: true location shift is less than 0

>
```

V-statistic is the sum of ranks assigned to the differences with positive signs. For the given example, $p<0.05$. Therefore, we can reject null hypothesis.
From the figure (2), we can see that sum of the positive ranks (2) and the sum of negative ranks (26) of the difference scores between young and old conditions was significantly different (as $p = 0.02$), it means that older children gave more inferences in talking about a story.

Kruskal-Wallis One Way ANOVA

This test will rank the data assuming there is no group differences (null hypothesis) and then produce a chi-square test statistic.
If the test statistic is significant, then the null hypothesis can be rejected.

R code:

```r
#Kruskal-Wallis One Way ANOVA

group <- factor(c(1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3))
score <- c(50,0,1,0,50,60,44,73,85,51,63,85,85,66,69,61,54,80,47)
kruskal.test(x = score, g = group)
```

```
>
>
> group <- factor(c(1,1,1,1,1,1,1,2,2,2,2,2,2,2,2,3,3,3,3))
> score <- c(50,0,1,0,50,60,44,73,85,51,63,85,85,66,69,61,54,80,47)
> kruskal.test(x = score, g = group)

        Kruskal-Wallis rank sum test

data:  score and group
Kruskal-Wallis chi-squared = 11.359, df = 2, p-value = 0.003415

>
```

Here, we can see p value = 0.003415, therefore there is an overall group effect.

Friedman's Rank Test for K correlated Samples

This is an alternative to one way repeated measures ANOVA.
If there are no systematic changes over a series of trials, then some participants would have their best scores on trial 1, some on trial 2, and some on trial 3. So we will rank the repeated measures from each participant and run the analysis.

For example:
Foertsch and Gernsbacher (1997) examined the substitution of the genderless words "they" for "he" or "she". They asked participants to read sentences such as "A truck driver should never drove when sleepy, even if (he/she/they) may be struggling to make a delivery on time…." On some trials the words in the parentheses were replaced by gender-stereotype expected pronoun, sometimes by gender-stereotype unexpected pronoun, sometimes by they. The dependent variable used here is the reading time in milliseconds.

Participant	1	2	3	4	5	6	7	8	9	10	11
Expect He/ See She	50	54	56	55	48	50	72	68	55	57	68
Expect She/ See He	53	53	55	58	52	53	75	70	67	58	67
Neutral/ See They	52	50	52	51	46	49	68	60	60	59	60

R code:

```r
#Friedman's Rank Test for K correlated Samples

data <- matrix(c(50,53,52,54,53,50,56,55,52,55,58,51,
                 48,52,46,50,53,49,72,75,68,68,70,60,
                 55,67,60,57,58,59,68,67,60),
                 byrow = TRUE, ncol=3)
friedman.test(data)
```

```
>
> data <- matrix(c(50,53,52,54,53,50,56,55,52,55,58,51,
+                  48,52,46,50,53,49,72,75,68,68,70,60,
+                  55,67,60,57,58,59,68,67,60),
+                  byrow = TRUE, ncol=3)
> friedman.test(data)

        Friedman rank sum test

data:  data
Friedman chi-squared = 8.9091, df = 2, p-value = 0.01163

>
```

Here, p value is 0.01163 which is less than 0.05, therefore null hypothesis will be rejected.

SO, WHEN TO USE PARAMETRIC AND NON_PARAMETRIC METHODS:

Use Parametric test whenever possible, except for an extreme violation of an assumption of the parametric test.
As, Parametric test are more powerful, versatile, and robust when violation of assumptions is not severe.
And Non-parametric methods are robust even when violation of assumption is severe, but it is less powerful and versatile.

For R code:

Visit [Github]