NONPARAMETRIC STATISTICS

1. Nonparametric Statistics (NPS)
   - Name nonparametric indicates – no assumption about parameters (means, variances)
   - Require very few assumptions; it is distribution free
   - Use median as a measure of central tendency
     - Applied when
       - The data being analysed is ordinal or nominal
       - In case of interval or ratio scale data when no assumption can be made about the population probability distribution
     - Appropriate foe small samples that are not normally distributed
     - Computationally easier
     - Less efficient than parametric counter parts
     - Loose information by substituting ranks in place of scales

Parametric test	**Nonparametric test**
One-sample t test	Sign test for one sample
Paired t-test	1. Sign test for paired samples 2. Wilcoxon Signed-Ranked test for pair samples
Two independent sample t-test	Man-Whitney test (Wilcoxon Rank Sum Test)
One-way ANOVA	Kruskal-Wallis Test
ANOVA (randomized block design)	Friedman’s Test
Pearson’s correlation coefficient	Spearman’s Rank correlation coefficient

2. Sign Test for Matched
   - Observation are matched pairs but assumption underlying the paired t-test are not met, or the measurement scale is weak then Sign Test can be applied
   - Hypothesis
     - H_0: $\Delta_d = 0$ (the median of differences is zero)
     - H_A: $\Delta_d \neq 0$
   - T.S.: smallest of n+ and n-
   - RR: Reject H_0 if p-value is less than α (assumed alpha)
   - Procedure
     - Exclude the observations for which the difference (d_i) is zero
     - For $d_i > 0$ assign (+sign) and for $d_i <$ assign (-sign)

3. Wilcoxon Signed-Rank Test for paired samples
   - It is sophisticated than Sign test
   - Sign test only tell whether the sign of a difference is positive or negative
   - This test makes use of both the signs and magnitudes of the differences
   - Thus for a strong measurement scale the sign test may be undesirable since it would not make full use of the information contained in the data.
   - Assumption
     - The distribution of difference is continuous
     - The distribution of differences is symmetric
   - Hypothesis
     - H_0: $\Delta_d = 0$ (the median of differences is zero)
     - H_A: $\Delta_d \neq 0$
   - T.S: $T = \min(T+, |T-|)$
   - Rejection region

- Reject H_0 if $T \leq$ critical value or
- Reject H_0 if p-value is less than α (assumed alpha)

- Procedure
  - Calculate the differences of each pair of observations (d_i)
  - Ignore the signs of these differences
  - Rank the absolute values from smallest to largest
  - Assign the signs of the corresponding differences to these ranks
  - A difference of zero is not ranked, it is eliminated from the analysis and the sample size is reduced by one
  - Tied observation are assigned an average rank (suppose two smallest differences; 4,4; each one will get average rank (1+2)/2 = 1.5)
  - Assign each rank either a (+) or (-) sign corresponding to the sign of the difference
  - Compute sum of +ve ranks (T)+ and sum of –ve ranks (T-)
  - Choose the test statistics (smallest of T+, |T-|)

4. Wilcoxon Rank Sum Test (Mann-Whitney-U test)

- Counter part of t-test for two independent samples
- Assumptions
  - The two samples have been drawn independently and randomly from their respective populations.
  - The measurement scale is at least ordinal.
  - The distributions of the two populations have the same general shape. They differ only with respect to their medians.
- Hypothesis
  - H_0: the two populations are identical ($\Delta_1 = \Delta_2$)
  - H_A: population 1 and 2 have different medians ($\Delta_1 \neq \Delta_2$)
- Rejection rule:
  - Reject H_0 if p-value less than 0.05 (assumed α)

- Procedure
  - Select independent random samples from each population
  - Combine the two samples
  - Jointly rank the combined samples. If tied observation, assign an average to all with the same value
    - For example: if two observations are tied for the rank 3 and 4 each is given 3.5.
    - Next higher value receives a rank of 5 and so on
  - Label sample smaller sample size as sample 1. test statistic is the SUM of RANKS for sample 1, denote region from the table
  - Determine rejection region from the table

5. Kruskal-Wallis Test

- Counter part of One Way Analysis of Variance (ANOVA: comparing means of more than two groups) if:
  - Normality assumption of ANOVA not justified
  - Or the data available is ordinal (consist ranks)
- Assumption:
  - The samples are independent and random
  - The measurement scale is at least ordinal
  - The distribution of the values is sampled populations are identical except for the possibility that one or more of the population are composed of values that tend to be larger than those of other populations.
- Hypothesis
  - H_0: the two populations are all identical
  - H_A: At least one of the population tend to exhibit larger values than others
- Procedure
  - If no ties or moderate number of ties the formula simplifies to:

- Rejection region
  - When the samples sizes are large ($n_i \geq 5$) the test statistic T is distributed approximately as x^2 (t – 1)
  - Reject H_0 if T > x^2_α (t – 1)

6. Spearman's Rank correlation coefficient
- Nonparametric alternative of Pearson's coefficient of correlation
- Relevant when the measurement scale is at least ordinal or the relationship between two variables is not linear
- It is denoted by r_s
- r_s = 1 implies strictly increasing monotonicity
- r_s = -1 implies strictly decreasing monotonicity