Non-Parametric Test

Introduction

- T-tests: tests for the means of continuous data
  - One sample $H_0: \mu = \mu_0$ versus $H_A: \mu \neq \mu_0$
  - Two sample $H_0: \mu_1 - \mu_2 = 0$ versus $H_A: \mu_1 - \mu_2 \neq 0$
- Underlying these tests is the assumption that the data arise from a normal distribution
- T-tests do not actually require normally distributed data to perform reasonably well in most circumstances
- Parametric methods: assume the data arise from a distribution described by a few parameters (Normal distribution with mean μ and variance σ^2).
- Nonparametric methods: do not make parametric assumptions (most often based on ranks as opposed to raw values)
- We discuss non-parametric alternatives to the one and two sample t-tests.

Examples of when the parametric t-test goes wrong

- T-statistic

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- For two sample tests

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- In the first dataset
  - $s_1^2 = 9.2$, $s_2^2 = 17.5$
- In the second dataset
  - $s_1^2 = 9.2$, $s_2^2 = 2335$

When to use non-parametric methods

- With correct assumptions (e.g., normal distribution), parametric methods will be more efficient / powerful than non-parametric methods but often not as much as you might think[1]
- If the normality assumption grossly violated, nonparametric tests can be much more efficient and powerful than the corresponding parametric test
- Non-parametric methods provide a well-foundationed way to deal with circumstance in which parametric methods perform poorly.

Non-parametric methods

- Many non-parametric methods convert raw values to ranks and then analyze ranks
- In case of ties, midranks are used, e.g., if the raw data were 105 120 120 121 the ranks would be 1 2.5 2.5 4

Parametric Test	Nonparametric Counterpart
1-sample t	Wilcoxon signed-rank
2-sample t	Wilcoxon 2-sample rank-sum
k-sample ANOVA	Kruskal-Wallis
Pearson r	Spearman ρ

One sample tests: Wilcoxon signed rank

- In the pre-post analysis
  - D = pre - post
  - Retain the sign of D (+/-)
  - Rank = rank of $|D|$ (absolute value of D)
  - Signed rank, SR = Sign * Rank
  - Base analyses on SR
- Observations with zero differences are ignored
- Example: A pre-post study

Post	Pre	D	Sign	Rank of $\|D\|$	Signed Rank
3.5	4	0.5	+	1.5	1.5
4.5	4	-0.5	-	1.5	-1.5
4	5	1.0	+	4.0	4.0
3.9	4.6	0.7	+	3.0	3.0

One sample tests

- A good approximation to an exact P-value (not discussed) may be obtained by computing

$$z = \frac{\sum SR_i}{\sqrt{\sum SR_i^2}},$$

  where the signed rank for observation i is SR_i.
- We can then compare $|z|$ to the normal distribution.
- Here, $z = \frac{7}{\sqrt{29.5}} = 1.29$ and by `surfstat` the 2-tailed P-value is 0.197
- If all differences are positive or all are negative, the exact 2-tailed P-value is $\frac{1}{2^{n-1}}$
  - This implies that n must exceed 5 for any possibility of significance at the $\alpha = 0.05$ level for a 2-tailed test

Two sample WMW test

- The Wilcoxon–Mann–Whitney (WMW) 2-sample rank sum test is for testing for equality of central tendency of two distributions (for unpaired data)
- Ranking is done by combining the two samples and ignoring which sample each observation came from
- Example:

Females	120	118	121	119
Males	124	120	133	
Ranks for Females	3.5	1	5	2
Ranks for Males	6	3.5	7	

Two sample WMW test

- Doing a 2-sample t-test using these ranks as if they were raw data and computing the P-value against 4+3-2=5 d.f. will work quite well
- Loosely speaking the WMW test tests whether the population medians of the two groups are the same
- More accurately and more generally, it tests whether observations in one population tend to be larger than observations in the other
- Letting x_1 and x_2 respectively be randomly chosen observations from populations one and two, WMW tests $H_0 : C = \frac{1}{2}$, where $C = \text{Prob}[x_1 > x_2]$

Two sample WMW test

- Wilcoxon rank sum test statistic

$$W = R - \frac{n_1(n_1+1)}{2}$$

  where R is the sum of the ranks in group 1

- Under H_0, $\mu_W = \frac{n_1 n_2}{2}$ and $\sigma_W = \sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}$, and

$$z = \frac{W - \mu_W}{\sigma_W}$$

  follow a N(0,1) distribution.

Two sample WMW test

- The C index (*concordance probability*) may be estimated by computing

$$C = \frac{\bar{R} - \frac{n_1+1}{2}}{n_2},$$

where $\bar{R}$ is the mean of the ranks in group 1

- For the above data $\bar{R} = 2.875$ and $C = \frac{2.875-2.5}{3} = 0.125$
- We estimate: probability that a randomly chosen female has a value greater than a randomly chosen male is 0.125.

Summary: non-parametric tests

- Wilcoxon signed rank test: alternative to the one sample t-test
- Wilcoxon Mann Whitney or rank sum test: alternative to the two sample t-test
- Attractive when parametric assumptions are believed to be violated
- Drawback: if based on ranks, tests do not provide insight into effect size
- Non-parametric tests are attractive if all we care about is getting a P-value