Test for two related (paired) samples

- **Compare two mean values**
- **Non-parametric test**

Compare mean values of two related samples

For related variables X and Y, the comparison of mean values is equivalent to the comparison the mean value of the difference variable $X - Y$ to value 0 → the problem reduces to ***one-sample model***. When sample size is large or when the difference of the two variable has Normal distribution, Student test (T test) can be appropriate.

SPSS

Non-parametric method for two paired samples - *Wilcoxon signed rank test*

T-tests are appropriate only for the cases when variables are Normal distributed or when sample sizes are large. For small sample size studies of non-Normal distributions, non-parametric methods must be used. Wilcoxon signed rank test: non-parametric method for two related (paired) samples.

Given two related samples

$$(X_1, X_2, \ldots, X_n)$$

$$(Y_1, Y_2, \ldots, Y_n)$$

Wilcoxon signed rank test checks Hypothesis

H : X and Y have common distribution

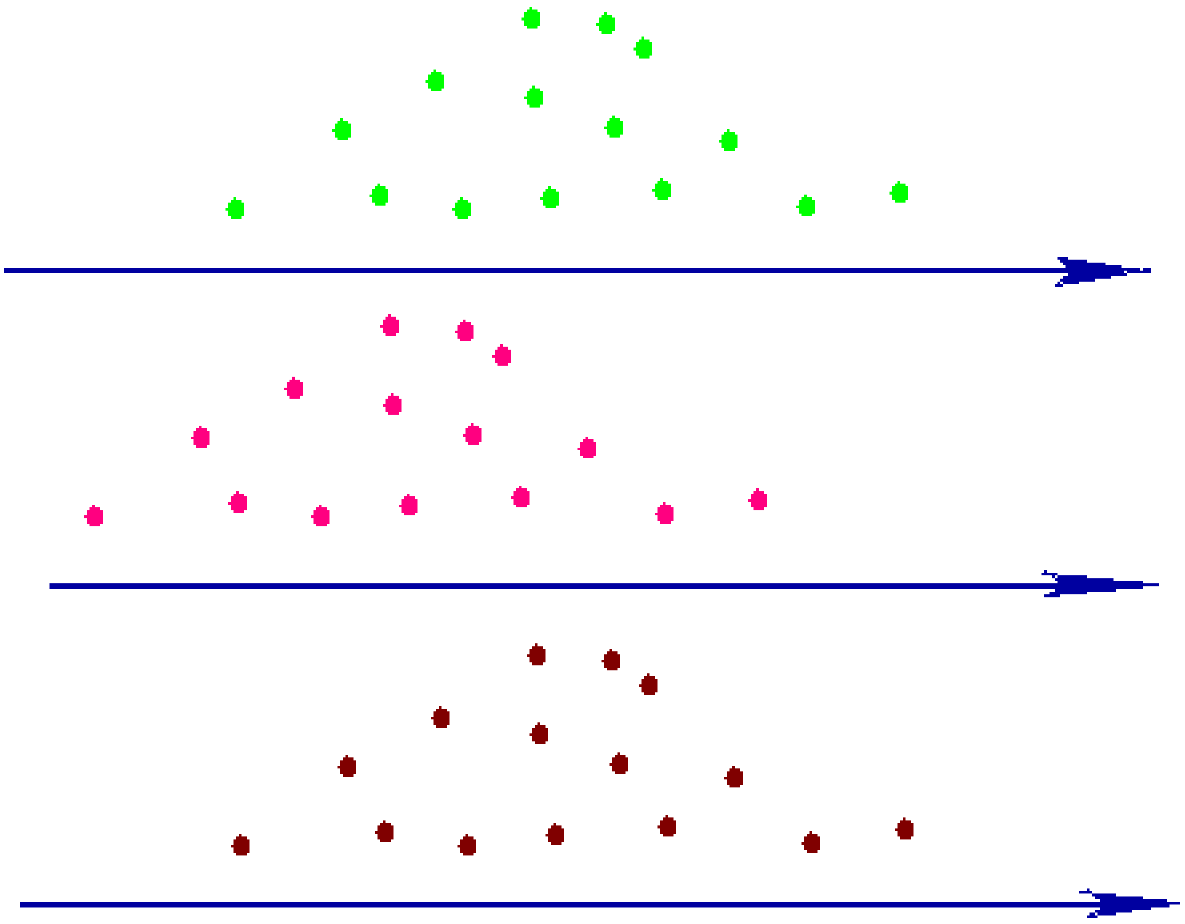
(two variables X and Y are identically distributed)

in constrain to Alternative Hypothesis

K: distributions of X and Y are different

(most of *X's* values can appear in positions higher than those of *Y's* values or inversely, most of *Y's* values have positions higher than those of *X's* values)

SPSS



Procedure:

Step 1. Calculate differences $d_i = X_i - Y_i$ and the ranks $h(|d_i|)$ omitting ties $d_i = 0$, $i = 1,2,\dots,n$

Step 2. Calculate quantiies

$$T^- = \sum_{m \in K^-} h(|d_m|)_i,\ T^+ = \sum_{m \in K^+} h(|d_m|) \text{ and } T = \min(T^-, T^+)$$

where $K^- = \{j : d_j < 0\}$, $K^+ = \{j : d_j > 0\}$;

$$n^+ = \#(K^-) + \#(K^+) \ ,\ \overline{T} = \frac{n^+.(n^+ + 1)}{4} \ ;$$

$$S_T^2 = \frac{n^+.(n^+ + 1).(2n^+ + 1)}{24} \ ,\ S_T = \sqrt{S_T^2}$$

LEMMA. Let $(X_1, X_2, \ldots, X_n)$ and $(Y_1, Y_2, \ldots, Y_n)$ be paired samples from two continuous varables X and Y. Suppose that hypothesis H is true. Then the distribution of variable T tends very fast to Normal distribution $N(\overline{T}, S_T^2)$ and then distribution of variable

$$t^* = \frac{T - \overline{T}}{S_T}$$

approximates the standard Normal distribution $N(0,1)$.

With the above lemma, for n >7 the testing can be continued as follows:

Step 3. (by computer) Taking standard Normal distribution N(0,1) to calculate the probability

$$b = P\{ |N(0,1)| > |t^*| \}$$

Step 4. Compare ***b*** significant level ***alpha***

* If ***b*** > ***alpha*** → accept Hypothesis ***H*** , cuclude ***X*** and ***Y*** to have common distribution

* If ***b*** <= ***alpha*** → reject Hypothesis ***H*** , confirm that ***X*** and ***Y*** have different distributions

Step3 B. Using critical value

For significant value *alpha=5%* take critical value (for standard Normal distribution) equal **1.96** and

Decide

- Reject Hypothesis **H** if

$$|t^*| \geq 1.96$$

- Accept Hypothesis **H** if

$$|t^*| < 1.96$$

- SPSS