

## Two-sample problem (location)

- **Assumptions:**

- Observations are independent.
- Observations come from populations with continuous distribution function.
- Distributions have the same shapes.

- **Testing: H:  $\eta_X = \eta_Y$ , Mann-Whitney-Wilcoxon**

$W_X$  = sum of the ranks for the Xs     $W_Y$  = sum of the ranks for the Ys.

$$U_X = W_X - \frac{m(m+1)}{2} \quad U_Y = W_Y - \frac{n(n+1)}{2}$$

Alternative	Test Statistic	P-value	Approximate P-value
A: $\eta_X > \eta_Y$	$U_X$	Table H: P for Right U	$\Pr\left(Z \geq \frac{U_X - 0.5 - \frac{m(n)}{2}}{\sqrt{\frac{m(n)(m+n+1)}{12}}}\right)$
A: $\eta_X < \eta_Y$	$U_Y$	Table H: P for Right U	$\Pr\left(Z \geq \frac{U_Y - 0.5 - \frac{n(n)}{2}}{\sqrt{\frac{m(n)(m+n+1)}{12}}}\right)$
A: $\eta_X \neq \eta_Y$	$U = \text{Larger}(U_X, U_Y)$	Table H: twice P for Right U	$2\Pr\left(Z \geq \frac{U - 0.5 - \frac{m(n)}{2}}{\sqrt{\frac{m(n)(m+n+1)}{12}}}\right)$

- **Estimation: ordered observations**

Point estimate of  $\eta_X - \eta_Y$  is the median of the  $\Delta_{ij} = X_i - Y_j$ .

Confidence interval for  $\eta_X - \eta_Y$  goes from  $\Delta_{(k+1)}$  to  $\Delta_{(N-k)}$ , where  $k$  is found using Table H with a P corresponding to  $(1 - \gamma)/2$ . Where  $\gamma$  is the level of confidence.

Table H	Left		Right
<u>n</u>	U	P	U
		m	
n	k	$\frac{1-\gamma}{2}$	n-k

For large values of n and/or m,

$$k = \frac{m(n)}{2} - 0.5 - z^* \sqrt{\frac{m(n)(m+n+1)}{12}}$$

where  $z^*$  corresponds to the standard normal z value for the given level of confidence.

## Two-sample problem (scale)

- **Assumptions:**

- Observations are independent.
- Observations come from populations with continuous distribution function.
- Distributions have the same medians (Siegel-Tukey).  
Distributions are symmetric (Fligner)

- **Testing: Siegel-Tukey and Fligner (see M-W-W)**

Siegel-Tukey assigns ranks alternating from largest to two smallest to the next two largest, etc.

Fligner assigns ranks to  $|X - M|$  and  $|Y - M|$  where  $M$  is the combined sample median.

## k-sample problem (location)

- **Assumptions:**

- Observations are independent.
- Observations come from populations with continuous distribution function.
- Distributions have the same shapes.

- **Testing:  $H_0: \eta_1 = \eta_2 = \dots = \eta_k$ , Kruskal-Wallis**

$R_i$  = sum of the ranks for the  $i^{\text{th}}$  group.

$$H = \frac{12}{N(N+1)} \left[ \sum \frac{R_i^2}{n_i} \right] - 3(N+1)$$

Compare  $H$  to a  $\chi^2$  distribution with  $k - 1$  degrees of freedom.

- **Multiple Comparisons**

Two groups are significantly different if

$$|\bar{R}_i - \bar{R}_j| > z^* \sqrt{\frac{N(N+1)}{12} \left( \frac{1}{n_i} + \frac{1}{n_j} \right)}$$

### Rank correlation

- **Spearman rank correlation coefficient,  $r_S$**

$$r_S = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$$

- **Test of significant monotonic trend**

Use Table I or for large sample sizes,  $z = \sqrt{(n-1)}r_S$  has an approximate standard normal distribution under the null hypothesis of no monotonic trend.