Statistics 431
Statistical Inference

Lecture XXIV: Distribution-Free Procedures

Mikhail Traskin

University of Pennsylvania

The Wharton School

Department of Statistics

Wilcoxon Signed Rank Test

- $X_1, \ldots, X_n$ are i.i.d. X.
- Distribution of r.v. X is continuous.
- Distribution of r.v. X is symmetric.
- Point of symmetry is median $\tilde{\mu}$.
- We are going to develop tests for $\tilde{\mu}$.
- If mean μ exists then all the methods we develop are applicable to it also.

Wilcoxon Signed Rank Test (cont.)

- Null hypothesis: $H_0 : \tilde{\mu} = \tilde{\mu}_0$.
- Rank absolute differences $|x_1 - \tilde{\mu}_0|, \ldots, |x_n - \tilde{\mu}_0|$ from smallest to largest.
- Test statistic
  $s_+ =$ the sum of the ranks associated with positive $(x_i - \tilde{\mu}_0)$s.
- Testing procedure

Alternative hypothesis	Rejection Region for Level α Test
$H_a : \tilde{\mu} > \tilde{\mu}_0$	$s_+ \geq c_1$
$H_a : \tilde{\mu} \neq \tilde{\mu}_0$	$s_+ \geq c$ or $s_+ \leq n(n+1)/2 - c$
$H_a : \tilde{\mu} < \tilde{\mu}_0$	$s_+ \leq c_2 = n(n+1)/2 - c_1$

Wilcoxon Signed Rank Test (cont.)

- Paired observations
  - $(X_1, Y_1), \ldots, (X_n, Y_n)$.
  - X's and Y's have continuous distributions that differ with respect to their means, then differences $D_i = X_i - Y_i$ have continuous symmetric distribution.
  - Do Wilcoxon signed-rank test for D_i's.
  - Null hypothesis $H_0 : \tilde{\mu}_D = \Delta_0$.
- Large sample approximation ($n > 20$): when H_0 is true, S_+ has approximately a normal distribution with

$$\mu_{S_+} = \frac{n(n+1)}{4} \qquad \sigma^2_{S_+} = \frac{n(n+1)(2n+1)}{24}.$$

- Analogous to one-sample tests.

Wilcoxon Rank-Sum Test

- $X_1, \ldots, X_m$ are i.i.d. X.
- $Y_1, \ldots, Y_n$ are i.i.d. Y.
- Xs and Ys are independent.
- R.v. X has a continuous distribution with mean μ_X.
- R.v. Y has a continuous distribution with mean μ_Y.
- $\mathbb{P}(X \leq z)\mathbb{P}(Y \leq z + (\mu_Y - \mu_X))$.
- We are going to develop test for equality of μ_1 and μ_2.

Wilcoxon Rank-Sum Test (cont.)

- Null hypothesis: $H_0 : \mu_X - \mu_Y = \Delta_0$.
- Compute $\tilde{x}_i = x_i - \Delta_0$
- Rank $\tilde{x}_1, \ldots, \tilde{x}_m, y_1, \ldots, y_n$ from smallest to largest.
- Test statistic $w = \sum_{i=1}^{m} r_i$, where r_i is the rank of $\tilde{x}_i$ in the combined sample of $\tilde{x}$s and ys.
- Testing procedure

Alternative hypothesis	Rejection Region for Level α Test
$H_a : \mu_X - \mu_Y > \Delta_0$	$w \geq c_1$
$H_a : \mu_X - \mu_Y \neq \Delta_0$	$w \geq c$ or $w \leq m(m+n+1) - c$
$H_a : \mu_X - \mu_Y < \Delta_0$	$w \leq c_2 = m(m+n+1) - c_1$

Wilcoxon Rank-Sum Test (cont.)

- Large sample approximation ($n > 8$ and $m > 8$): when H_0 is true, W has approximately a normal distribution with

$$\mu_W = \frac{m(m+n+1)}{2} \qquad \sigma_W^2 = \frac{mn(m+n+1)}{12}.$$

- Analogous to two-sample pooled test.