Nonparametric Statistics

Large-sample Wilcoxon sign test

Bruce M. Boghosian

Tufts UNIVERSITY | School of Arts and Sciences

Department of Mathematics

Tufts University

Outline

Bruce M. Boghosian

Example of application of Wilcoxon test

- Energy expenditures for women from heart rate (in kcal)
- During summer months and winter months
- Test difference D with $H_0: \mu = 0$ and $H_1: \mu \neq 0$

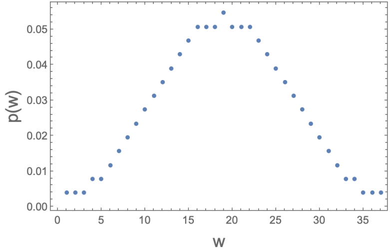
Subject	Summer, x_j	Winter, y_j	$d_j = y_j - x_j$	r_j	z_j
1	1458	1424	-34	1	0
2	1353	1501	148	5	1
3	2209	1495	-714	8	0
4	1804	1739	-65	2	0
5	1912	2031	119	4	1
6	1366	934	-432	7	0
7	1598	1401	-197	6	0
8	1406	1339	-67	3	0

- Wilcoxon rank statistic: $w = 5(1) + 4(1) = 9$

Example of application of Wilcoxon test

Bruce M. Boghosian

- Wilcoxon rank statistic: $w = 5(1) + 4(1) = 9$
- $\sum_{w=0}^{n(n+1)/2} p(w)e^{wt} = \frac{1}{2^n} \prod_{j=1}^{n} \left(1 + e^{jt}\right)$



- $\sum_{w=0}^{7} p(w) = \sum_{w=29}^{36} p(w) = \frac{19}{256} \approx 0.0742$
- Test is two-sided, so $2 \times \frac{19}{256} = \frac{19}{128} \approx 0.148$
- So for $\alpha = 0.15$, since $7 < w < 29$, we fail to reject H_0.

Example of application of Wilcoxon test

- $\sum_{w=0}^{n(n+1)/2} p(w)e^{wt} = \frac{1}{2^n}\prod_{j=1}^{n}\left(1+e^{jt}\right)$
- Tables of cutoffs for $n = 4, \ldots, 12$ are in Larsen & Marx Appendix A, Table A.6.
- For $n = 8$ one can look up that

$$P(W \leq w_1^*) = P(W \geq w_2^*) = 0.074$$

corresponds to $w_1^* = 7$ and $w_2^* = 29$, as we calculated on the previous slide.

Some discussion

- Note that the distribution of the Wilcoxon statistic is symmetric.
- The test should be used only if you have a priori reason to believe that your statistic is symmetric and comes from a continuous pdf.
- If it comes from normally distributed data, you might as well use a T test. If you have reason to believe that it is not normally distributed, the Wilcoxon tests may be useful.
- If you are not sure that is the case, use the sign test.

Example:

- Recall that the Wilcoxon statistic W has the same distribution as $U_j = \begin{cases} 0 & \text{with probability } 1/2 \\ j & \text{with probability } 1/2 \end{cases}$
- The expectation value is then

$$E(W) = E(U) = E\left(\sum_{j=1}^{n} U_j\right) = \sum_{j=1}^{n} E\left(U_j\right)$$

$$= \sum_{j=1}^{n}\left(\frac{1}{2}\cdot 0 + \frac{1}{2}\cdot j\right) = \sum_{j=1}^{n}\frac{j}{2} = \frac{n(n+1)}{4}.$$

- The variance is then

$$\text{Var}(W) = \text{Var}(U) = \text{Var}\left(\sum_{j=1}^{n} U_j\right) = \sum_{j=1}^{n}\text{Var}\left(U_j\right)$$

$$= \sum_{j=1}^{n}\left[\frac{i^2}{2} - \left(\frac{i}{2}\right)^2\right] = \sum_{j=1}^{n}\frac{i^2}{4} = \frac{n(n+1)(2n+1)}{24}$$

Large-sample Wilcoxon tests

- For large samples, e.g., ($n > 12$), that come from a continuous and symmetric pdf, form the statistic

$$z = \frac{w - E(W)}{\sqrt{\mathrm{Var}(W)}} = \frac{w - n(n+1)/4}{\sqrt{n(n+1)(2n+1)/24}}$$

  - To test $H_0 : \ \mu = \mu_0$ versus $H_1 : \ \mu > \mu_0$ at the α level of significance, reject H_0 if $z \geq +z_\alpha$.
  - To test $H_0 : \ \mu = \mu_0$ versus $H_1 : \ \mu < \mu_0$ at the α level of significance, reject H_0 if $z \leq -z_\alpha$.
  - To test $H_0 : \ \mu = \mu_0$ versus $H_1 : \ \mu \neq \mu_0$ at the α level of significance, reject H_0 if either $z \geq +z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$.

Summary

- For paired data, x_j and y_j, form the difference $d_j = x_j - y_j$.
- Let r_j be the rank of $|d_j|$.
- Define $z_j = \begin{cases} 1 & \text{if } d_j > 0 \\ 0 & \text{if } d_j < 0 \end{cases}$
- Calculate the Wilcoxon statistic as before, $w = \sum_{j=1}^{n} r_j z_j$.

Summary

Bruce M. Boghosian

- We have formulated a large-sample Wilcoxon test.
- We have looked at two-sample Wilcoxon tests