STAT 518 --- Chapter 5: Methods Based on Ranks

• The **rank** of an observation is its position in the **ordered** sample (ordered from smallest to largest).

• For example, in the sample

25.7, -1.3, 0.3, 4.2, -2.1

the corresponding **ranks** are

5, 2, 3, 4, 1

• Many of the methods in Chapter 5 rely on the **ranks** of the sample observations rather than the actual data values.

• This tends to reduce the effect of outliers, which makes rank-based methods ideal for data coming from heavy-tailed distributions.

Section 5.7: One-Sample or Matched-Pairs Data

• The t-test is a common parametric procedure to test whether the mean of a population equals some specified number.

• When we have paired data, we can test whether the two random variables have the same mean using the **paired t-test**.

• These classical tests require the sample (or the sample of differences) to come from a normal distribution.

• **If the normality assumption is questionable, but if our data (or differences) come from a <u>symmetric</u> distribution, and are measured on at least an <u>interval</u> scale, then the Wilcoxon Signed Ranks Test may be used.**

• **Note that earlier we learned the** <u>Sign</u> **Test as a procedure that could be used in this situation.**
• **The** <u>Sign</u> **Test did not require a** <u>symmetric</u> **distribution.**

Checking for Normality/Symmetry

• **A quick graphical check for whether data are normally distributed is the normal Q-Q plot.**

• **Sample quantiles are plotted against expected quantiles if the data were truly normal.**

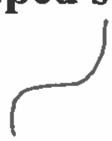
Using the R function `qqnorm`:

A roughly straight pattern → approx. normal

Bow-shaped curved pattern → skewed

right:  left: 

S-shaped snaking pattern → either heavy or light tails

heavy:  light: 

• **See examples on the course web page.**

Wilcoxon Signed Ranks Test

• Suppose we have n' paired observations (X_1, Y_1), (X_2, Y_2), …, $(X_{n'}, Y_{n'})$.

• The differences D_i are defined as

$$D_i = Y_i - X_i \quad \text{for} \quad i = 1, \ldots, n'$$

• The absolute differences are then

$$|D_i| = |Y_i - X_i|$$

• We remove any observations for which $D_i = 0$, leaving n untied pairs, where $n \leq n'$.

• We rank the n remaining pairs from 1 to n from the smallest absolute difference to the largest.

• If several pairs have the same absolute difference, we use midranks (average the ranks that would have been assigned) for these.

• For $i = 1, \ldots, n$, the signed ranks are

$$R_i = \begin{cases} \text{rank assigned to } (X_i, Y_i) \text{ if } D_i > 0 & (\text{if } Y_i > X_i) \\ \text{negative of rank assigned to } (X_i, Y_i), \text{ if } D_i < 0 & (\text{if } Y_i < X_i) \end{cases}$$

• The test statistic is T^+ = the sum of the positive signed ranks.

• If T^+ is large, it is evidence that the Y_i tends to be larger than the X_i.
• If T^+ is small, it is evidence that the Y_i tends to be smaller than the X_i.

• The null distribution (when $E(D_i)=0$ for all i) has been calculated.

• Table A12 gives lower quantiles of the null distribution when none of the absolute differences are tied and when $n \leq 50$.

• Upper quantiles can be found via the formula:

$$w_p = \frac{n(n+1)}{2} - w_{1-p}$$

• If there are many ties or if $n > 50$, the test statistic using all the signed ranks

$$T = \frac{\sum_{i=1}^{n} R_i}{\sqrt{\sum_{i=1}^{n} R_i^2}}$$

is used.
• This T has an approximately standard normal null distribution, so Table A1 is used.

3 Sets of Hypotheses

Two-tailed	Lower-tailed	Upper-tailed
H_0: $E(D) = 0$ (or $E(Y_i) = E(X_i)$)	H_0: $E(D) \geq 0$ (or $E(Y_i) \geq E(X_i)$)	H_0: $E(D) \leq 0$ (or $E(Y_i) \leq E(X_i)$)
H_1: $E(D) \neq 0$	H_1: $E(D) < 0$	H_1: $E(D) > 0$

• **The corresponding decision rules in each case are:**

Reject H_0 if $T^+ < W_{\alpha/2}$ or if $T^+ > W_{1-\alpha/2}$	Reject H_0 if $T^+ < W_\alpha$	Reject H_0 if $T^+ > W_{1-\alpha}$

• **If the large-sample tests are used, the decision rule is based on *T* and the normal quantiles from Table A1.**

• **P-values can be approximated by interpolating within Table A12 or by the normal approximation.**

• **In practice, software is the best way to get p-values.**

Example 1: 12 sets of twins were scored for "aggressiveness" on a personality test. The aggressiveness data are given on page 355. Can we conclude that the firstborn twins have greater median aggressiveness than the second twins? (Use $\alpha = .05$.)

Let X = first Y = second $D = Y - X$

$H_0: E(D) \geq 0$, $H_1: E(D) < 0$

D_i:	+2	+6	−1	−4	+5	0	−12	−1	−5	+9	−7	−15
$\lvert D_i\rvert$:	2	6	1	4	5	–	12	1	5	9	7	15
Rank of $\lvert D_i\rvert$:	3	7	1.5	4	5.5	–	10	1.5	5.5	9	8	11
R_i:	3	7	−1.5	−4	5.5	–	−10	−1.5	−5.5	9	−8	−11

Decision rule: Reject H_0 if $T^+ < W_\alpha = W_{.05} = 14 \leftarrow$ from Table A12 with $n = 11$

$T^+ = 3 + 7 + 5.5 + 9 = \boxed{24.5}$, so we fail to reject H_0.

• **See the R function `wilcox.test` for performing this test in R.**

```
wilcox.test(second, first, paired=T, alternative = "less")
```

what book calls "y"   what book calls "X"

From R:
Approx. P-value = .2382

Example 2: Problem 1 on page 365 describes a study in which the reaction time of 20 randomly sampled individuals was measured, both before and after an alcoholic drink for each person. Is there evidence that the alcohol alters reaction time? (Use $\alpha = .05$.) $\rightarrow \alpha/2 = .025$

Let X = before Y = after $D = Y - X$

$H_0: E(D) = 0$ vs. $H_1: E(D) \neq 0$

$n = 18$ (2 pairs tied)

From Table A12 $\rightarrow$ Reject H_0 if $T^+ < \underbrace{41}_{w_{.025}}$

or if $T^+ > \underbrace{171}_{\frac{n(n+1)}{2}} - \underbrace{41}_{w_{.975}} = 130$

From R: $T^+ = 154$, so we reject H_0. Approx P-value $= \underbrace{.003}_{\text{from R}}$

We can conclude that alcohol alters reaction time, on average

One-sample data

- **Suppose instead of paired data (X_i, Y_i), we had a single sample $Y_1, Y_2, \ldots, Y_n$.**

- **Suppose we wish to test whether the median of Y equals some constant number m.**

- **Our set of hypotheses would be one of:**

$H_0: \text{med}(Y) = m$	$H_0: \text{med}(Y) \geq m$	$H_0: \text{med}(Y) \leq m$
$H_1: \text{med}(Y) \neq m$	$H_1: \text{med}(Y) < m$	$H_1: \text{med}(Y) > m$

• We can let all $X_i = m$, i.e., form the pairs (m, Y_1), (m, Y_2), …, (m, Y_n) and carry out the signed-rank test as before.

• If T^+ is large, it is evidence that the Y_i's tend to be larger than m.

• If T^+ is small, it is evidence that the Y_i's tend to be smaller than m.

Example 3: Pollution measurements were taken on a site downstream from an industrial plant. Is there evidence that the median pollution level is less than 34 ppm? $\alpha = .05$

$H_0: \text{med}(Y) \geq 34$ $\quad H_1: \text{med}(Y) < 34$

$n = 15$. Table A12 → Reject H_0 if $T^+ < 31$ ← $w_{.05}$.

$T^+ = 5$, so reject H_0. Conclude the median pollution is less than 34 ppm. Approx P-value from R: .00098

Confidence Interval for the Median Difference

• Consider the sample $D_1, D_2, \ldots D_n$ (where $D_i = Y_i - X_i$, or where the D_i's are merely a single sample).

• If the D_i's are mutually independent, have the same median, have symmetric distributions, and are measured on at least an interval scale, the following method produces a confidence interval for the median of D_i's with coverage probability at least $1 - \alpha$.

• **Consider all possible averages** $\dfrac{D_i + D_j}{2}$ ← there are $\dfrac{n(n+1)}{2}$ of these.

(including when $i = j$)

• **Find** $W_{\alpha/2}$ **from Table A12. Then the**
$W_{\alpha/2}$th smallest and $W_{\alpha/2}$th largest of these averages
form the lower and upper bounds for the confidence interval for the median difference.

• **This is fairly computationally intensive, and the CI is best found using software: See the `WSRci` function on the course web page.**

Example 2 again: Find a 95% CI for the median difference in reaction time with alcohol and without.

Using WSRci function:
95% CI: [0.015, 0.06]. We are 95% confident that the true median reaction time is between 0.015 and 0.06 seconds <u>more with</u> alcohol than without.

Example 3 again: Find a 95% CI for the median pollution level.

95% CI: [27.85, 32.0]

– We are 95% confident that the true median pollution level is between 27.85 and 32.0 ppm.

Some Notes

• While the t-test (or paired t-test) requires normally distributed data, the Wilcoxon signed-rank test requires only symmetric data.

• Therefore the signed-rank test may be used with data having various types of distribution:

discrete
heavy-tailed
light-tailed

as long as we have symmetry

• In particular, the t-test lacks power when outliers exist in the data.

• The sign test is appropriate for any data measured on an ordinal or stronger scale.

• If the Y_i and X_i have identical symmetric distributions except for a shift in the means, the <u>Wilcoxon signed-rank test</u> has AT WORST an A.R.E. of 0.864 relative to the t-test.

• At BEST, the <u>Wilcoxon signed-rank test</u> has an A.R.E. of ∞ relative to the t-test.

Efficiency of the Signed Rank Test

Population	A.R.E.(signed-rank vs. t)	A.R.E.(signed-rank vs. sign)
Normal	0.955	1.5
Uniform (light tails)	1.00	3.00
Double exponential (heavy tails)	1.50	0.75

- **However, note that the A.R.E. is specific to large samples.**

- **For small samples from the double exponential distribution, the signed-rank test is** more **efficient than the sign test.**