STAT 141 **Non-Parametric tests** **11/02/04**

Here's a summary of the tests we will look at:

Setting	**Normal test**	**NonParametric Test**
One sample	One-sample *t*-test	Sign Test Wilcoxon signed-rank test
Matched pairs	*Apply one-sample test to differences within pairs*	
Two independent samples	Two-sample *t*-test	Wilcoxon rank sum test

Wilcoxon Rank Sum or Mann-Whitney Test – Chapter 7.11

Nonparametric comparison of two groups

Main Idea: If two groups come from the same distribution, but you've just randomly assigned labels to them, values in the two different groups should have values somewhat equally distributed between the two.

Group A: $X_1, \ldots, X_{n_1} \sim F_A$ Group B: $Y_1, \ldots, Y_{n_2} \sim F_B$

Null Hypothesis $H_0 : F_A = F_B$

Artificial Example:

gpA:	1.3	3.4		$(n_1 = 2)$
gpB:	4.9	10.3	3.3	$(n_2 = 3)$

Order all observations in the combined sample & assign ranks: (gp A data underlined)

Order	$\underline{1.3}$	3.3	$\underline{3.4}$	4.9	10.3
Assign ranks	$\underline{1}$	2	$\underline{3}$	4	5

Test statistic $R_1 =$ sum of ranks attached to group A $= 1 + 3 = 4$.

Under H_0, each 2-subset of the ranks $\{1, 2, 3, 4, 5\}$ is equally likely to occur as the ranks of X_1, X_2.

Possible ranks for X_1, X_2 :

sum =	R_1
$\{1,2\}$	3
$\{1,3\}$	4
$\{1,4\}$	5
$\{1,5\}$	6
$\{2,3\}$	5
$\{2,4\}$	6
$\{2,5\}$	7
$\{3,4\}$	7
$\{3,5\}$	8
$\{4,5\}$	9

Hence the distribution of R_1 under H_0 is given by

$r =$	3	4	5	6	7	8	9
$P_{H_0}(R_1 = r)$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{1}{10}$

In our toy example, $R_1 = 4$, the one-sided P-value

$$P = P_{H_0}(R_1 \leq 4) = P(\text{seeing a value as small or smaller than observed}) = \frac{1}{5}$$

Notes:

- values of R_1, p-value do not depend on the exact details of X_1, X_2 , only on their ranks in the combined sample.

- the distribution of R_1 under H_0 doesn't depend on the distribution of X or Y - it is a fixed distribution (which does, however, depend on n_1 and n_2). [for this reason, such methods are called distribution-free, or nonparametric.]
- Ranks can be sensitive to rounding. If we had additional data point $3.32 > 3.3$ we would give the new data point a rank of 3. If we rounded, however, we would have a tie because we would have 2 points with 3.3. We would then split the two possible ranks these two data points take up $(2\&3)$ and divide it between the two, so each would have rank 2.5

More Generally,

$X_1, \ldots, X_{n_1}$ in group A, distribution F_A; $Y_1, \ldots, Y_{n_2}$ in group B, distribution F_B

1. Combine the samples into one sample of W_i's. Order data in the combined sample $W_{(1)} \leq W_{(2)} \leq \ldots \leq W_{(n_1+n_2)}$
2. Assign rank i to the i^{th} smallest observation (in the case of ties, assign the average rank to each observation)
3. Let $R_1^{obs} =$ sum of ranks attached to observations in sample 1
4. $K_1 = R_1^{obs} - \frac{n_1(n_1+1)}{2}$
5. $U_s^{obs} = max(K_1, n_1n_2 - K_1)$
6. Find distribution of U_s under H_0. Reject if

$$P(U_s \geq U_s^{obs}) \leq \alpha$$

   More precisely: $U_{\alpha[n_1,n_2]}$ is largest u such that $P(U_s \geq u) \geq \alpha$

A Couple of Notes

- if $R_2 =$ sum of ranks attached to observations in sample 2, then

$$R_1 + R_2 = \sum_{i=1}^{n_1+n_2} i = (n_1+n_2)(n_1+n_2+1)/2$$

  So knowing R_1 is equivalent to knowing R_2

- Equivalently you can find $K_1 = \sum_{j=1}^{n_1} \#\{Y's < X_j\}$ and $K_2 = n_1n_2 - K_1 = \sum_{i=1}^{n_2} \#\{X's < Y_j\}$ as is done in the book [Can show algebraically these are the same value] so get $U_s = max(K_1, K_2)$
- By definition, $U_s \geq \frac{n_1n_2}{2}$
- $\frac{n_1(n_1+1)}{2}$ is the lowest possible value of R_1 and $n_1n_2 + \frac{n_1(n_1+1)}{2}$ is the largest value possible. $U_s = max(R_1 - \frac{n_1(n_1+1)}{2}, n_1n_2 + \frac{n_1(n_1+1)}{2} - R_1$.
- Traditionally would choose R_1 to go with smaller group because doing it by hand, but can show U_s the same whichever group you choose.
- Because of different sample sizes, the distribution of U_s is not symmetric.

$$\text{From Previous Example:}\quad \begin{array}{ll} R_1 & U_s \\ 3 & 6 \\ 4 & 5 \\ 5 & 4 \\ 6 & 3 \\ 7 & 4 \\ 8 & 5 \\ 9 & 6 \end{array} \quad \Rightarrow \quad \begin{array}{rcccc} u & 3 & 4 & 5 & 6 \\ P_{H_0}(U_s = u) & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} \end{array}$$

Alternative Hypothesis

One sided alternatives:

For $H_1 : F_A > F_B$, expect R_1 (and hence U_s) to be larger than H_0

For $H_1 : F_A < F_B$, expect R_1 to be smaller than under H_0 but U_s will still be larger, so still in the right tail.

Two sided alternative:

For $H_1 : F_B \neq F_A$, need to allow for both left and right tails of R_1.

It is possible to show that if the null hypothesis is verified,

$$E_{H_0} K_1 = \frac{n_1 n_2}{2} \qquad \text{Var}_{H_0} K_1 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

$$Z = \frac{K_1 - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}}} \sim \mathcal{N}(0,1) \left\{ n_1 \& n_2 \geq 20 \right.$$

Example Hollander & Wolfe (1973), 69f. Permeability constants of the human chorioamnion (a placental membrane) at term (x) and between 12 to 26 weeks gestational age (y). The alternative of interest is greater permeability of the human chorioamnion for the term pregnancy.

```
> term <- c(0.80, 0.83, 1.89, 1.04, 1.45,1.38, 1.91, 1.64, 0.73, 1.46)
> mid <- c(1.15, 0.88, 0.90, 0.74,1.21)
> rank(c(term,mid))
 [1]  3  4 14  7 11 10 15 13  1 12  8  5  6  2  9
> sum(rank(c(term,mid))[1:10])
[1] 90
> sum(rank(c(term,mid))[1:10])-(10*11/2)
[1] 35
> 1-pwilcox(34,10,5) # Beware this is a discrete random variable....
[1] 0.1272061 Output from the test:
>    wilcox.test(term, mid, alternative = "g")        # greater
    Wilcoxon rank sum test
data:  term and mid W = 35, p-value = 0.1272
alternative hypothesis: true mu is greater than 0
```

Paired Samples

$$\text{Paired samples } (X_i, Y_i) \rightarrow D_i = Y_i - X_i \qquad i = 1, \ldots, n$$

Sign Test – Chapter 9.4

Main Idea: If there is no difference between the pairs (i.e. true difference is 0), then equally likely to observe differences on either side of 0.

Equally likely to be on either side – a $Bin(0.5, n)$:

Wilcoxon signed rank test

Main Idea: If neither condition has an effect, then not only should the differences be equally distributed on either side of 0, but also how far away the differences are from 0 should be the same on either side.

1. let R_i be the rank of $|D_i|$ (absolute value of difference)
2. restore signs of D_i to the ranks → signed ranks
3. calculate either $W_+ =$ sum of ranks with positive signs of $W_- =$ sum of ranks with negative signs)

Idea
-if distribution of $x(F_x)$ is same as that of $Y(F_x)$ then D_i are equally likely to be positive as negative. So about half ranks are positive, and half are negative.
-if F_y is larger than F_x, then expect most ranks to have positive signs and so W_+ will be large.

More precisely,
H_0: distribution of D_i is symmetric about 0
Can show (if no ties)
$E_{H_0}W_+ = \frac{n(n+1)}{4}$
$Var_{H_0}W_+ = \frac{n(n+1)(2n+1)}{24}$

$$Z = \frac{W_+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim^{approx} N(0,1)$$

.

Notes:

- $$W_+ + W_- = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}, \text{ so either one of } W_+, W_- \text{determines the other.}$$
- Differences $D_i = 0$ are usually omitted
- If there are ties in the $|D_i|$, average the corresponding ranks in step 1 (like the Rank-Sum test). If many ties, need to adjust variances - see books on non-parametric such as Hollander-Wolfe (1973), Lehman(1975).
- Null distribution of W_+ is tabulated for $n \leq 50$. For $n \geq 20$, however, can use normal approximation
- Can also use for a one-sample to test for θ if you assume the distribution is symmetric around θ. In that case, you would subtract off θ from each data value, and then perform the same test.

Example: Measuring mercury levels in Fish (of Rice, 11.33); see calc. of $W_+ = 194.5; W_- = 105.5$ Here $n = 24$(one $D_i = 0$ discarded)

$$\text{Hence: } \frac{n(n+1)}{4} = \frac{24}{4} \cdot 25 = 150; \qquad \frac{n(n+1)(2n+1)}{24} = 1225$$

Two Ways to measure mercury levels in fish (ppm of mercury in 25 juvenile black marlin)

Sel Red	Permang	Diff.	SignedRank	Pos Ranks	Neg Ranks
0.32	0.39	0.07	15.5	15.5	
0.4	0.47	0.07	15.5	15.5	
0.11	0.11	0		«-note!	
0.47	0.43	-0.04	-11		-11
0.32	0.42	0.1	19	19	
0.35	0.3	-0.05	-13.5		-13.5
0.32	0.43	0.11	20	20	
0.63	0.98	0.35	23	23	
0.5	0.86	0.36	24	24	
0.6	0.79	0.19	22	22	
0.38	0.33	-0.05	-13.5		-13.5
0.46	0.45	-0.01	-2.5		-2.5
0.2	0.22	0.02	6.5	6.5	
0.31	0.3	-0.01	-2.5		-2.5
0.62	0.6	-0.02	-6.5		-6.5
0.52	0.53	0.01	2.5	2.5	
0.77	0.85	0.08	17.5	17.5	
0.23	0.21	-0.02	-6.5		-6.5
0.3	0.33	0.03	9	9	
0.7	0.57	-0.13	-21		-21
0.41	0.43	0.02	6.5	6.5	
0.53	0.49	-0.04	-11		-11
0.19	0.2	0.01	2.5	2.5	
0.31	0.35	0.04	11	11	
0.48	0.4	-0.08	-17.5		-17.5

```
Hg=scan() 0.32 0.39 0.4 0.47 0.11 0.11 ...........
way=factor(rep(1:2,25))
```

Based on $T_s = min(W_-, W_+)$ (105.5 here) [only need to tabulate left hand tail]

Normal approximation 2 sided P-value

$$2P(W_+ \leq 105.5) = 2P(\frac{W_+ - 150}{\sqrt{1225}} \leq \frac{105.5 - 150}{\sqrt{1225}}) = 2P(Z \leq -1.271) = .203$$

[Compare with t-test, gives 2 sided P = .094 - rather different but neither is statistically significant evidence for a difference].CompareT-TestwithWilcoxon:

Paired t-test

mean diff $= \bar{d} = 0.0404$

SD diff $= s_{\bar{d}} = 0.116$

SE diff $= SE_{\bar{d}} = 0.023$

$t_s = 1.745$

df = 24

```
> pt(-1.745,24)
[1] 0.04688927
2 sided P   0.094
t.test(Hg~way,paired=T)
Paired t-test
data:  Hg by way
t = -1.7448, df = 24,
p-value =0.0938
alternative hypothesis:
true difference
in means is not equal to 0
95 percent confidence interval:
-0.088189837  0.007389837
sample estimates:
mean of the differences
      -0.0404
```

Wilcoxon

signed rank sums $W_+ = 194.5,\ W_- = -105.5$

n=24

normal approx mean 150

sd of $W_+ = 35$

Z-score -1.272

```
> pnorm(-1.27)
[1] 0.1020423
 2 sided P       0.203
> wilcox.test(Hg~way,paired=T)
Wilcoxon signed rank test with correction
data:  Hg by way
V =107,
p-value = 0.2242
alternative hypothesis: true mu is
not equal to 0
```

Sign Test: $N_- = 10$ $\quad N_+ = 14$. So for 2-tailed test, we have:

$P(\text{get 10 or less successes}) + P(\text{get 14 or more successes}) =$ `pbinom(10,prob=.5,size=24)` $+$
`1- pbinom(13,prob=.5,size=24)`
$= 0.707$