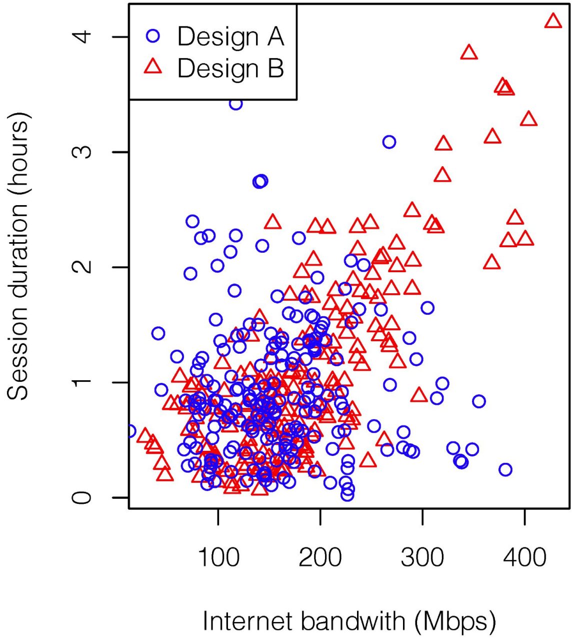
Two independent samples

11.2.3 of Rice - The non-parametric Mann-Whitney test

07/14/2021

Berkeley
UNIVERSITY OF CALIFORNIA

In the previous lecture,



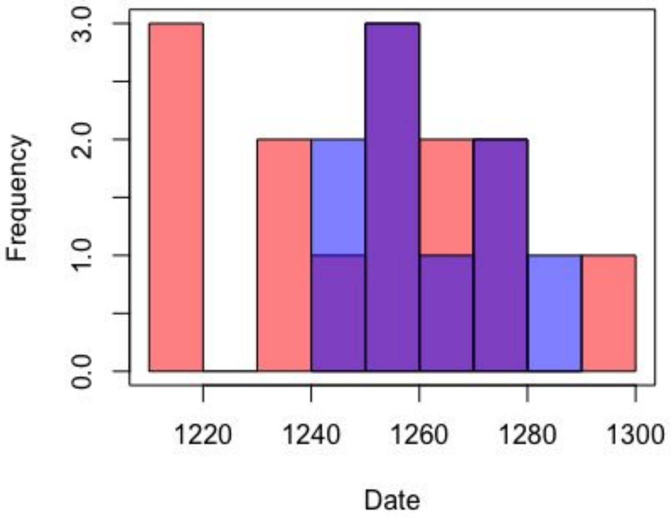
- Examples of generalized LRT:
  - χ^2_ν fits the sampling distribution of $-2\log\lambda(\mathbf{X}_n)$ well;
  - Greatly simplifies the derivation of the rejection region.

- Independent samples under Normal populations :
  - Equal variance assumption;
  - $\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 + \sum_{i=1}^{m}\left(Y_i - \bar{Y}_m\right)^2 \sim \sigma^2\chi^2_{m+n-2}$,
  - $\dfrac{\left(\bar{X}_n - \bar{Y}_m\right) - (\mu_X - \mu_Y)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{n+m-2}.$
  - Hypothesis tests & confidence intervals.

What if the Normal assumption is not true?

Example 2. Samples of wood were obtained from the core and periphery of a Byzantine church. The date of the wood were determined.

Core		Periphery	
1294	1251	1284	1274
1279	1248	1272	1264
1274	1240	1256	1256
1264	1232	1254	1250
1263	1220	1242	
1254	1218		
1251	1210		

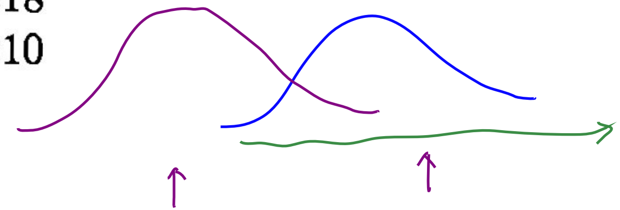


- Core (red)
- Periphery (blue)

The ranks of the observations

Example 2. Samples of wood were obtained from the core and periphery of a Byzantine church. The date of the wood were determined.

Core		Periphery	
1294	1251	1284	1274
1279	1248	1272	1264
1274	1240	1256	1256
1264	1232	1254	1250
1263	1220	1242	
1254	1218		
1251	1210		



low

1	2	3	4	5	6	7	8	9
1210	1218	1220	1232	1240	1242	1248	1250	1251
10	11.5	11.5	13.5	13.5	15	16.5	16.5	18
1251	1254	1254	1256	1256	1263	1264	1264	1272
1274	1274	1279	1284	1294				
19.5	19.5	21	22	23				

high

$$R_1 = 1+2+3+4+5+7+9+10 + 11.5 + 15 + 16.5 + 19.5 + 21 + 23 = 147.5.$$

$$R_2 = 6+8+11.5+\cdots = 128.5$$

The ranks of the observations

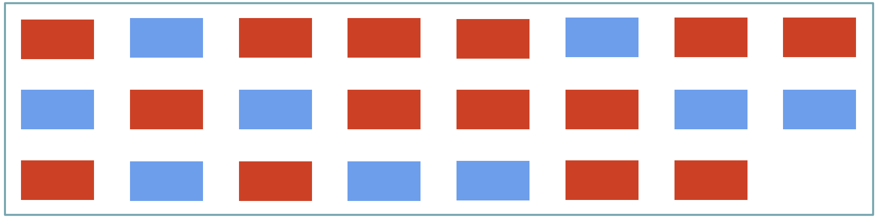
$\binom{14+9}{14} = \binom{14+9}{9} \geq (k$

Example 2. Samples of wood were obtained from the core and periphery of a Byzantine church. The date of the wood were determined.

$\frac{1}{\binom{m+n}{m}}$

14 **core** observations $\sim F$

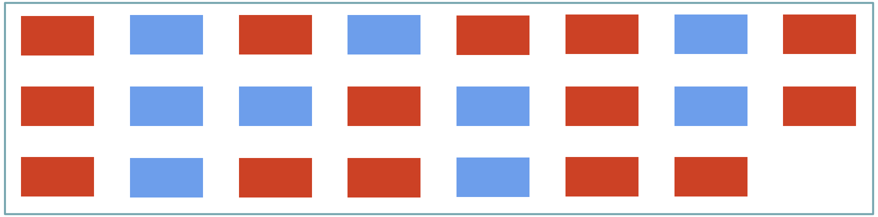
9 **periphery** observations $\sim G$



u

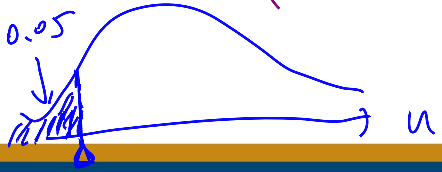
$$p = \frac{1}{\binom{14+9}{9}}$$

If F=G, every assignment of ranks to the 14 + 9 observations is equally likely,



u

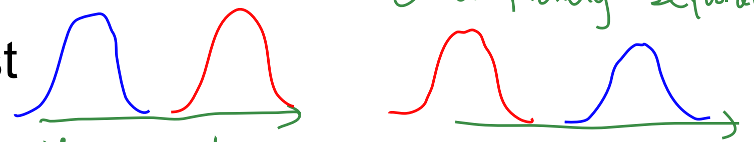
$$p = \frac{1}{\binom{14+9}{9}}$$



$P(U < c \mid H_0) = 0.05$

Mann-Whitney U test

A.k.a. Wilcoxon rank-sum test



$U_1 = 0$, $U_2 = mn$

① Completely separated :

$R_1 = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$

$R_2 = \sum_{i=n+1}^{m+n} i = \sum_{i=1}^{m+n} i - \sum_{i=1}^{n} i$

$= \frac{(m+n)(m+n+1)}{2} - \frac{n(n+1)}{2}$

$\Rightarrow U_1 = mn$

$U_2 = mn + \frac{m(m+1)}{2} - \frac{(m+n)(m+n+1)}{2} + \frac{n(n+1)}{2}$

$$H_0: F = G \quad \text{versus} \quad H_1: F \neq G$$

1. Calculate the rank sum R_1 and R_2 from the first and second sample respectively.
   We know $R_1 + R_2 = N(N+1)/2$. $= \sum_{i=1}^{N} i$ where $N = m+n$.

2. The Mann-Whitney U statistic is given by $U = \min\{U_1, U_2\}$:

$$U_1 = mn + \frac{n(n+1)}{2} - R_1 \qquad \frac{n(n+1)}{2} = \sum_{i=1}^{n} i$$

$$U_2 = mn + \frac{m(m+1)}{2} - R_2$$

   We know $U_1 + U_2 = mn$.

$U_1 + U_2 = 2mn + \frac{n(n+1)}{2} + \frac{m(m+1)}{2} - (R_1 + R_2)$

$= 2mn + \frac{n(n+1)}{2} + \frac{m(m+1)}{2} - \frac{(m+n)(m+n+1)}{2}$

$= mn$.

3. Find the rejection region using *the Table of Critical Values*.

② $F = G$, U_1 and U_2 should have closer values .

$R = \{ U < c \}$

$U \searrow 0$, more evidence against H_0.

The ranks of the observations

Example 2. Samples of wood were obtained from the core and periphery of a Byzantine church. The date of the wood were determined.

$R = \{ U < 31 \}$

We fail to reject Ho, and conclude that the date of the wood is not significantly different between the core & periphery.

Core		Periphery	
1294	1251	1284	1274
1279	1248	1272	1264
1274	1240	1256	1256
1264	1232	1254	1250
1263	1220	1242	
1254	1218		
1251	1210		

1210	1218	1220	1232	1240	1242	1248	1250	1251
1	2	3	4	5	6	7	8	9.5
1251	1254	1254	1256	1256	1263	1264	1264	1272
9.5	11.5	11.5	13.5	13.5	15	16.5	16.5	18
1274	1274	1279	1284	1294				
19.5	19.5	21	22	23				

$\Rightarrow U = \min\{U_1, U_2\}$

$= \boxed{42.5}$.

$R_2 = 128.5$

$R_1 = 147.5$

$U_1 = 14 \times 9 + \frac{14(15)}{2} - 147.5 = 83.5$

$U_2 = 14 \times 9 + \frac{9(10)}{2} - 128.5 = 42.5$.

```
> x=c(1294,1251,1279,1248,1274, 1240, 1264, 1232, 1263, 1220, 1254, 1218, 1251, 1210)
> y=c(1284, 1274, 1272, 1264, 1256, 1256, 1254, 1250, 1242)
> wilcox.test(x,y)

        Wilcoxon rank sum test with continuity correction

data:  x and y
W = 42.5, p-value = 0.2072
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(x, y) : cannot compute exact p-value with ties
```

Compared with p-value=0.21109.

Mann-Whitney U test

Example 3. Consider a Phase II clinical trial designed to investigate the effectiveness of a new drug to reduce symptoms of asthma in children.

Placebo	7	5	6	4	12
New Drug	3	6	4	2	1

*the number of episodes of shortness of breath over a 1 week period following receipt of assigned treatment.

$H_0: F = G$ vs. $H_1: F \neq G$

Obs: 1 2 3 4 4 5 6 6 7 12

Rank: 1 2 3 4.5 4.5 6 7.5 7.5 9 10

$R_1 = 4.5 + 6 + 7.5 + 9 + 10 = 37.$

$R_2 = 1 + 2 + 3 + 4.5 + 7.5 = 18.$

$U_1 = 5 \times 5 + \frac{5 \times 6}{2} - 37 = 3.$

$U_2 = 5 \times 5 + \frac{5 \times 6}{2} - 18 = 22$

$U = \min\{U_1, U_2\} = 3.$

$R = \{U < 2\}.$ We fail to reject H_0.

Comparing paired samples

11.3 of Rice

07/14/2021

Paired samples

$X_1, \ldots X_n$
$Y_1, \ldots Y_m$ } indep

Example 4. Blood samples from $n=10$ people were sent to each of two laboratories (Lab 1 and Lab 2) for cholesterol determinations. Is there a statistically significant difference at the α=0.01 level, in the (population) mean cholesterol levels reported by Lab 1 and Lab 2?

Subject	Lab1	Lab2	Diff
1	296	318	-22
2	268	287	-19
⋮	⋮	⋮	⋮
10	262	285	-23
	$\bar{x}_1 = 260.6$	$\bar{x}_2 = 275$	$\bar{d} = -14.4$
			$s_d = 6.77$

Can't assume independence between two samples.

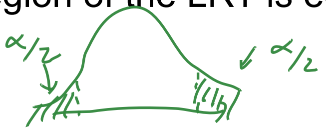
Paired samples based on Normal populations

$$\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} \right)$$

Theorem C. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d observations from a bivariate Normal distribution with parameters $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$. Consider

$$H_0 : \mu_X = \mu_Y \quad \text{versus} \quad H_1 : \mu_X \neq \mu_Y.$$

The rejection region of the LRT is equivalent to



$$R = \left\{ \frac{|\bar{D}_n|}{\sqrt{\frac{1}{n} S_D^2}} \geq c \right\}.$$

$$\alpha = P(R \mid H_0) = P\left(\frac{|\bar{D}_n|}{\sqrt{\frac{1}{n} S_D^2}} \geq c \,\middle|\, H_0 \right)$$

$$\Rightarrow \quad c = t_{n-1}(\alpha/2).$$

$$D_i = X_i - Y_i, \quad \bar{D}_n = \frac{1}{n} \sum_{i=1}^{n} D_i,$$

$$S_D^2 = \frac{1}{n-1} \sum_{i=1}^{n} (D_i - \bar{D}_n)^2.$$

Under H_0, $T(\mathbf{X}_n) = \dfrac{\bar{D}_n}{\sqrt{\frac{1}{n} S_D^2}} \sim t_{n-1}$.

$D_i = X_i - Y_i \sim N(\mu_X - \mu_Y, \ \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)$

$ED_i = EX_i - EY_i = \mu_X - \mu_Y$

$Var(D_i) = cov(X_i - Y_i, X_i - Y_i)$

$= cov(X_i, X_i) + cov(Y_i, Y_i) - 2\,cov(X_i, Y_i)$

$= \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y$

D_i's should be i.i.d. $\Rightarrow \bar{D}_n \sim N\left(\mu_X - \mu_Y, \dfrac{\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y}{n}\right)$

$\dfrac{1}{\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y} S_D^2 \sim \chi^2_{n-1}$

Paired samples based on Normal populations

Theorem C. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be i.i.d observations from a bivariate Normal distribution with parameters $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$. Consider

$$H_0 : \mu_X = \mu_Y \quad \text{versus} \quad H_1 : \mu_X \neq \mu_Y.$$

The rejection region of the LRT is equivalent to

$$R = \left\{ \frac{|\bar{D}_n|}{\sqrt{\frac{1}{n} S_D^2}} \geq c \right\}.$$

$$D_i = X_i - Y_i,\ \bar{D}_n = \frac{1}{n}\sum_{i=1}^{n} D_i,$$

$$S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(D_i - \bar{D}_n\right)^2.$$

Under H_0, $T(\mathbf{X}_n) = \dfrac{\bar{D}_n}{\sqrt{\frac{1}{n} S_D^2}} \sim t_{n-1}$.

Compare the relative efficiency between
$\bar{X}_n - \bar{Y}_n$ *(independent samples)* *and* $\bar{D}_n$ *(dependent samples)*.

$$\frac{var(\bar{D}_n)}{var(\bar{X}_n - \bar{Y}_n)} = \frac{\frac{1}{n}(\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)}{var(\bar{X}_n) + var(\bar{Y}_n)} = \frac{\frac{1}{n}(\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y)}{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{n}}$$

$$= \frac{\sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y}{\sigma_X^2 + \sigma_Y^2} \leq 1$$

Paired samples based on Normal populations

Corollary C. Let $(X_1, Y_1), \ \ldots \ , (X_n, Y_n)$ be i.i.d observations from a bivariate Normal distribution with parameters $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$. The LRTs with significance level α for the following hypotheses can be derived.

$$H_0 : \mu_X = \mu_Y \quad \leftrightarrow \quad H_1 : \mu_X \neq \mu_Y. \quad \Longleftrightarrow \quad R = \left\{ \frac{\left|\bar{D}_n\right|}{\sqrt{\frac{1}{n} S_D^2}} \geq t_{n-1}(\alpha/2) \right\}$$

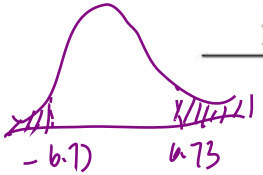
$$H_0 : \mu_X = \mu_Y \quad \leftrightarrow \quad H_1 : \mu_X > \mu_Y. \quad \Longleftrightarrow \quad R = \left\{ \frac{\bar{D}_n}{\sqrt{\frac{1}{n} S_D^2}} \geq t_{n-1}(\alpha) \right\}$$

$$H_0 : \mu_X = \mu_Y \quad \leftrightarrow \quad H_1 : \mu_X < \mu_Y. \quad \Longleftrightarrow \quad R = \left\{ \frac{\bar{D}_n}{\sqrt{\frac{1}{n} S_D^2}} \overset{\leq}{\cancel{\geq}} -t_{n-1}(\alpha) \right\}$$

Paired samples based on Normal populations

Example 4 *cont'd*. Blood samples from $n=10$ people were sent to each of two laboratories (Lab 1 and Lab 2) for cholesterol determinations. Is there a statistically significant difference at the α=0.01 level, in the (population) mean cholesterol levels reported by Lab 1 and Lab 2?

Subject	Lab1	Lab2	Diff
1	296	318	-22
2	268	287	-19
⋮	⋮	⋮	⋮
10	262	285	-23
	$\overline{x}_1 = 260.6$	$\overline{x}_2 = 275$	$\overline{d} = -14.4$
			$s_d = 6.77$



$$H_0 : \mu_X = \mu_Y \quad \leftrightarrow \quad H_1 : \mu_X \neq \mu_Y.$$

$$T = \frac{\sqrt{10} \times (-14.4)}{6.77} = -6.73$$

$$t_9(\alpha/2) = t_9(0.01/2) = 3.2498$$

$R = \{ |T| > 3.2498 \}$

Therefore, we reject H_0 and conclude that there is enough evidence to support H_1.

p-value $= P(|T| > 6.73 \mid H_0)$ $\quad S_p = \sqrt{S_D^2}$

$= 2 * pt(-6.73, df = 9)$

$= 8.6 \times 10^{-5} \ll 0.01$

Paired samples based on Normal populations

p-value = P(|T| > 1.747 | Ho)
= 2 * pt(-1.747 , 14)
= 0.1025
α = 0.05   > α

Example 5. A study is run to evaluate the effectiveness of an exercise program in reducing systolic blood pressure in patients with pre-hypertension (*defined as a systolic blood pressure between 120-139 mmHg or a diastolic blood pressure between 80-89 mmHg*).

Patient	Systolic Blood Pressure Before Exercise Program	Systolic Blood Pressure After Exercise Program	Difference (Before-After)
1	125	118	7
2	132	134	-2
3	138	130	8
4	120	124	-4
5	125	105	20
6	127	130	-3
7	136	130	6
8	139	132	7
9	131	123	8
10	132	128	4
11	135	126	9
12	136	140	-4
13	128	135	-7
14	127	126	1
15	130	132	-2

$\bar{D}_n = 3.2$, $S_D = 7.09$

$$H_0 : \mu_X = \mu_Y \quad \leftrightarrow \quad H_1 : \mu_X \neq \mu_Y .$$

$$T = \frac{\sqrt{15} \times 3.2}{7.09} = 1.747$$

$$t_{14}(\alpha/2) = t_{14}(0.05/2) = 2.145$$

R = { |T| > 2.145 }

We fail to reject Ho, and conclude that there is not enough evidence of ~~mean~~ difference.

$W_+ + W_- = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$; $H_0: F = G$ vs. $H_1: F \neq G$

$R = \{ W_+ > c_1$ or $W_+ < c_2 \}$

$= \{ W_- < c_3$ or $W_+ < c_2 \}$

$= \{ W_- < c$ or $W_+ < c \}$

Paired samples - *Nonparametric method*

$H_0: F = G$ vs. $H_1: F < G$

$R = \{ W_+ > c_1 \} = \{ W_- < c_2 \}$

Before	After	Difference	|Difference|	Rank	Signed Rank
25	27	2	2	2	2
29	25	−4	4	3	−3
60	59	−1	1	1	−1
27	37	10	10	4	4

```
wilcox.test(x,y, paired=TRUE)
```

After − Before

→ 1 2 . . . n

+ − + − + −

W_+

$\alpha/2$ $\alpha/2$

2^n

Wilcoxon signed rank test:

1. Calculate the differences, $Y_i - X_i$, and the absolute values of the differences and rank the latter.
2. Restore the signs of the differences to the ranks, obtaining signed ranks.
3. Calculate W_+, the sum of those ranks that have positive signs. For the table, this sum is $W_+ = 2 + 4 = 6$.

w_+ c_1 c_2

If $F = G$, we expect half of the differences to be positive half to be negative, and W_+ won't be too small or too large.

w_+ w_-

Paired samples - *Nonparametric method*

Example 6. Consider a clinical investigation to assess the effectiveness of a new drug designed to reduce repetitive behaviors in children affected with autism.

Child	Before Treatment	After 1 Week of Treatment	Difference (Before-After)
1	85	75	10
2	70	50	20
3	40	50	-10
4	65	40	25
5	80	20	60
6	75	65	10
7	55	40	15
8	20	25	-5

* A score of 0 = no repetitive behavior, while a score of 100 = constant repetitive behavior.

|diff|	5	10	10	10	15	20	25	60
rank	1	3	3	3	5	6	7	8

$$W_+ = 3+3+5+6+7+8 = 32,$$
$$W_- = 1+3 = 4,$$
$$R = \{W_- \le 6\}.$$

no location shift

We reject the null hypothesis and conclude that the mean after treatment

H_1 is greater than the mean before treatment.

$$W_+ + W_- = \frac{n(n+1)}{2}$$

$$W_- = \frac{n(n+1)}{2} - W_+$$

Paired samples - *Nonparametric method*

Example 6. Consider a clinical investigation to assess the effectiveness of a new drug designed to reduce repetitive behaviors in children affected with autism.

Child	Before Treatment	After 1 Week of Treatment	Difference (Before-After)
1	85	75	10
2	70	50	20
3	40	50	-10
4	65	40	25
5	80	20	60
6	75	65	10
7	55	40	15
8	20	25	-5

* A score of 0 = no repetitive behavior, while a score of 100 = constant repetitive behavior.

$$W_+ = 3 + 3 + 5 + 6 + 7 + 8 = 32,$$
$$W_- = 1 + 3 = 4,$$
$$R = \{W_- \leq 6\}.$$

$F < G$

```
> wilcox.test(x,y,pair=TRUE, alternative='greater')

        Wilcoxon signed rank test with continuity correction

data:  x and y
V = 32, p-value = 0.02874
alternative hypothesis: true location shift is greater than 0

Warning message:
In wilcox.test.default(x, y, pair = TRUE, alternative = "greater") :
  cannot compute exact p-value with ties
```

Paired samples - *Nonparametric method*

$H_0: F = G$ vs. $H_1: F \neq G$

Example 5 *cont'd*. A study is run to evaluate the effectiveness of an exercise program in reducing systolic blood pressure in patients with pre-hypertension (*defined as a systolic blood pressure between 120-139 mmHg or a diastolic blood pressure between 80-89 mmHg*).

Patient	Systolic Blood Pressure Before Exercise Program	Systolic Blood Pressure After Exercise Program	Difference (Before-After)
1	125	118	7
2	132	134	-2
3	138	130	8
4	120	124	-4
5	125	105	20
6	127	130	-3
7	136	130	6
8	139	132	7
9	131	123	8
10	132	128	4
11	135	126	9
12	136	140	-4
13	128	135	-7
14	127	126	1
15	130	132	-2

$\alpha = 0.05$

$$W_+ = 89, \leftarrow$$
$$W_- = 31, \leftarrow$$
$$R = \{W_+ \leq 25 \text{ or } W_- \leq 25\}.$$

We fail to reject H_0.

```
> x<-c(125,132,138,120,125,127,136,139,131,132,135,136,128,127,130)
> y<-c(118,134,130,124,105,130,130,132,123,128,126,140,135,126,132)
> wilcox.test(x,y,pair=TRUE)

        Wilcoxon signed rank test with continuity correction

data:  x and y
V = 89, p-value = 0.1048
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(x, y, pair = TRUE) :
  cannot compute exact p-value with ties
```

$> \alpha = 0.05$

Compared with p-value=0.1025.

Tomorrow ...

- Midterm review
- Go over a Practice Midterm