

Chushua Yu.

92 0  
100

STT 6310/4310

Exam 1

Start Date: Oct.

16, 2018

Instructor: Gengxin Li

(Take home exam - 100 pts.)

Due Date: Oct 23, 2018

1. What are phases of clinical trials? Please compare these phases.

Clinical trials which are the experimentation with human classified into 4 phases.

Phase I: To explain possible toxic effect of drugs and determine a tolerated dose for future experimentation

Phase II: Screening and feasibility by initial assessment for the therapeutic effect further assessment of toxicities.

Phase III: Comparison of new intervention (drug and therapy) to the current standard of treatment both with respect to efficacy and toxicity.

Phase IV: Observational study of morbidity adverse effects after a drug is on the market.

2. Which issues do we need to design a clinical trial?

In order to know whether there is enough evidence of efficacy to make it worth further study in larger and more costly clinical trial. There are two kinds of issues we need to design a clinical trial.

(1) Surrogate markers: used to evaluate the treatment efficacy

(2) Surrogate endpoint: to determine whether the proposed treatment should be further studied.

3. In phase II study,  $X$  denotes the number of patients who completely respond, and it follows the binomial distribution with the total number of patient ( $n$ ) and the probability of getting complete response ( $p$ ). What's the point estimator of  $p$ , and the distribution of this estimator? Please compute 99% confidence interval for true  $p$  when the estimator of response rate is 0.25 and total number of patients is 25.

$$\hat{p} = \bar{x} = \frac{X}{n}$$

$$\hat{p} \sim N\left(n, \frac{n(1-p)}{n^2}\right)$$

$$\text{Since } X \sim \text{Bin}(n, p) \text{ then } \begin{cases} E(X) = np \\ \text{Var}(X) = np(1-p) \end{cases}$$

$$n \stackrel{\text{iid}}{\sim} \text{Beta distribution with } \begin{cases} E(\lambda) = E\left(\frac{X}{n}\right) = \frac{1}{n} E(X) = p \\ \text{Var}(\lambda) = \text{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2} \text{Var}X = \frac{p(1-p)}{n} \end{cases}$$

$$\text{③ 99% C.I for } P \text{ is } \bar{x} \pm Z_{\alpha/2} \sqrt{\frac{\bar{x}(1-\bar{x})}{n}}$$

$$\begin{cases} \bar{x} = 0.25 \\ n = 25 \\ \alpha = 0.01 \\ Z_{\alpha/2} = 2.575 \end{cases}$$

$$0.25 \pm 2.575 \times \sqrt{\frac{0.25 \times 0.75}{25}} = 0.25 \pm 2.575 \times 0.087 = 0.25 \pm 0.223$$
$$\text{C.I} = [0.027, 0.473]$$

(-14)

4. In a phase II trial, 4 of 20 patients respond to a new drug for lung cancer. Please find the exact confidence 95% interval for the response rate ( $p$ ) for  $X=k$ ,  $k=4$ , and  $n=20$ . What about 95% confidence interval of the response rate ( $p$ )? Compare the exact confidence 95% interval with 95% confidence interval.

Solution: ①  $\begin{cases} A: P_{PL(4)}(X \geq 4) = \frac{\alpha}{2} = 0.025 \\ B: P_{PU(4)}(X \leq 4) = \frac{\alpha}{2} = 0.025 \end{cases} \Rightarrow \begin{cases} P_{PL}(X \leq 3) = 0.975 \\ P_{PU}(X \leq 4) = 0.025 \end{cases}$  Check the tables of the Binomial Cumulative dist.  
 From  $\begin{cases} \text{equation } A \Rightarrow PL = 0.06 \\ \text{equation } B \Rightarrow PU = 0.41 \end{cases}$  hence the exact 95% C.I for  $P$  is  $[0.06, 0.41]$   
 ②  $\hat{P} = \frac{4}{20} = 0.2$ ;  $\alpha = 0.05 \Rightarrow Z \frac{\alpha}{2} = Z_{0.025} = 1.96$ . 95% C.I of  $P: 0.2 \pm 1.96 \times \sqrt{\frac{0.2 \times 0.8}{20}}$   
 95% C.I of  $P: 0.2 \pm 0.1753 \Rightarrow [0.0247, 0.3753]$

- ③ From the results in the above calculation, we get:  
 Both the upper and lower confidence limit of exact 95% C.I is greater than normal approximation 95% C.I.

- ① 5. Please briefly describe the Gehan's two stage design. Suppose the minimal response rate ( $p_0$ ) is 0.3. How large must  $n$  be so that if there are 0 responses among  $n$  patients we are relatively confident that the response rate is not 30% or better for  $X \sim \text{Bin}(n, p)$  and  $\alpha = 0.025$ ? We want the 99% confidence interval for the response rate to be within  $\pm 16\%$  when a new drug is considered minimally effective at  $p_0 = 30\%$ , what about the sample size necessary for this degree of precision is? Please use these values to interpret the Gehan's two stage design.

① If the minimal acceptable response rate is given as  $\pi_0$ , then choose  $n_0$  patients in the 1st stage, such that  $n_0 = \frac{\ln \alpha}{\ln(1-\pi_0)}$ . If there are 0 responses among  $n_0$  patients, then stop and declare the treatment a failure. If there was at least 1 patient response, another  $n_0$  patients would be treated likewise.

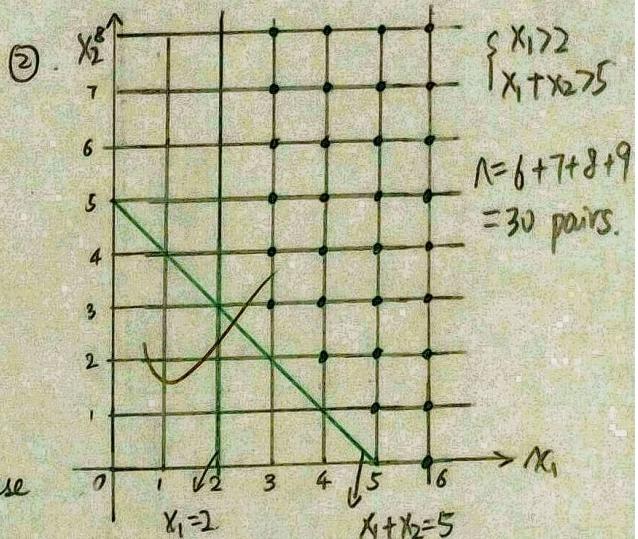
$$\text{② } n = \frac{\ln \alpha}{\ln(1-\pi_0)} = \frac{\ln 0.025}{\ln 0.7} = \frac{-3.689}{-0.3507} \approx 11. \quad \text{③ } \frac{\alpha}{2} = 0.005 \quad Z \frac{\alpha}{2} = 2.575 \quad \pi_0 = 0.3$$

$$Z \frac{\alpha}{2} \cdot \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = 0.16 \Rightarrow n = 54.5 \approx 55.$$

- ④ In this example, Gehan's design would treat 11 patients in the first stage. If none responded, the treatment would be declared a failure and study stop. If at least 1 response, another 11 patients would be treated, and a 99% C.I for  $P$  would be computed using 55 patients.

① 6. Please briefly describe the Simon's two stage design. Suppose  $X_1$  is the number of patients who respond in the first stage and  $X_2$  is the number of patients who respond in the second stage where  $n_1 = 6$ ,  $n=14$ .  $X_1 > 2$  and  $X_1 + X_2 > 5$ . Please draw the figure to show the combinations that satisfy the two conditions. How many pairs (combinations) satisfy these conditions?

② Let  $r_1$  and  $r$  denote some given number related to the clinical trial design, if  $X_1 \leq r_1$  in the 1st stage, declare the treatment a failure. If  $X_1 > r_1$  in the 1st stage and after two stages  $X_1 + X_2 > r$ , the treatment is declared a success. If  $X_1 > r$  in the 1st stage, declare the treatment success and there is no need to proceed to 2nd stage.



There are 30 pairs satisfy these conditions

7. For a Hierarchical Model,  $X_i | n_i, \pi_i \sim \text{Bin}(n_i, \pi_i)$ ,  $i = 1, \dots, N$  and

$\pi_1, \dots, \pi_N$  are iid ( $\mu_\pi, \delta_\pi^2$ ). Let  $p_i = \frac{x_i}{n_i}$  denote the sample proportion that respond from the  $i$ th study. What about conditional expectation and conditional variance of  $p_i$  given  $n_i, \pi_i$  (i.e.,  $E(p_i | n_i, \pi_i)$  and  $\text{Var}(p_i | n_i, \pi_i)$ )? What about the expectation and variance of  $p_i$  (i.e.,  $E(p_i)$  and  $\text{Var}(p_i)$ )? What's the estimate for  $\delta_\pi^2$ ?

$$\textcircled{1} \quad E(p_i | n_i, \pi_i) = E\left(\frac{X_i}{n_i}\right) = \frac{\mathbb{E}(X_i)}{n_i} = \frac{n_i \pi_i}{n_i} = \pi_i$$

$$\text{Var}(p_i | n_i, \pi_i) = \text{Var}\left(\frac{X_i}{n_i}\right) = \frac{1}{n_i^2} \text{Var}X_i = \frac{1}{n_i^2} \cdot n_i \pi_i (1 - \pi_i) = \frac{\pi_i(1 - \pi_i)}{n_i}$$

$$\textcircled{2} \quad \text{Var}(p_i | n_i, \pi_i) = E(p_i^2 | n_i, \pi_i) - E^2(p_i | n_i, \pi_i) = E(p_i^2) - p_i | n_i, \pi_i + E(p_i | n_i, \pi_i) - E^2(p_i | n_i, \pi_i)$$

$$E(p_i^2) - p_i | n_i, \pi_i = \text{Var}(p_i | n_i, \pi_i) + E^2(p_i | n_i, \pi_i) - E(p_i | n_i, \pi_i) = \frac{\pi_i(1 - \pi_i)}{n_i} + \pi_i^2 - \pi_i$$

$$E(p_i^2) - p_i | n_i, \pi_i = \frac{n_i - 1}{n_i} \pi_i (1 - \pi_i) \Rightarrow E\left(\frac{p_i(p_i - 1)}{n_i - 1} | n_i, \pi_i\right) = \frac{\pi_i(1 - \pi_i)}{n_i} \quad \dots \text{equation } \textcircled{A}$$

$$E(p_i) = E\{E(p_i | n_i, \pi_i)\} = E(\pi_i) = \mu_\pi$$

$$\text{Var}(p_i) = E\{\text{Var}(p_i | n_i, \pi_i)\} + \text{Var}\{E(p_i | n_i, \pi_i)\} = E\left\{\frac{\pi_i(1 - \pi_i)}{n_i}\right\} + \text{Var}\pi_i$$

$$\text{Applying } \textcircled{A} = E\{E\left(\frac{p_i(p_i - 1)}{n_i - 1}\right) | n_i, \pi_i\} + \text{Var}\pi_i = E\left(\frac{p_i(p_i - 1)}{n_i - 1}\right) + \sigma_\pi^2$$

$$= \frac{\sum_{i=1}^N p_i(p_i - 1)}{N} + \sigma_\pi^2$$

$$\textcircled{3} \quad \sigma_\pi^2 = \text{Var}(p_i) - E\left\{\frac{\pi_i(1 - \pi_i)}{n_i}\right\} = \text{Var}(p_i) - \frac{\sum_{i=1}^N \frac{p_i(1 - \pi_i)}{n_i}}{N}$$

$$\sigma_\pi^2 = E(\text{Var}p_i) - \frac{\sum_{i=1}^N \frac{p_i(1 - \pi_i)}{n_i - 1}}{N} = S_p^2 - \frac{\sum_{i=1}^N \frac{p_i(1 - \pi_i)}{n_i - 1}}{N} = \frac{\sum_{i=1}^N (p_i - \bar{p})^2}{N-1} - \frac{\sum_{i=1}^N p_i(1 - \bar{p})}{N}$$

8. Example for results of one particular treatment for lung cancer.

Study	# of patients ( $n_i$ )	% response rate ( $p_i - \bar{p}$ ) <sup>2</sup>	$P_i(1-p_i)/(n_i - 1)$
1. A study	12	65	0.136
2. B study	52	12	0.026
3. C study	26	5	0.054
4. D study	39	42	0.019
5. E study	148	35	0.005
6. F study	10	17	0.124
7. G study	47	21	0.051
			$\sum = 0.251$
			$\sum = 0.0519$

Using this example to estimate  $\delta^2_{\pi}$ .

$$\hat{\sigma}_{\pi}^2 = S_p^2 - \frac{\sum_{i=1}^N p_i(1-p_i)}{N} = \frac{\sum_{i=1}^N (p_i - \bar{p})^2}{N-1} - \frac{\sum_{i=1}^N p_i(1-p_i)}{N} = \frac{0.251}{7-1} - \frac{0.0519}{7}$$

$$= 0.0428 - 0.0074 = 0.0354$$

From the result, we conclude that the data is not good because the major part is from  $\hat{\sigma}_{\pi}^2$ .

9-10. There is a clinical trial regarding the Hypertension disease: two treatments are involved in this study (Treatment A and Treatment B), and the DBP values of 30 subjects are recorded at time point 1 and time point 5 for two treatments.

Subject	TRT	DBP1	DBP5	Age	Sex
1	A	114	105	43	F
2	A	116	101	51	M
3	A	119	98	48	F
4	A	115	101	42	F
5	A	116	105	49	M
6	A	117	102	47	M
7	A	118	99	50	F
8	A	120	102	61	M
9	A	114	103	43	F
10	A	115	97	51	M
11	A	117	101	47	F
12	A	116	102	45	M
13	A	119	104	54	F
14	A	118	99	52	M
15	A	115	102	42	F
16	B	114	113	39	M
17	B	116	110	40	F
18	B	114	109	39	F
19	B	114	115	38	M
20	B	116	109	39	F

(5)

21	B	114	110	41	M
22	B	119	115	56	F
23	B	118	112	56	M
24	B	114	108	38	F
25	B	120	113	57	M
26	B	117	115	47	F
27	B	118	110	48	M
28	B	121	115	61	F
29	B	116	111	49	M
30	B	118	112	52	M

9. Please design the hypothesis and test the mean difference of DBP5 between treatment A and treatment B, and write down the formula, results and your codes? (Please write down all equations and statistics)

$$\text{Test: } \begin{cases} H_0: \mu_1 = \mu_2 \\ H_a: \mu_1 \neq \mu_2 \end{cases}$$

Formula:

$$SD_1 = \sqrt{\frac{\sum_{j=1}^n (y_{1j} - \bar{y}_{1.})^2}{n_1 - 1}}$$

$$F^* = \frac{SS_{\text{between}} / (2-1)}{SS_{\text{within}} / (n_1 + n_2 - 2)}$$

$$SD_2 = \sqrt{\frac{\sum_{j=1}^n (y_{2j} - \bar{y}_{2.})^2}{n_2 - 1}}$$

$$SS_{\text{between}} = \sum_{i=1}^2 n_i (\bar{y}_{i.} - \bar{y})^2$$

$$SS_{\text{within}} = \sum_{i=1}^2 (n_i - 1) SD_i^2$$

should be verified the assumption  
of ANOVA

- 3

From the R output:

$$\begin{cases} \bar{y}_{1.} = 101.4 \\ \bar{y}_{2.} = 111.8 \end{cases} \quad \begin{cases} SD_1^2 = 5.686 \\ SD_2^2 = 6.0286 \end{cases}$$

$$\bar{y} = 106.6$$

$$SS_{\text{between}} = 811.2$$

$$SS_{\text{within}} = 164$$

$$\begin{cases} F^* = 138.5 \\ F_{(0.95, 1, 28)} = 4.2 \end{cases} \Rightarrow F^* > F_{(0.95, 1, 28)}$$

In this case, we conclude  $H_a$ .

10. For each treatment, DBP1 and DBP5 will be paired data, please use three methods we have learned in the class to test the difference between DBP1 and DBP5 for treatment A?

Please see the attachment.

(-B)

## Q9-R code

```
> data=matrix(0,15,3)
> data=matrix(0,30,3)
> data[,1]=c(114,116,119,115,116,117,118,120,114,115,117,116,119,118,115,114,
116,114,116,114,119,118,114,120,117,118,121,116,118)
> data[,2]=c(105,101,98,101,105,102,99,102,103,97,101,102,104,99,102,113,110,
109,115,109,110,115,112,108,113,115,110,115,111,112)
> data[,3]=c('A','A','A','A','A','A','A','A','A','A','A','A','A','A','A','A','B',
'B','B','B','B','B','B','B','B','B','B','B','B','B','B')
> write.table(data,'data.midterm.txt',quote = F)
> data=read.table('data.midterm.txt')
> data
   V1  V2 V3
1 114 105 A
2 116 101 A
3 119  98 A
4 115 101 A
5 116 105 A
6 117 102 A
7 118  99 A
8 120 102 A
9 114 103 A
10 115  97 A
11 117 101 A
12 116 102 A
13 119 104 A
14 118  99 A
15 115 102 A
16 114 113 B
17 116 110 B
18 114 109 B
19 114 115 B
20 116 109 B
21 114 110 B
22 119 115 B
23 118 112 B
24 114 108 B
25 120 113 B
26 117 115 B
27 118 110 B
28 121 115 B
29 116 111 B
30 118 112 B
> id.a=which(data[,3]=='A')
> id.b=which(data[,3]=='B')
> id.a
[1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
> id.b
[1] 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
> data.a=data[id.a,]
> data.b=data[id.b,]
> data.a;data.b
   V1  V2 V3
1 114 105 A
2 116 101 A
3 119  98 A
4 115 101 A
```

```

5 116 105 A
6 117 102 A
7 118 99 A
8 120 102 A
9 114 103 A
10 115 97 A
11 117 101 A
12 116 102 A
13 119 104 A
14 118 99 A
15 115 102 A
    V1 V2 V3
16 114 113 B
17 116 110 B
18 114 109 B
19 114 115 B
20 116 109 B
21 114 110 B
22 119 115 B
23 118 112 B
24 114 108 B
25 120 113 B
26 117 115 B
27 118 110 B
28 121 115 B
29 116 111 B
30 118 112 B
> mean.a=mean(data.a[,2])
> mean.b=mean(data.b[,2])
> mean.a
[1] 101.4
> mean.b
[1] 111.8
> var.a=var(data.a[,2])
> var.b=var(data.b[,2])
> var.a
[1] 5.685714
> var.b
[1] 6.028571
> na=nrow(data.a)
> nb=nrow(data.b)
> ss.in=(na-1)*var.a+(nb-1)*var.b
> mean=mean(data[,2])
> mean
[1] 106.6
> ss.bt=na*(mean.a-mean)*(mean.a-mean)+nb*(mean.b-mean)*(mean.b-mean)
> ss.bt
[1] 811.2
> F=(ss.bt/(2-1))/(ss.in/(na+nb-2))
> F
[1] 138.4976
> ss.in
[1] 164
> F.critical=qf(.95, df1=1, df2=na+nb-2)
> F.critical
[1] 4.195972

```

②

## Q10-METHOD I:

```

> data.aa=cbind(data.a[,1:2],data.a[,1]-data.a[,2])
> data.aa
   V1   V2 data.a[, 1] - data.a[, 2]
1 114 105          9
2 116 101         15
3 119 98          21
4 115 101         14
5 116 105         11
6 117 102         15
7 118 99          19
8 120 102         18
9 114 103         11
10 115 97          18
11 117 101         16
12 116 102         14
13 119 104         15
14 118 99          19
15 115 102         13
> n=nrow(data.aa)
> n
[1] 15
> mean.a.d=mean(data.aa[,3])
> mean.a.d
[1] 15.2
> var.a.d=var(data.aa[,3])
> var.a.d
[1] 11.45714
> sem.d=sqrt(var.a.d/n)
> sem.d
[1] 0.873962
> T.d=mean.a.d/sem.d
> T.d
[1] 17.39206
> alpha=0.05
> t.critical=qt(1-alpha/2,n-1)
> t.critical
[1] 2.144787

```

Test:  $\begin{cases} H_0: d = 0 \\ H_a: d \neq 0 \end{cases}$   $d = \bar{d}_{11} - \bar{d}_{12}$

RR:  $t_{obs} > t_{0.975, 14}$

In this case,

$t_{obs} = 17.4 > t_{0.975, 14} = 2.145$

We conclude  $H_a$ .

From the output:  $SD_d = 11.457$

$$\bar{d} = 15.2$$

$$SEM = 0.874$$

$$t_{obs} = 17.4$$

$$t_{0.975, 14} = 2.201$$

$$2.145$$

E1

Q10-Method II:

```
> data.a  
  V1  V2  V3  
1  114 105 A  
2  116 101 A  
3  119 98 A  
4  115 101 A  
5  116 105 A  
6  117 102 A  
7  118 99 A  
8  120 102 A  
9  114 103 A  
10 115 97 A  
11 117 101 A  
12 116 102 A  
13 119 104 A  
14 118 99 A  
15 115 102 A  
  
> mean.x=mean(data.a[,1])  
> mean.y=mean(data.a[,2])  
> mean.x  
[1] 116.6  
> mean.y  
[1] 101.4  
> var.x=var(data.a[,1])  
> var.y=var(data.a[,2])  
> var.x  
[1] 3.542857  
> var.y  
[1] 5.685714  
> sd.x=sd(data.a[,1])  
> sd.y=sd(data.a[,2])  
> sd.x  
[1] 1.882248  
> sd.y  
[1] 2.384474  
> corr.a=cor(data.a[,1],data.a[,2])  
> corr.a  
[1] -0.2482717  
> n=nrow(data.a)  
> n  
[1] 15  
> sem=sqrt((var.x+var.y-2*corr.a*sd.x*sd.y)/n)  
> sem  
[1] 0.873962  
> T.2=(mean.x-mean.y)/sem  
> T.2  
[1] 17.39206  
> t.critical  
[1] 2.144787
```

Test:  $\begin{cases} H_0: d=0 \\ H_a: d \neq 0 \end{cases}$

Formula:  $r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$

$$SEM = \sqrt{\frac{SD_x^2 + SD_y^2 - 2rSD_xSD_y}{n}}$$
$$t_{obs} = \frac{\bar{X} - \bar{Y}}{SEM}$$

From the output.

$$\bar{X} = 116.6$$

$$\bar{Y} = 101.4$$

$$r = -0.248$$

$$SD_x^2 = 3.54$$

$$SD_y^2 = 5.686$$

$$SD_x = 1.88$$

$$SD_y = 2.38$$

$$SEM = 0.874$$

$$t_{obs} = 17.397$$

$$t_{0.975, 14} = 2.145$$

RR:  $t_{obs} > t_{0.975, 14}$

In this case, since  $t_{obs} > t_{0.975, 14}$   
we conclude  $H_a$ .

(3)

```

Q10-Method III
> mean.within=apply(data.a[,1:2],1,mean)
> mean.within
  1   2   3   4   5   6   7   8   9   10  11  12
109.5 108.5 108.5 108.0 110.5 109.5 108.5 111.0 108.5 106.0 109.0 109.0
  13  14  15
111.5 108.5 108.5
> mean.group=apply(data.a[,1:2],2,mean)
> mean.group
  V1   V2
116.6 101.4
> ss.within=sum((data.a[,1:2]-mean.within)^2)
> ss.within
[1] 1813
> mean.all=mean(mean.within)
> mean.all
[1] 109
> n=nrow(data.a)
> n
[1] 15
> ss.trt=sum(n*(mean.group-mean.all)^2)
> ss.trt
1732.8
> ss.res=ss.within-ss.trt
> ss.res
[1] 80.2
> F.obs=(ss.trt/1)/(ss.res/(n-1))
> F.obs
[1] 302.4838
> F.critical=qf(0.95,1,n-1)
> F.critical
[1] 4.60011

```

Test  $\left\{ \begin{array}{l} H_0: d = 0 \\ H_a: d \neq 0 \end{array} \right.$

Formula:

$$SS_{\text{within subject } i} = (X_i - \bar{X}_{\cdot i})^2 + (Y_i - \bar{Y}_{\cdot i})^2$$

$$SS_{\text{within}} = \sum_{i=1}^n SS_{\text{within subject } i}$$

$$SS_{\text{treatment}} = n(\bar{X} - \bar{T})^2 + n(\bar{Y} - \bar{T})^2$$

where  $\bar{T}$  denotes the overall mean.

$$SS_{\text{residual}} = SS_{\text{within}} - SS_{\text{treatment}}$$

$$F_{\text{obs}} = \frac{SS_{\text{treatment}} / df_{\text{treatment}}}{SS_{\text{residual}} / df_{\text{residual}}}$$

$$\left\{ \begin{array}{l} df_{\text{treatment}} = 1 \\ df_{\text{residual}} = n-1 \end{array} \right.$$

$n$  is # of pairs

From the R output

$$SS_{\text{within}} = 1813$$

$$SS_{\text{treatment}} = 1732.8$$

$$SS_{\text{residual}} = 80.2$$

~~$F_{\text{obs}} = 302.48$~~

~~$F_{\text{obs}} \rightarrow F_{0.95, 1, 14}$~~

In this case. Since  
 $F_{\text{obs}} = 302.48 \rightarrow F_{0.95, 1, 14} = 4.6$   
we conclude  $H_a$ .