

# STA305/1004 - Homework #3 - Solutions

*Due Date: March 20, 2017, 22:00*

If you work with other students on this assignment then:

- indicate the names of the students on your solutions;
- your solutions must be written up independently (i.e., your solutions should not be the same as another student's solutions).

1. Suppose that twenty independent two-sample t tests are conducted at the 0.05 level, and that all the null hypotheses are true.

- (a) What is the distribution of the total number of tests that will be significant? How many tests do you expect would produce a significant result? Briefly explain.

Let  $X$  be the total number of significant tests. Since  $H_0$  is true in every test, the probability that a test rejects is 0.05, and the tests are independent. This implies that  $X \sim \text{Bin}(20, 0.05)$ .  $E(X) = 20(0.05) = 1$ .

- (b) The following R code simulates two independent t-tests based on sample sizes of 100 when  $H_0$  is true and calculates the probability at least one  $H_0$  is rejected.

```
y1 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y2 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
sum(y1<=0.05|y2<=0.05)/1000
```

NB: The R code `sum(y1<=0.05 | y2 <=0.05)` means add up all the elements of `y1` or `y2` that are less than or equal to 0.05. So for example the code below checks if each element of `y1` or `y2` is less than or equal to 0.05. If at least one of the elements is less than or equal to 0.05 then it is counted in the sum. Here is a simple example:

```
y1 <- c(0.01,0.06,0.03)
y2 <- c(0.05,0.09,0.07)
y1<=0.05|y2<=0.05
```

```
## [1] TRUE FALSE TRUE
```

```
sum(y1<=0.05|y2<=0.05)
```

```
## [1] 2
```

Extend the simulation above to calculate the probability that at least one null hypothesis is rejected among the twenty studies. Does this probability correspond to a type I or type II error? Briefly explain.

```
set.seed(303)
y1 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y2 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y3 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y4 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y5 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y6 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y7 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y8 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y9 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y10 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y11 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y12 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y13 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y14 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y15 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y16 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y17 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y18 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y19 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
y20 <-replicate(1000,t.test(rnorm(100,0,1),rnorm(100,0,1))$p.value)
```

```
sum(y1<=0.05|y2<=0.05|y3<=0.05|y4<=0.05|y5<=0.05|y6<=0.05|y7<=0.05|y8<=0.05|y9<=0.05|y10<=0.05|y11<=0.05|y12<=0.05|y13<=0.05|y14<=0.05|y15<=0.05|y16<=0.05|y17<=0.05|y18<=0.05|y19<=0.05|y20<=0.05)
```

```
## [1] 0.625
```

This probability is a type I error rate since it is the probability of rejecting when  $H_0$  is true.

(c) Suppose that  $k$  independent t-tests are conducted at level  $\alpha$ . Show that

$$P(\text{at least one } H_0 \text{ rejected}) = 1 - (1 - \alpha)^k.$$

This is called the experimentwise error rate. Verify the results of part (b) by choosing the appropriate values of  $\alpha$  and  $k$ .

$$\begin{aligned} P(\text{at least one } H_0 \text{ rejected}) &= 1 - P(\text{none of } H_0 \text{ rejected}) \\ &= 1 - [P(H_{0_1} \text{ not rejected}) P(H_{0_2} \text{ not rejected}) \cdots P(H_{0_k} \text{ not rejected})] \\ &= 1 - [(1 - \alpha)(1 - \alpha) \cdots (1 - \alpha)] \\ &= 1 - (1 - \alpha)^k \end{aligned}$$

In part (c) we have  $k = 20$  and  $\alpha = 0.05$  so

```
1-(1-.05)^(20)
```

```
## [1] 0.6415141
```

The simulation yields 0.625. The accuracy of the simulation would increase if the number of simulations were increased.

(d) Does the experimentwise error rate increase or decrease as  $k$  increases for a fixed  $\alpha$ ? Interpret this probability in terms of conducting multiple hypothesis tests.

The probability increases as  $k$  increases. The experimentwise error rate is the probability that at least one of the tests will incorrectly reject  $H_0$ . There is a 0.64 probability that this will be the case.

(e) Consider a study of the effects of a person's sex (e.g., male or female) on salary at a university. In such a case, it would be impossible to assign subjects to be men or women, but a nonrandomized study might be conducted by carefully matching males to females on factors such as age and years of experience. In one such study, fifty variables were recorded for each subject such as promotion, external awards, articles cited, and number of children born. After termination of the study, the male and female salaries were compared on each of these fifty variables, and it was found that there was a "significant difference (at the 1% level)" in the salary between men and women that had a large number of highly cited articles during their career. Explain why there might be a problem with this "significant finding".

The probability that at least one one null hypothesis will be incorrectly rejected is at most

```
1-(1-0.01)^50
```

```
## [1] 0.3949939
```

There is a 0.39 probability of mistakenly declaring that at least one of these fifty results is significant when it is not. So, there is a reasonable probability that the difference in salary between men and women with large numbers of highly cited articles is due to chance.

2. Suppose that we would like to compare sales from four different versions of the landing page of a company's website. The company would like to test if there is evidence of a difference in sales between different landing pages. Assume that sales on each landing page are normally distributed with means of  $\mu_1 = 55, \mu_2 = 58, \mu_3 = 50$ , and  $\mu_4 = 60$  ( $\mu_i$  is the mean sales of landing page  $i = 1, \dots, 4$ ),  $\alpha = 0.05$  and that a reasonable estimate of the error variance is  $\sigma^2 = 50$ .

(a) What are the null and alternative hypotheses?

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$  versus  $H_1 : \mu_i \neq \mu_j$ , for at least one pair  $i, j$  such that  $i \neq j$ .

(b) Suppose that each page can generate ten sales. Calculate the power of the test directly using the non-central F distribution. Make sure to specify how you calculated the non-centrality parameter and degrees of freedom. Hand in your R output.

In this case there are  $k = 4$  means and  $N = 4(10) = 40$  observations in total. So the numerator df is  $4-1=3$  and the denominator df is  $40-4=36$ . The rejection region is defined by  $F_{3,36,.05}$  which is given by:

```
qf(p = .95, df1 = 3, df2 = 36)
```

```
## [1] 2.866266
```

The non-centrality parameter is given by:

```
mu <- c(55,58,50,60)
mubar <- mean(mu)
delta <- sum(10*(mu-mubar)^2)/50
delta
```

```
## [1] 11.35
```

The power of the test is given by:

$$P(F_{3,36}(11.35) > 2.87)$$

```
1-pf(q = 2.87, df1 = 3, df2 = 36, ncp = 11.35)
```

```
## [1] 0.7686471
```

(c) What is the effect size that this study is trying to detect? How many observations in each group would be required to detect an effect size that is twice as large, as this effect size, to achieve the same power that you calculated in part (b) (for this calculation you may use one of the power functions built into R). Explain why the required sample size (per group) increases or decreases.

```
mu <- c(55,58,50,60)
mubar <- mean(mu)
eff.size <- sqrt(sum((mu-mubar)^2/4)/50)
library(pwr)
pwr.anova.test(k = 4, power=0.77, f = 2*eff.size)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##              k = 4
##              n = 3.395032
##              f = 1.065364
##      sig.level = 0.05
##              power = 0.77
##
## NOTE: n is number in each group
```

The sample size per group decreases since the effect size is the ratio of the standard deviation of the between group population means to the standard deviation of the within group population means. Larger values correspond to stronger evidence that between group variation is larger than within group variation therefore a smaller sample is needed to detect a larger effect size.

3. (Adapted from Box, Hunter and Hunter) A study was conducted to determine the effects of individual bathers on the fecal and total coliform bacterial populations in water. The variables of interest were the time since the subject's last bath, the vigor of the subject's activity in the water, and the subject's sex. The experiments were performed in a 100-gallon polyethylene tub using dechlorinated tap water at 38°C. The bacterial contribution of each bather was determined by subtracting the bacterial concentration measured at 15 and 30 minutes from that measured initially.

A replicated  $2^3$  factorial design was used for this experiment. The data obtained are presented below. (Note: Because the measurement of bacterial populations in water involves a dilution technique, the experimental errors do not have constant variance. Rather, the variation increases with the value of the mean.) Perform analysis using a logarithmic transformation of the data (the natural logarithm in R is obtained from the function `log()`).

The data can be read into R using the code:

```
prb0506 <- read.table(file = "prb0506.dat", header = T)
```

run	x1	x2	x3	y1	y2	y3	y4
1	-1	-1	-1	1	1	3	7
2	1	-1	-1	12	15	57	80
3	-1	1	-1	16	10	323	360
4	1	1	-1	4	6	183	193
5	-1	-1	1	153	170	426	590
6	1	-1	1	129	148	250	243
7	-1	1	1	143	170	580	450
8	1	1	1	113	217	650	735
9	-1	-1	-1	2	4	10	27
10	1	-1	-1	37	39	280	250
11	-1	1	-1	21	21	33	53
12	1	1	-1	2	5	10	87
13	-1	-1	1	96	67	147	193
14	1	-1	1	390	360	1470	1560
15	-1	1	1	300	377	665	810
16	1	1	1	280	250	675	795

Code	Name	Low Level	High Level
$x_1$	Time since last bath	1 hour	24 hours
$x_2$	Vigor of bathing activity	Lethargic	Vigorous
$x_3$	Sex of bather	Female	Male

Code	Name
$y_1$	Fecal coliform contribution after 15 minutes (organisms/100mL)
$y_2$	Fecal coliform contribution after 30 minutes (organisms/100mL)
$y_3$	Total coliform contribution after 15 minutes (organisms/100mL)
$y_4$	Total coliform contribution after 30 minutes (organisms/100mL)

- (a) Briefly explain why this is a  $2^3$  factorial design.

There are 3 factors each at 2 levels. The experiment was run at all factor-level combinations.

- (b) Calculate the factorial effects (main and interaction effects) on total coliform populations after 15 minutes. Hand in your R output. Interpret the main effect of  $x_1$ , and the interaction between  $x_1$  and

$x_3$ .

```
fact2 <- lm(log(y3)~x1*x2*x3,data=prb0506)
summary(fact2)

##
## Call:
## lm(formula = log(y3) ~ x1 * x2 * x3, data = prb0506)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4534 -0.6505  0.0000  0.6505  1.4534
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.9737     0.2927  16.994 1.46e-07 ***
## x1              0.4008     0.2927   1.370  0.20804
## x2              0.3564     0.2927   1.218  0.25801
## x3              1.2405     0.2927   4.239  0.00284 **
## x1:x2          -0.6050     0.2927  -2.067  0.07256 .
## x1:x3          -0.1635     0.2927  -0.559  0.59168
## x2:x3          -0.1070     0.2927  -0.366  0.72413
## x1:x2:x3       0.3999     0.2927   1.366  0.20901
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.171 on 8 degrees of freedom
## Multiple R-squared:  0.7772, Adjusted R-squared:  0.5823
## F-statistic: 3.987 on 7 and 8 DF,  p-value: 0.03551
#Factorial Effects - multiply regression parameter estimates by two
2*fact2$coefficients

## (Intercept)          x1          x2          x3          x1:x2          x1:x3
##   9.9473900   0.8016289   0.7127928   2.4810727  -1.2099336  -0.3270098
##           x2:x3      x1:x2:x3
##  -0.2140054   0.7997494
```

The main effect of  $x_1$  is mean difference between log total coliform after 15 minutes when time since last bath is 24 hours versus 1 hour.

The interaction between time since last bath and sex of bather is the difference of differences. That is the difference between:

- mean difference in log total coliform after 15 minutes between male and female bathers when time since last bath is 24 hours; and
- mean difference in log total coliform after 15 minutes between male and female bathers when time since last bath is 1 hour.

(c) Let  $w_1, w_2$  be the results from two individual runs under the same conditions from this  $2^3$  factorial design. For example, the responses for Fecal coliform contribution after 15 minutes ( $y_1$ ) under the condition  $x_1, x_2, x_3$  set at -1 -1 -1 are 1 and 2 respectively. (runs 1 and 9).

Show that the estimated variance  $s_i^2, i = 1, \dots, 8$  at the  $i^{th}$  conditions is:

$$s_i^2 = \frac{(w_1 - w_2)^2}{2}.$$

The variance of  $w_1, w_2$  is  $s_i^2 = \sum_{i=1}^2 \frac{(w_i - \bar{w})^2}{2-1}$ .

$$\begin{aligned}
s_i^2 &= \left( w_1 - \frac{(w_1 + w_2)}{2} \right)^2 + \left( w_2 - \frac{(w_1 + w_2)}{2} \right)^2 \\
&= \left( \frac{w_1}{2} - \frac{w_2}{2} \right)^2 + \left( \frac{w_2}{2} - \frac{w_1}{2} \right)^2 \\
&= \left( \frac{w_1 - w_2}{2} \right)^2 + \left( (-1) \frac{(w_1 - w_2)}{2} \right)^2 \\
&= 2 \left( \frac{w_1 - w_2}{2} \right)^2 \\
&= \frac{(w_1 - w_2)^2}{2}.
\end{aligned}$$

- (d) Use the result in part (c) to show that the variance of any of the factorial effects in a  $2^3$  factorial design with replicated runs is

$$V(effect) = \left( \frac{1}{8} + \frac{1}{8} \right) s^2,$$

where  $s^2 = \frac{\sum_{i=1}^8 s_i^2}{8}$ .

Let  $effect = \bar{y}_+ - \bar{y}_-$ , where for the first effect of a  $2^3$  design  $\bar{y}_+ = \frac{y_2 + y_4 + y_6 + y_8}{4}$  and  $\bar{y}_- = \frac{y_1 + y_3 + y_5 + y_7}{4}$ . Since the design is replicated each observation is the average of the two duplicate runs,  $y_i = \frac{y_{i1} + y_{i2}}{2}$ , and

$$\begin{aligned}
Var(y_i) &= Var\left(\frac{y_{i1} + y_{i2}}{2}\right) \\
&= \frac{1}{4} (\sigma^2 + \sigma^2) \\
&= \frac{\sigma^2}{2}
\end{aligned}$$

Therefore,

$$\begin{aligned}
Var(effect) &= Var\left(\frac{y_2 + y_4 + y_6 + y_8}{4} + \frac{y_1 + y_3 + y_5 + y_7}{4}\right) \\
&= \frac{1}{16} \left(4 \frac{\sigma^2}{2}\right) + \frac{1}{16} \left(4 \frac{\sigma^2}{2}\right) \\
&= \frac{1}{8} \sigma^2 + \frac{1}{8} \sigma^2
\end{aligned}$$

$\sigma^2$  can be estimated by  $s^2$ . Therefore  $Var(effect) = \left( \frac{1}{8} + \frac{1}{8} \right) s^2$ .