University of Toronto
Faculty of Arts and Science
April Examinations, 2016
Preview of STA305H1S/STA1004H
Duration - 3 hours

Name:

Student Number:

Preview note: All questions with R output are previewed. Question 9 does not contain output but a preview is provided. Questions 1,2,10 do not contain R output and are not previewed.

Aids allowed: 8.5'x11' sheet with hand writing (not typed) on both sides and a calculator.

Instructions: Complete all parts of the 10 questions. Show all your work. The value of each question is shown in the table below.

Question	Total
1	20
2	10
3	15
4	25
5	20
6	30
7	20
8	20
9	30
10	15
Total	205

1. **Answer the following questions (5 marks each, 20 marks).**
   This question does not have a preview

a)

b)

c)

d)

2. **Answer the following questions (5 marks each, 10 marks).**
   This question does not have a preview

a)

b)

3. **Describe the design in the following scenarios. Explain the rationale behind your description of the design. (5 marks each, 15 marks)**
   The design scenarios are NOT previewed

a)

b)

c)

4. Six burn treatments A, B, C, D, E, F were tested on six subjects. Each subject had six sites on which a burn could be applied for testing: on each arm two sites were below the elbow and one above. A standard burn was administered at each site and the six treatments were arranged so that each treatment occurred once with every subject once in every position. After treatment each burn was covered by a clean gauze: treatment C was a control with clean gauze but without other treatment. The data are the number of hours for a clearly defined degree of partial healing to occur. (5 marks each, 25 marks)

	Subjects					
Position on arm	1	2	3	4	5	6
I	A 32	B 40	C 72	D 43	E 35	F 50
II	B 29	A 37	F 59	E 53	D 32	C 53
III	C 40	D 56	A 53	B 48	F 37	E 43
IV	D 29	F 59	E 67	A 56	C 38	B 42
V	E 28	C 50	B 100	F 46	A 29	D 56
VI	F 37	E 42	D 67	C 50	B 33	A 48

Answer the following questions using the R output below.

a)

b)

c)

d)

e)

R output for burn treatments question.

```
#Mean and SD for Treatment A
mean(y[treatment=="A"])
```

```
## [1] 42.5
```

```
sd(y[treatment=="A"])
```

```
## [1] 11.36222
```

```
#Mean and SD for Treatment B
mean(y[treatment=="B"])
```

```
## [1] 48.66667
```

```r
sd(y[treatment=="B"])
```

```
## [1] 26.02819
```

```r
#Mean and SD for Treatment C
mean(y[treatment=="C"])
```

```
## [1] 50.5
```

```r
sd(y[treatment=="C"])
```

```
## [1] 12.12848
```

```r
#Mean and SD for Treatment D
mean(y[treatment=="D"])
```

```
## [1] 47.16667
```

```r
sd(y[treatment=="D"])
```

```
## [1] 15.01222
```

```r
#Mean and SD for Treatment E
mean(y[treatment=="E"])
```

```
## [1] 44.66667
```

```r
sd(y[treatment=="E"])
```

```
## [1] 13.77921
```

```r
#Mean and SD for Treatment F
mean(y[treatment=="F"])
```

```
## [1] 48
```

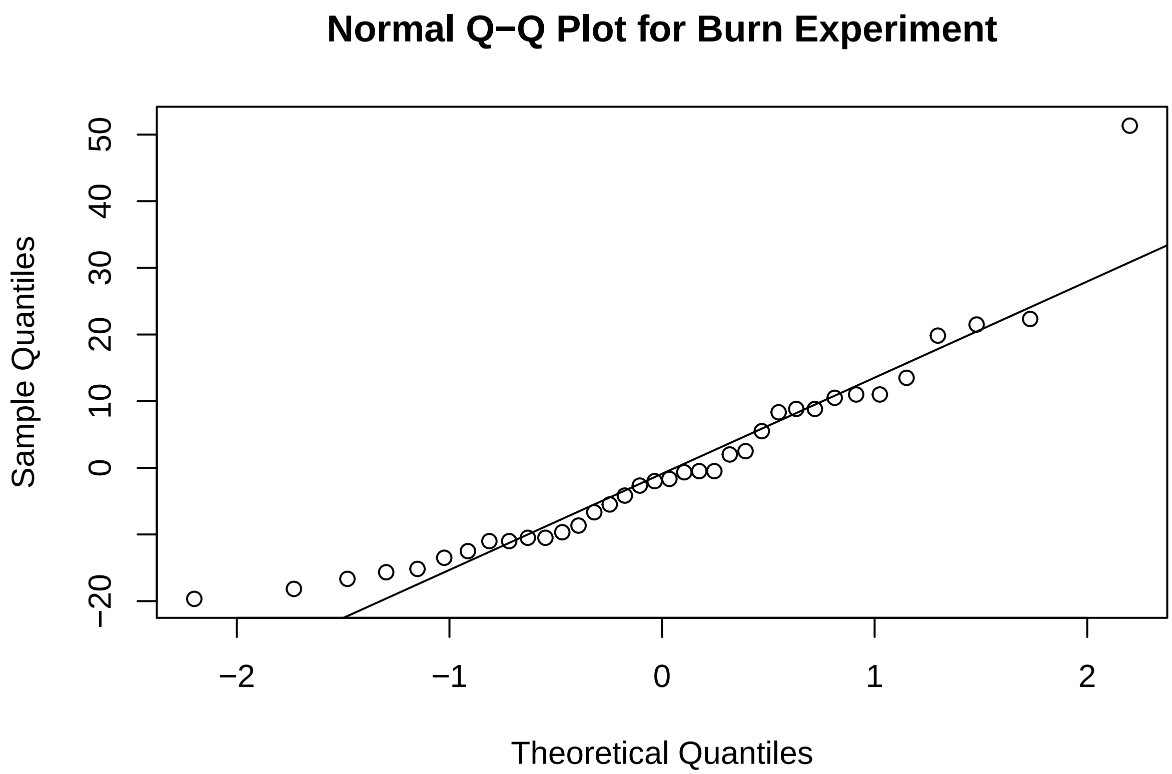
```r
sd(y[treatment=="F"])
```

```
## [1] 9.919677
```

```r
model1 <- lm(y~treatment,dat=prb0402)
anova(model1)
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value Pr(>F)
## treatment  5  250.3   50.05  0.2047  0.958
## Residuals 30 7336.5  244.55
```

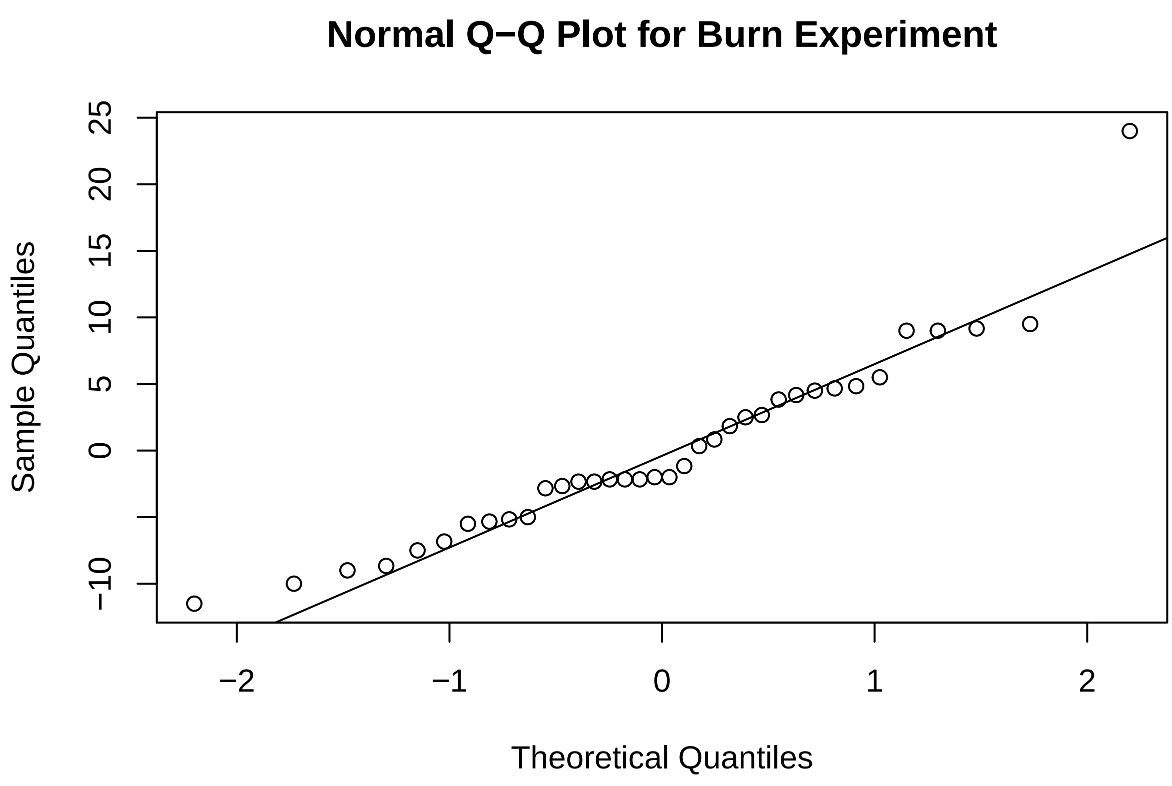
```r
qqnorm(model1$residuals,main="Normal Q-Q Plot for Burn Experiment")
qqline(model1$residuals)
```



```r
model2 <- lm(y~treatment+as.factor(position)+as.factor(subject),dat=prb0402)
anova(model2)
```

```
## Analysis of Variance Table
##
## Response: y
##                     Df Sum Sq Mean Sq F value    Pr(>F)
## treatment            5  250.3   50.05  0.5858    0.7107
## as.factor(position)  5  219.9   43.98  0.5148    0.7619
## as.factor(subject)   5 5407.9 1081.58 12.6600 1.248e-05 ***
## Residuals           20 1708.7   85.43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
qqnorm(model2$residuals,main="Normal Q-Q Plot for Burn Experiment")
qqline(model2$residuals)
```



5. Kowalski and Potcner (2003) described an experiment involving the water resistant property of wood. The two experimental factors are the type of wood pretreatment (A) and the type of stain (B). Two types of pretreatment and four types of stain are used in the experiment. An experimental unit is a wood panel to which a specific type of pretreatment is applied before a specific type of stain is applied. It was very inconvenient to apply the pretreatment to a small wood panel. So the experimenters selected two large panels and applied pretreatment A1 to one panel and pretreatment A2 to the other. Next each panel was cut into four smaller pieces and the different stain types (B1, B2, B3, B4) were applied to them. The assignment of pretreatment and stain type was done randomly.
   Three replications of the wood experiment were conducted, and some of the data are shown below.

Wood	Pretreatment.A.	Stain.B.	Replication.Rep.	Resistance
4	2	2	1	53.5
⋮	⋮	⋮	⋮	⋮
4	2	3	1	35.4
5	2	4	2	44.6
⋮	⋮	⋮	⋮	⋮
5	2	2	2	48.3
1	1	3	1	40.8
⋮	⋮	⋮	⋮	⋮
1	1	4	1	45.5
2	1	1	2	57.4
⋮	⋮	⋮	⋮	⋮
2	1	1	2	57.4
6	2	1	3	32.1
⋮	⋮	⋮	⋮	⋮
6	2	3	3	32.2
3	1	1	3	52.8
⋮	⋮	⋮	⋮	⋮
3	1	2	3	59.2

Answer the following questions using the R output below (5 marks each, 20 marks).

a)

b)

c)

d)

R output for wood experiment.

```r
A <- as.factor(Pretreatment.A.)
B <- as.factor(Stain.B.)
Replication <- as.factor(Replication.Rep.)

model1 <- lm(Resistance~A*B,data=WaterResistance)
anova(model1)
```

```
## Analysis of Variance Table
##
## Response: Resistance
##           Df Sum Sq Mean Sq F value Pr(>F)
## A          1    782     782   13.49 0.0021 **
## B          3    266      89    1.53 0.2454
## A:B        3     63      21    0.36 0.7820
## Residuals 16    928      58
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
model2 <- aov(Resistance~Replication+A+Replication:A+B+A:B+Error(A/Replication),data=WaterResis
summary(model2)
```

```
##
## Error: A
##   Df Sum Sq Mean Sq
## A  1    782     782
##
## Error: A:Replication
##               Df Sum Sq Mean Sq
## Replication    2    377     188
## Replication:A  2    398     199
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## B          3  266.0    88.7    6.98 0.0057 **
## A:B        3   62.8    20.9    1.65 0.2309
## Residuals 12  152.5    12.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# The 2.5th percentile of the F(1,2) distribution
qf(.025,1,2)
```

```
## [1] 0.001250782
```

```r
# The 5th percentile of the F(1,2) distribution
qf(.05,1,2)
```

```
## [1] 0.005012531
```

```r
# The 95th percentile of the F(1,2) distribution
qf(.95,1,2)
```

```
## [1] 18.51282
```

```r
# The 97.5th percentile of the F(1,2) distribution
qf(.975,1,2)
```

```
## [1] 38.50633
```

6. In studying a chemical reaction, a chemist performed a 2^3 factorial design and obtained the following results:

run	temp	conc	rate	yield
1	50	6	60	54
2	60	6	60	57
3	50	10	60	69
4	60	10	60	70
5	50	6	100	55
6	60	6	100	54
7	50	10	100	80
8	60	10	100	81

Answer the following questions using the R output below (5 marks each, 30 marks).

a)

b)

c)

d)

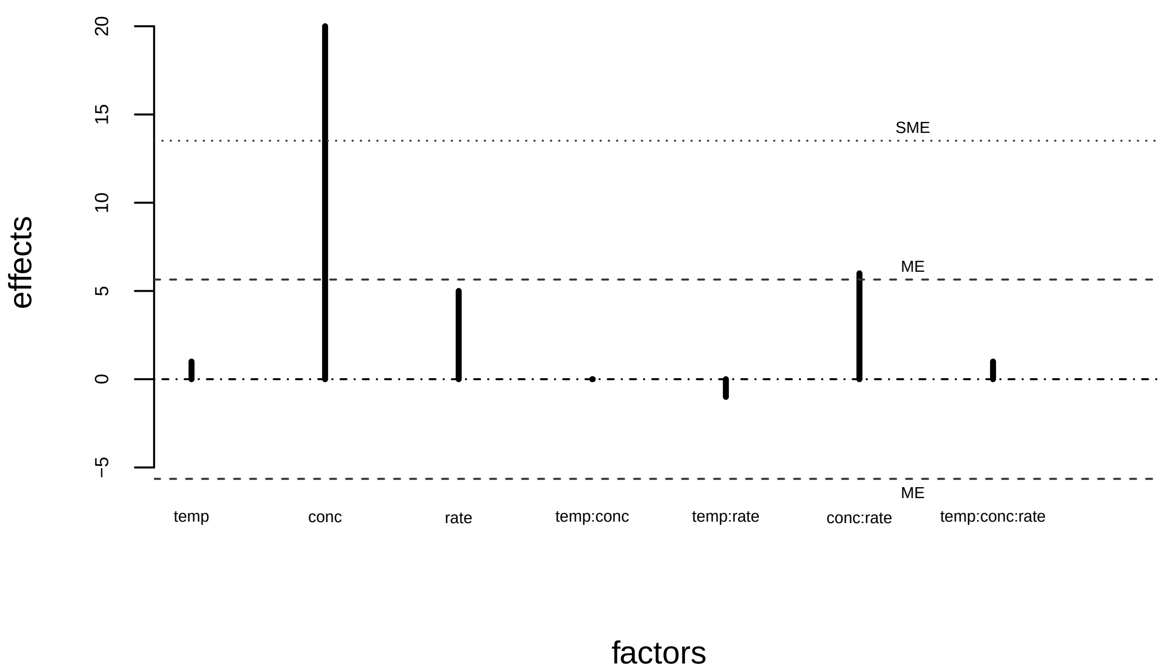
e)

f)

R output for chemical reaction data. NB: The notation `6.50e+01` means 6.5×10^1, `-2.08e-15` means -2.08×10^{-15}. Other values using this notation in the output below are defined in a similar manner.

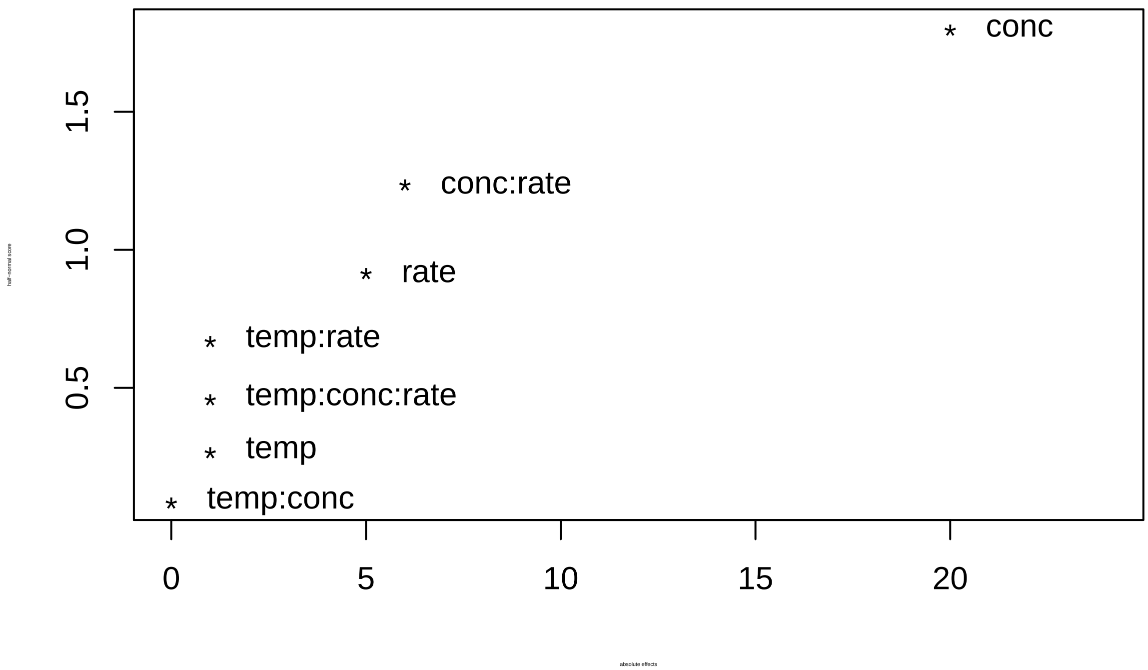
```
##
## Call:
## lm.default(formula = yield ~ temp * conc * rate, data = prb0510)
##
## Residuals:
## ALL 8 residuals are 0: no residual degrees of freedom!
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.50e+01         NA      NA       NA
## temp             5.00e-01         NA      NA       NA
## conc             1.00e+01         NA      NA       NA
## rate             2.50e+00         NA      NA       NA
## temp:conc       -2.08e-15         NA      NA       NA
## temp:rate       -5.00e-01         NA      NA       NA
## conc:rate        3.00e+00         NA      NA       NA
## temp:conc:rate   5.00e-01         NA      NA       NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:    NaN
## F-statistic:   NaN on 7 and 0 DF,  p-value: NA
```

```
LenthPlot(fact.prob4,cex.fac=0.5,cex.axis=.6,adj=0.8)
```



```
##      alpha        PSE         ME        SME
##   0.050000   1.500000   5.646185 13.512461
```

```r
DanielPlot(fact.prob4,half=T,cex.fac=.1,cex.lab=0.2)
```



7. The data for a 2^{k-p} fractional factorial design are shown below.

run	A	B	C	D	E	Obsy
1	-1	-1	-1	-1	1	14.8
2	1	-1	-1	-1	-1	14.5
3	-1	1	-1	-1	-1	18.1
4	1	1	-1	-1	1	19.4
5	-1	-1	1	-1	-1	18.4
6	1	-1	1	-1	1	15.7
7	-1	1	1	-1	1	27.3
8	1	1	1	-1	-1	28.2
9	-1	-1	-1	1	-1	16.0
10	1	-1	-1	1	1	15.1
11	-1	1	-1	1	1	18.9
12	1	1	-1	1	-1	22.0
13	-1	-1	1	1	1	19.8
14	1	-1	1	1	-1	18.9
15	-1	1	1	1	-1	29.9
16	1	1	1	1	1	27.4

Answer the following questions using the R output below (5 marks each, 20 marks)

a)

b)

c)

d)

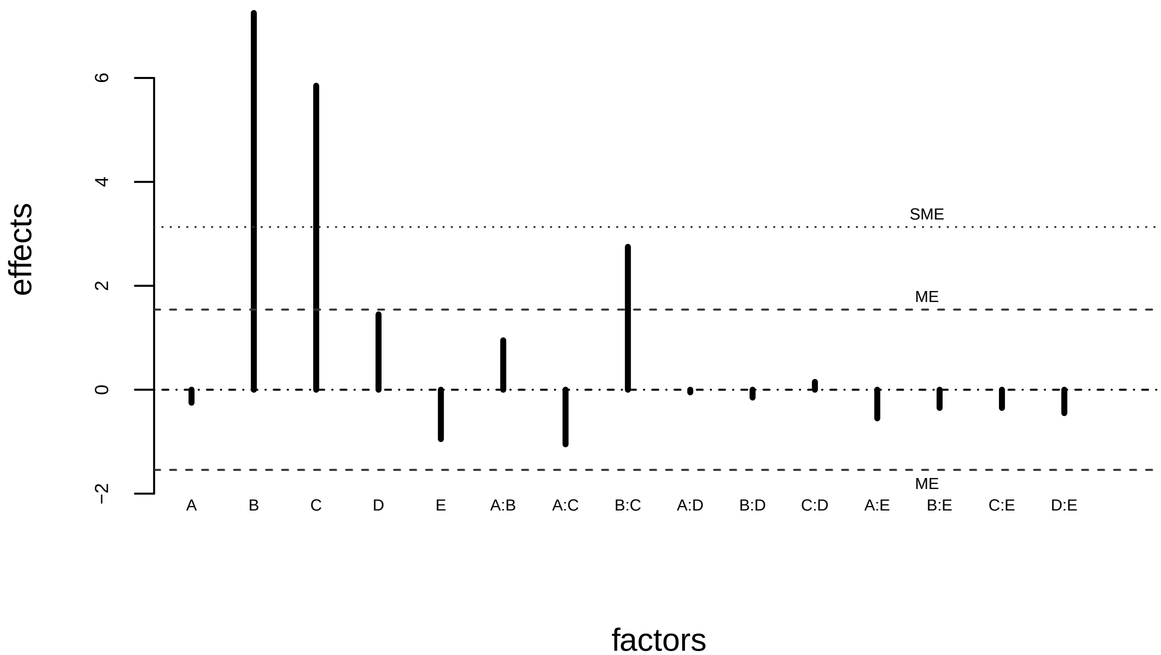
R output for fractional factorial experiment.

```
fact.final1 <- lm(Obsy~A*B*C*D*E,data=prb0602)
summary(fact.final1)
```

```
##
## Call:
## lm.default(formula = Obsy ~ A * B * C * D * E, data = prb0602)
##
## Residuals:
## ALL 16 residuals are 0: no residual degrees of freedom!
##
## Coefficients: (16 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    20.275         NA      NA       NA
## A              -0.125         NA      NA       NA
```

```
## B              3.625         NA       NA        NA
## C              2.925         NA       NA        NA
## D              0.725         NA       NA        NA
## E             -0.475         NA       NA        NA
## A:B            0.475         NA       NA        NA
## A:C           -0.525         NA       NA        NA
## B:C            1.375         NA       NA        NA
## A:D           -0.025         NA       NA        NA
## B:D           -0.075         NA       NA        NA
## C:D            0.075         NA       NA        NA
## A:E           -0.275         NA       NA        NA
## B:E           -0.175         NA       NA        NA
## C:E           -0.175         NA       NA        NA
## D:E           -0.225         NA       NA        NA
## A:B:C             NA         NA       NA        NA
## A:B:D             NA         NA       NA        NA
## A:C:D             NA         NA       NA        NA
## B:C:D             NA         NA       NA        NA
## A:B:E             NA         NA       NA        NA
## A:C:E             NA         NA       NA        NA
## B:C:E             NA         NA       NA        NA
## A:D:E             NA         NA       NA        NA
## B:D:E             NA         NA       NA        NA
## C:D:E             NA         NA       NA        NA
## A:B:C:D           NA         NA       NA        NA
## A:B:C:E           NA         NA       NA        NA
## A:B:D:E           NA         NA       NA        NA
## A:C:D:E           NA         NA       NA        NA
## B:C:D:E           NA         NA       NA        NA
## A:B:C:D:E         NA         NA       NA        NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1,  Adjusted R-squared:    NaN
## F-statistic:   NaN on 15 and 0 DF,  p-value: NA
```

```
fact.final1 <- aov(Obsy~A*B*C*D*E,data=prb0602)
LenthPlot(fact.final1,cex.fac=0.5,cex.axis=.6,adj=0.8)
```




```
##    alpha      PSE       ME      SME
## 0.050000 0.600000 1.542349 3.131191
```

8. A randomized clincal trial was conducted by randomizing subjects into three treatment groups: surgery(`trt=1`); surgery plus chemotherapy (`trt=2`); surgery plus chemotherapy plus radiation therapy (`trt=3`). The study was designed to have 80% power to detect differences in a continuous outcome (`y`) at the 5% significance level between:
(1) surgery and surgery plus chemotherapy;
(2) surgery and surgery plus chemotherapy plus raditaion therapy.

Answer the following questions using the R output below (20 marks) .

(a)

(b)

(c)

(d)

The R output for the clinical trial is below. The dataframe `rctdat` contains the outcome `y` and treatment variable `trt`

```
# Mean of y in each group
t(lapply(split(rctdat$y,rctdat$trt),mean))
```

```
##      Surg  SurgChemo SurgChemoRad
## [1,] 96.18 111.7     102.1
```

```
# SD of y in each group
t(lapply(split(rctdat$y,rctdat$trt),sd))
```

```
##      Surg  SurgChemo SurgChemoRad
## [1,] 7.223 11.12     10.58
```

`mod1`

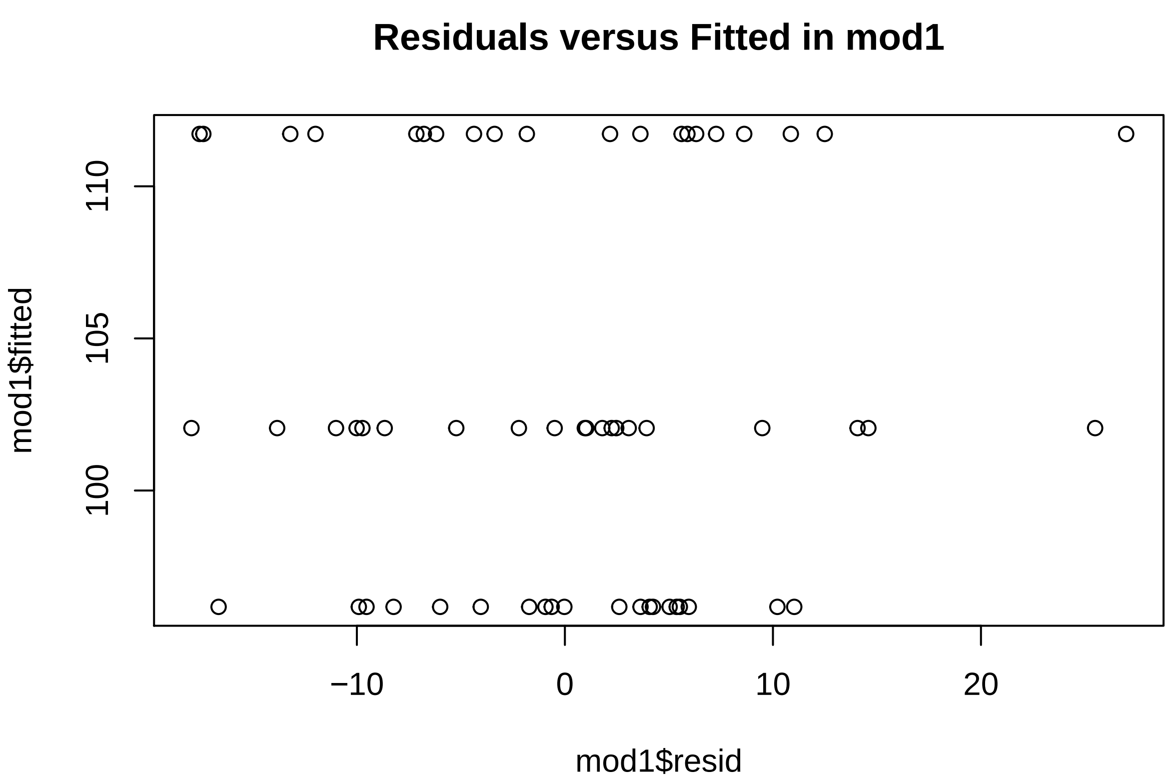
```
contrasts(rctdat$trt) <- contr.sum(3)
contr.sum(3)
```

```
##   [,1] [,2]
## 1    1    0
## 2    0    1
## 3   -1   -1
```

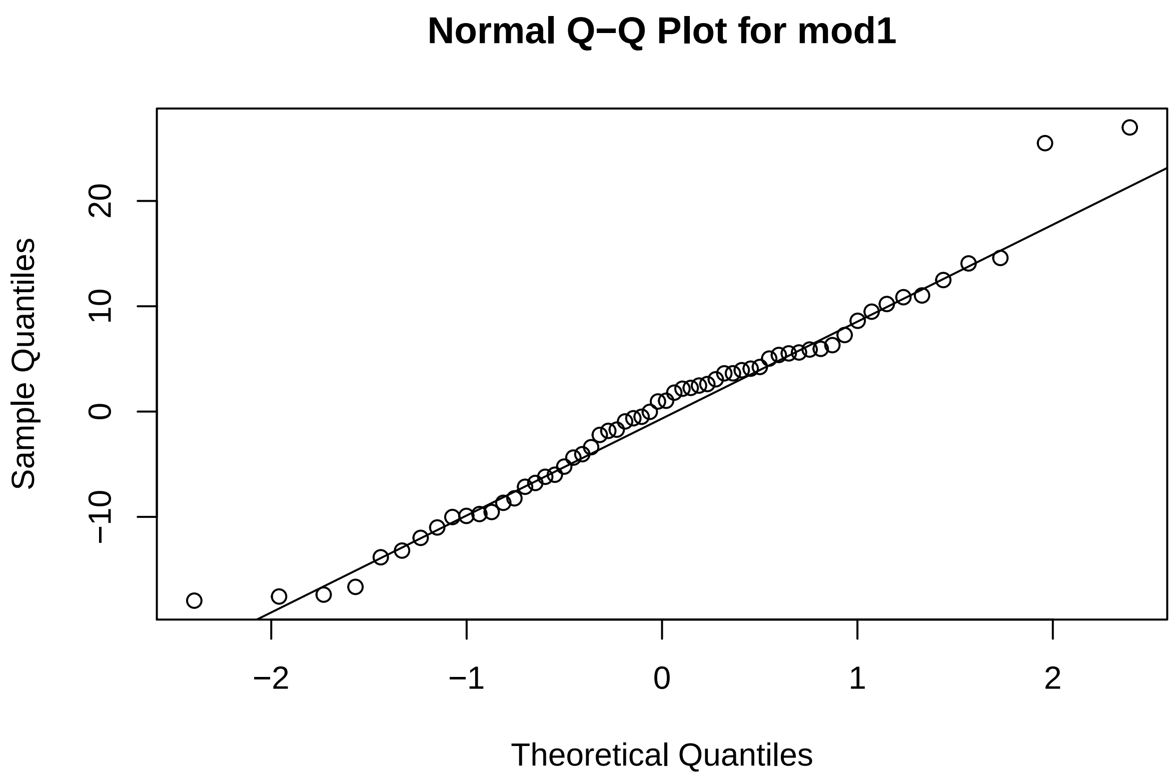
```
mod1 <- lm(y~trt,data=rctdat)
summary(mod1)
```

```
##
## Call:
## lm.default(formula = y ~ trt, data = rctdat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.953  -6.870   0.993   5.545  26.979
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   103.32       1.26   81.69  < 2e-16 ***
## trt1           -7.14       1.79   -3.99  0.00019 ***
## trt2            8.40       1.79    4.70  1.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.8 on 57 degrees of freedom
## Multiple R-squared:  0.311,  Adjusted R-squared:  0.286
## F-statistic: 12.8 on 2 and 57 DF,  p-value: 2.49e-05
```

```r
plot(mod1$resid,mod1$fitted,main="Residuals versus Fitted in mod1")
```




```r
qqnorm(mod1$resid,main="Normal Q-Q Plot for mod1");qqline(mod1$resid)
```




```
mod11 <- aov(y~trt,data=rctdat)
summary(mod11)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)    
## trt          2   2464    1232    12.8 2.5e-05 ***
## Residuals   57   5470      96                    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

mod2

```
contrasts(rctdat$trt) <- contr.treatment(3)
contr.treatment(3)
```

```
##   2 3
## 1 0 0
## 2 1 0
## 3 0 1
```

```
mod2 <- lm(y~trt,data=rctdat)
summary(mod2)
```

```
##
```

```
## Call:
## lm.default(formula = y ~ trt, data = rctdat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.953  -6.871   0.993   5.545  26.979
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    96.18       2.19   43.91  < 2e-16 ***
## trt2           15.54       3.10    5.02  5.4e-06 ***
## trt3            5.88       3.10    1.90    0.063 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.8 on 57 degrees of freedom
## Multiple R-squared:  0.311,  Adjusted R-squared:  0.286
## F-statistic: 12.8 on 2 and 57 DF,  p-value: 2.49e-05
```

```
mod22 <- aov(y~trt,data=rctdat)
summary(mod22)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## trt          2   2464    1232    12.8 2.5e-05 ***
## Residuals   57   5470      96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

mod3

```
contrasts(rctdat$trt) <- contr.treatment(3)
contr.helmert(3)
```

```
##   [,1] [,2]
## 1   -1   -1
## 2    1   -1
## 3    0    2
```

```
mod3 <- lm(y~trt,data=rctdat)
summary(mod3)
```

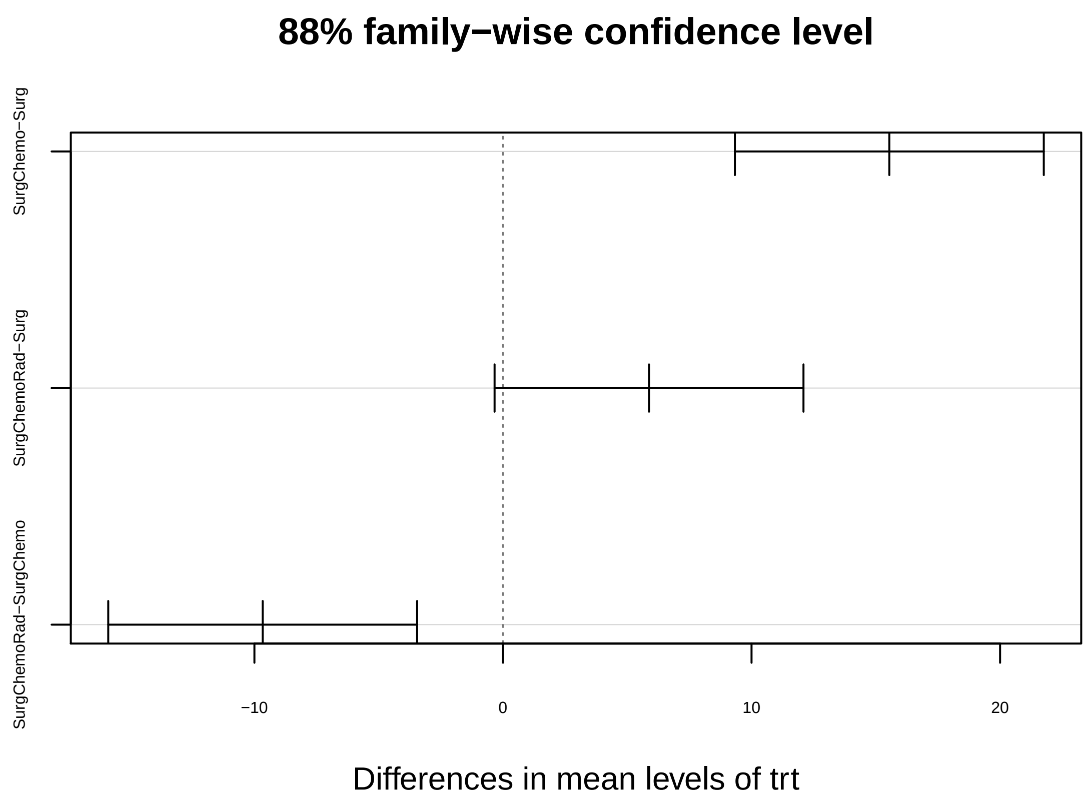
```
##
## Call:
## lm.default(formula = y ~ trt, data = rctdat)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.953  -6.871   0.993   5.545  26.979
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    96.18       2.19   43.91  < 2e-16 ***
## trt2           15.54       3.10    5.02  5.4e-06 ***
## trt3            5.88       3.10    1.90    0.063 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.8 on 57 degrees of freedom
## Multiple R-squared:  0.311,  Adjusted R-squared:  0.286
## F-statistic: 12.8 on 2 and 57 DF,  p-value: 2.49e-05
```

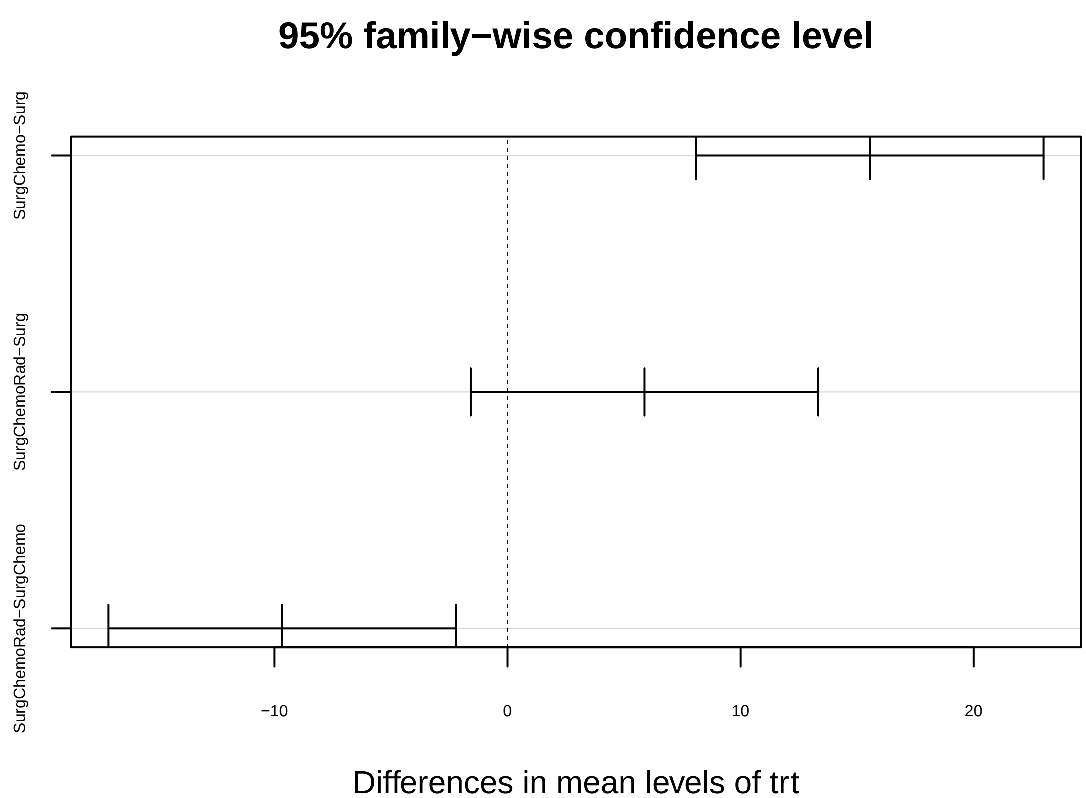
```r
mod33 <- aov(y~trt,data=rctdat)
summary(mod33)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## trt          2   2464    1232    12.8 2.5e-05 ***
## Residuals   57   5470      96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

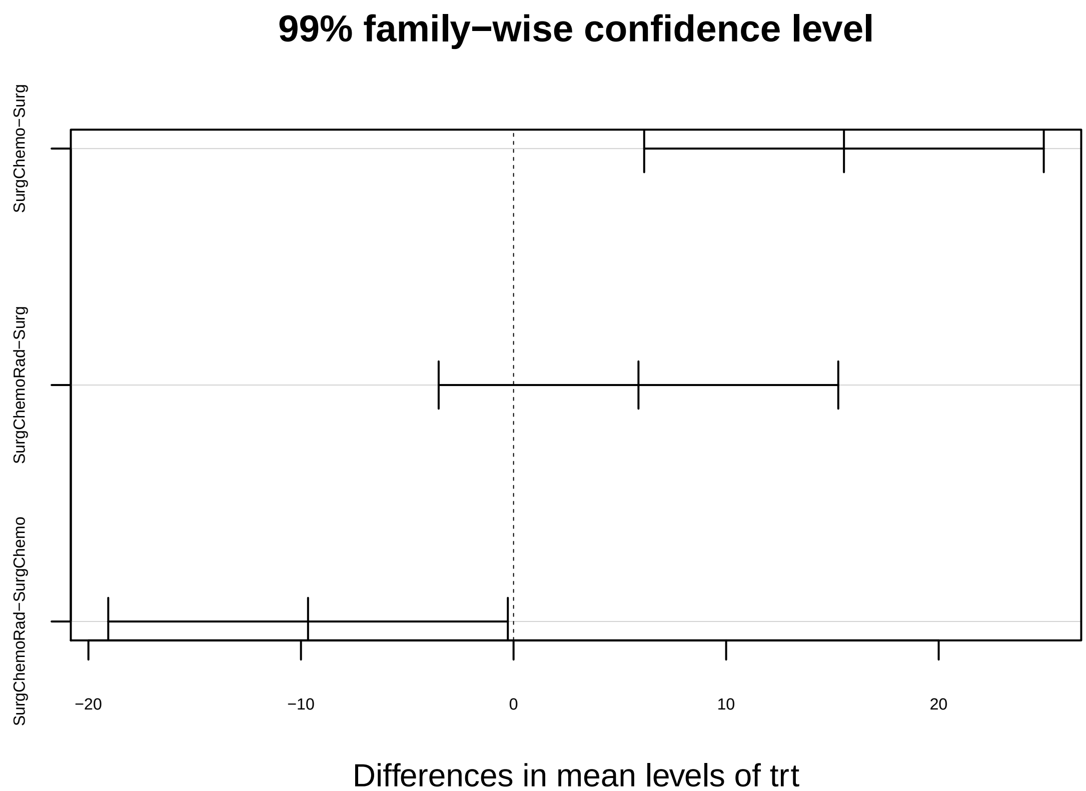
```r
#Adjustments for multiple comparisons

plot(TukeyHSD(mod11,conf.level=0.88),cex.axis = 0.5)
```




```r
plot(TukeyHSD(mod11,conf.level=0.95),cex.axis = 0.5)
```



```r
plot(TukeyHSD(mod11,conf.level=0.99),cex.axis = 0.5)
```




```r
pairwise.t.test(rctdat$y,rctdat$trt,p.adj="bonf")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  rctdat$y and rctdat$trt
##
##             Surg    SurgChemo
## SurgChemo   1.6e-05 -
## SurgChemoRad 0.1887 0.0085
##
## P value adjustment method: bonferroni
```

9. An injection molding process that is producing an unacceptable percentage of burned parts is being studied by two engineers (Sally and Zhenhua). There are three two-level factors to be studied: injection pressure (1200-1400 psi) (P); screw RPM control (0.3-0.6 turns counterclockwise) (R); and injection speed (slow-fast) (S). Sally conducts a factorial experiment and obtains the following results:

P	R	S	Percent Burned
1200	0.3	Slow	11
1200	0.3	Fast	17
1200	0.6	Slow	25
1200	0.6	Fast	29
1400	0.3	Slow	2
1400	0.3	Fast	9
1400	0.6	Slow	37
1400	0.6	Fast	40

Zhenhua claims that by designing a study to investigate one factor at a time it will require less observations to achieve the same precision (variance) compared to Sally's factorial experiment. Zhenhua proposed the following experimental design based on Sally's factorial experiment above.

Step 1. Factor P is thought to be the most important. By fixing the other factors at standard conditions (R=0.6, S=fast), two levels of P at 1200 and 1400 are compared. Here, P=1200 is chosen as it gives a smaller percent burned than at P=1400.

Step 2. The next most important factor is thought to be R. By fixing P=1200 from step 1 and S = fast (standard condition), the two levels of R at 0.3 and 0.6 are compared. Here, R=0.3 is chosen as it gives a smaller percent burned than at R=0.6.

Step 3. Vary the remaining factor S. Two levels of S are compared with P=1200 and R =0.3 based on steps 1 and 2. The S=slow level is chosen as it gives a smaller percent burned than at S=fast.

Suppose that the observations for each factorial run are independent and normally distributed with variance σ^2.

Answer the following questions (30 marks).

(a)

(b)

(c)

(d)

10. **This question does not have a preview**

a)
b)
c)

THE END