1 2^2 Factorial Design

Q1.1: The data set below is similar to the one we considered in class. Please remember that the labels "low" and "high" are just hypothetical, they reflect upon the geometrical structure of the factorial design.

A	B	Yates notation	Replicates I	II	III	Total
$-$	$-$	(1)	28	25	27	80
$+$	$-$	(a)	36	32	32	100
$-$	$+$	(b)	18	19	23	60
$+$	$+$	(ab)	31	30	29	90

(a) Fill out the interaction table:

Effect	I	A	B	AB
(1)	$+$	$-$	$-$	$+$
(a)	$+$	$+$	$-$	$-$
(b)	$+$	$-$	$+$	$-$
(ab)	$+$	$+$	$+$	$+$

(b) Show that the main effect of A is equal to

$$A = \frac{1}{2n}((ab) + (a) - (1) - (b)),$$

and compute its value for the given data.

This expression can be derived from the column of A in the interaction table. The corresponding effects are taken with plus or minus, and then divided by the number of observations in each treatment group:

$$A = (-(1) + (a) - (b) + (ab))/(2n) = ((ab) + (a) - (1) - (b))/(2n).$$

There were $2n$ combinations with A and $2n$ combinations without A.

(c) Show that:

$$A = \bar{y}_{A+} - \bar{y}_{A-}, \qquad B = \bar{y}_{B+} - \bar{y}_{B-}.$$

Because $\bar{y}_{A+}$ is the average of the observations which were treated with A and $\bar{y}_{A-}$ is the average of the observations which were treated without A, the equation is simply a rewriting of A effect in Yates notation.

(d) Using the columns of the interaction table, compute the interaction effect AB for the given data.

The equation for AB in Yates notation was given in Lecture 11. It can also be derived using the last column of the interaction table above:

$$AB = ((1) + (ab))/(2n) - ((a) - (b))/(2n).$$

For the given data: $(1) = 80, (ab) = 90, (a) = 100, (b) = 60$, $n = 3$. Therefore,

$$AB = (80 + 90 - 100 - 60)/6 = 10/6/approx1.67.$$

(e) Fit the interaction model in R and compare the coefficient estimates with your calculations.

(f) Change the contrasts to *options(contrasts=c("contr.sum", "contr.poly"))* and fit the model again. Compare the effects estimates from the output with the effects estimates you obtained from Yates formulas. What do you notice?

```
my.data <- NULL
my.data$A <-as.factor(rep(c(-1,1,-1,1), each = 3))
my.data$B <- as.factor(rep(c(-1,-1,1,1), each = 3))
my.data$yield <- c(28,25,27,36,32,32,18,19,23,31,30,29)
my.data <- as.data.frame(my.data)

lm1 <- lm(yield ~ A + B + A*B, data = my.data)
anova(lm1)

## Analysis of Variance Table
##
## Response: yield
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## A          1 208.333 208.333 53.1915 8.444e-05 ***
## B          1  75.000  75.000 19.1489  0.002362 **
## A:B        1   8.333   8.333  2.1277  0.182776
## Residuals  8  31.333   3.917
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(lm1)

##
## Call:
## lm(formula = yield ~ A + B + A * B, data = my.data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.000 -1.333 -0.500  1.083  3.000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   26.667      1.143  23.338 1.21e-08 ***
## A1             6.667      1.616   4.126  0.00332 **
## B1            -6.667      1.616  -4.126  0.00332 **
## A1:B1          3.333      2.285   1.459  0.18278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.979 on 8 degrees of freedom
## Multiple R-squared:  0.903,Adjusted R-squared:  0.8666
## F-statistic: 24.82 on 3 and 8 DF,  p-value: 0.0002093

## to change the contrasts:

options(contrasts=c("contr.sum", "contr.poly"))

lm2 <- lm(yield ~ A + B + A*B, data = my.data)
anova(lm2)
```

```
## Analysis of Variance Table
##
## Response: yield
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## A          1 208.333 208.333 53.1915 8.444e-05 ***
## B          1  75.000  75.000 19.1489  0.002362 **
## A:B        1   8.333   8.333  2.1277  0.182776
## Residuals  8  31.333   3.917
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(lm2)
```

```
##
## Call:
## lm(formula = yield ~ A + B + A * B, data = my.data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.000 -1.333 -0.500  1.083  3.000
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  27.5000     0.5713  48.135 3.84e-11 ***
## A1           -4.1667     0.5713  -7.293 8.44e-05 ***
## B1            2.5000     0.5713   4.376  0.00236 **
## A1:B1         0.8333     0.5713   1.459  0.18278
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.979 on 8 degrees of freedom
## Multiple R-squared:  0.903,Adjusted R-squared:  0.8666
## F-statistic: 24.82 on 3 and 8 DF,  p-value: 0.0002093
```

The option ”contrasts” specifies the form in which estimated model effects are presented in the model output. For example, in the default R coding (treatment coding), the intercept is the average value of observations at the baseline (reference level) of each variable in the model. In our context, it is equal to $(1)/3 = 80/3 \approx 26.6667$. Each effect is the difference in averages with respect to the reference category. For example, in the output, the A-effect is the average difference $((a) - (1))/3 = (100 - 80)/3 \approx 6.6667$ and so on.

If the ”contr.sum” contrast is selected, the intercept is the total mean through all observations: $(80+100+60+90)/12 = 27.5$. Each effect is the difference with respect to this total. For example, $A = (100 + 90)/6 - 27.5 \approx 4.166667$.

Notice that the ANOVA table is the same in both cases.

2 2^3 Factorial Design

Q2.1:

Consider the 2^3 full factorial design, with two replicates per treatment combination.

(a) Let A, B, C denote the treatment indicators. The observed data for all treatment combinations are given in the R script below.

(b) Compute the interaction table for this design.

(c) By hand (calculator allowed), using the columns of this table, obtain the estimates of all seven effects according to Yates.

(d) Fit a linear regression model with A, B,C as explanatory variables, with all interactions included.

(e) Compare the estimates for model coefficients with the effect estimates according to Yates.

The interaction table for the $2 \times 2 \times 2$ design was shown and explained in the lecture notes.

```
## Here are the observed data

obs.data <- matrix(c( 550,  604,  669,  650, 633, 601,
                      642,  635, 1037, 1052, 749, 868,
                     1075, 1063,  729,  860), byrow = TRUE, ncol=2)

dimnames(obs.data) <- list(c("(1)","a","b","ab","c","ac","bc", "abc"),
                           c("Rep1","Rep2"))

## the column of totals:

total <- apply(obs.data,1,sum);
total

##  (1)    a    b   ab    c   ac   bc  abc
## 1154 1319 1234 1277 2089 1617 2138 1589

## Define the columns for the main effects as

A <- rep(c(-1,1),4)
B <- rep(c(-1,-1,1,1),2)
C <- c(rep(-1,4),rep(1,4))


##Using the column multiplication obtain the design table

my.data <- cbind(A,B,C,obs.data,total)
my.data

##      A  B  C Rep1 Rep2 total
## (1) -1 -1 -1  550  604  1154
## a    1 -1 -1  669  650  1319
## b   -1  1 -1  633  601  1234
## ab   1  1 -1  642  635  1277
## c   -1 -1  1 1037 1052  2089
## ac   1 -1  1  749  868  1617
## bc  -1  1  1 1075 1063  2138
## abc  1  1  1  729  860  1589
```

```r
## Columns corresponding to the interaction  effects:

AB <- A*B
AC <- A*C
BC <- B*C
ABC <- A*B*C

## the big matrix we obtained during the lecture (the column for the intercept not included)
M <- cbind(A, B, C, AB, AC, BC, ABC)

n <- 2 # the number of replicates

## Effect estimates are equal to the differences of averages of 4 means
eff.A <- (total %*% A)/(4*n)
eff.B <- (total %*% B)/(4*n)
eff.C <- (total %*% C)/(4*n)
eff.AB <- (total %*% AB)/(4*n)
eff.AC <- (total %*% AC)/(4*n)
eff.BC <- (total %*% BC)/(4*n)
eff.ABC <- (total %*% ABC)/(4*n)

all.effects <- t(total)%*% cbind(A,B,C,AB,AC,BC,ABC)/(4*n)
M.complete <- rbind(M,all.effects)
M.complete

##                A      B       C      AB       AC     BC    ABC
##  [1,]   -1.000 -1.000  -1.000   1.000    1.000  1.000 -1.000
##  [2,]    1.000 -1.000  -1.000  -1.000   -1.000  1.000  1.000
##  [3,]   -1.000  1.000  -1.000  -1.000    1.000 -1.000  1.000
##  [4,]    1.000  1.000  -1.000   1.000   -1.000 -1.000 -1.000
##  [5,]   -1.000 -1.000   1.000   1.000   -1.000 -1.000  1.000
##  [6,]    1.000 -1.000   1.000  -1.000    1.000 -1.000 -1.000
##  [7,]   -1.000  1.000   1.000  -1.000   -1.000  1.000 -1.000
##  [8,]    1.000  1.000   1.000   1.000    1.000  1.000  1.000
##  [9,] -101.625  7.375 306.125 -24.875 -153.625 -2.125  5.625

## to fit a linear regression model:

M1 <- rbind(M[,1:3],M[,1:3])
dat <- cbind(M1,as.vector(obs.data))
dat <- as.data.frame(dat)
colnames(dat) <- c("A","B", "C", "Response")

options(contrasts=c("contr.treatment","contr.poly"))
mod1 <- lm(Response ~ A*B*C, data = dat)
summary(mod1)

##
## Call:
## lm(formula = Response ~ A * B * C, data = dat)
##
## Residuals:
```

```
##    Min     1Q Median     3Q    Max
## -65.50 -11.12   0.00  11.12  65.50
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  776.062     11.865  65.406 3.32e-12 ***
## A            -50.812     11.865  -4.282 0.002679 **
## B              3.688     11.865   0.311 0.763911
## C            153.062     11.865  12.900 1.23e-06 ***
## A:B          -12.437     11.865  -1.048 0.325168
## A:C          -76.812     11.865  -6.474 0.000193 ***
## B:C           -1.062     11.865  -0.090 0.930849
## A:B:C          2.813     11.865   0.237 0.818586
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 47.46 on 8 degrees of freedom
## Multiple R-squared:  0.9661,Adjusted R-squared:  0.9364
## F-statistic: 32.56 on 7 and 8 DF,  p-value: 2.896e-05
```

```
options(contrasts=c("contr.sum","contr.poly"))

mod <- lm(Response ~ A*B*C, data = dat)
summary(mod)
```

```
##
## Call:
## lm(formula = Response ~ A * B * C, data = dat)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -65.50 -11.12   0.00  11.12  65.50
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  776.062     11.865  65.406 3.32e-12 ***
## A            -50.812     11.865  -4.282 0.002679 **
## B              3.688     11.865   0.311 0.763911
## C            153.062     11.865  12.900 1.23e-06 ***
## A:B          -12.437     11.865  -1.048 0.325168
## A:C          -76.812     11.865  -6.474 0.000193 ***
## B:C           -1.062     11.865  -0.090 0.930849
## A:B:C          2.813     11.865   0.237 0.818586
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 47.46 on 8 degrees of freedom
## Multiple R-squared:  0.9661,Adjusted R-squared:  0.9364
## F-statistic: 32.56 on 7 and 8 DF,  p-value: 2.896e-05
```

```
mean(total)/2; coef(mod)[1]
```

```
## [1] 776.0625
```

```
## (Intercept)
##    776.0625
```