Response Surface Methodology, Stat 579 Homework 10 Name: Answer Key
Fall 2014 Prof. Erik B. Erhardt

Part I. (50 points) I recommend reading through all the parts of the HW (with my adjustments) before starting; this may save you some work.
MMA-RSM Chapter 10: 10.2, 10.4, 10.7, 10.12.

- For 10.4, (a) use the "pick the winner" strategy in the Taguchi analysis, (b) do a second analysis using $\bar{y}$ and $\ln(s^2)$. Comment on any differences between the two analyses in the conclusions you make.

General: Try to do all calculations in R. All R code for the assignment should be included with the part of the problem it addresses (for code and output use a fixed-width font, such as Courier). Code is used to calculate result; text is used to report and interpret results – do not report or interpret results in the code.

(10^{pts}) **1. 10.2** Consider Example 4.2 (p. 143) in Chapter 4. Suppose, in this process for manufacture of integrated circuits, that the temperature is very difficult to control: There is some concern over variability in wafer resistivity due to uncontrolled variability in temperature. Consider the other factors as control variables.

(a) (5 pts) From the analysis given in Chapter 4, can any of the control variables be used to exert some influence over this variability? Explain.

Solution: Yes, control variables can be used to exert influence over the variability, since there is a significant control-by-noise interaction (AB = implant dose by temperature).

(b) (5 pts) It is of interest to maximize wafer resistivity and still minimize variability produced by changes in temperature. Can this be done? Explain.

Solution: The interaction plot (Fig. 4.9, p. 147) shows that the low setting of A will keep variation lower (spread between high and low values of B) but also keep the mean response lower, so what is desired can not be done.

(15^{pts}) **2. 10.4** Consider Exercise 10.3. Because E (furnace position) and D (oxide thickness) are unimportant, reduce the design to a 2^3 factorial with duplicate runs. Using temperature as a single noise variable, determine the optimal values of implant dose and time. Use a Taguchi analysis with the appropriate SNR ratio for a larger-thebetter situation.
For 10.4, (a) use the "pick the winner" strategy in the Taguchi analysis, (b) do a second analysis using $\bar{y}$ and $\ln(s^2)$. Comment on any differences between the two analyses in the conclusions you make.

(a) (5 pts) Use the "pick the winner" strategy in the Taguchi analysis.

Solution: The Taguchi approach using "pick a winner" would select the high levels of SNR for each factor: $+A$ and $+C$.

```
#### 10.4
df.10.4 <- read.table(text="
 a  c    y1    y2    y3    y4
-1 -1  68.7  62.1  15.1  11.3
-1  1  87.7  87.5  32.9  27.1
 1 -1 103.2 101.0  20.6  19.6
 1  1 128.3 119.0  46.1  40.3
", header=TRUE)
df.10.4

##    a  c    y1    y2   y3   y4
## 1 -1 -1  68.7  62.1 15.1 11.3
## 2 -1  1  87.7  87.5 32.9 27.1
## 3  1 -1 103.2 101.0 20.6 19.6
## 4  1  1 128.3 119.0 46.1 40.3

# calculate some summary statistics, especially SNR
df.10.4a       <- df.10.4
df.10.4a$m     <- apply(df.10.4a[,c("y1","y2","y3","y4")], 1, mean)
df.10.4a$s     <- apply(df.10.4a[,c("y1","y2","y3","y4")], 1, sd)
df.10.4a$s2    <- apply(df.10.4a[,c("y1","y2","y3","y4")], 1, var)
df.10.4a$logs2 <- log(df.10.4a$s2)
df.10.4a$SNR   <- apply(df.10.4a[,c("y1","y2","y3","y4")], 1
```

```
                    , function(x) {
                      -10 * log10( sum(1 / x^2) / length(x) )
                    })

# average SNR over each condition, independently
library(plyr)
df.10.4a.SNR <- rbind(
  ddply(df.10.4a, .(a), function(.X) {
    data.frame(
        var   = "a"
      , level = .X$a[1]
      , m.SNR = mean(.X$SNR)
      , s.SNR = sd(.X$SNR)
      , m.SNR = min(.X$SNR)
      , M.SNR = max(.X$SNR)
      )
    })[,-1]
    ,
  ddply(df.10.4a, .(c), function(.X) {
    data.frame(
        var   = "c"
      , level = .X$c[1]
      , m.SNR = mean(.X$SNR)
      , s.SNR = sd(.X$SNR)
      , m.SNR = min(.X$SNR)
      , M.SNR = max(.X$SNR)
      )
    })[,-1]
  )
df.10.4a.SNR

##   var level m.SNR s.SNR m.SNR.1 M.SNR
## 1   a    -1 28.47 4.932   24.99 31.96
## 2   a     1 32.03 4.429   28.90 35.17
## 3   c    -1 26.94 2.768   24.99 28.90
## 4   c     1 33.56 2.265   31.96 35.17
```

(b) (10 pts) Do a second analysis using $\bar{y}$ and $\ln(s^2)$.

Solution: Create response variables.

```
# reshape data into long format
library(reshape2)
df.10.4b <- melt(df.10.4, id.vars = c("a", "c"), variable.name = "rep", value.name = "y")
str(df.10.4b)

## 'data.frame': 16 obs. of  4 variables:
##  $ a  : int  -1 -1 1 1 -1 -1 1 1 -1 -1 ...
##  $ c  : int  -1 1 -1 1 -1 1 -1 1 -1 1 ...
##  $ rep: Factor w/ 4 levels "y1","y2","y3",..: 1 1 1 1 2 2 2 2 3 3 ...
##  $ y  : num  68.7 87.7 103.2 128.3 62.1 ...

library(plyr)
df.10.4b.ybar.lns2 <-
  ddply(df.10.4b, .(a,c), function(.X) {
    data.frame(
        ybar = mean(.X$y)
      , lns2 = log(var(.X$y))
      )
    })
df.10.4b.ybar.lns2

##    a  c  ybar  lns2
## 1 -1 -1 39.30 6.822
## 2 -1  1 58.80 7.013
## 3  1 -1 61.10 7.715
## 4  1  1 83.42 7.686

library(rsm)
```

```
lm.10.4b.ybar.FOac <- rsm(ybar ~ FO(a, c), data = df.10.4b.ybar.lns2)
# externally Studentized residuals
lm.10.4b.ybar.FOac$studres <- rstudent(lm.10.4b.ybar.FOac)
summary(lm.10.4b.ybar.FOac)
```

```
##
## Call:
## rsm(formula = ybar ~ FO(a, c), data = df.10.4b.ybar.lns2)
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   60.656      0.706    85.9   0.0074 **
## a             11.606      0.706    16.4   0.0387 *
## c             10.456      0.706    14.8   0.0429 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Multiple R-squared:  0.998,Adjusted R-squared:  0.994
## F-statistic:  245 on 2 and 1 DF,  p-value: 0.0452
##
## Analysis of Variance Table
##
## Response: ybar
##             Df Sum Sq Mean Sq F value Pr(>F)
## FO(a, c)     2    976     488     245  0.045
## Residuals    1      2       2
## Lack of fit  1      2       2
## Pure error   0      0
##
## Direction of steepest ascent (at radius 1):
##      a      c
## 0.7430 0.6693
##
## Corresponding increment in original units:
##      a      c
## 0.7430 0.6693
```

```
library(rsm)
lm.10.4b.lns2.FOac <- rsm(lns2 ~ FO(a, c), data = df.10.4b.ybar.lns2)
# externally Studentized residuals
lm.10.4b.lns2.FOac$studres <- rstudent(lm.10.4b.lns2.FOac)
summary(lm.10.4b.lns2.FOac)
```

```
##
## Call:
## rsm(formula = lns2 ~ FO(a, c), data = df.10.4b.ybar.lns2)
##
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.3092     0.0552  132.46   0.0048 **
## a             0.3914     0.0552    7.09   0.0892 .
## c             0.0405     0.0552    0.73   0.5969
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Multiple R-squared:  0.981,Adjusted R-squared:  0.942
## F-statistic: 25.4 on 2 and 1 DF,  p-value: 0.139
##
## Analysis of Variance Table
##
## Response: lns2
##             Df Sum Sq Mean Sq F value Pr(>F)
## FO(a, c)     2  0.619  0.3096    25.4   0.14
## Residuals    1  0.012  0.0122
## Lack of fit  1  0.012  0.0122
## Pure error   0  0.000
##
## Direction of steepest ascent (at radius 1):
##      a      c
## 0.9947 0.1029
##
```

```
## Corresponding increment in original units:
##      a      c
## 0.9947 0.1029
```

The separate analysis of $\bar{y}$ and $\ln(s^2)$ shows that A and C are both significant at 0.05 for $\bar{y}$, but neither are for $\ln(s^2)$. The result is that we choose to set A and C at the high levels and then we can set the other factors at whatever levels are most convenient or cost effective.

(10[pts]) **3. 10.7** Consider Example 10.2. The design is criticized for not allowing interaction among the control variables to be studied.

(a) (5 pts) What design would be a good candidate to replace the crossed array in this example? Do not allow your design to admit a run size any greater than the design listed in the example. Use a design that allows quadratic effects in the control factors and allows construction of a response model that can be used to generate the process mean and variance models.

Solution: **For the same number of runs:** a 2^{7-1}_{VII} with $I = x_1x_2x_3x_4z_1z_2z_3$ plus 8 star points will allow unambiguous estimation of all terms of form x_i, x_ix_j, x_i^2, z_i, and x_iz_j. Degrees of freedome for error would have to be obtained from higher-order interactions.
For fewer runs: a 2^{7-2}_{IV} can be designed to estimate all main effects and control-by-noise interactions. Adding star points and center runs allows estimation of quadratic terms and gives a pure error estimate.

(b) (5 pts) Explain why your design is better than that used in the example.

Solution: see (a)

(15[pts]) **4. 10.12** Suppose a study is conducted with three control variables and two noise variables. A 2^{5-1} fractional factorial is used as the experimental design. Only the main effects and the x_1z_1, x_2z_1, and x_2z_2 interactions are found to be important. The interaction plots are shown in Fig. E10.1. The main effect plots are shown in Fig. E10.2. The purpose of the experiment is to determine condition on the control variables that minimize mean response.

(a) (5 pts) Give approximate conditions on x_1, x_2, and x_3 that result in minimum mean response.

Solution: From Fig E10.2: $x_1 = -1$, x_2 doesn't matter, and $x_3 = +1$.

(b) (5 pts) Give approximate conditions on x_1, x_2, and x_3 that give approximate minimum process variance.

Solution: From Fig E10.1: (a) $x_1 = -1$ minimizes variation due to z_1, (b) $x_2 = +1$ has slightly less variation due to z_1, (c) $x_2 = -1$ minimizes variation due to z_2, so $x_2 = +1$ should have a lower variance, and $x_3 = +1$ does not have an interaction so should be set to the level with lower response.

(c) (5 pts) Is there a tradeoff between optimum conditions? Explain.

Solution: There is almost no trade-off. $x_1 = -1$ has lower response and less variability wrt z_1. $x_2 = -1$ doesn't affect the response and has slightly higher variability wrt z_1 but much less variability wrt z_2. $x_3 = +1$ has lower response, and doesn't interact with the noise variability.