Experimental Design and Analysis
HW06 Solution

Problem 1

(a)

The regression equations for location and dispersion, respectively, are

$$\begin{cases} \hat{y}_{\mathbf{x}} = \quad 7.6360 + 0.1106x_B + 0.0519x_E + 0.0881x_C - 0.1298x_Q + 0.0423x_Bx_Q - 0.0827x_Cx_Q & (5.8) \\ ln(\hat{\sigma}^2_{\mathbf{x}}) = -4.9313 + 0.9455x_B + 0.5556x_Dx_Q - 0.5445x_Bx_Cx_Q & (5.9) \end{cases}$$

We then choose factor settings $\mathbf{x} \in [-1, 1]^5$ *to minimize the predicted mean squared error*

$$\widehat{MSE}_{\mathbf{x}} = (\hat{y}_{\mathbf{x}} - 8)^2 + exp\Big(ln(\hat{\sigma}^2_{\mathbf{x}})\Big)$$

Using the ***optim*** *function, we obtain*

$$\underset{\mathbf{x}\in[-1,1]^5}{arg\,min}\ \widehat{MSE}_{\mathbf{x}} = (-0.2723377, 1, 1, 1, -1)$$

(b)

Case	B	C	D	E	Q	$\hat{y}_{\mathbf{x}}$	$\hat{\sigma}^2_{\mathbf{x}}$	$\widehat{MSE}_{\mathbf{x}}$
Section 5.3	-1	1	1	2.5376	-1	8.000001	-6.976900	0.000933191
Exercise 1	-1	2.0749	1	-1	-1	7.999993	-7.562183	0.000519739
Exercise 2	-0.2723	1	1	1	-1	7.969902	-5.892627	0.003665612

To further reduce the predicted mean square error, we need to relax the range of the factor levels.

Factor Q is a good candidate to relax first because its main effect and its interaction terms have relatively large coefficients, and Q appears in both the location and dispersion models.

(c)

When we relax the range of the factor Q and allow x_Q to go below -1 (specifically, $x_Q \in [-1.6, 1]$ while keeping the other factors in $[-1, 1]$), the optimized setting improves further. Using the ***optim*** *again, the we obtain a new minimizer:*

$$\mathbf{x} = (-1, 1, 1, -0.4080925, -1.6)$$

which yields a smaller predicted mean square error $\widehat{MSE}_{\mathbf{x}} \approx 0.0004822924$, outperforming the solutions obtained using the other methods.

However, this improvement comes from setting $x_Q = -1.6$, which lies outside the original design region $[-1, 1]$. Therefore, the result relies on extrapolation of the fitted regression models beyond the range of the experimental data, and the predicted gain should be interpreted cautiously (ideally confirmed by follow-up runs in the expanded region).

Problem 2

(a)

The defining words are $\mathbf{I = 134 = 235 = 1245}$*, so the resolution is III.*

(b)

All the aliasing relations for this design are

$\mathbf{I}$	$= \mathbf{134}$	$= \mathbf{235}$	$= \mathbf{1245}$
$\mathbf{1}$	$= \mathbf{34}$	$= \mathbf{1235}$	$= \mathbf{245}$
$\mathbf{2}$	$= \mathbf{1234}$	$= \mathbf{35}$	$= \mathbf{145}$
$\mathbf{3}$	$= \mathbf{14}$	$= \mathbf{25}$	$= \mathbf{12345}$
$\mathbf{4}$	$= \mathbf{13}$	$= \mathbf{2345}$	$= \mathbf{125}$
~~$\mathbf{5}$~~	$= \sout{\mathbf{1345}}$	$= \mathbf{23}$	$= \sout{\mathbf{124}}$
$\mathbf{12}$	$= \sout{\mathbf{234}}$	$= \sout{\mathbf{135}}$	$= \sout{\mathbf{45}}$
~~$\mathbf{15}$~~	$= \sout{\mathbf{345}}$	$= \sout{\mathbf{123}}$	$= \mathbf{24}$

We can see that the 2-factor interaction effects $\mathbf{13}$*,* $\mathbf{14}$ *and* $\mathbf{34}$ *are aliased with main effects. Therefore, only the interaction effects* $\mathbf{12}$*,* $\mathbf{23}$ *and* $\mathbf{24}$ *can be estimated, provided that variable* 5 *is inert, all 2-factor interactions involving variable* 5 *and all higher order interactions are negligible.*

Problem 3

(a)

We first write down the defining contrast subgroup for each choice of generators

$$\begin{cases} A: & \mathbf{I = 12345 = 1236 = 456} \\ B: & \mathbf{I = 1235 = 2346 = 1456} \end{cases}$$

From the defining relations, we find that design A is of resolution III, whereas design B is of resolution IV. Under the maximum resolution criterion, we recommend design B.

However, choosing the design with the highest resolution is not the only criterion. To gain a more detailed understanding of the alias structure, we next list the aliasing relations for designs A and B in turn, and then use these results to further justify our comparison and recommendation.

For design A, all the aliasing relations are

$\mathbf{I}$	$= \mathbf{12345}$	$= \mathbf{1236}$	$= \mathbf{456}$
$\mathbf{1}^{SC}$	$= \mathbf{2345}$	$= \mathbf{236}$	$= \mathbf{1456}$
$\mathbf{2}^{SC}$	$= \mathbf{1345}$	$= \mathbf{136}$	$= \mathbf{2456}$
$\mathbf{3}^{SC}$	$= \mathbf{1245}$	$= \mathbf{126}$	$= \mathbf{3456}$
$\mathbf{4}$	$= \mathbf{1235}$	$= \mathbf{12346}$	$= \mathbf{56}$
$\mathbf{5}$	$= \mathbf{1234}$	$= \mathbf{12356}$	$= \mathbf{46}$
$\mathbf{6}$	$= \mathbf{123456}$	$= \mathbf{123}$	$= \mathbf{45}$
$\mathbf{12}$	$= \mathbf{345}$	$= \mathbf{36}$	$= \mathbf{12456}$
$\mathbf{13}$	$= \mathbf{245}$	$= \mathbf{26}$	$= \mathbf{13456}$
$\mathbf{14}^{C}$	$= \mathbf{235}$	$= \mathbf{2346}$	$= \mathbf{156}$
$\mathbf{15}^{C}$	$= \mathbf{234}$	$= \mathbf{2356}$	$= \mathbf{146}$
$\mathbf{16}$	$= \mathbf{23456}$	$= \mathbf{23}$	$= \mathbf{145}$
$\mathbf{24}^{C}$	$= \mathbf{135}$	$= \mathbf{1346}$	$= \mathbf{256}$
$\mathbf{25}^{C}$	$= \mathbf{134}$	$= \mathbf{1356}$	$= \mathbf{246}$
$\mathbf{34}^{C}$	$= \mathbf{125}$	$= \mathbf{1246}$	$= \mathbf{356}$
$\mathbf{35}^{C}$	$= \mathbf{124}$	$= \mathbf{1256}$	$= \mathbf{346}$

For design B, all the aliasing relations are

$$
\begin{array}{llll}
\mathbf{I} & = \mathbf{1235} & = \mathbf{2346} & = \mathbf{1456} \\
\mathbf{1}^{SC} & = \mathbf{235} & = \mathbf{12346} & = \mathbf{456} \\
\mathbf{2}^{SC} & = \mathbf{135} & = \mathbf{346} & = \mathbf{12456} \\
\mathbf{3}^{SC} & = \mathbf{125} & = \mathbf{246} & = \mathbf{13456} \\
\mathbf{4}^{SC} & = \mathbf{12345} & = \mathbf{236} & = \mathbf{156} \\
\mathbf{5}^{SC} & = \mathbf{123} & = \mathbf{23456} & = \mathbf{146} \\
\mathbf{6}^{SC} & = \mathbf{12356} & = \mathbf{234} & = \mathbf{145} \\
\mathbf{12} & = \mathbf{35} & = \mathbf{1346} & = \mathbf{2456} \\
\mathbf{13} & = \mathbf{25} & = \mathbf{1246} & = \mathbf{3456} \\
\mathbf{14} & = \mathbf{2345} & = \mathbf{1236} & = \mathbf{56} \\
\mathbf{15} & = \mathbf{23} & = \mathbf{123456} & = \mathbf{46} \\
\mathbf{16} & = \mathbf{2356} & = \mathbf{1234} & = \mathbf{45} \\
\mathbf{24} & = \mathbf{1345} & = \mathbf{36} & = \mathbf{1256} \\
\mathbf{26} & = \mathbf{1356} & = \mathbf{34} & = \mathbf{1245} \\
\mathbf{124} & = \mathbf{345} & = \mathbf{136} & = \mathbf{256} \\
\mathbf{126} & = \mathbf{356} & = \mathbf{134} & = \mathbf{245}
\end{array}
$$

After listing the aliasing relations, we summarize the clear and strongly clear effects for the two designs:

$$
\textit{Design } A \textit{ has} \begin{cases} \textit{3 clear main effects:} & \mathbf{1},\ \mathbf{2},\ \mathbf{3} \\ \textit{6 clear 2-factor interactions:} & \mathbf{14},\ \mathbf{15},\ \mathbf{24},\ \mathbf{25},\ \mathbf{34},\ \mathbf{35} \end{cases}
$$

$$
\textit{Design } B \textit{ has} \begin{cases} \textit{6 strongly clear main effects:} & \mathbf{1},\ \mathbf{2},\ \mathbf{3},\ \mathbf{4},\ \mathbf{5},\ \mathbf{6} \\ \textit{0 clear 2-factor interaction} & \end{cases}
$$

Therefore, if the objective is to obtain clear estimates of as many main effects as possible, design B is preferable (consistent with maximum resolution criterion). However, in applications where 2-factor interactions are of primary interest, design A can be more attractive, even though its overall resolution is lower.

(b)

A 2^{6-2} design has 16 runs, so the total degrees of freedom available for estimating factorial effects is 15. Now suppose, for contradiction, that there exists a 2_V^{6-2} design with the following two properties:

$$
\begin{cases} \textbf{1. Main effects are clear} \\ \textbf{2. Two-factor interactions are clear} \end{cases}
$$

Consequently, each main effect and each two-factor interaction must belong to a distinct alias set. Hence, to estimate all main effects and all two-factor interactions separately, we would need at least

$$
\binom{6}{1} + \binom{6}{2} = 21
$$

degrees of freedom, which is impossible because the design provides only 15 degrees of freedom.

Therefore, a 2_V^{6-2} design does not exist.

Problem 4

(a)

All the aliasing relations of this experiment are

$$
\begin{aligned}
\mathbf{I} &= -ABCDE \\
A^{SC} &= -BCDE \\
B^{SC} &= -ACDE \\
C^{SC} &= -ABDE \\
D^{SC} &= -ABCE \\
E^{SC} &= -ABCD \\
AB^{C} &= -CDE \\
AC^{C} &= -BDE \\
AD^{C} &= -BCE \\
AE^{C} &= -BCD \\
BC^{C} &= -ADE \\
BD^{C} &= -ACE \\
BE^{C} &= -ACD \\
CD^{C} &= -ABE \\
CE^{C} &= -ABD \\
DE^{C} &= -ABC
\end{aligned}
$$

From the aliasing relations above, we observe that all the main effects are strongly clear and all the 2-factor interactions are clear.

(b)

We list $\overline{y}$ and $ln(s^2)$ under each level combination below (If $s^2 = 0$, then we set $s = 0.001$).

```
data <- cbind(unique(data[,-6]), matrix(data[,6], ncol = 3, byrow = TRUE))
data[data == "-"] <- -1
data[data == "+"] <- 1
data <- as.data.frame(apply(data, 2, as.numeric))

y_bar <- apply(data[,6:8], 1, mean)
s_square <- apply(data[,6:8], 1, var); s_square[s_square == 0] <- 0.001
ln_s_square <- log(s_square)
data <- cbind(data[,1:5], y_bar, ln_s_square)

colnames(data) = c("A", "B", "C", "D", "E", "$\\bar{y}$", "ln$s^2$")
kable(data, digits = 3, row.names = F)
```

A	B	C	D	E	$\bar{y}$	$\ln s^2$
-1	-1	-1	-1	-1	1275.000	9.113
1	1	-1	-1	-1	1916.667	6.916
1	-1	1	-1	-1	1770.000	-6.908
-1	1	1	-1	-1	1275.000	-6.908
1	-1	-1	1	-1	1898.333	6.916
-1	1	-1	1	-1	1440.000	8.015
-1	-1	1	1	-1	1275.000	9.961
1	1	1	1	-1	2118.333	6.916
1	-1	-1	-1	1	1696.667	6.916
-1	1	-1	-1	1	1495.000	9.113
-1	-1	1	-1	1	1220.000	11.059
1	1	1	-1	1	1990.000	-6.908
-1	-1	-1	1	1	1495.000	10.580
1	1	-1	1	1	1641.667	6.916
1	-1	1	1	1	1916.667	6.916
-1	1	1	1	1	1256.667	6.916

In part (a), we know that all the main effects and all the 2-factor interactions are clear and hence estimable.

Thus, we fit the location model and the dispersion model, respectively, as follows:

$$\begin{cases} \overline{y} = \mathbf{X}\boldsymbol{\beta} + \mathcal{E} \\ ln(s^2) = \mathbf{X}\boldsymbol{\gamma} + \mathcal{E}^* \end{cases}$$

where

$$\mathbf{X} = \begin{array}{c} \begin{array}{cccccccccccccccc} \text{(Int.)} & A & B & C & D & E & AB & AC & AD & AE & BC & BD & BE & CD & CE & DE \end{array} \\ \begin{pmatrix} 1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ 1 & +1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 \\ 1 & +1 & -1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ 1 & -1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 \\ 1 & +1 & -1 & -1 & +1 & -1 & -1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\ 1 & -1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 \\ 1 & -1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\ 1 & +1 & +1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 \\ 1 & +1 & -1 & -1 & -1 & +1 & -1 & -1 & -1 & +1 & +1 & +1 & -1 & +1 & -1 & -1 \\ 1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 & -1 & -1 & +1 & +1 & -1 & -1 \\ 1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & -1 & +1 & -1 & -1 & +1 & -1 \\ 1 & +1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 & -1 & +1 & -1 & +1 & -1 \\ 1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 & -1 & -1 & +1 & -1 & -1 & -1 & -1 & +1 \\ 1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 & -1 & +1 & +1 & -1 & -1 & +1 \\ 1 & +1 & -1 & +1 & +1 & +1 & -1 & +1 & +1 & +1 & -1 & -1 & -1 & +1 & +1 & +1 \\ 1 & -1 & +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 & +1 & +1 & +1 & +1 \end{pmatrix} \end{array}$$

consists of all the main effects and all the 2-factor interactions.

Then apply regression analysis to estimate the factorial effects:

```
y_bar <- data[,6]
ln_s_square <- data[,7]

location_model <- lm(
  y_bar ~ A + B + C + D + E +
    A:B + A:C + A:D + A:E +
    B:C + B:D + B:E +
    C:D + C:E +
    D:E,
  data = data
)

dispersion_model <- lm(
  ln_s_square ~ A + B + C + D + E +
    A:B + A:C + A:D + A:E +
    B:C + B:D + B:E +
    C:D + C:E +
    D:E,
  data = data
)

effect_name <- c("A","B","C","D","E","AB","AC","AD","AE","BC","BD","BE","CD","CE","DE")

location_effect   <- 2 * coef(location_model)[-1]
dispersion_effect <- 2 * coef(dispersion_model)[-1]

names(location_effect)   <- gsub(":", "", names(location_effect))
names(dispersion_effect) <- gsub(":", "", names(dispersion_effect))
location_effect   <- location_effect[effect_name]
dispersion_effect <- dispersion_effect[effect_name]

table_effect <- cbind(
  effect_name,
  round(location_effect, 3),
  round(dispersion_effect, 3)
)

sig_loc <- c("A", "AC")
sig_dis <- c("C", "D", "CD")

fmt_cell <- function(val, is_sig, digits = 3){
  s <- sprintf(paste0("%.", digits, "f"), val)
  if (is_sig) paste0("$\\textcolor{red}{", s, "}$")
  else        paste0("$\\textcolor{black}{", s, "}$")
}

loc_col <- mapply(fmt_cell, location_effect, effect_name %in% sig_loc,
                  MoreArgs = list(digits = 3))
dis_col <- mapply(fmt_cell, dispersion_effect, effect_name %in% sig_dis,
                  MoreArgs = list(digits = 3))

table_effect_colored <- cbind(effect_name, loc_col, dis_col)
```

```r
colnames(table_effect_colored) <- c("Effect", "$\\bar y$", "$\\ln(s^2)$")

knitr::kable(table_effect_colored, row.names = FALSE, escape = FALSE)
```

Effect	$\bar{y}$	$\ln(s^2)$
A	527.083	−3.771
B	73.333	−2.947
C	−4.583	−5.430
D	50.417	5.093
E	−32.083	2.186
AB	22.917	2.947
AC	165.000	−1.482
AD	0.000	1.819
AE	−82.500	−2.186
BC	41.250	−2.306
BD	−105.417	1.545
BE	−59.583	−1.911
CD	27.500	5.001
CE	18.333	1.545
DE	−73.333	−2.306

Then check the significance of location and dispersion effects by half-normal plots:

```r
stopifnot(exists("location_effect"), exists("dispersion_effect"))

if (is.null(names(location_effect))) {
  names(location_effect) <- paste0("E", seq_along(location_effect))
}
if (is.null(names(dispersion_effect))) {
  names(dispersion_effect) <- paste0("E", seq_along(dispersion_effect))
}

halfnorm_pink <- function(x, outlier = 1, main = "Half-normal Plot",
                          hi_col = "deeppink",
                          xlab = "Half-normal Quantiles",
                          ylab = "Absolute Effects") {

  n <- length(x)

  ord <- order(abs(x))
  y   <- abs(x)[ord]
  lab <- names(x)[ord]

  quan <- qnorm(0.5 + 0.5 * ((1:n) - 0.5) / n)

  xpad <- 0.08 * diff(range(quan))
  ypad <- 0.10 * diff(range(y))
  xlim <- c(min(quan) - xpad, max(quan) + xpad)
  ylim <- c(min(y) - ypad, max(y) + ypad)

  plot(quan, y, pch = 16, col = "black",
```

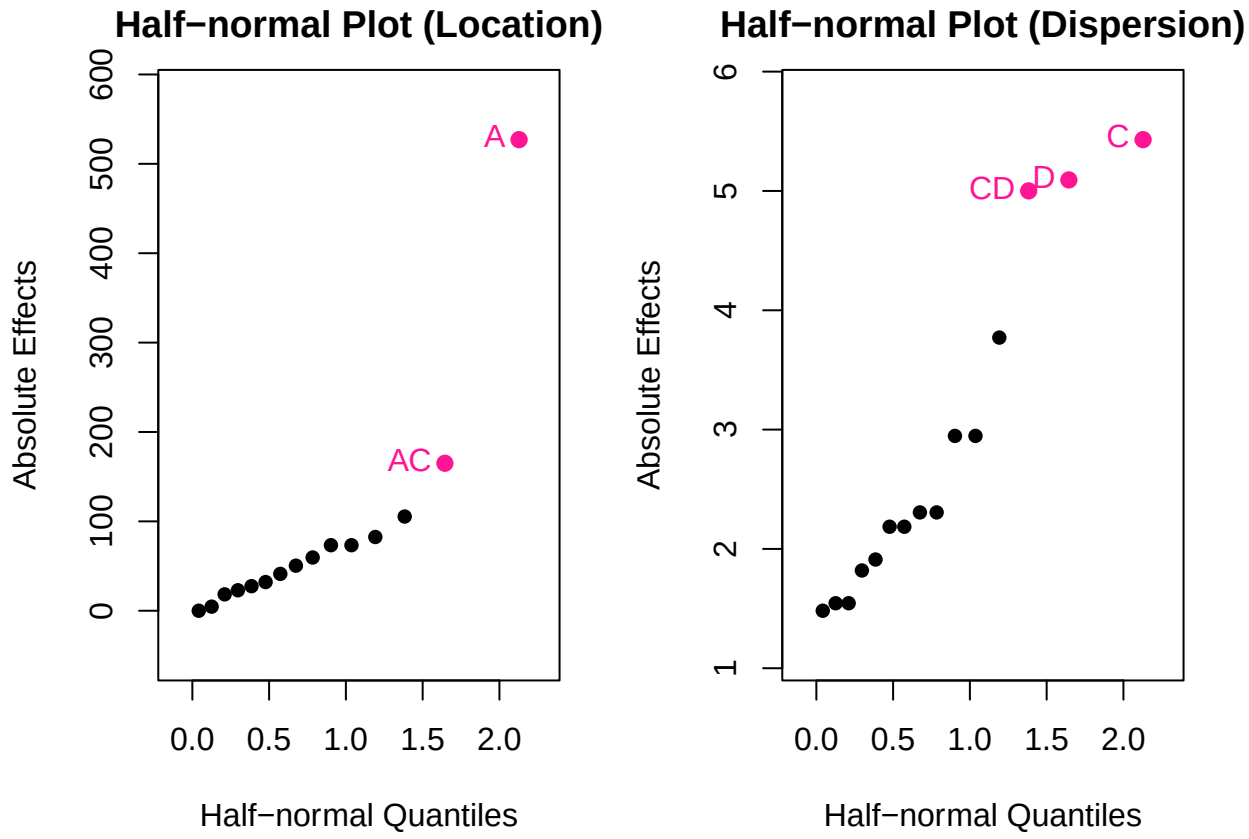
```
       main = main, xlab = xlab, ylab = ylab,
       xlim = xlim, ylim = ylim)

  outlier <- min(outlier, n)
  idx <- (n - outlier + 1):n

  points(quan[idx], y[idx], pch = 16, col = hi_col, cex = 1.2)

  pos_vec <- ifelse(quan[idx] > mean(xlim), 2, 4)
  text(quan[idx], y[idx], labels = lab[idx],
       col = hi_col, pos = pos_vec, offset = 0.35, cex = 1, font = 1)
}

op <- par(mfrow = c(1, 2), mar = c(4.2, 4.2, 2.4, 1.6))
halfnorm_pink(location_effect,   outlier = 2, main = "Half-normal Plot (Location)")
halfnorm_pink(dispersion_effect, outlier = 3, main = "Half-normal Plot (Dispersion)")
```



From the significance screening, effects A and AC are retained in the location model, while C, D, and CD are retained in the dispersion model. Therefore, the fitted models are:

$$\begin{cases} \hat{\bar{y}} = 1605 + 263.5415A + 82.5AC \\ ln(\hat{s}^2) = 5.3456 - 2.715025C + 2.546284D + 2.500318CD \end{cases}$$

(c)

By inspecting the signs of the fitted dispersion model, we choose factor levels that minimize $ln(\hat{s}^2)$. We set $C = +$ because its coefficient is negative, set $D = -$ because its coefficient is positive, and this choice also makes $CD = -$. This also makes $CD = -$ simultaneously, so the interaction term contributes negatively and further decreases $ln(\hat{s}^2)$.

(d)

Since the coefficient of A is positive, we set $A = +$.
Since the coefficient of AC is also positive, we want $AC = +$, which (given $A = +$) implies $C = +$.

(e)

Because D is an adjustment factor and the response is larger-the-better, we adopt the following two-step procedure:

- *Step 1: select $(A, C) = (+, +)$ so that*

$$\hat{\bar{y}} = 1605 + 263.5415 + 82.5 = 1951.$$

- *Step 2: select $(C, D) = (+, -)$ so that*

$$\hat{s}^2 = \exp\Big(5.3456 - 2.715025 - 2.546284 - 2.500318\Big) = 0.08927561.$$

That is, we first choose the factor levels to maximize the fitted mean, and then tune the adjustment factor to minimize the fitted variance.