Answer Keys to Homework#11

Problem 1

(a) A 2^{6-2} design will have $2^2 - 1 = 3$ defining words. We already have two defining words $ABE = CDF = I$ for d_1, then the remaining one is simply the product of these two. So the complete defining relation is

$$I = ABE = CDF = ABCDEF.$$

The alias structure of a 2^{6-2} design should have $2^{6-2} = 16$ effect groups, each having $2^2 = 4$ effects. The alias structure for d_1 is then

$I = ABE = CDF = ABCDEF$	$AD = ACF = BDE = ACEF$
$A = BE = ACDF = BCDEF$	$AF = ACD = BEF = BCDE$
$B = AE = BCDF = ACDEF$	$BC = ACE = BDF = ADEF$
$C = DF = ABCE = ABDEF$	$BD = ADE = BCF = ACEF$
$D = CF = ABDE = ABCEF$	$BF = AEF = BCD = ACDE$
$E = AB = CDEF = ABCDF$	$CE = ABC = DEF = ABDF$
$F = CD = ABEF = ABCDE$	$DE = ABD = CEF = ABCF$
$AC = ADF = BCE = BDEF$	$EF = ABF = CDE = ABCD$

The resolution is the minimum length of defining words in the complete defining relation, so the resolution of d_1 is III.

The wordlength pattern $W = (W_0, W_1, \ldots, W_6)$, where W_i is the number of defining words involving i factors. Hence the wordlength pattern for d_1 is $(1, 0, 0, 2, 0, 0, 1)$.

(b) If effects of order three or higher are negligible, then a main effect or two-factor interaction is clearly estimable if and only if it isn't aliased with another main effect or two-factor interaction. According to the alias structure for d_1 in (a), the estimable main effects and two-factor interactions are $AC, AD, AF, BC, BF, CE, DE, EF, BD$.

(c) Similar to (a) and (b), the complete defining relation for d_2 is

$$I = ABCE = ADEF = BCDF.$$

The alias structure for d_2 is

$I = ABCE = ADEF = BCDF$	$AC = BE = ABDF = CDEF$
$A = BCE = DEF = ABCDF$	$AD = EF = ABCF = BCDE$
$B = ACE = CDF = ABDEF$	$AE = BC = DF = ABCDEF$
$C = ABE = BDF = ACDEF$	$AF = DE = ABCD = BCEF$
$D = AEF = BCF = ABCDE$	$BD = CF = ABEF = ACDE$
$E = ABC = ADF = BCDEF$	$BF = CD = ABDE = ACEF$
$F = ADE = BCD = ABCEF$	$ABD = ACF = BEF = CDE$
$AB = CE = ACDF = BDEF$	$ABF = ACD = BDE = CEF$

The resolution of d_2 is IV. The wordlength pattern of d_2 is $(1, 0, 0, 0, 3, 0, 0)$.

The estimable main effects and two-factor interactions are A, B, C, D, E, F. That is, all the main effects are clearly estimable, whereas all the two-factor interactions are unestimable.

(d) I will choose design d_2, since it has higher resolution than d_1 and less aberration than d_1 ($W_3(d_2) = 0 < 2 = W_3(d_1)$). Furthermore, all the main effects, which are generally the most important effects, are clearly estimable in d_2.

Design d_1 could have some advantages over d_2 in some cases. For example, if all the two-factor interactions aliased with the main effects are known to be negligible as well as the higher order (≥ 3) effects, then d_1 can also estimate the main effects, and it can estimate more two-factor interactions than d_2. However, this kind of assumptions on the two-factor interactions are not common in practice.

(e) $E = AB$ in design d_1 and $E = ABC$ in design d_2. Clearly, $E = ABC$ in the given design matrix. So design d_2 has been used.

(f) The estimates of the factorial effects are listed in the table below and the normal Q-Q plot for the estimates is in Figure 1

Factor	Estimate	Factor	Estimate
A	50.5	AC	-2.5
B	-1.0	AD	4.0
C	-13.0	AE	1.0
D	37.0	AF	-22.0
E	34.5	BD	4.5
F	4.5	BF	14.5
AB	-4.0	ABD	0.5
		ABF	6.0

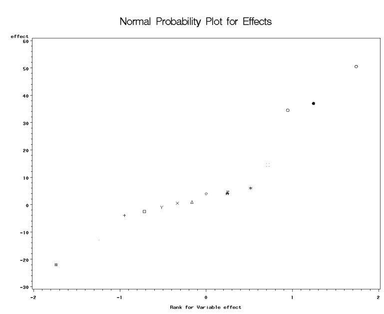

Figure 1: Normal Q-Q Plot of Factorial Effects

Based on the table and the plot, I have identified the following potentially important effects: A, C, D, E, AF and BF.

(g) The really estimated (important) effects are:

$$
\begin{aligned}
\mathcal{L}_A = 50.5 &\rightarrow A + BCE + DEF + ABCDF, \\
\mathcal{L}_C = -13.0 &\rightarrow C + ABE + BDF + ACDEF, \\
\mathcal{L}_D = 37.0 &\rightarrow D + AEF + BCF + ABCDE, \\
\mathcal{L}_E = 34.5 &\rightarrow E + ABC + ADF + BCDEF, \\
\mathcal{L}_{AF} = -22.0 &\rightarrow AF + DE + ABCD + BCEF, \\
\mathcal{L}_{BF} = 14.5 &\rightarrow BF + CD + ABDE + ACEF.
\end{aligned}
$$

(h) If the effects of order 3 or higher are negligible, then from (g) we know effects A, C, D and E are significant, but we can't distinguish effect AF from DE (and $AF + DE$), or BF from CD (and $BF + CD$). Hence we can have the 4 possible models with the sets of included effects listed in the following table.

Index	Factors Included in Model
1	(A, C, D, E, AF, BF)
2	(A, C, D, E, AF, CD)
3	(A, C, D, E, BF, DE)
4	(A, C, D, E, CD, DE)

By the Effect Heredity Principle, in order for an interaction to be significant, at least one of its parent factors should be significant. So the interaction effect BF is not likely to be significant, so candidate models 1 and 3 are eliminated. The AF interaction term in model 2 may force us to introduce another insignificant main effect F, which generally is not recommended. Hence the most likely model is model 4 which includes the effects (A, C, D, E, CD, DE).

(i) The regression result from SAS is given below.

```
                                  The REG Procedure
                                    Model: MODEL1
                                Dependent Variable: y

                                Analysis of Variance

                                       Sum of           Mean
Source                      DF        Squares         Square    F Value    Pr > F

Model                        6          23891     3981.83333      76.57    <.0001
Error                        9      468.00000       52.00000
Corrected Total             15          24359


            Root MSE              7.21110    R-Square     0.9808
            Dependent Mean      135.75000    Adj R-Sq     0.9680
            Coeff Var             5.31205


                                Parameter Estimates

                         Parameter       Standard
    Variable     DF       Estimate          Error    t Value    Pr > |t|

    Intercept     1      135.75000        1.80278      75.30      <.0001
    A             1       25.25000        1.80278      14.01      <.0001
    C             1       -6.50000        1.80278      -3.61      0.0057
    D             1       18.50000        1.80278      10.26      <.0001
    E             1       17.25000        1.80278       9.57      <.0001
    CD            1        7.25000        1.80278       4.02      0.0030
    DE            1      -11.00000        1.80278      -6.10      0.0002
```

All the effects in the model are significant and the model has a $R^2 = 0.9808$ (replacing CD by BF, or DE by AF will give the same results due to the alias structure). So the model fits well.

(j) The diagnostic plots in Figure 2 are: normal probability Q-Q plot, plots of residuals versus, respectively, predicted values, factor A (furnace temperature), factor C (carbon concentration), factor D (duration of the carbonizing cycle), and factor E (carbon concentration of the diffuse cycle). Also, the results of formal normality tests are listed below.

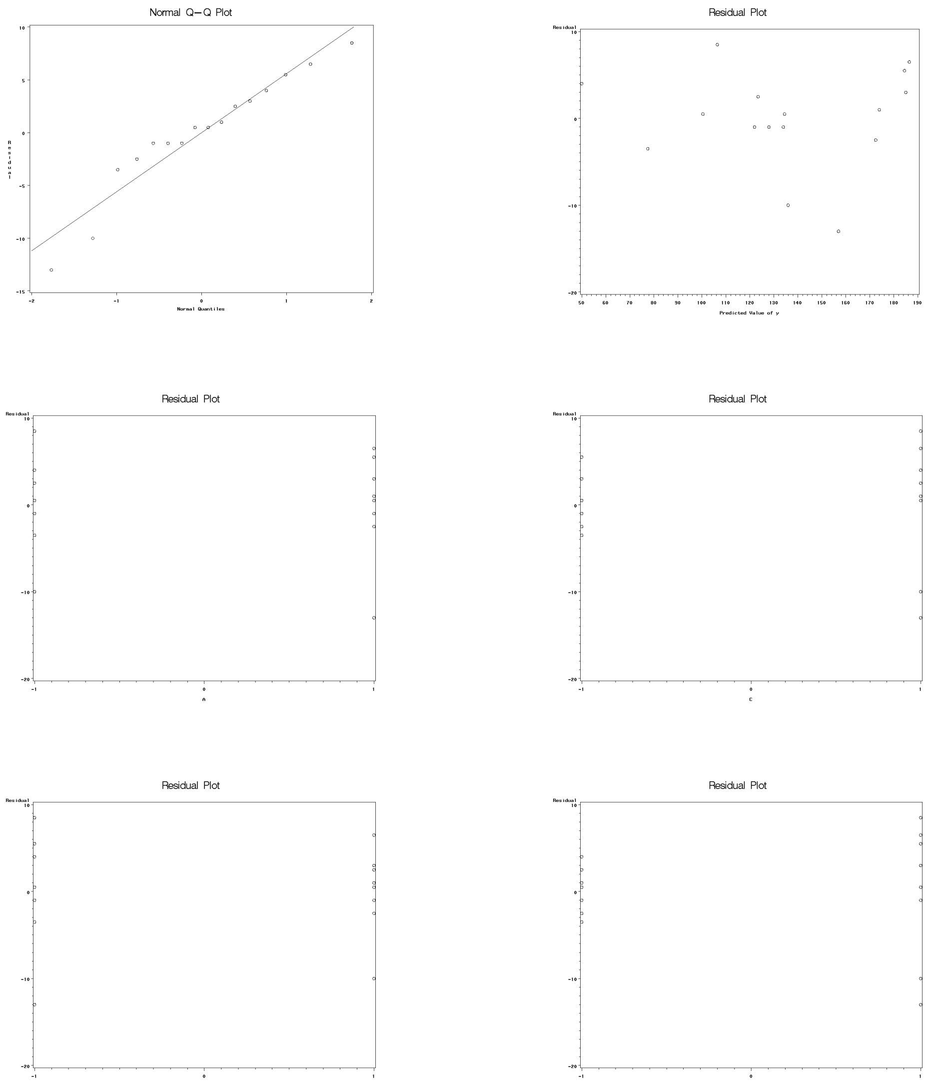

Figure 2: Diagnostic Plots

Tests for Normality

Test	--Statistic---		-----p Value------	
Shapiro-Wilk	W	0.931558	Pr < W	0.2580
Kolmogorov-Smirnov	D	0.178958	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.069953	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.453107	Pr > A-Sq	0.2408

The normal Q-Q plot and the normality tests shows that the normality assumption is valid. None of the four residual plots has shown unequal variances or potential outliers. ∎