Recommended Exercise 9-10 in Statistical Linear Models, Spring 2021

alexaoh

10 april, 2021

Problem 1 Exam 2015 Spring, Problem 2

a)

The estimator for the parameters are given as $\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \boldsymbol{Y}$. When the columns of X are orthogonal, it means that $X^T X = CI$, where C is some constant depending on the values in X. Hence, $(X^T X)^{-1} = \frac{1}{C} I$ and $\hat{\boldsymbol{\beta}} = \frac{1}{C} X^T \boldsymbol{Y}$. This means that a component of the parameter vector has the value $\hat{\beta}_j = \frac{1}{C} \sum_{i=0}^{n} x_{ij} y_j = \frac{1}{\boldsymbol{x}_j^T \boldsymbol{x}_j} \boldsymbol{x}_j^T \boldsymbol{Y}$. Hence, the j^{th} entry of $\hat{\boldsymbol{\beta}}$ only depends on the j^{th} column of X and $\boldsymbol{Y}$.

b)

We are given the data

```
A <- c(-1, 1, -1, 1)
B <- c(-1, -1, 1, 1)
y <- c(6, 4, 10, 7)
df <- data.frame("A" = A, "B" = B, "y" = y)
df
```

```
#>    A  B  y
#> 1 -1 -1  6
#> 2  1 -1  4
#> 3 -1  1 10
#> 4  1  1  7
```

The interaction effect of the two factors can be calculated by

$$
\begin{aligned}
&\frac{1}{2}(\text{main effect of A when B is 1}) - \frac{1}{2}(\text{main effect of A when B is -1}) \\
&= \frac{1}{2}\left(\frac{7-10}{2}\right) - \frac{1}{2}\left(\frac{4-6}{2}\right) = \frac{1}{4}\left(7+6-10-4\right) \\
&= -\frac{1}{4} = \frac{1}{4}(1 \quad -1 \quad -1 \quad 1)^T y = \text{interaction vector} \times y.
\end{aligned}
$$

Problem 2 Factorial experiments

a)

```
library(FrF2)
y <- c(14.6, 24.8, 12.3, 20.1, 13.8, 22.3, 12.0, 20.0, 16.3, 23.7, 13.5, 19.4, 11.3, 23.6, 11.2, 21.8)
```

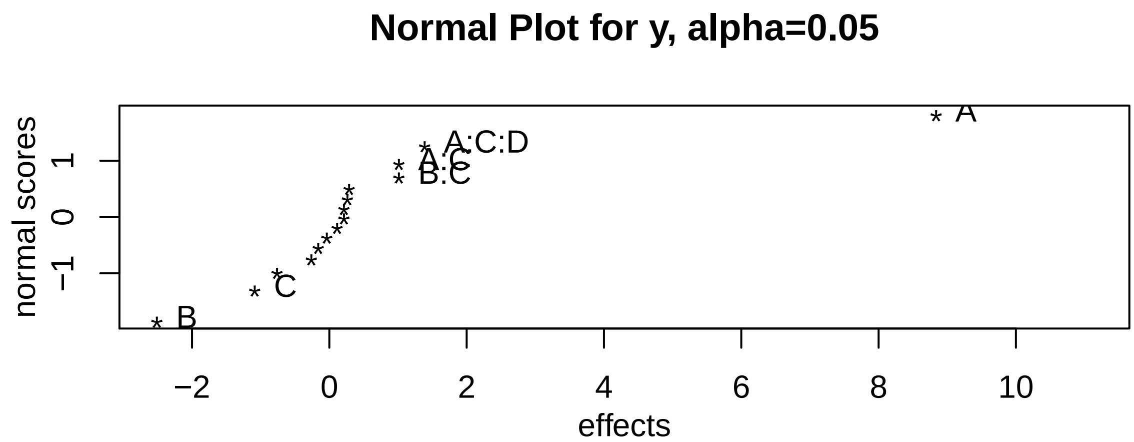
```r
plan <- FrF2(nruns=16,nfactors=4,randomize=FALSE)
plan <- add.response(plan,y)

full.fit <- lm(y~.^4, data = plan)
summary(full.fit)
```

```
#>
#> Call:
#> lm.default(formula = y ~ .^4, data = plan)
#>
#> Residuals:
#> ALL 16 residuals are 0: no residual degrees of freedom!
#>
#> Coefficients:
#>             Estimate Std. Error t value Pr(>|t|)
#> (Intercept) 17.54375         NA      NA       NA
#> A1           4.41875         NA      NA       NA
#> B1          -1.25625         NA      NA       NA
#> C1          -0.54375         NA      NA       NA
#> D1           0.05625         NA      NA       NA
#> A1:B1       -0.38125         NA      NA       NA
#> A1:C1        0.50625         NA      NA       NA
#> A1:D1        0.10625         NA      NA       NA
#> B1:C1        0.50625         NA      NA       NA
#> B1:D1        0.13125         NA      NA       NA
#> C1:D1       -0.08125         NA      NA       NA
#> A1:B1:C1     0.10625         NA      NA       NA
#> A1:B1:D1    -0.01875         NA      NA       NA
#> A1:C1:D1     0.69375         NA      NA       NA
#> B1:C1:D1     0.14375         NA      NA       NA
#> A1:B1:C1:D1 -0.13125         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
```

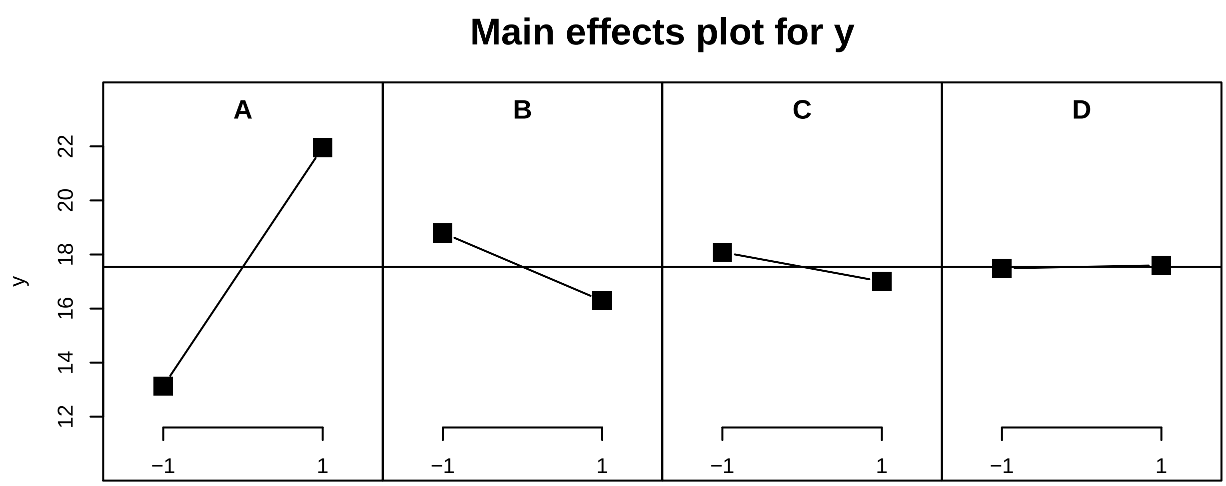
```r
effects <- full.fit$coefficients*2
effects
```

```
#> (Intercept)          A1          B1          C1          D1       A1:B1
#>     35.0875      8.8375     -2.5125     -1.0875      0.1125     -0.7625
#>       A1:C1       A1:D1       B1:C1       B1:D1       C1:D1    A1:B1:C1
#>      1.0125      0.2125      1.0125      0.2625     -0.1625      0.2125
#>    A1:B1:D1    A1:C1:D1    B1:C1:D1 A1:B1:C1:D1
#>     -0.0375      1.3875      0.2875     -0.2625
```

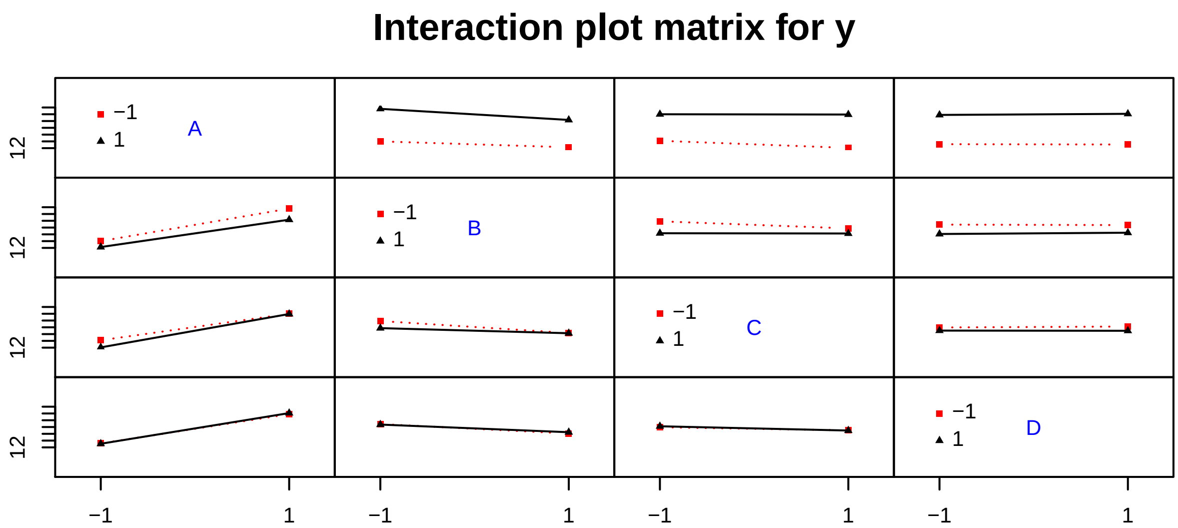
```r
DanielPlot(full.fit)
```




```r
MEPlot(full.fit)
```




```r
IAPlot(full.fit)
```



b)

The regression model that corresponds to this analysis is

$$
\begin{aligned}
Y = \beta_0 &+ \beta_A A + \beta_B B + \beta_C C + \beta_D D \\
&+ \beta_{AB} AB + \beta_{AC} AC + \beta_{AD} AD + \beta_{BC} BC + \beta_{BD} BD + \beta_{CD} CD \\
&+ \beta_{ABC} ABC + \beta_{ABD} ABD + \beta_{BCD} BCD + \beta_{ACD} ACD + \beta_{ABCD} ABCD + \epsilon,
\end{aligned}
$$

where $A, B, C, D \in \{-1, 1\}$.

c)

There are no standard deviation estimates in the output above, since we have $2^4 = 16 = p$ and the model is perfectly fit to the data.

Assume $\sigma^2 = 4$. We know that $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \frac{\sigma^2}{16} I)$ in two-level factorial designs. Thus, we have the following confidence interval for an element $\hat{\beta}_J$

$$
\begin{aligned}
1 - \alpha &= P\left(-z_{\alpha/2} < \frac{\hat{\beta}_j - \beta_i}{\sigma/4} < z_{\alpha/2}\right) \\
&= P\left(\hat{\beta}_j - \frac{\sigma}{4} z_{\alpha/2} < \beta_i < \hat{\beta}_j + \frac{\sigma}{4} z_{\alpha/2}\right)
\end{aligned}
$$

Calculated in R, the 95 % confidence interval for each of the effects (twice the coefficients) are

```
cbind(2*full.fit$coefficients-qnorm(0.025, lower.tail = F), 2*full.fit$coefficients+qnorm(0.025, lower.
```

```
#>                   [,1]      [,2]
#> (Intercept) 33.127536 37.047464
#> A1           6.877536 10.797464
#> B1          -4.472464 -0.552536
#> C1          -3.047464  0.872464
#> D1          -1.847464  2.072464
#> A1:B1       -2.722464  1.197464
```

```
#> A1:C1       -0.947464  2.972464
#> A1:D1       -1.747464  2.172464
#> B1:C1       -0.947464  2.972464
#> B1:D1       -1.697464  2.222464
#> C1:D1       -2.122464  1.797464
#> A1:B1:C1    -1.747464  2.172464
#> A1:B1:D1    -1.997464  1.922464
#> A1:C1:D1    -0.572464  3.347464
#> B1:C1:D1    -1.672464  2.247464
#> A1:B1:C1:D1 -2.222464  1.697464
```

From the output, it becomes apparent that the main effects for A and B are the only effects that are significantly different from 0, since 0 is not in their confidence intervals.

d)

When assuming that all three-way and four-way interactions are zero, the variance σ^2 can be estimated by excluding these higher order interactions using the estimate $\hat{\sigma}^2 = \frac{\text{SSE}_{\text{red}}}{2^k-m-1}$, where SSE_{red} is the SSE in the reduced model and m are the columns in the design matrix that are kept in the reduced model. Note that SSE_{red} can be calculated by the formula $\text{SSE}_{\text{red}} = 2^k \sum_{j=m+1}^{2^k-1} \hat{\beta}_j^2$, i.e. a constant times the sum of the squares of the removed parameter estimates. This means that the variance of the effect estimators can be estimated by $\widehat{\text{Var}(2\hat{\beta}_j)} = \frac{\hat{\sigma}^2}{2^k} = \frac{1}{2^k-m-1} \sum_{j=m+1}^{2^k-1} \hat{\beta}_j^2$, i.e. the average of the square of the removed coefficient parameters. The significant effects can be found via a t-test with the test statistic $\frac{\hat{\beta}_j}{\hat{\sigma}/2^{k/2}} \sim t_{2^k-m-1}$.

```r
reduced.fit <- lm(y~.^2, data = plan)
summary(reduced.fit)
```

```
#>
#> Call:
#> lm.default(formula = y ~ .^2, data = plan)
#>
#> Residuals:
#>       1       2       3       4       5       6       7       8       9      10
#> -1.0562  0.7687 -0.3313  0.6188  1.0937 -0.8062  0.2938 -0.5813  0.8437 -0.5562
#>      11      12      13      14      15      16
#>  0.5438 -0.8312 -0.8812  0.5938 -0.5063  0.7937
#>
#> Coefficients:
#>             Estimate Std. Error t value Pr(>|t|)
#> (Intercept) 17.54375    0.32583  53.844 4.18e-08 ***
#> A1           4.41875    0.32583  13.562 3.91e-05 ***
#> B1          -1.25625    0.32583  -3.856   0.0119 *
#> C1          -0.54375    0.32583  -1.669   0.1560
#> D1           0.05625    0.32583   0.173   0.8697
#> A1:B1       -0.38125    0.32583  -1.170   0.2947
#> A1:C1        0.50625    0.32583   1.554   0.1810
#> A1:D1        0.10625    0.32583   0.326   0.7576
#> B1:C1        0.50625    0.32583   1.554   0.1810
#> B1:D1        0.13125    0.32583   0.403   0.7037
#> C1:D1       -0.08125    0.32583  -0.249   0.8130
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 1.303 on 5 degrees of freedom
```

```
#> Multiple R-squared:  0.9765, Adjusted R-squared:  0.9296
#> F-statistic: 20.81 on 10 and 5 DF,  p-value: 0.001849
```

The output gives that an estimate of σ^2 is $1.303^2 \approx 1.70$. how can you the estimate the variance of the error and the variance of the effect estimators??

e)

f)

Problem 3 Process development - from Exam TMA4255 2012 Summer