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

alexaoh

11 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 $j^{\text{th}}$ entry of $\hat{\boldsymbol{\beta}}$ only depends on the $j^{\text{th}}$ 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

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

# 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)
```

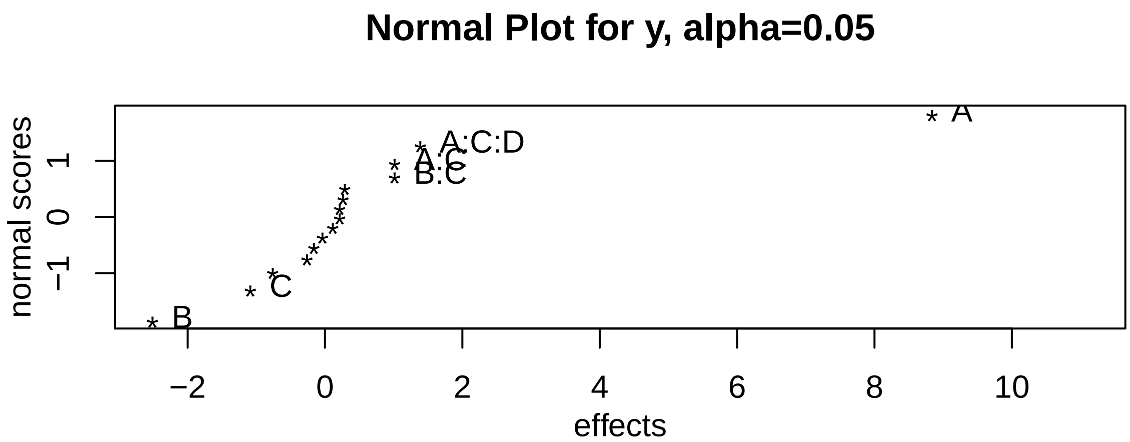
```r
plan <- FrF2(nruns=16,nfactors=4,randomize=FALSE)
plan <- add.response(plan,y)

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
```

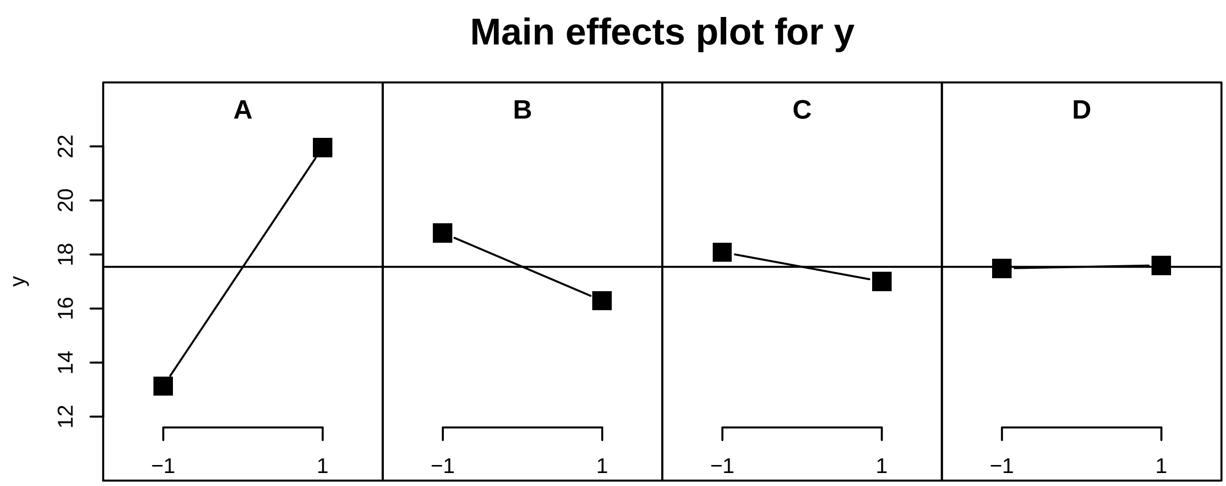
```r
effects <- full.fit$coefficients*2
effects
```

```
#> (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
```

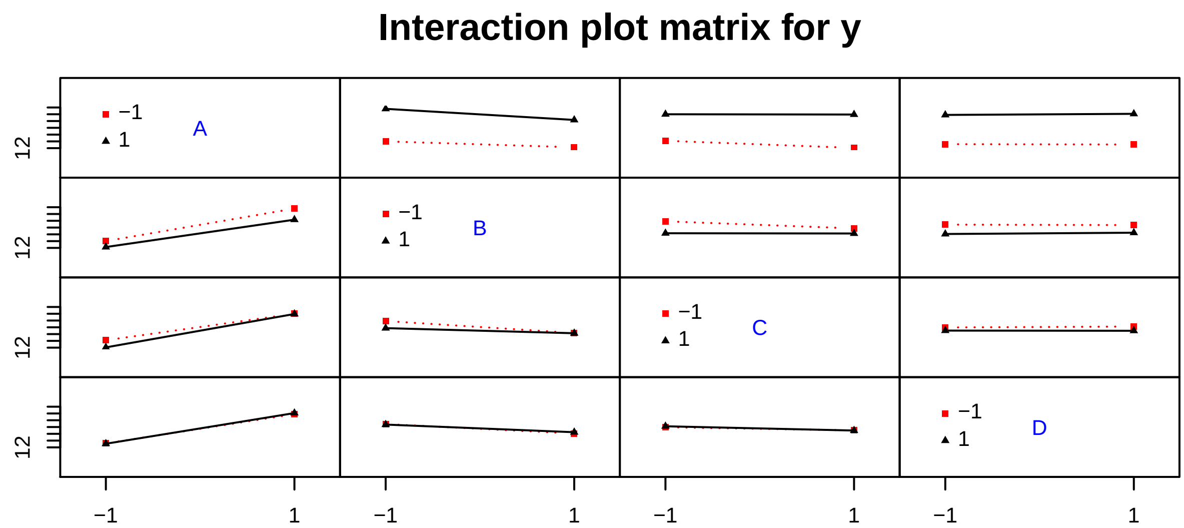
```r
DanielPlot(full.fit)
```




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




```
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.t
```

```
#>                    [,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 $\sigma^2$ 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 $\text{SSE}_{\text{red}}$ 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 $\text{SSE}_{\text{red}}$ 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 $\widehat{\text{Var}(\hat{\beta}_j)} = \frac{\hat{\sigma}^2}{2^k} = \frac{1}{2^k-m-1} \sum_{j=m+1}^{2^k-1} \hat{\beta}_j^2$, 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 $\sigma^2$ is $1.303^2 \approx 1.70$. The estimate of the variance of the coefficients is hence $\frac{\hat{\sigma}^2}{16} \approx \frac{1.70}{16} \approx 0.106$. This means that the estimate of the variance of the effects is $\approx 4 \cdot 0.106 \approx 0.425$. These values can also be approximated by the formulas written above, as seen in the R output below.

```
sdbetahatest<-sqrt(summary(reduced.fit)$sigma^2/16)# st. errors shown in summary table.
sdbetahatest
```

```
#> [1] 0.3258283
```

```
# Three different ways to estimate SSE.
sse <- sum((y-reduced.fit$fitted.values)^2)
sse
```

```
#> [1] 8.493125
```

```
summary(reduced.fit)$sigma^2*(16-11) # the same
```

```
#> [1] 8.493125
```

```
# this function of removed covariates of full model is equal to sse of reduced model
16*sum(full.fit$coeff[12:16]^2)
```

```
#> [1] 8.493125
```

```
# Estimate of sigma^2, same as found above.
sigma.sq <- sse/(16-11)
sigma.sq
```

```
#> [1] 1.698625
```

```
# Estimate of variance of effects.
4*sigma.sq/16
```

```
#> [1] 0.4246562
```

```
4/(16-11)*sum(full.fit$coefficients[12:16]^2)
```

```
#> [1] 0.4246562
```

## e)

```
design1 <- FrF2(16, 4, blocks=2, randomize=FALSE)
summary(design1)
```

```
#> Call:
#> FrF2(16, 4, blocks = 2, randomize = FALSE)
#>
#> Experimental design of type  FrF2.blocked
#> 16  runs
#> blocked design with  2  blocks of size  8
#>
#> Factor settings (scale ends):
#>    A  B  C  D
#> 1 -1 -1 -1 -1
#> 2  1  1  1  1
#>
#> Design generating information:
```

```
#> $legend
#> [1] A=A B=B C=C D=D
#>
#> $`generators for design itself`
#> [1] full factorial
#>
#> $`block generators`
#> [1] 15
#>
#>
#> no aliasing of main effects or 2fis  among experimental factors
#>
#> Aliased with block main effects:
#> [1] none
#>
#> The design itself:
#>   run.no run.no.std.rp Blocks  A  B  C  D
#> 1      1          2.1.1      1 -1 -1 -1  1
#> 2      2          3.1.2      1 -1 -1  1 -1
#> 3      3          5.1.3      1 -1  1 -1 -1
#> 4      4          8.1.4      1 -1  1  1  1
#> 5      5          9.1.5      1  1 -1 -1 -1
#> 6      6         12.1.6      1  1 -1  1  1
#> 7      7         14.1.7      1  1  1 -1  1
#> 8      8         15.1.8      1  1  1  1 -1
#>    run.no run.no.std.rp Blocks  A  B  C  D
#> 9       9          1.2.1      2 -1 -1 -1 -1
#> 10     10          4.2.2      2 -1 -1  1  1
#> 11     11          6.2.3      2 -1  1 -1  1
#> 12     12          7.2.4      2 -1  1  1 -1
#> 13     13         10.2.5      2  1 -1 -1  1
#> 14     14         11.2.6      2  1 -1  1 -1
#> 15     15         13.2.7      2  1  1 -1 -1
#> 16     16         16.2.8      2  1  1  1  1
#> class=design, type= FrF2.blocked
#> NOTE: columns run.no and run.no.std.rp  are annotation,
#>  not part of the data frame
```

ABCD is the only effect confounded with the block effect.

## f)

```r
design2 <- FrF2(16, 4, blocks=4, alias.block.2fis=TRUE, randomize=FALSE)
summary(design2)
```

```
#> Call:
#> FrF2(16, 4, blocks = 4, alias.block.2fis = TRUE, randomize = FALSE)
#>
#> Experimental design of type  FrF2.blocked
#> 16  runs
#> blocked design with  4  blocks of size  4
#>
#> Factor settings (scale ends):
#>    A  B  C  D
```

```
#> 1 -1 -1 -1 -1
#> 2  1  1  1  1
#>
#> Design generating information:
#> $legend
#> [1] A=A B=B C=C D=D
#>
#> $`generators for design itself`
#> [1] full factorial
#>
#> $`block generators`
#> b1 b2
#> 13 14
#>
#>
#> no aliasing of main effects or 2fis  among experimental factors
#>
#> Aliased with block main effects:
#> [1] AB
#>
#> The design itself:
#>   run.no run.no.std.rp Blocks  A  B  C  D
#> 1      1         1.1.1      1 -1 -1 -1 -1
#> 2      2         4.1.2      1 -1 -1  1  1
#> 3      3        14.1.3      1  1  1 -1  1
#> 4      4        15.1.4      1  1  1  1 -1
#>   run.no run.no.std.rp Blocks  A  B  C  D
#> 5      5         5.2.1      2 -1  1 -1 -1
#> 6      6         8.2.2      2 -1  1  1  1
#> 7      7        10.2.3      2  1 -1 -1  1
#> 8      8        11.2.4      2  1 -1  1 -1
#>    run.no run.no.std.rp Blocks  A  B  C  D
#> 9       9         6.3.1      3 -1  1 -1  1
#> 10     10         7.3.2      3 -1  1  1 -1
#> 11     11         9.3.3      3  1 -1 -1 -1
#> 12     12        12.3.4      3  1 -1  1  1
#>    run.no run.no.std.rp Blocks  A  B  C  D
#> 13     13         2.4.1      4 -1 -1 -1  1
#> 14     14         3.4.2      4 -1 -1  1 -1
#> 15     15        13.4.3      4  1  1 -1 -1
#> 16     16        16.4.4      4  1  1  1  1
#> class=design, type= FrF2.blocked
#> NOTE: columns run.no and run.no.std.rp  are annotation,
#>  not part of the data frame
```

Covariates 13 ($ACD$) and 14 ($BCD$) are proposed as block generators. This means that both these effects are confounded with the blocks, in addition to the effect $ACD \cdot BCD = AB$.

## Problem 3 Process development - from Exam TMA4255 2012 Summer

### a)

The effect estimate for factor $B$ is $\frac{633+642+1075+729}{4} - \frac{550+669+1037+749}{4} = 18.5$. The main effects plot was drawn by hand. The estimated main effect of $B$ is interpreted as the expected increase in the etch rate when the flow of gas increases from 125 to 200. This increase in etch rate corresponds to 18.5.

### b)

The "Std. Error" gives the estimated standard deviation of each of the regression coefficients. These are the same for all covariates, since the design matrix is orthogonal, which means that $(X^T X)^{-1}$ is diagonal. More precisely, all estimated standard deviations are $\hat{\sigma}/\sqrt{n}$. Since the residual standard error is 47.46, the estimated standard deviation of each of the coefficients is $47.46/\sqrt{16} = 11.865$, which matches the output.

The estimated effect for $B$ is twice the estimated coefficient for $B$. The missing number for the $t$-statistic is $\frac{\hat{B}}{\hat{\sigma}/2^{k/2}} = \frac{\hat{B}}{\text{Std. Error}} = \frac{3.688}{11.865} = 0.3108$. The hypotheses underlying the $p$-value are $H_0 : B = 0$ vs. $H_1 : B \neq 0$. The conclusion is that there is not enough evidence to conclude that $Bneq0$, hence $H_0$ is kept. The significant covariates are $A, B$ and $AC$, at significance level 0.05.

### c)

The estimated coefficients for $A, B$ and $AC$ are the same in both the models since the coefficient estimates only depend on the response and each respective column in the design matrix. The standard errors have changed, since the $SSE$ changes when the amount of covariates used changes.

Based on the estimates, I would suggest that $A$ and $AC$ should be low and $C$ should be high, in order to yield the largest etch rate. However, this is not possible, since $AC$ is negative when the signs of $A$ and $C$ are different. Hence, the optimal values are $A$ low, $C$ high, and, thus, $AC$ low.

Have a look at the solution for a 95% prediction interval based on the chosen levels of $A$ and $C$. The expression for the prediction interval is the same as always. We use the vector $\boldsymbol{x}_0^T = (1 \quad -1 \quad 1 \quad -1)$, since we have an intercept, $A$ low, $C$ high and $AC$ low, as the new predictor.

### d)

This is a fractional factorial design. More specifically, a half fraction of a $2^3$ experiment, i.e. a $2^{3-1}$ experiment.

The generator for the experiment is $C = -AB$.

The defining relation for the experiment is $I = -ABC$, since these are the only rows in the original plan that are performed. This can also be seen from the above noted generator, by multiplying both sides by $C$.

The alias structure of the experiment is found by multiplying the effects by the defining relation. This gives the aliases $A = -BC$, $B = -AC$ and $C = -AB$.

The resolution of the experiment is III, since the defining relation has three letters.

## Problem 4 Blocking

No software is used.

This experiment can be blocked as explained in the following. Choose the block factors $BC, BD$ and $DE$. Then the block factors are confunded with these three effects and $BC \cdot BD \cdot DE = CE$, $BC \cdot BD = CD$, $BC \cdot DE = BCDE$ and $BD \cdot DE = BE$. Hence, the main effect $A$ and all two-factor interactions involving $A$ are not confunded with the block factors.

## Problem 5 Fractional factorial design

No software is used.

### a)

The generator for the design is $D = ABC \implies I = ABCD$ is the defining relation. Hence, the resolution is IV.

### b)

After calculating all the aliases, the resolution is seen to be IV.

### c)

Denne skjønner jeg ikke :( Hvordan finner jeg defining relation her?