Untitled

```
data <- read.delim('q6.txt', header = TRUE, sep = ',')

# feature engineering
f1 <- data$x1 - data$x2 - data$x3
f2 <- data$x4 + (1/5)*data$x2 + (8/5)*data$x3

# model with restriction
m.r <- lm(data$y~ f1 + f2)
summary(m.r)
```

```
##
## Call:
## lm(formula = data$y ~ f1 + f2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7349 -1.6190 -0.4981  1.0728  4.7363
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 39.15850   12.48929   3.135   0.0106 *
## f1           1.25022    0.18563   6.735 5.14e-05 ***
## f2           0.08553    0.34976   0.245   0.8118
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.752 on 10 degrees of freedom
## Multiple R-squared:  0.8283, Adjusted R-squared:  0.7939
## F-statistic: 24.11 on 2 and 10 DF,  p-value: 0.0001494
```

```
# model without restriction
m <- lm(data$y ~ data$x1 + data$x2 + data$x3 + data$x4)
summary(m)
```

```
##
## Call:
## lm(formula = data$y ~ data$x1 + data$x2 + data$x3 + data$x4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2629 -1.0501 -0.5268  0.2542  3.3737
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  49.6766    12.2737   4.047 0.003697 **
## data$x1       0.9202     0.2163   4.254 0.002785 **
## data$x2      -1.9936     0.3012  -6.618 0.000166 ***
## data$x3      -0.7015     0.7219  -0.972 0.359615
## data$x4      -0.2839     0.2979  -0.953 0.368496
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.119 on 8 degrees of freedom
## Multiple R-squared:  0.9186, Adjusted R-squared:  0.8778
## F-statistic: 22.56 on 4 and 8 DF,  p-value: 0.0002057
```

```r
# compare models
anova(m.r, m)
```

```
## Analysis of Variance Table
##
## Model 1: data$y ~ f1 + f2
## Model 2: data$y ~ data$x1 + data$x2 + data$x3 + data$x4
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     10 75.710
## 2      8 35.905  2    39.805 4.4345 0.05058 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p value is not significant we cannot reject the restricted model