STK2100
Oblig 1

Egil Furnes
Studentnummer: 693784

Oppgave 1

a)

i)

The only model that can be written as a linear model as is, is model 2.

$$2. \quad Y = \beta_0 + \frac{\beta_1}{x} + \beta_2 x^2 + \epsilon$$

ii)

For model 4 we can fix β_2 to a constant, say c and get

$$4. \quad \beta_0 + \beta_1 x^c + \epsilon$$

iii)

For model 5 we can log-transform such as this:

$$Y = \beta_0 x^{\beta_1} \varepsilon$$

$$\log(Y) = \log(\beta_0 x^{\beta_1} \varepsilon)$$
$$\log(Y) = \log(\beta_0) + \log(x^{\beta_1}) + \log(\varepsilon)$$
$$\log(Y) = \log(\beta_0) + \beta_1 \log(x) + \log(\varepsilon)$$

b)

1. $X = \begin{bmatrix} \frac{1}{1+x_i} & x_i^{1/2} \end{bmatrix},$ $\beta = \begin{bmatrix} \beta_0 \\ \beta_2 \end{bmatrix}$
2. $X = \begin{bmatrix} 1 & \frac{1}{x_i} & x_i^2 \end{bmatrix},$ $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$
3. $X = \begin{bmatrix} 1 & x_i & x_i^2 \end{bmatrix},$ $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$
4. $X = \begin{bmatrix} 1 & x_i \end{bmatrix},$ $\beta = \begin{bmatrix} \beta_0 \\ \beta_1' \end{bmatrix}$
5. $X = \begin{bmatrix} 1 & \log(x_i) \end{bmatrix},$ $\beta = \begin{bmatrix} \beta_0' \\ \beta_1 \end{bmatrix}$

Oppgave 2

a)

```
# set seed for replication!
set.seed(1814)

# loading packages
library(tidyverse)

# reading data
nuclear <- read_delim("nuclear.dat")

# a)

# fitting the data using linear regression
lm1 <- lm(log(cost)~., nuclear)
lm1 %>% summary()

# creating a 95 percent confidence interval
confint(lm1, level = 0.95)[c("t1","t2","bw"),]
```

A 95% confidence interval is then found to be the following.

```
          2.5 %     97.5 %
t1 -0.041123331 0.05162679
t2 -0.003949911 0.01516185
bw -0.184211843 0.25780411
```

b)

```
# b)

# creating a new dataframe
df <- tibble(
  date = 70.0,
  t1 = 13,
  t2 = 50,
  cap = 800,
  pr = 1,
  ne = 0,
  ct = 0,
  bw = 1,
  cum.n = 8,
  pt = 1
)

```

```
# predicting on data using linear regression
pred1 <- predict(lm1, newdata = df, interval = "prediction", level =
    0.95)

# retrieving the coefficients
yhat <- pred1[1,"fit"]
lwry <- pred1[1,"lwr"]
upry <- pred1[1,"upr"]

# transforming y to find z
zfit <- exp(yhat)
zlwr <- exp(lwry)
zupr <- exp(upry)

# saving coefficients for z
predz <- data.frame(
  fit = zfit,
  lwr = zlwr,
  upr = zupr
)

# presenting my findings
print(pred1)
print(predz)
```

The 95% prediction interval with the cost Z is then found to be the following.

```
> print(pred1)
       fit      lwr      upr
1 5.964135 5.394248 6.534022
> print(predz)
       fit      lwr      upr
1 389.2163 220.1366 688.1607
```

c)

```
# c)

# individual t-tests
sum_lm1 <- summary(lm1)
print(sum_lm1$coefficients[c("t1", "t2", "bw"),
                           c("Estimate", "Pr(>|t|)"),
                           drop = FALSE])

# joint F-tests
lm1red <- lm(log(cost)~date+cap+pr+ne+ct+cum.n+pt, data = nuclear)
anova(lm1red, lm1)
```

Find the output for the individual t-test, where we find that the p-value is larger than 0.5 for all predictors `t1`, `t2`, and `bw`, such that we fail to reject the null-hypothesis $H_0 : \beta_j = 0$ that the predictors are significant.

```
     Estimate  Pr(>|t|)
t1 0.005251730 0.8160981
t2 0.005605968 0.2359862
bw 0.036796131 0.7326075
```

For the joint F-test, we find a p-value of $0.5173 > 0.5$ where we fail to reject the H_0 at a 5% confidence level.

```
Analysis of Variance Table

Model 1: log(cost) ~ date + cap + pr + ne + ct + cum.n + pt
Model 2: log(cost) ~ date + t1 + t2 + cap + pr + ne + ct + bw + cum.n +
    pt
  Res.Df     RSS Df Sum of Sq     F Pr(>F)
1     24 0.67195
2     21 0.60443  3   0.06752 0.782 0.5173
```

d)

The output of the order matrix is below, which gives the following order of variables included: `pt` $\rightarrow$`cap` $\rightarrow$`date` $\rightarrow$`ne` $\rightarrow$`ct` $\rightarrow$`cum.n` $\rightarrow$`bw t2` $\rightarrow$`t1`

```
Selection Algorithm: forward
          date t1  t2  cap pr  ne  ct  bw  cum.n pt
1  ( 1 )  " "  " " " " " " " " " " " " " " " "   "*"
2  ( 1 )  " "  " " " " "*" " " " " " " " " " "   "*"
3  ( 1 )  "*"  " " " " "*" " " " " " " " " " "   "*"
4  ( 1 )  "*"  " " " " "*" " " "*" " " " " " "   "*"
5  ( 1 )  "*"  " " " " "*" " " "*" "*" " " " "   "*"
6  ( 1 )  "*"  " " " " "*" " " "*" "*" " " "*"   "*"
7  ( 1 )  "*"  " " " " "*" " " "*" "*" "*" "*"   "*"
8  ( 1 )  "*"  " " " " "*" "*" "*" "*" "*" "*"   "*"
9  ( 1 )  "*"  " " "*" "*" "*" "*" "*" "*" "*"   "*"
10  ( 1 ) "*"  "*" "*" "*" "*" "*" "*" "*" "*"   "*"
```

e)

f)

g)

h)

i)