Modelleren2A

```
library(knitr)
library(MASS)
library(car)
```

```
## Loading required package: carData
```

```
library(readxl)
```

```
x<-read_excel("TU Delft publications.xlsx")
x
```

```
## # A tibble: 2,412 x 9
##          ut     cs      js p_int_collab p_industry n_authors n_countries
##       <dbl>  <dbl>  <dbl>        <dbl>      <dbl>     <dbl>       <dbl>
##  1 3.37e11    11   53.6            1          0         7           3
##  2 3.41e11    34   94.2            1          0        10           2
##  3 3.30e11    24  122.             1          0         6           3
##  4 3.36e11   149  122.             1          0         2           3
##  5 3.29e11    34  116.             0          0         2           1
##  6 3.31e11   137  116.             1          1        15           5
##  7 3.37e11    83  116.             1          0        10           2
##  8 3.43e11    48  116.             1          0        13           4
##  9 3.31e11   110  151.             0          0         1           1
## 10 3.40e11   123. 121.             1          1        11           2
## # ... with 2,402 more rows, and 2 more variables: n_pages <dbl>,
## #   n_refs <dbl>
```

```
summary(x)
```

```
##        ut                  cs                js             p_int_collab
##  Min.   :2.096e+11   Min.   :  0.00   Min.   :  0.1364   Min.   :0.0000
##  1st Qu.:3.339e+11   1st Qu.:  3.00   1st Qu.:  4.8593   1st Qu.:0.0000
##  Median :3.385e+11   Median :  6.00   Median :  8.0245   Median :1.0000
##  Mean   :3.381e+11   Mean   : 11.74   Mean   : 10.7170   Mean   :0.5659
##  3rd Qu.:3.431e+11   3rd Qu.: 12.00   3rd Qu.: 11.9910   3rd Qu.:1.0000
##  Max.   :3.657e+11   Max.   :484.00   Max.   :154.1385   Max.   :1.0000
##    p_industry        n_authors        n_countries        n_pages
##  Min.   :0.0000   Min.   :  1.00   Min.   : 1.000   Min.   : 2.00
##  1st Qu.:0.0000   1st Qu.:  3.00   1st Qu.: 1.000   1st Qu.: 8.00
##  Median :0.0000   Median :  4.00   Median : 2.000   Median :11.00
##  Mean   :0.1049   Mean   :  4.72   Mean   : 1.835   Mean   :12.04
##  3rd Qu.:0.0000   3rd Qu.:  6.00   3rd Qu.: 2.000   3rd Qu.:14.00
##  Max.   :1.0000   Max.   :114.00   Max.   :15.000   Max.   :60.00
##      n_refs
##  Min.   :  0.00
##  1st Qu.: 23.00
##  Median : 32.00
##  Mean   : 38.92
##  3rd Qu.: 46.00
##  Max.   :404.00
```

```
model <- lm(cs~js+p_int_collab+p_industry+n_authors+n_countries+n_pages+n_refs, data=x)
stepd <- stepAIC(model, direction ='backward',trace=FALSE)
```

```
stepf <- stepAIC(model, direction ='forward',trace=FALSE)
stepdb <- stepAIC(model, direction ='both',trace=FALSE)
summary(stepd)
```

```
##
## Call:
## lm(formula = cs ~ js + n_authors + n_refs, data = x)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -91.33  -5.81  -2.05   2.14 445.07
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.12810    0.87900  -5.834 6.14e-09 ***
## js           0.76539    0.03495  21.899  < 2e-16 ***
## n_authors    0.55406    0.11437   4.845 1.35e-06 ***
## n_refs       0.15555    0.01508  10.312  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.87 on 2408 degrees of freedom
## Multiple R-squared:  0.2493, Adjusted R-squared:  0.2483
## F-statistic: 266.5 on 3 and 2408 DF,  p-value: < 2.2e-16
```