**Statistical Modelling Techniques ~ Elif Kartal**

Analyze MROZ and AIDS2 datasets with truncated and censored regression, respectively.

**Install and load needed packages:**

```
install.packages("AER")
```

```
Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
(as 'lib' is unspecified)
```

```
install.packages("censReg")
```

```
Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
(as 'lib' is unspecified)
```

```
install.packages("truncreg")
```

```
Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
(as 'lib' is unspecified)
```

```
install.packages("Ecdat")
```

```
Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
(as 'lib' is unspecified)
```

```
install.packages("MASS")
```

```
Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
(as 'lib' is unspecified)
```

```
library(MASS)
library(Ecdat)
```

```
Loading required package: Ecfun
```

```
Attaching package: 'Ecfun'
```

```
The following object is masked from 'package:base':

    sign


Attaching package: 'Ecdat'

The following object is masked from 'package:MASS':

    SP500

The following object is masked from 'package:datasets':

    Orange
```

```
library(AER)
```

```
Loading required package: car

Loading required package: carData


Attaching package: 'carData'

The following object is masked from 'package:Ecdat':

    Mroz

Loading required package: lmtest

Loading required package: zoo


Attaching package: 'zoo'

The following objects are masked from 'package:base':

    as.Date, as.Date.numeric
```

```
Loading required package: sandwich

Loading required package: survival
```

```
library(censReg)
```

```
Loading required package: maxLik

Loading required package: miscTools


Attaching package: 'miscTools'

The following object is masked from 'package:Ecfun':

    checkNames


Please cite the 'maxLik' package as:
Henningsen, Arne and Toomet, Ott (2011). maxLik: A package for maximum likelihood estimation

If you have questions, suggestions, or comments regarding the 'maxLik' package, please use a
https://r-forge.r-project.org/projects/maxlik/


Please cite the 'censReg' package as:
Henningsen, Arne (2017). censReg: Censored Regression (Tobit) Models. R package version 0.5.

If you have questions, suggestions, or comments regarding the 'censReg' package, please use a
https://r-forge.r-project.org/projects/sampleselection/
```

```
library(truncreg)
```

**Truncated Model:**

```
data("Mroz", package = "Ecdat")

# Veriyi gözlemleyelim
summary(Mroz)
```

```
 work         hoursw           child6          child618          agew
no :325   Min.   :   0.0   Min.   :0.0000   Min.   :0.000   Min.   :30.00
yes:428   1st Qu.:   0.0   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:36.00
          Median : 288.0   Median :0.0000   Median :1.000   Median :43.00
          Mean   : 740.6   Mean   :0.2377   Mean   :1.353   Mean   :42.54
          3rd Qu.:1516.0   3rd Qu.:0.0000   3rd Qu.:2.000   3rd Qu.:49.00
          Max.   :4950.0   Max.   :3.0000   Max.   :8.000   Max.   :60.00
    educw           hearnw           wagew          hoursh          ageh
Min.   : 5.00   Min.   : 0.000   Min.   :0.00   Min.   : 175   Min.   :30.00
1st Qu.:12.00   1st Qu.: 0.000   1st Qu.:0.00   1st Qu.:1928   1st Qu.:38.00
Median :12.00   Median : 1.625   Median :0.00   Median :2164   Median :46.00
Mean   :12.29   Mean   : 2.375   Mean   :1.85   Mean   :2267   Mean   :45.12
3rd Qu.:13.00   3rd Qu.: 3.788   3rd Qu.:3.58   3rd Qu.:2553   3rd Qu.:52.00
Max.   :17.00   Max.   :25.000   Max.   :9.98   Max.   :5010   Max.   :60.00
    educh           wageh            income          educwm
Min.   : 3.00   Min.   : 0.4121   Min.   : 1500   Min.   : 0.000
1st Qu.:11.00   1st Qu.: 4.7883   1st Qu.:15428   1st Qu.: 7.000
Median :12.00   Median : 6.9758   Median :20880   Median :10.000
Mean   :12.49   Mean   : 7.4822   Mean   :23081   Mean   : 9.251
3rd Qu.:15.00   3rd Qu.: 9.1667   3rd Qu.:28200   3rd Qu.:12.000
Max.   :17.00   Max.   :40.5090   Max.   :96000   Max.   :17.000
    educwf         unemprate       city       experience
Min.   : 0.000   Min.   : 3.000   no :269   Min.   : 0.00
1st Qu.: 7.000   1st Qu.: 7.500   yes:484   1st Qu.: 4.00
Median : 7.000   Median : 7.500             Median : 9.00
Mean   : 8.809   Mean   : 8.624             Mean   :10.63
3rd Qu.:12.000   3rd Qu.:11.000             3rd Qu.:15.00
Max.   :17.000   Max.   :14.000             Max.   :45.00
```

```
str(Mroz)
```

```
'data.frame':   753 obs. of  18 variables:
 $ work      : Factor w/ 2 levels "no","yes": 2 2 2 2 2 2 2 2 2 2 ...
 $ hoursw    : int  1610 1656 1980 456 1568 2032 1440 1020 1458 1600 ...
 $ child6    : int  1 0 1 0 1 0 0 0 0 0 ...
 $ child618  : int  0 2 3 3 2 0 2 0 2 2 ...
 $ agew      : int  32 30 35 34 31 54 37 54 48 39 ...
 $ educw     : int  12 12 12 12 14 12 16 12 12 12 ...
 $ hearnw    : num  3.35 1.39 4.55 1.1 4.59 ...
 $ wagew     : num  2.65 2.65 4.04 3.25 3.6 4.7 5.95 9.98 0 4.15 ...
 $ hoursh    : int  2708 2310 3072 1920 2000 1040 2670 4120 1995 2100 ...
 $ ageh      : int  34 30 40 53 32 57 37 53 52 43 ...
```

```
$ educh     : int  12 9 12 10 12 11 12 8 4 12 ...
$ wageh     : num  4.03 8.44 3.58 3.54 10 ...
$ income    : int  16310 21800 21040 7300 27300 19495 21152 18900 20405 20425 ...
$ educwm    : int  12 7 12 7 12 14 14 3 7 7 ...
$ educwf    : int  7 7 7 7 14 7 7 3 7 7 ...
$ unemprate : num  5 11 5 5 9.5 7.5 5 5 3 5 ...
$ city      : Factor w/ 2 levels "no","yes": 1 2 1 1 2 2 1 1 1 1 ...
$ experience: int  14 5 15 6 7 33 11 35 24 21 ...
```

```
truncated <- subset(Mroz, wageh > 0)
```

```
model <- truncreg(wageh ~ educw + experience + I(experience^2) + ageh,
                  data = truncated, point = 0, direction = "left")
```

```
summary(truncated)
```

```
 work         hoursw          child6          child618          agew
no :325   Min.   :   0.0   Min.   :0.0000   Min.   :0.000   Min.   :30.00
yes:428   1st Qu.:   0.0   1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:36.00
          Median : 288.0   Median :0.0000   Median :1.000   Median :43.00
          Mean   : 740.6   Mean   :0.2377   Mean   :1.353   Mean   :42.54
          3rd Qu.:1516.0   3rd Qu.:0.0000   3rd Qu.:2.000   3rd Qu.:49.00
          Max.   :4950.0   Max.   :3.0000   Max.   :8.000   Max.   :60.00
    educw           hearnw           wagew          hoursh          ageh
Min.   : 5.00   Min.   : 0.000   Min.   :0.00   Min.   : 175   Min.   :30.00
1st Qu.:12.00   1st Qu.: 0.000   1st Qu.:0.00   1st Qu.:1928   1st Qu.:38.00
Median :12.00   Median : 1.625   Median :0.00   Median :2164   Median :46.00
Mean   :12.29   Mean   : 2.375   Mean   :1.85   Mean   :2267   Mean   :45.12
3rd Qu.:13.00   3rd Qu.: 3.788   3rd Qu.:3.58   3rd Qu.:2553   3rd Qu.:52.00
Max.   :17.00   Max.   :25.000   Max.   :9.98   Max.   :5010   Max.   :60.00
    educh           wageh            income          educwm
Min.   : 3.00   Min.   : 0.4121   Min.   : 1500   Min.   : 0.000
1st Qu.:11.00   1st Qu.: 4.7883   1st Qu.:15428   1st Qu.: 7.000
Median :12.00   Median : 6.9758   Median :20880   Median :10.000
Mean   :12.49   Mean   : 7.4822   Mean   :23081   Mean   : 9.251
3rd Qu.:15.00   3rd Qu.: 9.1667   3rd Qu.:28200   3rd Qu.:12.000
Max.   :17.00   Max.   :40.5090   Max.   :96000   Max.   :17.000
    educwf         unemprate       city       experience
Min.   : 0.000   Min.   : 3.000   no :269   Min.   : 0.00
1st Qu.: 7.000   1st Qu.: 7.500   yes:484   1st Qu.: 4.00
Median : 7.000   Median : 7.500             Median : 9.00
```

```
 Mean   : 8.809   Mean   : 8.624             Mean   :10.63
 3rd Qu.:12.000   3rd Qu.:11.000             3rd Qu.:15.00
 Max.   :17.000   Max.   :14.000             Max.   :45.00
```

Elde edilen sonuçlar, bağımsız değişkenlerin saatlik ücretine etkisini açıkça göstermektedir. Eğitim düzeyi ve iş deneyimi, saatlik ücrete pozitif ve istatistiksel olarak anlamlı bir etkiye sahipken, yaş değişkeni negatif bir etkiye işaret etmektedir. Bu, yaş arttıkça kadınların saatlik ücretinin azaldığına işaret edebilir.

**Censored Model:**

```r
data(Aids2) #loading data
```

```r
model <- survreg(Surv(death, death >= 9000) ~ age + sex, data = Aids2, dist = "gaussian")

summary(model)
```

```
Call:
survreg(formula = Surv(death, death >= 9000) ~ age + sex, data = Aids2,
    dist = "gaussian")
               Value Std. Error      z      p
(Intercept)  1.10e+04   7.83e+01 140.51 <2e-16
age         -3.07e+00   1.15e+00  -2.68 0.0074
sexM         1.17e+02   6.59e+01   1.78 0.0749
Log(scale)   6.42e+00   1.33e-02 482.96 <2e-16

Scale= 612

Gaussian distribution
Loglik(model)= -22196.3   Loglik(intercept only)= -22201.5
    Chisq= 10.47 on 2 degrees of freedom, p= 0.0053
Number of Newton-Raphson Iterations: 2
n= 2843
```

Sonuçlar, yaş ve cinsiyetin ölüm riski üzerindeki etkilerini göstermektedir. Cinsiyet değişkeninin, ölüm riskinde anlamlı bir etkiye sahip olduğu görülmüştür. Ayrıca, yaşın etkisinin de pozitif ve anlamlı olduğu tespit edilmiştir.