**Statistical Modelling Techniques ~ Elif Kartal**

Create a single script with the appropriate and different built-in datasets for Regression, ANOVA, ANCOVA, Generalized Linear Model with any specific link function, Linear Probability Model, Logit OR Probit Model, Truncated Regression, Censored Regression, Poisson Model, Negative-Binomial Model, Zero-inflated Model, Quantile Regression. Briefly write dataset properties (Which properties of this dataset cause you to use X analysis method?) and comment on all results (with#), separately.

**Install and load necessary libraries:**

```
install.packages("datasets")
```

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

```
Warning: package 'datasets' is a base package, and should not be updated
```

```
install.packages("car")
```

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

```
install.packages("lmtest")
```

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

```
install.packages("sandwich")
```

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

```
install.packages("survival")
```

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

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

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

```r
install.packages("pscl")
```

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

```r
install.packages("quantreg")
```

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

```r
install.packages("VGAM")
```

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

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

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

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

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

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

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

```r
library(AER)
```

```
Loading required package: car
```

```
Loading required package: carData
```

```
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
```

```r
library(datasets)
library(car)
library(lmtest)
library(sandwich)
library(survival)
library(MASS)
library(pscl)
```

```
Classes and Methods for R originally developed in the
Political Science Computational Laboratory
Department of Political Science
Stanford University (2002-2015),
by and under the direction of Simon Jackman.
hurdle and zeroinfl functions by Achim Zeileis.
```

```
library(quantreg)
```

```
Loading required package: SparseM
```

```

Attaching package: 'SparseM'
```

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

    backsolve
```

```

Attaching package: 'quantreg'
```

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

    untangle.specials
```

```
library(VGAM)
```

```
Loading required package: stats4
```

```
Loading required package: splines
```

```

Attaching package: 'VGAM'
```

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

    tobit
```

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

    lrtest
```

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

    logit
```

```r
library(truncreg)
```

```
Loading required package: maxLik
```

```
Loading required package: miscTools
```

```
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/
```

```r
library(censReg)
```

```
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/
```

**Linear Regression:**

```r
data(mtcars)
linear_model <- lm(mpg ~ ., data = mtcars)
summary(linear_model)
```

```
Call:
lm(formula = mpg ~ ., data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max
-3.4506 -1.6044 -0.1196  1.2193  4.6271

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 12.30337   18.71788   0.657   0.5181
```

```
cyl         -0.11144    1.04502  -0.107   0.9161
disp         0.01334    0.01786   0.747   0.4635
hp          -0.02148    0.02177  -0.987   0.3350
drat         0.78711    1.63537   0.481   0.6353
wt          -3.71530    1.89441  -1.961   0.0633 .
qsec         0.82104    0.73084   1.123   0.2739
vs           0.31776    2.10451   0.151   0.8814
am           2.52023    2.05665   1.225   0.2340
gear         0.65541    1.49326   0.439   0.6652
carb        -0.19942    0.82875  -0.241   0.8122
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.65 on 21 degrees of freedom
Multiple R-squared:  0.869, Adjusted R-squared:  0.8066
F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

```r
#Properties: Continuous response variable (mpg), multiple predictors
```

**ANOVA:**

```r
data(iris)
anova_model <- aov(Sepal.Length ~ Species, data = iris)
summary(anova_model)
```

```
             Df Sum Sq Mean Sq F value Pr(>F)
Species       2  63.21  31.606   119.3 <2e-16 ***
Residuals   147  38.96   0.265
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Properties: Categorical independent variable (Species) and
# continuous response variable (Sepal.Length)
```

ANCOVA:

```r
data(ToothGrowth)
ancova_model <- lm(len ~ supp + dose + supp:dose, data = ToothGrowth)
summary(ancova_model)
```

```
Call:
lm(formula = len ~ supp + dose + supp:dose, data = ToothGrowth)

Residuals:
    Min      1Q  Median      3Q     Max
-8.2264 -2.8462  0.0504  2.2893  7.9386

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   11.550      1.581   7.304 1.09e-09 ***
suppVC        -8.255      2.236  -3.691 0.000507 ***
dose           7.811      1.195   6.534 2.03e-08 ***
suppVC:dose    3.904      1.691   2.309 0.024631 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.083 on 56 degrees of freedom
Multiple R-squared:  0.7296,    Adjusted R-squared:  0.7151
F-statistic: 50.36 on 3 and 56 DF,  p-value: 6.521e-16
```

```r
# Properties: Continuous response variable, continuous and
# categorical predictors
```

Generalized Linear Model (Poisson regression):

```r
data(warpbreaks)
glm_model <- glm(breaks ~ wool + tension, family = poisson, data = warpbreaks)
summary(glm_model)
```

```
Call:
glm(formula = breaks ~ wool + tension, family = poisson, data = warpbreaks)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.69196    0.04541  81.302  < 2e-16 ***
woolB       -0.20599    0.05157  -3.994 6.49e-05 ***
tensionM    -0.32132    0.06027  -5.332 9.73e-08 ***
tensionH    -0.51849    0.06396  -8.107 5.21e-16 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 297.37  on 53  degrees of freedom
Residual deviance: 210.39  on 50  degrees of freedom
AIC: 493.06

Number of Fisher Scoring iterations: 4
```

```r
# Properties:The warpbreaks dataset is used because it has a
#count response variable (breaks), suitable for Poisson regression.
```

Linear Probability Model:

```r
lpm_model <- lm(am ~ wt + hp + qsec, data = mtcars)
summary(lpm_model)
```

```

Call:
lm(formula = am ~ wt + hp + qsec, data = mtcars)

Residuals:
     Min       1Q   Median       3Q      Max
-0.47595 -0.24600 -0.01487  0.28334  0.50096

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.476e+00  1.076e+00   3.232 0.003143 **
wt          -3.831e-01  9.617e-02  -3.984 0.000439 ***
hp          -6.007e-05  1.914e-03  -0.031 0.975187
qsec        -1.025e-01  5.612e-02  -1.826 0.078532 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3293 on 28 degrees of freedom
Multiple R-squared:  0.6065,    Adjusted R-squared:  0.5644
F-statistic: 14.39 on 3 and 28 DF,  p-value: 7.347e-06
```

```r
# Properties:The mtcars dataset has a binary response variable (am) and
# several predictors, suitable for Linear Probability Model.
```

Probit Model:

```
probit_model <- glm(am ~ wt + hp + qsec, family = binomial(link = "probit"), data = mtcars)
```

```
Warning: glm.fit: algorithm did not converge
```

```
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(probit_model)
```

```
Call:
glm(formula = am ~ wt + hp + qsec, family = binomial(link = "probit"),
    data = mtcars)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  6.808e+02  1.890e+05   0.004    0.997
wt          -6.612e+01  1.737e+04  -0.004    0.997
hp          -2.137e-01  9.226e+01  -0.002    0.998
qsec        -2.515e+01  7.283e+03  -0.003    0.997

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4.3230e+01  on 31  degrees of freedom
Residual deviance: 8.3545e-09  on 28  degrees of freedom
AIC: 8

Number of Fisher Scoring iterations: 25
```

```
# Properties:The mtcars dataset has a binary response variable (am)
# and several predictors, suitable for Probit regression.
```

Truncated Regression:,

```
data("Affairs", package="AER")
fit <- truncreg(affairs ~ age + yearsmarried + religiousness + rating, data=Affairs, point=0
summary(fit)
```

```
Call:
truncreg(formula = affairs ~ age + yearsmarried + religiousness +
    rating, data = Affairs, point = 0, direction = "left")

BFGS maximization method
75 iterations, 0h:0m:0s
g'(-H)^-1g = 0.301



Coefficients :
              Estimate Std. Error t-value  Pr(>|t|)
(Intercept)    30.6816    41.4946  0.7394   0.45966
age            -2.3899     1.5007 -1.5926   0.11126
yearsmarried   17.0571     8.1109  2.1030   0.03547 *
religiousness -57.6926    26.7378 -2.1577   0.03095 *
rating        -55.4266    25.2016 -2.1993   0.02785 *
sigma          17.7881     4.4007  4.0421 5.297e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -681.72 on 6 Df
```

```r
# Properties: The Affairs dataset contains a continuous response variable
# (affairs) that is truncated, suitable for truncated regression.
```

Censored Regression (Tobit Model):

```r
fit <- censReg(affairs ~ age + yearsmarried + religiousness + rating, data=Affairs, left=0)
summary(fit)
```

```
Call:
censReg(formula = affairs ~ age + yearsmarried + religiousness +
    rating, left = 0, data = Affairs)

Observations:
         Total  Left-censored     Uncensored Right-censored
           601            451            150              0
```

```
Coefficients:
              Estimate Std. error t value  Pr(> t)
(Intercept)    9.08289    2.65881   3.416 0.000635 ***
age           -0.16034    0.07772  -2.063 0.039095 *
yearsmarried   0.53890    0.13417   4.016 5.91e-05 ***
religiousness -1.72337    0.40471  -4.258 2.06e-05 ***
rating        -2.26735    0.40813  -5.556 2.77e-08 ***
logSigma       2.11310    0.06712  31.482  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Newton-Raphson maximisation, 7 iterations
Return code 1: gradient close to zero (gradtol)
Log-likelihood: -706.4048 on 6 Df
```

```r
# Properties: The Affairs dataset has a continuous response variable (affairs)
#that is censored, suitable for a Tobit model.
```

Poisson Model:

```r
model <- glm(breaks ~ wool + tension, data=warpbreaks, family=poisson)
summary(model)
```

```
Call:
glm(formula = breaks ~ wool + tension, family = poisson, data = warpbreaks)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.69196    0.04541  81.302  < 2e-16 ***
woolB       -0.20599    0.05157  -3.994 6.49e-05 ***
tensionM    -0.32132    0.06027  -5.332 9.73e-08 ***
tensionH    -0.51849    0.06396  -8.107 5.21e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 297.37  on 53  degrees of freedom
Residual deviance: 210.39  on 50  degrees of freedom
AIC: 493.06
```

```
Number of Fisher Scoring iterations: 4
```

```
# Properties: The warpbreaks dataset has a count response variable (breaks),
#suitable for a Poisson model.
```

Negative-Binomial Model:

```
data("quine", package="MASS")
model2 <- glm.nb(Days ~ Sex + Age + Lrn + Eth, data=quine)
summary(model2)
```

```
Call:
glm.nb(formula = Days ~ Sex + Age + Lrn + Eth, data = quine,
    init.theta = 1.274892646, link = log)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.89458    0.22842  12.672  < 2e-16 ***
SexM         0.08232    0.15992   0.515 0.606710
AgeF1       -0.44843    0.23975  -1.870 0.061425 .
AgeF2        0.08808    0.23619   0.373 0.709211
AgeF3        0.35690    0.24832   1.437 0.150651
LrnSL        0.29211    0.18647   1.566 0.117236
EthN        -0.56937    0.15333  -3.713 0.000205 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial(1.2749) family taken to be 1)

    Null deviance: 195.29  on 145  degrees of freedom
Residual deviance: 167.95  on 139  degrees of freedom
AIC: 1109.2

Number of Fisher Scoring iterations: 1

              Theta:  1.275
          Std. Err.:  0.161

 2 x log-likelihood:  -1093.151
```

```
# Properties: The quine dataset has a count response variable (Days), suitable
# for a negative-binomial model due to overdispersion.
```

Zero-inflated Model:

```
data("bioChemists", package="pscl")
model3 <- zeroinfl(art ~ fem + mar + kid5 + phd + ment, data=bioChemists, dist="poisson")
summary(model3)
```

```
Call:
zeroinfl(formula = art ~ fem + mar + kid5 + phd + ment, data = bioChemists,
    dist = "poisson")

Pearson residuals:
    Min      1Q  Median      3Q     Max
-2.3253 -0.8652 -0.2826  0.5404  7.2976

Count model coefficients (poisson with log link):
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.640838   0.121307   5.283 1.27e-07 ***
femWomen    -0.209145   0.063405  -3.299 0.000972 ***
marMarried   0.103751   0.071111   1.459 0.144565
kid5        -0.143320   0.047429  -3.022 0.002513 **
phd         -0.006166   0.031008  -0.199 0.842378
ment         0.018098   0.002294   7.888 3.07e-15 ***

Zero-inflation model coefficients (binomial with logit link):
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.577059   0.509386  -1.133  0.25728
femWomen     0.109746   0.280082   0.392  0.69518
marMarried  -0.354014   0.317611  -1.115  0.26502
kid5         0.217097   0.196482   1.105  0.26919
phd          0.001274   0.145263   0.009  0.99300
ment        -0.134114   0.045243  -2.964  0.00303 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Number of iterations in BFGS optimization: 21
Log-likelihood: -1605 on 12 Df
```

```
# Properties: The bioChemists dataset can be used with zero-inflated models
# due to the nature of count data with potential excess zeros.
```

Quantile Regression:

```
fit <- rq(medv ~ ., data=Boston, tau=0.5)
summary(fit)
```

```
Call: rq(formula = medv ~ ., tau = 0.5, data = Boston)

tau: [1] 0.5

Coefficients:
            coefficients lower bd  upper bd
(Intercept)  14.85002      6.34690  25.81307
crim         -0.14446     -0.15448  -0.02408
zn            0.03703      0.01835   0.06029
indus         0.02166     -0.05594   0.06244
chas          1.30227      0.60041   2.52568
nox          -9.18412    -14.87307  -2.76601
rm            5.32517      4.09059   6.44724
age          -0.03135     -0.04396  -0.00409
dis          -1.04478     -1.24708  -0.75657
rad           0.18003      0.07576   0.28132
tax          -0.00994     -0.01472  -0.00559
ptratio      -0.73731     -0.91502  -0.56798
black         0.01125      0.00820   0.01662
lstat        -0.29766     -0.40363  -0.21898
```

```
# Proporties: The Boston housing dataset is used because it has a continuous
# response variable and several predictors, suitable for quantile regression.
```