HW01: Exercise 3 on page 12 of the text:  The dataset prostate is from a study on 97 men with prostate cancer who were due to receive a radical prostatectomy.  Make a numerical and graphical summary of the data.

|  |  |
| --- | --- |
| This data frame contains the following columns: |  |
| lcavol | log(cancer volume) |
| lweight | log(prostate weight) |
| age | age at time of medical exam (years) |
| lbph | log(benign prostatic hyperplasia amount) |
| svi | seminal vesicle invasion |
| lcp | log(capsular penetration) |
| gleason | Gleason Score: The higher the Gleason Score, the more likely that the cancer will grow and spread quickly |
| pgg45 | Percentage Gleason scores 4 or 5.  This percentage of highgrade carcinoma (Gleason score 4 or 5 or both) is considered a prognostic factor of survival. |
| lpsa | log(prostate specific antigen) |

For numerical summary: the summary statement of ALL variables is sufficient.

For graphical summary: provide histogram plot and density plot of lpsa.  Also provide a scatter plot with x=lpsa and y=lcavol.

These four R code lines below should get you started:

library(faraway)

 data(prostate, package="faraway")

 head(prostate)

 summary(prostate)

-------------------------------------------------------------------------------------------------------------

HW02: First, generate 1000 observations from a binomial distribution with n=30 and p=0.2

Use the 1000 observations you generated:

a) Generate poisson, binomial, negative binomial Diagnostic Distribution Plots using distplot

             See PoissonnessPlotMaterial.pdf in Lecture 01 21JAN2020 folder.

b) Generate a histogram and overlay a kernel estimator of the density

-------------------------------------------------------------------------------------------------------------

HW03: First, generate 1000 observations from a Poisson distribution with lambda=2.4

Use the 1000 observations you generated to:

a) Generate poisson, binomial, negative binomial Diagnostic Distribution Plots using distplot

b) Generate a histogram and overlay a kernel estimator of the density

You can use

poisson <- rpois(n=1000,lambda=2.4)

-------------------------------------------------------------------------------------------------------------

HW04: The National Institute of Diabetes and Digestive and Kidney Diseases conducted a study on 768 adult female Pima Indians living near Phoenix. One variable that was collected is bmi (body mass index) which represents weight in kg/ (height in meters squared). The pima data set is available in the Faraway library. Compute the mean, 0.05 trimmed mean, 0.10 trimmed mean, 0.05 winsorized mean, and 0.10 winsorized mean.  
Hint: #need psych library. The variable of interest is pima$bmi.

library(psych)

-------------------------------------------------------------------------------------------------------------

HW05: Here are the first six observations from the prostate data set found in the faraway library. Use help(prostate) to describe the dataset and the variables in the data sets.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| obs | lcavol | lweight | age | lbph | svi | lcp | gleason | pgg45 | lpsa |
| 1 | -0.579819 | 2.7695 | 50 | -1.38629 | 0 | -1.38629 | 6 | 0 | -0.43078 |
| 2 | -0.994252 | 3.3196 | 58 | -1.38629 | 0 | -1.38629 | 6 | 0 | -0.16252 |
| 3 | -0.510826 | 2.6912 | 74 | -1.38629 | 0 | -1.38629 | 7 | 20 | -0.16252 |
| 4 | -1.203973 | 3.2828 | 58 | -1.38629 | 0 | -1.38629 | 6 | 0 | -0.16252 |
| 5 | 0.7514161 | 3.4324 | 62 | -1.38629 | 0 | -1.38629 | 6 | 0 | 0.37156 |
| 6 | -1.049822 | 3.2288 | 50 | -1.38629 | 0 | -1.38629 | 6 | 0 | 0.76547 |

Perform a simple linear regression with lpsa as the response and lcavol as the predictor. Show the ANOVA table and provide a histogram of the residuals.

Hint: If your linear model name is “lmod” then

> residuals(lmod) #prints out the residuals

-------------------------------------------------------------------------------------------------------------

HW06

Exercise #1 on page 30 of the text.

Do only parts a, b, c, d, e

Hint: If your linear model name is “lmod”

r <- cor(predict(lmod),lpsa) #correlation yhat with yi

-------------------------------------------------------------------------------------------------------------

HW07

Exercise #4 on page 30. See HW05.

You only need to show the summary output for each model, no need to do any plotting as requested in the exercise.

-------------------------------------------------------------------------------------------------------------

 HW08: The three commands below will provide a listing of the stackloss data.

 library(faraway)  
 data(stackloss)  
 stackloss

This command will provide an explanation of the data.

help(stackloss)

Perform an OLS regression of stackloss on the 3 predictors in the data set and obtain 95% CIs on the regression coefficients.

Perform a  LAD regression of stackloss on the 3 predictors in the data set and obtain 95% bootstrapped CIs on the regression coefficients.

HINT:  Follow the RobustGala.R program discussed in class.

-------------------------------------------------------------------------------------------------------------

HW09: Permutation test applied to a single coefficient in Linear Regression

Use the "happy" data set in the faraway library consisting of data on 39 MBA students who were asked about happiness in their work.  
Response variable: happy  
Predictors: money sex love work

First 6 observations from happy:

happy money sex love work  
1 10 36 0 3 4  
2 8 47 1 3 1  
3 8 53 0 3 5  
4 8 35 1 3 3  
5 4 88 1 1 2  
6 9 175 1 3 4

happy is a data frame with 39 observations on the following 5 variables.

**happy**

Happiness on a 10 point scale where 10 is most happy

**money**

family income in thousands of dollars

**sex**

1 = satisfactory sexual activity, 0 = not

**love**

1 = lonely, 2 = secure relationships, 3 = deep feeling of belonging and caring

**work**

5 point scale where 1 = no job, 3 = OK job, 5 = great job

Execute a permutation test as an alternative to the t-test H0: B4=0 versus H1: B4 not equal to 0 for the coefficient B4 of the variable work in the model.

To begin to answer the question use

library(faraway)  
data(happy)  
head(happy)  
lmod <- lm(happy ~ money + sex + love + work, data=happy)  
summary(lmod)  
summary(lmod)$coefficients[5,3] #extracts the t-statistic you need for work

Use nreps <- 5000 and the program permutation3.R as a guide.