Warm-up exercise

altitude	tree_presence	tree_nr
2	0	0
3	0	0
5	0	0
6	0	0
7	0	0
6	0	0
7	0	0
5	0	0
4	0	0
12	1	3
15	1	5
16	1	6
20	1	5
22	1	8

Explanatory variable:

altitude = altitude (m asl) of the 100-m2 study plot

tree_presence = occurrence of trees in the study plot, 1 = yes, 0 = no

tree_nr = number of trees in the study plot

1) Insert data in R, use c(….) –command and check that the data are valid
2) What is the number of trees at the altitude of 1 m and 10 m based on a linear model? Follow the lecture notes from week 2.
3) What is the probability of tree presence at the altitude of 1 m and 10 m based on a linear model? Follow the lecture notes from week 2.
4) What is the interpretation of the models?
5) Are these results realistic?

Insert and check the data (task 1)

```
altitude <- c(2,3,5,6,7,6,7,5,4,12,15,16,20,22)
tree_presence <- c(0,0,0,0,0,0,0,0,0,1,1,1,1,1)
tree_nr <- c(0,0,0,0,0,0,0,0,0,3,5,6,5,8)
data_tree <- data.frame(altitude,tree_presence, tree_nr)
str(data_tree)
summary(data_tree)
plot(data_tree)
```

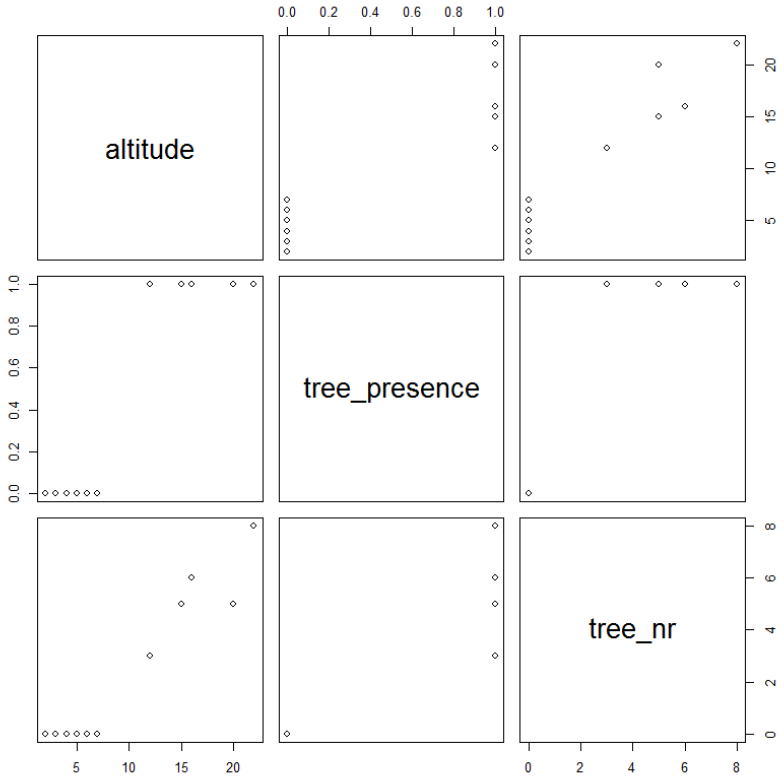
str and summary (task)

```
> str(data_tree)
'data.frame':   14 obs. of  3 variables:
 $ altitude     : num  2 3 5 6 7 6 7 5 4 12 ...
 $ tree_presence: num  0 0 0 0 0 0 0 0 0 1 ...
 $ tree_nr      : num  0 0 0 0 0 0 0 0 0 3 ...
```

```
> summary(data_tree)
    altitude      tree_presence       tree_nr
 Min.   : 2.000   Min.   :0.0000   Min.   :0.000
 1st Qu.: 5.000   1st Qu.:0.0000   1st Qu.:0.000
 Median : 6.500   Median :0.0000   Median :0.000
 Mean   : 9.286   Mean   :0.3571   Mean   :1.929
 3rd Qu.:14.250   3rd Qu.:1.0000   3rd Qu.:4.500
 Max.   :22.000   Max.   :1.0000   Max.   :8.000
```

multipanel plot

> plot(data_tree)



Linear model and prediction (task 2)

```
> tree_nr_lm <- lm(tree_nr ~ altitude)
> summary(tree_nr_lm)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.98703    0.41298  -4.811 0.000425 ***
altitude     0.42168    0.03685  11.442 8.21e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.865 on 12 degrees of freedom
Multiple R-squared:  0.916,     Adjusted R-squared:  0.909
F-statistic: 130.9 on 1 and 12 DF,  p-value: 8.206e-08
```

```
Altitude of 1 m
> -1.98703 + 0.42168*1
[1] -1.56535
Altitude of 10 m
> -1.98703 + 0.42168*10
[1] 2.22977
```

OR

```
> altitudev <- data.frame(altitude= c(1,10))
> prediction_tree_nr <- predict(tree_nr_lm,altitudev)
> prediction_tree_nr
        1         2
-1.565353  2.229772
```

Linear model and prediction (task 3)

```
> tree_nr_lm2 <- lm(tree_presence ~ altitude)
> summary(tree_nr_lm2)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.293050   0.098764  -2.967   0.0118 *
altitude     0.070021   0.008814   7.945 4.04e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2069 on 12 degrees of freedom
Multiple R-squared:  0.8402,    Adjusted R-squared:
0.8269
F-statistic: 63.12 on 1 and 12 DF,  p-value: 4.037e-06
```

Altitude of 1 m

```
> -0.293050 + 0.070021 *1
[1] -0.223029
```

Altitude of 10 m

```
> -0.293050 + 0.070021 *10
[1] 0.40716
```

OR

```
> altitudev <- data.frame(altitude= c(1,10))
> prediction_tree_nr <- predict(tree_nr_lm2,altitudev)
> prediction_tree_nr
         1          2
-0.2230290  0.4071577
```

Tasks 4 and 5

Task 4: what is the interpretation of the models?

- Negative values are difficult to explain, what is the logic, idea and realism of -1.56535 trees per study plot at the altitude of 1 m? (task2)

- Negative probability value is problematic to understand because it is infeasible, what is the probability of -0.223029 trees per plot at the altitude of 1 m? (task3)

Task 5: are these results realistic?

- They are not realistic, count data and binomial data can not have negative values
- The results are based on wrong statistical assumptions and misuse of statistics

Statistics are supposed to make something easier to understand but when used in a misleading fashion can trick the casual observer into believing something other than what the data shows.

Task 3 based on GLM

```
> tree_nr_glm <- glm(tree_nr ~ altitude, family="poisson")
> altitudev <- data.frame(altitude= c(1,10))
> altitudev
  altitude
1        1
2       10
> prediction_tree_nr <- predict(tree_nr_glm,altitudev, type="response")
> prediction_tree_nr
        1         2
0.1365785 0.8739976
```

Task 3 based on GLM

```
> tree_nr_glm2 <- glm(tree_presence ~ altitude, family="binomial")
> altitudev <- data.frame(altitude= c(1,10))
> altitudev
  altitude
1        1
2       10
> prediction_tree_pre <- predict(tree_nr_glm2,altitudev, type="response")
> prediction_tree_pre
           1           2
2.220446e-16 9.879527e-01
```