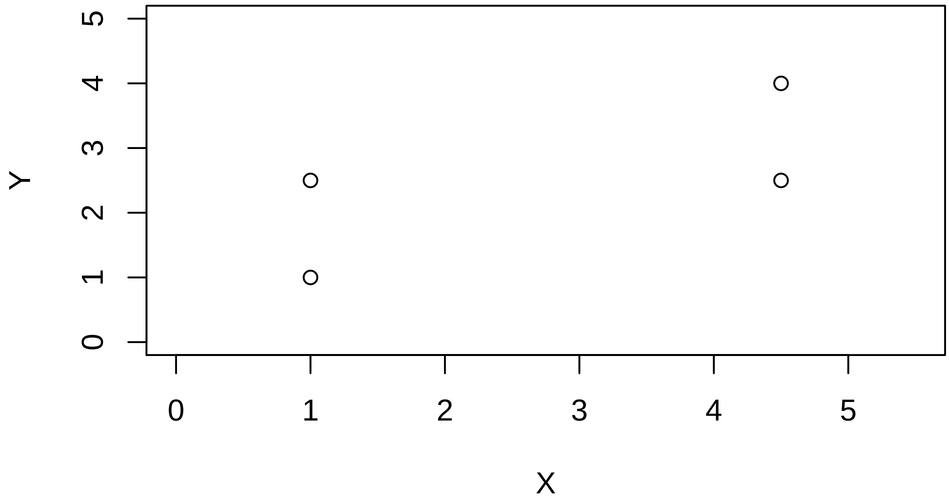
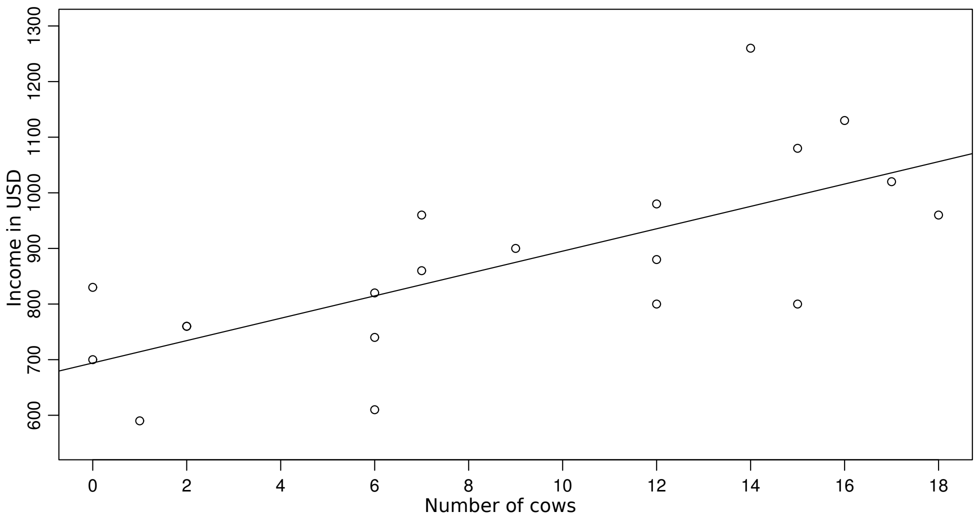
Series 1

1. **a)** In the plot below, draw the regression line for Y being the dependent and X being the independent variable and vice versa.



b) In the plot below, we depicted for several farms the yearly income in Dollars versus the number of cows. What are the values for the intercept and the slope?



c) What is your estimate for the average deviation of the points with respect to the regression line?

d) Estimate the income of a farm with 15 cows and of a farm with 100 cows? Are these estimates meaningful?

2. The following table contains some functions which can be linearized by a suitable transformation. Complete the table by inserting the needed transformations of x and y, and the resulting linear forms.

Function	Transformation	Linear form
$y = \alpha x^{\beta}$	$y' = \log(y),\ x' = \log(x)$	$y' = \log(\alpha) + \beta \cdot x'$
$y = \alpha e^{\beta \cdot x}$	....	....
$y = \alpha + \beta \cdot \log(x)$	....	....
$y = 1/(\alpha + \beta e^{-x})$	....	....

3. The behaviour of the least squares estimator can be investigated by a small simulation study. Here are the R-commands for linear regression:

```
> ## simple linear regression
> set.seed(21)                              # initializes the random number generator
> x   <- seq(1, 40, 1)                      # generates equidistant x-values
> y   <- 1 + 2 * x + 5 * rnorm(length(x)) # y-values = linear function(x) + error
> reg <- lm(y ~ x)                          # fit of the linear regression
> summary(reg)                              # output of selected results
> plot(x, y)                                # scatter plot
> abline(reg)                               # draws regression line
```

a) Write a sequence of R-commands which randomly generates 100 times a vector of y-values according to the above model with the given x-values and generates a vector of estimated slopes $\hat{\beta}_2$ of the regression lines. Plot one of the simulated responses against the x-values and look as well at the Tukey-Anscombe plot for a given data set.
Hint:

- Look at the help file of the function `for`, i.e. `?for`.
- Use the following R syntax to create a Tukey-Anscombe plot.
  `plot(reg, which = 1)`

b) Compute the mean and empirical standard deviation of the estimated slopes.

c) Compute the theoretical variance of $\hat{\beta}_2$.
Hint: To compute the inverse of a matrix use `solve()`.

d) Draw a histogram of the 100 estimated slopes and add the normal density of the theoretical distribution of $\hat{\beta}_2$ to the histogram. What do you observe? Does it fit well?
Hints: The histogram must be drawn with parameter `freq = FALSE`, so that the y-axis is suitably scaled for drawing the density. The density can be added by
`lines(seq(1.8, 2.3, by = 0.01), dnorm(seq(1.8, 2.3, by = 0.01), mean= ?, sd = ?))`,
where you have to find the correct values for the arguments `mean` and `sd`.

4. Repeat the simulation from exercise 3 with different error distributions that violate some of the assumptions. Repeat part a), b), and d) and add the same normal density from exercise 3. Answer the following questions for all the tasks. Which (if any) assumptions are violated? What properties of the distribution of $\hat{\beta}_2$ are affected by this? Which part of the R output do you trust?

a) Replace the second line of the R code in the previous exercise by

```
y   <- 1 + 2 * x + 5 * (1 - rchisq(length(x), df = 1)) / sqrt(2)
```

Hints: To get an idea of the error distribution, you may look at the following histogram and values

```
errors <- 5 * (1 - rchisq(40, df = 1)) / sqrt(2)
hist(errors)
mean(errors)
var(errors)
```

b) Replace the second line of the R code in the previous exercise by

```
y   <- 1 + 2 * x + 5 * rnorm(length(x), mean = x^2 / 40 - 1, sd = 1)
```

c) Replace the second line of the R code in the previous exercise by

```
require(MASS)
Sigma <- toeplitz(c(seq(from = 1, to = 0, by = -0.1), rep(0, 29)))
y   <- 1 + 2 * x + 5 * mvrnorm(n = 1, mu = rep(0, length(x)), Sigma = Sigma)
```

d) Replace the second line of the R code in the previous exercise by

```
y   <- 1 + 2 * x + 5 * rnorm(length(x), mean = 0, sd = x / 20)
```

Preliminary discussion: Friday, March 02.

Deadline: Friday, March 09.