TMA4255 Applied Statistics
Exercise 1

Problem 1

The deposit of Cobalt (Co) in Oppland county is being investigated. 100 soil samples were chosen from a larger area, and the content of Co (measured in mg Co per kg soil) in every sample was determined.

The results are given in Table 1.

17	12	17	40	24	14	13	16	21	43
16	28	20	134	184	23	34	22	31	23
17	258	35	45	49	243	300	83	76	13
42	72	30	52	32	18	56	135	66	16
16	6	87	90	21	126	40	21	113	15
51	47	38	14	99	48	79	30	39	87
66	162	27	32	141	23	67	16	78	22
26	15	63	42	45	16	55	19	16	209
7	14	14	9	21	14	39	10	75	50
15	31	37	58	21	17	30	81	10	46

Table 1: Content of Co

a) Reading data and calculating summery statistics.

MINITAB   Import the dataset into MINITAB. This is done either by opening the data file on the webpage (called data1.MTW), then MINITAB will start automatically, or by saving the file, then start MINITAB and open the file by selecting Open Worksheet on the File menu. (If you have imported the data set correctly, the mean of the data should be 51.76)

Describe the dataset by

Stat $\rightarrow$ Basic Statistics $\rightarrow$ Display Descriptive Statistics

R   Read the data from the www-page (data1.csv) into R by `read.csv`. This will make the data into a data frame. Since the data set only contain one column, it would be wise to construct a vector instead of a data frame, e.g. by just selecting column 1, `cobolt=read.csv(``data1.csv'')[,1]` Describe the data using `summary(cobolt)`.

Most of the statistics displayed should be well-known. Write down the expressions and try to figure out what the ones you do not know mean.

b) We want to investigate if it is reasonable to assume that the data are independent and normally distributed.

To visually check if there is a trend in the data, select

MINITAB   Graph → Time Series Plot.

R `plot`

To make a normal plot, select

MINITAB   Graph → Probability Plot

R   `qqnorm` and `qqline`.

c) Do the following transformations of the data:

$$y_i = \sqrt{x_i} \tag{1}$$
$$y_i = \ln x_i \tag{2}$$
$$y_i = 1/x_i \tag{3}$$

To do this, select

MINITAB   Calc → Calculator

Make a histogram and a normal plot of the transformed data. Which of these transformations should be used to transform the data into normally distributed random variables?

Problem 2

a) MINITAB can be used to simulate data from various distributions by the menu choices

Calc → Random Data.

Use this to simulate two datasets consisting of 50 observations each from the $N(2, 4^2)$ in two columns (C1 and C2). Describe the dataset by descriptive statistics and plots.

To simulate normal data in R use `rnorm`.

b) MINITAB can also be used to look up the "tables values" for many distributions by the menu choices Calc → Probability Distributions.

R use `dbinom` and `pbinom`.

It is possible to find the probability density function, the cumulative distribution function and the inverse cumulative distribution function.

Assume that X is binomially distributed, $X \sim bin(100, 0.3)$.

i) Find $P(X = 32)$
ii) Find $P(28 \leq X \leq 32)$
iii) X is approximately normally distributed. Use this to find the probabilities in i) and ii).

Problem 3

Let $X_1, ..., X_{10}$ be independent and $N(200, 5^2)$.

We define

$V = \max(X_j), \ j = 1, ..., 10.$

We cannot find the exact distribution of V, but we can estimate the desired statistics by simulation.

Simulate 10 observations from $N(200, 5^2)$ and find the maximum value for these 10 observations. Do this 500 times, i.e. 500 rows with 10 observations in each row.

a) Find an estimate of $E(V)$.

b) Find an estimate of $P(V < 205)$.