TMA 4255 Applied Statistics, Exercise 1

Problem 1

The deposit of Cobolt (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.

Table 1: Content of Co

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

Importing the data set:

The data are in a Excel file and can be loaded and stored on your own PC by clicking on data1.csv on the wiki-page (Exercises).
You may also load them directly into MINITAB by clicking on File, than click on Open and use: https://www.math.ntnu.no/emner/TMA4255/2023v/files/data1.csv as File name. Press OK, and they will be in a MINITAB work sheet.

The data can be read into R by the command:
cobolt=read.csv("https://www.math.ntnu.no/emner/TMA4255/2023v/files/data1.csv")[,1]

a) Import the data set and calculate summary statistics.

MINITAB: Stat → Basic Statistics → Display Descriptive Statistics

R: summary(cobolt)

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

MINITAB: Graph → Time Series Plot

Graph → Probability Plot

R: coboltts=ts(cobolt)

plot(coboltts)

qqnorm(cobolt) and qqline(cobolt)

c) Do the following transformations of the data:

$$y_i = \sqrt{x_i}$$
$$y_i = \ln x_i$$
$$y_i = 1/x_i$$

MINITAB: Calc $\rightarrow$ Calculator
R: sqrt() , log() , 1/x

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

MINITAB: Graph $\rightarrow$ Histogram

R: hist()

Problem 2

a) Simulate two datasets consisting of 50 observations each from a $N(2, 4^2)$ distribution. Describe the datasets by descriptive statistics and plots.

MINITAB: Calc $\rightarrow$ Random Data. Put the data in C1 and C2.

R: rnorm(50,2,4)

b) MINITAB and R can also be used to find values for the probability density function, the cumulative distribution function and the inverse distribution function.
MINITAB: Calc $\rightarrow$ Probability Distributions.
R: dbinom(x,n,p), pbinom(x,n,p), pnorm(x, μ, σ) μ, σ

Assume that X is binomially distributed with n=100 and p=0.3.

i) Find $P(X = 32)$

ii) Find $P(28 \le X \le 32)$

iii) X is approximately normally distributed. Use this to find the probabilities in i) and ii).

Problem 3

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

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

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

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

a) Find an estimate of E(V)
b) Find an estimate of P(V < 205)