# Univariate Data and Modelling

## Exercise Session 2 : Distributions and confidence intervals

## Exercise 1

Observe the throwing of a fair dice

a) How would the most obvious random variable be defined?

b) Is it a discrete or continuous random variable?

c) Draw the probability density function of this random variable. (hint: create a vector of probabilities and use the function **barplot**)

d) Draw the cumulative distribution function of this random variable. (hint: create a vector of probabilities and use the function **plot**)

## Exercise 2

Observe 10 tosses of a fair coin.

a) How would an obvious random variable be defined?

b) Is it discrete or continuous?

c) How is this random variable distributed?

d) What is the probability of tossing 6 or fewer heads in 10 tosses of a fair coin?(hint: use the function **pbinom**)

e) Calculate the probability of 7 or more heads in 10 tosses of a fair coin?(hint: use the function **pbinom**)

f) What is the median number of heads in 10 tosses of a fair coin?(hint: use the function **qbinom**)

g) What is the 3th quartile?(hint: use the function **qbinom**)

## Exercise 3

A junior software developer writes on average 2 bugs every ten minutes while programming in PYTHON

a) How would the most obvious random variable be defined?

b) How is this random variable distributed?

c) What is the probability that the developer writes no bugs in ten minutes?(hint: use the function **ppois**)

d) What is the probability that the developer at least one bug in ten minutes?(hint: use the function **ppois**)

e) What is the probability that the developer writes no bugs after 50 minutes of coding?(hint: use the function **ppois**)

f) Plot the density distribution of the number of bugs per ten minutes coding with a mean number of 4 bugs per ten minutes of coding.(hint: use the functions **density** and **barplot**)

## Exercise 4

Take as random variable the bodyweight of adult men with a mean of 85 kg and a variance of 500 $kg^2$

a) Is this a discrete or continuous random variable?

b) How is this random variable most likely distributed?

c) What is the probability that the bodyweight is exactly 100 kg?(hint: use the function **pnorm**)

d) What is the probability that the bodyweight is less than 100 kg?(hint: use the function **pnorm**)

e) What is the probability that the bodyweight is less than 80 kg?(hint: use the function **pnorm**)

f) What is the probability that the bodyweight exceeds 60 kg?(hint: use the function **pnorm**)

g) Draw the probability from (f) (hint: use the functions **dnorm** and **plot**)

h) What s the 97.5% quantile?(hint: use the function **quantile**)

## Exercise 5

Given a t-distribution T with 15 degrees of freedom

a) What is the probability of T bigger than 1?

b) Determine t when $\mathbb{P}(T > t) = 0.05$. (hint: use the function **qt**)

c) Draw the probability from (b). (hint: use the functions **dt**, **plot** and **polygon**)

## Exercise 6

Given a Chi-square distribution with 10 degrees of freedom

a) Determine c when $\mathbb{P}(X > c) = 0.05$.(hint: use the function **qchisq**)

## Exercise 7

Given an $F$ distribution with 4 numerator and 9 denominator degrees of freedom

a) Determine $\mathbb{P}(5 < F < 10)$. (hint: use the function **pf**)

b) Determine $f^0$ if $\mathbb{P}(F \leq f^0) = (1 - \alpha)$ and $\alpha = 0.05$. (hint: use the function **qf**)

## Exercise 8

The standard deviation of the thermal conductivity at 38°C and 550W is assumed to be 0.3. Ten measurements were taken with a mean of 41.924. Let us assume that thermal conductivity is normally distributed

a) Construct, using the method shown in class, a 95% confidence interval around the mean conductivity.(hint: apply the formula for the confidence interval of a normal distribution, using the function **qnorm**)

b) Construct, using the **zsum.test** function from the BSDA package, a 95% confidence interval around the mean conductivity. (hint: type ?zsum.test in the R console for help)

## Exercise 9

20 measurements of the testosterone level of healthy men resulted in a mean value of 750 ng/dl. Assume that testosterone level follows a normal distribution with a sample standard deviation of 30 ng/dl.

a) Find the 95% confidence interval on the mean testosterone level.(hint: use the function **tsum.test**)

## Exercise 10

Import the dataset BLOOD as blood.df

a) Construct, using the **z.test** function from the BSDA package, a 90% confidence interval around the mean of the variable "age" assuming that the true standard deviation is 5 years.

b) Construct a 95% confidence interval around the mean of the variable "prolactn".(hint: use the function **t.test**)

c) What is the proportion of persons with an age between 50 and 60?(hint: use the function **subset** to subset the observations with age between 50 and 60, then divide by the total number of observations).

d) Construct a 95% confidence interval for this proportion.(hint: use the function **prop.test**)

e) Make a subset "subset.df" from blood.df of those with an age lower than 50.(hint: use the function subset)

f) Construct in this subset a 99% confidence interval around the mean of the variable "testost".(hint: use the functions **shapiro.test** and **t.test**)

## Remark

When a function is mentioned in the hints, it is useful to read on the input arguments and output values of the function, by using the keyword "?**function**". For example, executing **?read.table** will give you information on the **read.table** function.

To install the packages "BSDA" and "PropCIs" use **install.packages("BSDA", "PropCIs")**. To load these libraries, use library(BSDA), library(PropCIs).