Univariate Data and Modelling – Exercises

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 discrete or continuous distributed?
c) Draw (by hand) the Probability Density Function of this random variable
d) Draw (by hand) the Cumulative Distribution Function of this random variable

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?
e) Calculate the probability of 7 or more heads in 10 tosses of a fair coin?
f) What is the median number of heads in 10 tosses of a fair coin?
g) What is the 3th quantile?

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 she writes no bugs in ten minutes?
d) What is the probability that she writes at least one bug in ten minutes?
e) What is the probability that she writes no bugs after 50 minutes of coding?
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.

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?
d) What is the probability that the bodyweight is less than 100 kg?
e) What is the probability that the bodyweight is less than 80 kg?
f) What is the probability that the bodyweight exceeds 60 kg?
g) Draw the probability from f) by hand
h) What is the 97.5% percentile?

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 $P(T > t) = 0.05$.
c) Draw the probability from b) by hand

Exercise 6

Given a Chi-square distribution with 10 degrees of freedom

a) Determine c when $P(X > c)=0.05$

Exercise 7

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

a) Determine $P(5 < F < 10)$
b) Determine f^0 if $P(F \leq f^0) = (1-\alpha)$ and $\alpha = 0.05$

INSTALL PACKAGES "BSDA" and ""PropCIs" IN R STUDIO BEFORE GOING FURTHER

```
install.packages("BSDA", "PropCIs")
library(BSDA)
library(PropCIs)
```

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
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.

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".
c) What is the proportion of persons with an age between 50 and 60?
d) Construct a 95% confidence interval for this proportion.
e) Make a subset "subset.df" from blood.df of those with an age lower than 50.
f) Construct in this subset a 99 % confidence interval around the mean of the variable "testost".