

EXAM ASSIGNMENT

Study Programme and level	MSc Logistics and Supply Chain Management + elective					
Term	V21-22r					
Course name and exam code(s)	Tools for Analytics				460202E016	
Exam form and duration	WOAI				6 hours	
Date and time	19 February 2022				9:00-15:00	
Supplementary material/aids	All	<input checked="" type="checkbox"/>	Specified	<input type="checkbox"/>	No	<input type="checkbox"/>
Other relevant information	<p>Avoid being suspected of exam cheating</p> <p>Remember to state references and use quotation marks, if you copy text from other sources or re-use parts of a previously submitted exam paper (plagiarism and self-plagiarism). Students must answer the exam assignment individually.</p> <p>All submitted exam papers are checked for plagiarism, so cheating and collaboration between students will be detected.</p>					
Hand-in of hand-written material allowed	Yes	<input type="checkbox"/>	No	<input checked="" type="checkbox"/>		
Anonymous exam?	Yes	<input checked="" type="checkbox"/>	No	<input type="checkbox"/>	Please do not write your name or student ID number anywhere. Use your flow-id number (find this on the cover sheet in WISEflow).	
Number of pages (incl. front page)	9					

Practical information

- This exam is open book, open internet, closed other people. You may use any on-line or book-based resource you would like, but you must include citations for any code that you use (directly or indirectly). You may not consult with anyone else during this exam. You cannot ask direct questions on the internet, or consult with each other, not even hypothetical questions.
- This assignment has an appendix available for download from WISEflow with files you may need during the exam.
- The exam has a VBA and R part with approx. equal weight.
- Please note that the weights on each assignment are only guideline weights, and that they only provide information regarding the relative weight of the assignments. The final evaluation will be given based on the total material handed in.
- If you find that some information is missing in the assignments, you may make the necessary assumptions and clearly specify these.
- Handing in: You must hand in a set of files (.Rmd, .xlsm and .html) as "Appendix material". Due to the system, you must also upload an empty pdf-document named `yourFlowId.pdf`.
- Your VBA code will be tested using Excel 2016 and your R code will be tested using R version 4.1.1 (2021-08-10).
- About Assignments 1-3: R
 - Your R code must be written up in an R Markdown (Rmd) file named `yourFlowId.Rmd`. Moreover, also hand in the rendered/knitted html file `yourFlowId.html`.
 - Your file must include your code and a (brief) comment for each question. For example, "The three companies with smallest profit are ..." or "The plot shows that ...".
 - An R markdown template file is given in the appendix that you may use as a starting point.
 - You may load and use the following packages:

```
library(tidyverse)
library(fitdistrplus)
library(skimr)
library(conflicted)
```
- About Assignments 4-6: VBA
 - Your VBA code should be contained in a single Excel file named `yourFlowId.xlsm`.
 - *Do not* protect your code with password or turn it into an Add-In.
 - Be very careful to save your file as an .xlsm file (and not a .xlsx file).
 - An Excel template file is given in the appendix that you may use as a starting point. This file also contains the data used in Assignments 4-6.

R - Assignment 1 (15%)

Consider the dataset in the file `income.csv` with the monthly income from a population. You can read the dataset file using

```
library(tidyverse)
dat <- read_csv("income.csv")
head(dat)
```

```
## # A tibble: 6 x 2
##   sex    income
##   <chr>  <dbl>
## 1 male  50336.
## 2 male  39712.
## 3 male  45124.
## 4 male  38807.
## 5 male  44001.
## 6 male  32347.
```

Question 1

Fit probability distributions to data for both men and women.

Question 2

Make a plot using `ggplot` of the two empirical densities with the following features:

- Different fill colours are used for each sex.
- Informative figure title and axis titles are given.

Based on the plot and Question 1, is there any difference in the income for men and women? If yes on average how much?

R - Assignment 2 (18%)

Consider COVID-19 data obtained from Our World in Data in the file `covid.csv`. The dataset contains data from different countries. Some of the columns/variables are:

- `cases`: New confirmed cases of COVID-19.
- `deaths`: New deaths attributed to COVID-19.
- `icu_patients`: Number of COVID-19 patients in intensive care units (ICUs) on a given day.
- `hosp_patients`: Number of COVID-19 patients in hospital on a given day.
- `tests`: Total tests for COVID-19.
- `positive_rate`: The share of COVID-19 tests that are positive, given as a rolling 7-day average.
- `vac`: Total number of people who received at least one vaccine dose.
- `fully_vac`: Total number of people who received all doses prescribed by the vaccination protocol.
- `population`: Country population.

Other columns are `date`, `country`, `month` and `year`. You can read the dataset file using

```
library(tidyverse)
dat <- read_csv("covid.csv")
```

Use the *dplyr* package in *tidyverse* to calculate relevant summary tables (tibbles/data frames) and answer the following questions.

Question 1

Which countries are considered and what is the timespan of the data?

Question 2

What is the number of new confirmed cases November 22nd, 2021 in Denmark?

Question 3

Calculate the total number of confirmed cases and deaths. Hint: you may use the `cumsum` function to add all cases up until a given date. You may here consider NA values in the `cases` and `deaths` columns as equal to zero (e.g. using `replace_na(cases, 0)`).

What is the total number of deaths in Norway up to October 10th, 2021?

Question 4

For each country calculate the number of tests done in each month in a given year. Which country had the highest number of tests per capita in March 2021?

Question 5

Consider United Kingdom. Which month had the highest number of ICU patients on a given day?

R - Assignment 3 (17%)

Answer this assignment using the *ggplot2* package in *tidyverse* (you may need *dplyr* for preparing the datasets you want to plot). We work with the dataset from Assignment 2 which can be read using:

```
library(tidyverse)
dat <- read_csv("covid.csv")
```

Question 1

Create a visualization showing the number of cases for each date in Denmark with the following features:

- A blue line is used to visualize the data.
- Informative figure title and axis titles are given.

Is the number of cases low or high in July 2020 in the plot?

Question 2

Create a visualization showing the total number of deaths per 100000 capita as a function of date with the following features:

- Different colours are used for each country.
- Lines are used to visualize the data.
- Legends are put at the bottom of the plot.
- Informative figure title and axis titles are given.

Hint: you may use the `cumsum` function to add all deaths up until a given date. You may here consider NA values in the `deaths` column as equal to zero (e.g. using `replace_na(deaths, 0)`).

Which country has the highest relative number of deaths in general?

Question 3

Create a visualization showing the percentage of persons vaccinated as a function of date with the following features:

- We consider 2021.
- Different colours are used to differ between vaccinated and fully vaccinated.
- The plot is divided using `country` (facet).
- Lines are used to visualize the data.
- Informative figure title and axis titles are given.

Hint: If you calculated the two percentages in two new columns `partly` and `full`, then the values can be joined to one column using

```
dat %>%
  pivot_longer(cols = c(partly, full))
```

Which country has the highest percentage of vaccinated people and the lowest gap between partly and fully vaccinated?

Question 4

Consider Germany. Create a visualization showing the variation in ICU patients with the following features:

- A sub-plot is given for each month and year (facet).
- Informative figure title and axis titles are given.

In which year did the ICU have the most patients when considering October?

Question 5

Create a visualization showing the total number of deaths per 100000 capita for each country and year with the following features:

- The numbers are shown using columns for each country
- Different fill colours are used for year. Hint: columns for each year can be shown beside each other using `position = position_dodge()`.
- Informative figure title and axis titles are given.

Which country had the lowest number of deaths in 2021?

VBA - Assignment 4 (15%)

Save all your VBA code in the module named *modA4*.

Temperatures in Fahrenheit can be converted to Celsius using

$$C = \frac{5(F - 32)}{9}$$

Similar temperatures in Celsius can be converted to Fahrenheit using

$$F = \frac{9C}{5} + 32$$

Question 1

Make functions:

- `CelsiusToFahrenheit` that takes a double `dblVal` in Celsius as input argument and returns the number converted to Fahrenheit.
- `FahrenheitToCelsius` that takes a double `dblVal` in Fahrenheit as input argument and returns the number converted to Celsius.

Question 2

Make a sub `ConvertTemp` with the following features:

- Input arguments are an array `ary` of doubles with numbers and a string `unit`.
- If `unit` equals "c" then `ary` is returned (ByRef) with the numbers converted to Fahrenheit.
- If `unit` equals "f" then `ary` is returned (ByRef) with the numbers converted to Celsius.
- If `unit` does not equals "f" or "c" then a message box is given with an error.

Question 3

Create a procedure with the following features:

- Reads the numbers in worksheet *DataA4* into an array.
- Copy the numbers to worksheet *A4*.
- Convert the numbers using `ConvertTemp` to Celsius.
- Add the converted numbers besides the original numbers in worksheet *DataA4*.
- Add a button to worksheet *A4* that run the procedure.
- The procedure should work for other datasets with a different size. You may assume that the data starts in cell A1.

VBA - Assignment 5 (15%)

Save all your VBA code in the module named *modA5*.

Consider worksheet *DataA5*, which contains a set of numbers.

Question 1

Create a procedure with the following features:

- Copy the numbers to worksheet *A5*.
- Scan all the numbers and remove (clear the cell) all the negative numbers.
- Highlight all the numbers above 20.
- Add a button to worksheet *A5* that run the procedure.
- The procedure should work for other datasets with a different size. You may assume that the data starts in cell *A1*.

Question 2

Create a procedure with the following features:

- Scan the numbers and find
 - the sum of all non-negative numbers,
 - the mean of all negative numbers.
- Use a message box to display the sum and mean calculated.
- Add a button to worksheet *A5* that run the procedure.
- The procedure should work for other datasets with a different size. You may assume that the data starts in cell *A1*.

VBA - Assignment 6 (20%)

Save all your VBA code in the module named *modA6*.

Consider worksheet *DataA6* which contains a table with three columns. The table contains data about payments for clients at a set of dates.

Create a procedure that searches the table with the following features:

- Read the *payment boolean* in cell *B1* (*TRUE* or *FALSE*) and the *grouping string* in cell *B2* (*None* or *Year*) on worksheet *A6*.
- Only consider payments where cells in the *Payment* column equals the payment boolean.
- If the grouping string equals *None* then count the number of payments for each client. Next, output the results on worksheet *A6*. See worksheet *A6Ex1* for an example.
- If the grouping string equals *Year* then count the number of payments for each client and year. Next, output the results on worksheet *A6*. See worksheet *A6Ex2* for an example.
- Add a button to worksheet *A6* that run the procedure.

The procedure should work for other datasets with different number of rows too.