Data Analysis
Week 9: Group Projects

For the Data Analysis Group Project you will be placed into groups and be tasked with writing a concise report summarising the results of a statistical analysis that you have conducted. The group report is worth 20% of your final grade for the course. Group reports should be written using R Markdown such that:

- all R output, including figures and tables, are appropriately labelled and presented;
- R code should not be included in the body of the report; and
- the report should be no longer than 6 pages.

Your group report should include:

- An appropriate **Title**;
- An **Introduction** section detailing the data set and question of interest;
- An **Exploratory Analysis** of the data;
- A **Formal Analysis** of the data; and
- Finish with your **Conclusions**.

The report is worth a total of **15 MARKS**, which are broken down as follows:

- **Title** and **Introduction**: **1 MARK**
- **Exploratory data analysis**: **5 MARKS**
- **Formal data analysis**: **6 MARKS**
- **Conclusions**: **2 MARKS**
- **General layout**: **1 MARK**

Group Project descriptions

NOTE: Ensure your group downloads and analyses the correct data set you have been assigned otherwise your report will be void and you will not receive a grade.

Modelling weight/obesity in Scotland

The weight and categorisation of weight (namely obesity) in Scotland has been monitored since the introduction of the Scottish Health Survey, which is designed to monitor the health of the Scottish population living in private households. The main aim of the survey is to keep an eye on health trends in Scotland. The Scottish Heath Survey data will be used to explore trends in weight/obesity in Scotland.

Project 1 - Obesity prevalence

Data are available on socio-economic and lifestyle factors from the 2008 - 2012 Scottish Health Surveys. The data are stored in `DAProject1.csv` and contain the following columns.

- `Age` - Age of individual
- `Sex` - Sex of individual (Male / Female)
- `Education` - Highest educational qualification of individual
- `Veg` - Consume recommended daily vegetable intake (Yes / No)
- `Fruit` - Consume recommended daily fruit intake (Yes / No)

在苏格兰建立体重/肥胖模型

自实施苏格兰健康调查以来，苏格兰的体重和体重分类(即肥胖)一直处于监测状态，该调查旨在监测居住在私人家庭中的苏格兰人口的健康状况。这项调查的主要目的是密切关注苏格兰的健康趋势。苏格兰健康调查的数据将用于探索苏格兰的体重/肥胖趋势。

- `Year` - Year of the Scottish Health Survey
- `Obese` - Indicator of individuals obesity classification (Yes / No)

Questions of interest
The main questions of interest are:

- Has the prevalence of obesity in Scotland changed over the given years of the Scottish Health Survey?
- Are there any differences in obesity by age, gender, socio-economic status or lifestyle factors?

Project 2 - BMI distribution

Data are available on socio-economic and lifestyle factors from the 2008 - 2012 Scottish Health Surveys. The data are stored in `DAProject2.csv` and contain the following columns.

- `Age` - Age of individual
- `Sex` - Sex of individual (Male / Female)
- `Education` - Highest educational qualification of individual
- `Veg` - Consume recommended daily vegetable intake (Yes / No)
- `Fruit` - Consume recommended daily fruit intake (Yes / No)
- `Year` - Year of the Scottish Health Survey
- `BMI` - Body Mass Index of individual

Questions of interest
The main questions of interest are:

- Has the body mass index (BMI) in Scotland changed over the given years of the Scottish Health Survey?
- Are there any differences in the BMI distribution by age, gender, socio-economic status or lifestyle factors?

Project 3 - Obesity prevalence

Data are available on socio-economic and lifestyle factors from the 2013 - 2016 Scottish Health Surveys. The data are stored in `DAProject3.csv` and contain the following columns.

- `AgeGroup` - Age range of individual
- `Sex` - Sex of individual (Male / Female)
- `Employment` - Employment status of individual
- `Veg` - Consume recommended daily vegetable intake (Yes / No)
- `Fruit` - Consume recommended daily fruit intake (Yes / No)
- `Year` - Year of the Scottish Health Survey
- `Obese` - Indicator of individuals obesity classification (Yes / No)

Questions of interest
The main questions of interest are:

- Has the prevalence of obesity in Scotland changed over the given years of the Scottish Health Survey?
- Are there any differences in obesity by age, gender, socio-economic status or lifestyle factors?

Project 4 - BMI distribution

Data are available on socio-economic and lifestyle factors from the 2013 - 2016 Scottish Health Surveys. The data are stored in `DAProject4.csv` and contain the following columns.

- `AgeGroup` - Age range of individual
- `Sex` - Sex of individual (Male / Female)
- `Employment` - Employment status of individual
- `Veg` - Consume recommended daily vegetable intake (Yes / No)
- `Fruit` - Consume recommended daily fruit intake (Yes / No)

- `Year` - Year of the Scottish Health Survey
- `BMI` - Body Mass Index of individual

Questions of interest
The main questions of interest are:

- Has the body mass index (BMI) in Scotland changed over the given years of the Scottish Health Survey?
- Are there any differences in the BMI distribution by age, gender, socio-economic status or lifestyle factors?

Project 5 - Obesity prevalence

Data are available on socio-economic and lifestyle factors from the 2013 - 2016 Scottish Health Surveys. The data are stored in `DAProject5.csv` and contain the following columns.

- `Age` - Age of individual
- `Sex` - Sex of individual (Male / Female)
- `Education` - Highest educational qualification of individual
- `Veg` - Consume recommended daily vegetable intake (Yes / No)
- `Fruit` - Consume recommended daily fruit intake (Yes / No)
- `Year` - Year of the Scottish Health Survey
- `Obese` - Indicator of individuals obesity classification (Yes / No)

Questions of interest
The main questions of interest are:

- Has the prevalence of obesity in Scotland changed over the given years of the Scottish Health Survey?
- Are there any differences in obesity by age, gender, socio-economic status or lifestyle factors?

Project 6 - BMI distribution

Data are available on socio-economic and lifestyle factors from the 2013 - 2016 Scottish Health Surveys. The data are stored in `DAProject6.csv` and contain the following columns.

- `Age` - Age of individual
- `Sex` - Sex of individual (Male / Female)
- `Education` - Highest educational qualification of individual
- `Veg` - Consume recommended daily vegetable intake (Yes / No)
- `Fruit` - Consume recommended daily fruit intake (Yes / No)
- `Year` - Year of the Scottish Health Survey
- `BMI` - Body Mass Index of individual

Questions of interest
The main questions of interest are:

- Has the body mass index (BMI) in Scotland changed over the given years of the Scottish Health Survey?
- Are there any differences in the BMI distribution by age, gender, socio-economic status or lifestyle factors?

Project 7 - Obesity prevalence

Data are available on socio-economic and lifestyle factors from the 2008 - 2012 Scottish Health Surveys. The data are stored in `DAProject7.csv` and contain the following columns.

- `AgeGroup` - Age range of individual
- `Sex` - Sex of individual (Male / Female)
- `Employment` - Employment status of individual
- `Veg` - Consume recommended daily vegetable intake (Yes / No)

- `Fruit` - Consume recommended daily fruit intake (Yes / No)
- `Year` - Year of the Scottish Health Survey
- `Obese` - Indicator of individuals obesity classification (Yes / No)

Questions of interest
The main questions of interest are:

- Has the prevalence of obesity in Scotland changed over the given years of the Scottish Health Survey?
- Are there any differences in obesity by age, gender, socio-economic status or lifestyle factors?

Project 8 - BMI distribution

Data are available on socio-economic and lifestyle factors from the 2008 - 2012 Scottish Health Surveys. The data are stored in `DAProject8.csv` and contain the following columns.

- `AgeGroup` - Age range of individual
- `Sex` - Sex of individual (Male / Female)
- `Employment` - Employment status of individual
- `Veg` - Consume recommended daily vegetable intake (Yes / No)
- `Fruit` - Consume recommended daily fruit intake (Yes / No)
- `Year` - Year of the Scottish Health Survey
- `BMI` - Body Mass Index of individual

Questions of interest
The main questions of interest are:

- Has the body mass index (BMI) in Scotland changed over the given years of the Scottish Health Survey?
- Are there any differences in the BMI distribution by age, gender, socio-economic status or lifestyle factors?

Project 9 - Obesity prevalence

Data are available on socio-economic and lifestyle factors from the 2008 - 2012 Scottish Health Surveys. The data are stored in `DAProject9.csv` and contain the following columns.

- `AgeGroup` - Age range of individual
- `Sex` - Sex of individual (Male / Female)
- `Employment` - Employment status of individual
- `Year` - Year of the Scottish Health Survey
- `BMIgroup` - Indicator of individuals weight classification group

Questions of interest
The main questions of interest are:

- Has the prevalence of obesity in Scotland changed over the given years of the Scottish Health Survey?
- Are there any differences in obesity by age, gender, socio-economic status or lifestyle factors?

Hint: You will need to create a new binary response variable for obesity classification from the `BMIgroup` variable.

Project 10 - Obesity prevalence

Data are available on socio-economic and lifestyle factors from the 2008 - 2012 Scottish Health Surveys. The data are stored in `DAProject10.csv` and contain the following columns.

- `Age` - Age of individual
- `Sex` - Sex of individual (Male / Female)
- `Education` - Highest educational qualification of individual

项目9 -肥胖流行率

2008 - 2012年苏格兰健康调查提供了社会经济和生活方式因素的数据。数据存储在DAProject9.csv中，并包

•年龄组别-个人年龄范围-个人性别(男/女)

- 就业-个人就业状况
- 年度-苏格兰健康调查年度
- bmigrgroup -个人体重分类组的指标

感兴趣的问题
值得关注的主要问题有:

- 在苏格兰健康调查的这些年里，苏格兰的肥胖患病率有变化吗?
- 肥胖是否因年龄、性别、社会经济地位或生活方式而有所不同?

提示:您需要从bmigrgroup变量为肥胖分类创建一个新的二元响应变量。

- `Year` - Year of the Scottish Health Survey
- `BMIgroup` - Indicator of individuals weight classification group

Questions of interest
The main questions of interest are:

- Has the prevalence of obesity in Scotland changed over the given years of the Scottish Health Survey?
- Are there any differences in obesity by age, gender, socio-economic status or lifestyle factors?

Hint: You will need to create a new binary response variable for obesity classification from the `BMIgroup` variable.

Project 11 - Obesity prevalence

Data are available on socio-economic and lifestyle factors from the 2013 - 2016 Scottish Health Surveys. The data are stored in `DAProject11.csv` and contain the following columns.

- `AgeGroup` - Age range of individual
- `Sex` - Sex of individual (Male / Female)
- `Employment` - Employment status of individual
- `Year` - Year of the Scottish Health Survey
- `BMIgroup` - Indicator of individuals weight classification group

Questions of interest
The main questions of interest are:

- Has the prevalence of obesity in Scotland changed over the given years of the Scottish Health Survey?
- Are there any differences in obesity by age, gender, socio-economic status or lifestyle factors?

Hint: You will need to create a new binary response variable for obesity classification from the `BMIgroup` variable.

Project 12 - Obesity prevalence

Data are available on socio-economic and lifestyle factors from the 2013 - 2016 Scottish Health Surveys. The data are stored in `DAProject12.csv` and contain the following columns.

- `Age` - Age of individual
- `Sex` - Sex of individual (Male / Female)
- `Education` - Highest educational qualification of individual
- `Year` - Year of the Scottish Health Survey
- `BMIgroup` - Indicator of individuals weight classification group

Questions of interest
The main questions of interest are:

- Has the prevalence of obesity in Scotland changed over the given years of the Scottish Health Survey?
- Are there any differences in obesity by age, gender, socio-economic status or lifestyle factors?

Hint: You will need to create a new binary response variable for obesity classification from the `BMIgroup` variable.

MMR vaccination in Scotland

The Scottish Childhood Immunisation Record System (SCIRS) holds the individual records of all childhood vaccinations in Scotland. These include measles, mumps, and rubella (MMR) vaccination uptake, which occurs when a child is 12-13 months old and again at 4-5 years of age. Data are available from the SCIRS database between 1998 and 2014. The beginning of this time period was when Wakefield et al. (1998) linked the MMR vaccine with an increased risk of autism, with the media coverage surrounding the article

resulting in vaccination rates dropping to around 80% in 2003 in parts of the United Kingdom. These reduced vaccination rates later resulted in large outbreaks of measles in the UK in 2013. The article by Wakefield et al. (2008) was partially retracted in 2004, before being discredited in 2010 after several epidemiological studies failed to find any association with an increased risk in autism.

Project 13 - Measles susceptibility in Glasgow

Data are available on measles susceptibility in pre-school children from the 133 intermediate zones (IZ) comprising Glasgow, which are small geographical units containing, on average, 4000 residents between 1998 and 2014. The data are stored in `DAProject13.csv` and contain the following columns.

- `Y` - The number of pre-school children susceptible to measles in a given IZ
- `N` - The total number of pre-school children in a given IZ
- `Year` - Year the data was collected

Questions of interest
The main questions of interest are:

- Did Glasgow exhibit a change in measles susceptibility following the retraction of the Wakefield article?
- Did the change, if any, in measles susceptibility occur in 2004 alongside the articles' retraction?

Hint: It is the proportion of pre-school children susceptible to measles that is modelled.

Project 14 - Measles susceptibility in Edinburgh

Data are available on measles susceptibility in pre-school children from the 101 intermediate zones (IZ) comprising Edinburgh, which are small geographical units containing, on average, 4000 residents between 1998 and 2014. The data are stored in `DAProject14.csv` and contain the following columns.

- `Y` - The number of pre-school children susceptible to measles in a given IZ
- `N` - The total number of pre-school children in a given IZ
- `Year` - Year the data was collected

Questions of interest
The main questions of interest are:

- Did Edinburgh exhibit a change in measles susceptibility following the retraction of the Wakefield article?
- Did the change, if any, in measles susceptibility occur in 2004 alongside the articles' retraction?

Hint: It is the proportion of pre-school children susceptible to measles that is modelled.

Project 15 - Measles susceptibility in Glasgow

Data are available on measles susceptibility in primary school children from the 133 intermediate zones (IZ) comprising Glasgow, which are small geographical units containing, on average, 4000 residents between 1998 and 2012. The data are stored in `DAProject15.csv` and contain the following columns.

- `Y` - The number of primary school children susceptible to measles in a given IZ
- `N` - The total number of primary school children in a given IZ
- `Year` - Year the data was collected

Questions of interest
The main questions of interest are:

- Did Glasgow exhibit a change in measles susceptibility following the retraction of the Wakefield article?
- Did the change, if any, in measles susceptibility occur in 2004 alongside the articles' retraction?

Hint: It is the proportion of primary school children susceptible to measles that is modelled.

Project 16 - Measles susceptibility in Edinburgh

Data are available on measles susceptibility in primary school children from the 101 intermediate zones (IZ) comprising Edinburgh, which are small geographical units containing, on average, 4000 residents between 1998 and 2012. The data are stored in `DAProject16.csv` and contain the following columns.

- `Y` - The number of primary school children susceptible to measles in a given IZ
- `N` - The total number of primary school children in a given IZ
- `Year` - Year the data was collected

Questions of interest
The main questions of interest are:

- Did Edinburgh exhibit a change in measles susceptibility following the retraction of the Wakefield article?
- Did the change, if any, in measles susceptibility occur in 2004 alongside the articles' retraction?

Hint: It is the proportion of primary school children susceptible to measles that is modelled.

Submission

Everything that you are required to submit for your group project should be submitted by **17:00 26th March 2021** via Moodle. Details of each submission are given below.

Group Report

Your report should summarise the statistical analysis you conducted and a thorough interpretation of your model in relation to the data analysed. You should elect one member of your group to be responsible for submitting the group report as a **.pdf html** file via the **Group Project Submission** link in section **Week 9: Group project** on the Data Analysis Moodle page. You should submit your group report as **GroupNumber_ProjectNumber.pdf html**. Even though one person will be elected to submit the report, all members of the group should be able to see what has been submitted on Moodle.

Contribution Form

Alongside the group report, each member will be need to separately submit a **Contribution Form** as a **.pdf** file via the **Group Project Contribution Form Submission** link in section **Week 9: Group project** on the Data Analysis Moodle page. You should submit your contribution form as **GroupNumber_MatriculationNumber_ContributionForm.pdf**. Here, you will provide information on how much you believe each member of your group contributed to the group project. Contribution of each member may consist of the following:

- attending arranged group meetings;
- analysis of the data set via writing `R` code and/or interpretation of the results;
- writing the groups findings and sections of the report in R Markdown; and
- being cooperative and supportive throughout the project.

If a meaningful discrepancy in the contributions of group members is observed then that may result in different grades being awarded to different group members. For example, no grade will be awarded to a group member who does not contribute anything to the group project.

Declaration of Originality Form

Each member should also submit a **Declaration of Originality Form** as a **.pdf** file via the **Group Project Declaration of Originality Form Submission** link in section **Week 9: Group project** on

the Data Analysis Moodle page. You should submit your Declaration of Originality form as **GroupNumber_MatriculationNumber_DeclarationForm.pdf**.

The **Group Report**, **Contribution Form** and **Declaration of Originality Form** should be submitted by **17:00 26th March 2021**.