Problem Sheet 1

Quantitative Methods

Preparation: Work through chapters 2 and 3 of Grolemund's *Hands-On Programming with R* (available here) and chapters 3–5 of Wickham and Grolemund's *R for Data Science* (available here). Do not just read — type in all the commands yourself, replicate the examples whenever possible, and try the exercises.

A. You roll three dice and record the sum of the number of eyes on each die.

1. Use the `expand.grid()` function in R to create a data frame `S` that shows all possible outcomes of the experiment (the sample space).
2. Note that `expand.grid()` has automatically created three variables called `Var1`, `Var2`, and `Var3`. Create a new variable called `value` that records the sum of `Var1`, `Var2`, and `Var3`.
3. Assuming the dice are fair, calculate the probability that the sum of eyes is equal to 12.
4. Now assume each die is biased, with a probability of rolling a 6 equal to $\frac{3}{8}$ and the probability of all other outcomes equal to $\frac{1}{8}$. Create a vector recording these probabilities and assign them to the relevant entries in `S`.
5. Calculate the updated probability that the sum of eyes is equal to 12.

B.[1] Download the following CSV files and save them in your working directory: `Kenya.csv`, `Sweden.csv`, and `World.csv`. These files contain the following variables:

- ⋆ `country`: abbreviated country name.
- ⋆ `period`: period during which data are collected.
- ⋆ `age`: age group.
- ⋆ `births`: number of children (in thousands) born to women in each age group.
- ⋆ `deaths`: number of deaths (in thousands).
- ⋆ `py.men`: person-years for men (in thousands).
- ⋆ `py.women`: person-years for women (in thousands).

[1] This exercise draws on Exercise 1.5.2 in Kosuke Imai's *Quantitative Social Science: An Introduction*, Chapter 1. The relevant materials are also available here.

The data are collected for a period of 5 years, where *person-year* is a measure of the time contribution of each person during the period. For example, a person who lives through the entire 5-year period contributes 5 person-years, whereas a person who dies after 2 years contributes only 2 person-years.

1. Read each data set into R using either `read.csv()` or `tidyverse::read_csv()`.

2. Use the functions `summary()`, `head()`, and `tail()` to inspect each data set. You can also look directly at the data via `print()` or by double-clicking on the data frame in the `Environment` tab in RStudio.

3. The *age-specifific fertility rate* (ASFR) within an age range $[x,\ x+n)$, where x is the starting age and n is the width of the age range (in years), is defined as

$$\text{ASFR}_{[x,\,x+n)} = \frac{\text{number of births to women of age } [x,\ x+n)}{\text{number of person-years lived by women of age } [x,\ x+n)}.$$

   Create a function called `asfr()` that computes the ASFR for each age group. Calculate the ASFR for Kenya, Sweden, and the whole world separately for each of the two time periods in the data.

4. Using the ASFR, the *total fertility rate* (TFR) is defined as the average number of children given birth to by women who live through their entire reproductive age:

$$\text{TFR} = 5 \times (\text{ASFR}_{[15,\,19)} + \text{ASFR}_{[20,\,24)} + \cdots + \text{ASFR}_{[45,\,49)}).$$

   We multiply the sum by 5 because each woman spends five years in each age range, during which time her annual fertility rate is the ASFR. Create a function `tfr()` to compute the TFR for Kenya, Sweden, and the whole world separately for each of the two time periods in the data.

5. Now calculate the *age-specific death rate* (ASDR), which is defined as

$$\text{ASDR}_{[x,\,x+n)} = \frac{\text{number of deaths of people of age } [x,\ x+n)}{\text{number of person-years of people of age } [x,\ x+n)}.$$

   Create a function called `asdr()` to calculate the ASDR seperately by geography, age group, and time period.

6. Use `ggplot()` to visualise the the ASFR and the ASDR by geography, age group, and time period. Briefly summarise your principal findings.

Deadline: Submit a tidy and annotated R script via email by 2PM on Wednesday 21 October.