Contraceptive Prescribing Project

## Data Sources

* The main source of data for this project is the English Prescribing Dataset (EPD), provided by the NHS Business Services Authority. You can read more about the data in the guidance documents provided here: <https://www.nhsbsa.nhs.uk/prescription-data/prescribing-data/english-prescribing-data-epd>. The dataset provided contains data from **June 2023.**
* The number of patients at each GP practice has been taken from NHS Digital: <https://digital.nhs.uk/data-and-information/publications/statistical/patients-registered-at-a-gp-practice> The number in the dataset provided are from **June 2023** and are the total number of **female patients between the ages of 18 and 60** at each GP as the types of contraception are prescribed only to women.
* The Index for Multiple Deprivation is taken from the Government’s Ministry of Housing, Communities, and Local Government from here: <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019> Deciles were calculated in 2019 but the EPD data is from 2023. The deciles were found using the IMD postcode lookup tool (<https://imd-by-postcode.opendatacommunities.org/imd/2019>). However, as postcode can change frequently, there may be some missing in the dataset. For any missing IMD records, you can use <https://checkmypostcode.uk/> to find the nearest active postcode and use the postcode lookup tool to find the IMD.
* IMD is usually available in deciles (1 referring to the most deprived 10%, 10 referring to the least deprived 10%). I have created a variable called IMD\_quintile that combines IMD deciles into groups of 20% (1 referring to the most deprived 20%, 5 referring to the least deprived 10%)
* The drugs have been aggregated and labelled according to the table of BNF codes described here: Walker, S.H. Effect of the COVID-19 pandemic on contraceptive prescribing in general practice: a retrospective analysis of English prescribing data between 2019 and 2020. *Contracept Reprod Med* **7**, 3 (2022). https://doi.org/10.1186/s40834-022-00169-w

## Software

* For the initial analysis, you will be using Microsoft Excel to create pivot tables. Potential resources are available at:
  + [Navigating within an Excel spreadsheet (video)](https://www.youtube.com/watch?v=qeMUjyOQ_-c)
  + [Filtering and Sorting in Excel (video)](https://www.youtube.com/watch?v=DgRDnqsAAzo)
  + [Pivot tables (Excel)](https://www.youtube.com/watch?v=m0wI61ahfLc)
* For the rest of the analysis, you will be using R in RStudio. You will need to download R (<https://cran.r-project.org/>) and RStudio(<https://posit.co/download/rstudio-desktop/>). Check that the download has been successful by opening RStudio. You may find these resources helpful:
  + [Basics of RStudio and R](https://www.youtube.com/watch?v=TQMAKGDIe_8) (video)
  + [RStudio Tutorial](https://www.datacamp.com/tutorial/r-studio-tutorial) (article)

## Tasks

1. Calculated the rates of prescribing of the different types of contraception by GP practice.
   1. Use an IFELSE function to populate the MOST\_DEPRIVED column in EPD\_DATA such that 1 represents the most deprived GPs and 0 otherwise.
   2. Using the filter function, populate the MOST\_DEPRIVED and NOT\_MOST\_DEPRIVED sheets with the relevant Practice names, practice codes, and total patients.
   3. Use pivot tables to populate the TOTAL\_OVERALL, TOTAL\_COMBINED\_CONTRA, TOTAL\_PROGESTERONE, and TOTAL\_DEVICE columns.
   4. Calculate the rates by dividing each by the number of patients. Usually rates are given per 1000 patients, so multiply each by 1000.
2. For each IMD quintile, calculated the average proportion of contraceptives prescriptions that are combined hormonal contraceptive methods, progesterone only methods, and contraceptive drives. This shows what preferences there are for certain types of contraceptives given that someone is going to be prescribed a contraceptive.
   1. Give as a %.
   2. Use pivot table to find the average % for each type by IMD quintile.
3. Visualised regional differences in contraceptive prescribing rates in the UK as a heatmap. Produced one map for each subgroup of contraceptives.
4. For each type of contraceptive, performed a t-test or Wilcoxon test comparing GPs in the most deprived practices with the rest of the GP practices.
   1. You will need to save each sheet as a CSV to be able to load it into R.
   2. Before conducting a t-test, you will want to remove values in the dataframe that have a zero for the rate of interest (e.g. RATE\_COMBINED\_CONTRA). I would recommend you save this as a separate dataframe that makes it clear which variable has had the zeroes removed. Therefore, in your t-test, you will be comparing the difference in prescribing rates between GPs that prescribe that type of contraceptive. Be careful not to remove all zeroes from the dataset, you only want to remove zeroes from the variable that you are putting into the t-test. Here is the R code to remove values from a dataframe where the variable has a value of 0:

A close up of a number

Description automatically generated

## Tips and Things to Check

* In R, you need to both install and load packages to use them. You can install any packages you need by going to Tools > Install Packages and you load them into your script using library(package name). The packages you will likely use are all in the tutorials so you should be prompted to install them when you open the tutorials and then you load them by running the chunk that features the library functions. You only need to install packages once but you need to load them into any script you want to use them in. If you are getting errors that a function cannot be found, it is probably because you haven’t run the code chunk that loads all of the libraries.
* When reading in CSV files, the path is relative. If, when you run that chunk, it cannot find the data or raises an error, it is because it cannot find the file. Make sure that the data and the Rmd file are in the same folder to avoid this problem.
* Some GP practices may share the same name (it’s more common than you might think!) so when creating pivot tables, use the **Practice code** as this is unique to each GP practice.