**REQUESTED TASK**

You have been asked to complete analyses for a paper on 1) the prevalence of hypertension among adults aged **20-64 years** visiting their GP from **01/01/2018 to 31/12/2018,** and 2) its association with sociodemographic characteristics and the diagnosis of depression/anxiety.

You will be responsible for cleaning, coding and analysing the dataset provided, and filling the table requested. Please carefully check the instructions below.

1. Open the dataset in **STATA** and create a new **do-file** document to save the different steps you followed to complete this task.
2. **The different steps used to complete this task (do-file), the filled in table and the dataset with the new variables need to be saved as a ZIP document.**
3. Check the format and distribution of the variables available in the dataset, so that you can create the required variables to complete the task.
4. Create a binary variable to identify patients with hypertension.

* You should extract information from the variable **“reason”** for consultation (String variable)
* Terms used to identify hypertension include **HPT** and **HYPERTENSION**;
* Terms to be excluded: **WHITE COAT** and **OCULAR HYPERTENSION**
* Remember that this is a free text field, so doctors may have recorded these terms using lower or upper case letters, as well as part of a sentence;
* Because this is a longitudinal dataset, the same patient may have multiple consultations in a year. For the analysis of prevalence, each patient should be counted only once. The variable **“patientid”** denotes the unique identifier per patient, and the variable **“visit date”** the date of consultation.

1. Create a variable **age** in 2018 based on the year of birth, and categorise it using the groups presented in Table 1.
2. Create a variable depression/anxiety, combining the variables depression and anxiety.
3. For your analyses, remember the study should only include adults aged 20-64 years visiting their GP from 01/01/2018 to 31/12/2018.
4. All analyses need to consider the practice (“**practiced”)** as a cluster, and remember that each patient should be counted only once.
5. Fill the table with:

* The total number of patients in each category of sociodemographic variables;
* The prevalence of hypertension according to sociodemographic characteristics;
* Results of a logistic regression model with no adjustments (Crude OR and 95%CI);
* Results of a logistic regression model with adjustments (Adjusted OR and 95%CI).
* Please specify the approach used for adjusted analyses as part of the table footnote.

**Table 1.** Prevalence of hypertension among adults aged 20-64 years and its association with sociodemographic variables and the diagnosis of depression/anxiety.

| **Variable** | **N** | **% of hypertension** | **Crude OR (95%CI)** | **Adjusted OR\* (95%CI)** |
| --- | --- | --- | --- | --- |
| **Sex** |  |  |  |  |
| Male |  |  |  |
| Female |  |  |  |  |
| **Age** |  |  |  |  |
| 20-34 |  |  |  |
| 35-44 |  |  |  |  |
| 45-54 |  |  |  |  |
| 55-64 |  |  |  |  |
| **Remoteness** |  |  |  |  |
| Major cities |  |  |  |  |
| Inner regional |  |  |  |  |
| Outer/ Remote/Very remote |  |  |  |  |
| **IRSAD quintile** |  |  |  |  |
| Very high |  |  |  |  |
| High |  |  |  |  |
| Middle |  |  |  |  |
| Low |  |  |  |  |
| Very Low |  |  |  |  |
| **Depression/Anxiety** |  |  |  |  |
| No |  |  |  |
| Yes |  |  |  |  |
| Total |  |  |  |  |

IRSAD is a macroeconomic indicator; very high means most advantage location and very low means least advantage location. \* Adjustments: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.