Attachment A

STUDY ON THE INCIDENCE OF TUMOR DIAGNOSIS IN THE PROVINCE OF MODENA

References:

[1] I NUMERI DEL CANCRO IN ITALIA 2018; Stefania G., Lucia M., Fabrizio N., Maria M. - Ed. 2018.

[2] I NUMERI DEL CANCRO IN ITALIA 2019; Stefania G., Massimo R., Fabrizio N., Maria M. - Ed. 2019.

[3] I NUMERI DEL CANCRO IN ITALIA 2020; Giordano B., Massimo R., Anna S. - Ed. 2020.

[4] I NUMERI DEL CANCRO IN ITALIA 2021; Giordano B., Stegania G., Anna S., Maria M. - Ed. 2021.

[5] https://www.tuttitalia.it/emilia-romagna/provincia-di-modena/statistiche/popolazione-eta-sesso-stato-civile-2020/

1. IMPORT OF THE DATABASE AND CONVERSION INTO DATAFRAME

```
In [ ]:
# import of the "pandas" and "numpy" packages
import pandas as pd
import numpy as np
```

```
In [ ]:
# we import the database
# saving it as a dataframe
df = pd.read_excel('incidence.xlsx')

# note: the 'incidence.xlsx' file is not publicly provided for privacy reasons
```

```
In [ ]:
# verification of the total number of diagnoses
# from July 2018 to June 2021
print('Total number of diagnoses from July 2018 to June 2021:  {} diagnoses'.format(df.shape[0]))
```

```
Total number of diagnoses from July 2018 to June 2021:  58969 diagnoses
```

2. BREAKDOWN INTO PERIODS

2.a Pre-Covid Period from Feb 2019 to Jen 2020

```
In [ ]:
# selection of only cases within the period specified for the pre-covid
# and creation of a new specific dataframe for pre-covid diagnoses
df_pre = df[((df.Year == 2019) & (df.Month != 1)) | ((df.Year == 2020) & (df.Month == 1))]
# verification of the total number of diagnoses
# from February 2019 to January 2020
print('Total number of diagnoses from February 2019 to January 2020 (pre-covid period):  {} diagnoses'.format(df_pre.shape[0]))
```

```
Total number of diagnoses from February 2019 to January 2020 (pre-covid period):  21288 diagnoses
```

2.b Post-Covid Period from Feb 2020 to Jen 2021

```
In [ ]:
# selection of only cases within the period specified for the post-covid
# and creation of a new specific dataframe for post-covid diagnoses
df_post = df[((df.Year == 2020) & (df.Month != 1)) | ((df.Year == 2021) & (df.Month == 1))]
# verification of the total number of diagnoses
# from February 2020 to January 2021
print('Total number of diagnoses from February 2020 to January 2021 (post-covid period):  {} diagnoses'.format(df_post.shape[0]))
```

```
Total number of diagnoses from February 2020 to January 2021 (post-covid period):  17538 diagnoses
```

3. RESTRICTION OF DIAGNOSES OF INTEREST

In this phase we eliminate the cases that are not of interest for the purposes of the research, therefore:

- BNP;
- NDIS;
- MD.

```python
# we create a function that does the operations automatically
def drop_cases(dataFrame):
    # elimination of diagnoses of BNP
    dataFrame.drop(dataFrame[dataFrame.ISTOLOGIA == 1].index, inplace=True)
    # elimination of diagnoses of NDIS
    dataFrame.drop(dataFrame[dataFrame.ISTOLOGIA == 2].index, inplace=True)
    # elimination of diagnoses of MD
    dataFrame.drop(dataFrame[dataFrame.ISTOLOGIA == 4].index, inplace=True)
```

```python
# apply the previously created function
# to the pre-covid dataframe
drop_cases(df_pre)
# verification of the total number of cancer diagnoses
# occurred in the pre-covid period
print('Total number of cancer diagnoses occurred in the pre-covid period:  {} diagnoses'.format(df_pre.shape[0]))
```

```
Total number of cancer diagnoses occurred in the pre-covid period:  9848 diagnoses
```

```python
# apply the previously created function
# to the post-covid dataframe
drop_cases(df_post)
# verification of the total number of cancer diagnoses
# occurred in the post-covid period
print('Total number of cancer diagnoses occurred in the post-covid period:  {} diagnoses'.format(df_post.shape[0]))
```

```
Total number of cancer diagnoses occurred in the post-covid period:  8195 diagnoses
```

note: the annual incidence is expressed without taking into account cancers arising in the skin excluding melanomas [1-4]. In view of the above, in order to compare the (raw) incidence rates obtained from the database, it is necessary to eliminate these cases.

```python
icd_melanomi = ['Melanoma maligno',
                'Melanoma a diffusione superficiale',
                'Melanoma a cellule epitelioidi',
                'Melanoma amelanotico',
                'Melanoma nodulare',
                'Melanoma a cellule fusate']

topo_zone = ['CUTE',
             'CUTE ED EPPENDICI CUTANEE',
             'REGIONI CUTANEE',
             'REGIONE CUTANEE']
```

```python
# elimination of cases belonging to the skin regions
# and which are not melanomas in the pre-covid period
df_pre = df_pre[~(df_pre['Zone_T'].isin(topo_zone) & ~df_pre['Descrizione_M_x'].isin(icd_melanomi))]
# verification of the total number of diagnoses of the pre-covid period
print('Total number of cancer diagnoses occurred in the pre-covid period without skin cancers but including melanomas:  {} diagnoses'.format(df_pre.shape[0]))
```

```
Total number of cancer diagnoses occurred in the pre-covid period without skin cancers but including melanomas:  6395 diagnoses
```

```python
# elimination of cases belonging to the skin regions
# and which are not melanomas in the pre-covid period
df_post = df_post[~(df_post['Zone_T'].isin(topo_zone) & ~df_post['Descrizione_M_x'].isin(icd_melanomi))]
# verification of the total number of diagnoses of the pre-covid period
print('Total number of cancer diagnoses occurred in the post-covid period without skin cancers but including melanomas:  {} diagnoses'.format(df_post.shape[0]))
```

```
Total number of cancer diagnoses occurred in the post-covid period without skin cancers but including melanomas:  5439 diagnoses
```

4. ELIMINATION OF DOUBLE DIAGNOSIS

From the analysis of the database it is evident that (on average) the patients received 2 histological diagnoses for each pathology. For this reason, since the interest in this case is the incidence, we eliminate the double cases for single patient.

```python
# we eliminate double diagnoses for single patients
# in the pre-covid period
df_pre.drop_duplicates(subset=['COD_PATIENT'], inplace=True)
# verification of the total number of cases of the pre-covid period
print('Total number of cancer cases occurred in the pre-covid period:  {} diagnoses'.format(df_pre.shape[0]))
```

```
Total number of cancer cases occurred in the pre-covid period:  5316 diagnoses
```

```python
# we eliminate double diagnoses for single patients
# in the pre-covid period
df_post.drop_duplicates(subset=['COD_PATIENT'], inplace=True)
# verification of the total number of cases of the pre-covid period
print('Total number of cancer cases occurred in the post-covid period:  {} diagnoses'.format(df_post.shape[0]))
```

```
Total number of cancer cases occurred in the post-covid period:  4515 diagnoses
```

5. EXPLORATION ESTIMATE OF THE ANNUAL INCIDENCE

The purpose of this section is to arrive at the incidence per 100.000 inhabitants of cancer diagnoses in the province of Modena.

note: the robustness of the estimate set out below was investigated in more detail in attachment B

```python
# [5]
POP_MODENA_F = 360433
POP_MODENA_M = 346686

# [3]
INCIDENCE_ITALY_2020_F = 512.0
INCIDENCE_ITALY_2020_M = 735.5
INCIDENCE_ITALY_2019_F = 509.4
INCIDENCE_ITALY_2019_M = 730.0

# [3]
NUMBER_CANCERS_ITALY_2020_F = 181857
NUMBER_CANCERS_ITALY_2020_M = 194754

# note: the Italian cancer report of 2021 does not show the incidences due to covid-19
```

Aggregation of data by number of cancers occurring in women or men in the pre and post covid periods in the Province of Modena

```python
# number of cancers in women before covid in the Province of Modena
CANCERS_YEAR_PRE_COVID_MODENA_F = df_pre[df_pre['SESSO'] == 1].shape[0]
# number of cancers in women after covid in the Province of Modena
```

```python
CANCERS_YEAR_PRE_COVID_MODENA_M = df_pre[df_pre['SESSO'] == 2].shape[0]
# number of cancers in men before covid in the Province of Modena
CANCERS_YEAR_POST_COVID_MODENA_F = df_post[df_post['SESSO'] == 1].shape[0]
# number of cancers in men after covid in the Province of Modena
CANCERS_YEAR_POST_COVID_MODENA_M = df_post[df_post['SESSO'] == 2].shape[0]
```

Calculation of the incidences of cancer in women or men in the pre and post covid periods in the Province of Modena

In [ ]:
```python
# incidence of cancer in women before covid in the Province of Modena
CANCER_INCIDENCE_MODENA_PRE_COVID_F = round(CANCERS_YEAR_PRE_COVID_MODENA_F / POP_MODENA_F * 100000, 2)
# incidence of cancer in women after covid in the Province of Modena
CANCER_INCIDENCE_MODENA_POST_COVID_F = round(CANCERS_YEAR_POST_COVID_MODENA_F / POP_MODENA_F * 100000, 2)
# incidence of cancer in man before covid in the Province of Modena
CANCER_INCIDENCE_MODENA_PRE_COVID_M = round(CANCERS_YEAR_PRE_COVID_MODENA_M / POP_MODENA_M * 100000, 2)
# incidence of cancer in man after covid in the Province of Modena
CANCER_INCIDENCE_MODENA_POST_COVID_M = round(CANCERS_YEAR_POST_COVID_MODENA_M / POP_MODENA_M * 100000, 2)
```

Differences in incidence occurred between the pre and post covid periods in the Province of Modena with respect to the female and male population

In [ ]:
```python
# difference in incidence of cancer in women before and after covid in the Province of Modena
INCIDENCE_DIFFERENCE_F = CANCER_INCIDENCE_MODENA_PRE_COVID_F - CANCER_INCIDENCE_MODENA_POST_COVID_F
print('Difference in incidence of cancer in women before and after covid in the Province of Modena:     {} diagnoses'.format(round(INCIDENCE_DIFFERENCE_F, 2)))
# difference in incidence of cancer in man before and after covid in the Province of Modena
INCIDENCE_DIFFERENCE_M = CANCER_INCIDENCE_MODENA_PRE_COVID_M - CANCER_INCIDENCE_MODENA_POST_COVID_M
print('Difference in incidence of cancer in man before and after covid in the Province of Modena:       {} diagnoses'.format(round(INCIDENCE_DIFFERENCE_M, 2)))
```

```
Difference in incidence of cancer in women before and after covid in the Province of Modena:     94.05 diagnoses
Difference in incidence of cancer in man before and after covid in the Province of Modena:       133.26 diagnoses
```

Percentage differences in incidence occurred between the pre and post covid periods in the Province of Modena with respect to the female and male population

In [ ]:
```python
# percentage difference in incidence of cancer in women before and after covid in the Province of Modena
PERCENTAGE_INCIDENCE_DIFFERENCE_F = round((1- (CANCER_INCIDENCE_MODENA_POST_COVID_F/CANCER_INCIDENCE_MODENA_PRE_COVID_F))*100, 2)
print('Percentage difference in incidence of cancer in women before and after covid in the Province of Modena:     {} %'.format(round(PERCENTAGE_INCIDENCE_DIFFERENCE_F, 2)))
# percentage difference in incidence of cancer in man before and after covid in the Province of Modena
PERCENTAGE_INCIDENCE_DIFFERENCE_M = round((1- (CANCER_INCIDENCE_MODENA_POST_COVID_M/CANCER_INCIDENCE_MODENA_PRE_COVID_M))*100, 2)
print('Percentage difference in incidence of cancer in man before and after covid in the Province of Modena:       {} %'.format(round(PERCENTAGE_INCIDENCE_DIFFERENCE_M, 2)))
```

```
Percentage difference in incidence of cancer in women before and after covid in the Province of Modena:     13.75 %
Percentage difference in incidence of cancer in man before and after covid in the Province of Modena:       16.2 %
```

6. NATIONAL SCREENING

It is interesting to project the reductions in incidence obtained at the national level to calculate, indicatively, the number of missed diagnoses compared to national projections

In [ ]:
```python
print('Number of missed diagnoses compared to national projections in women:           {} diagnoses'.format(round(NUMBER_CANCERS_ITALY_2020_F/100*PERCENTAGE_INCIDENCE_DIFFERENCE_F, 2)))
print('Number of missed diagnoses compared to national projections in man:             {} diagnoses'.format(round(NUMBER_CANCERS_ITALY_2020_M/100*PERCENTAGE_INCIDENCE_DIFFERENCE_M, 2)))
```

```
Number of missed diagnoses compared to national projections in women:           25005.34 diagnoses
Number of missed diagnoses compared to national projections in man:             31550.15 diagnoses
```

In conclusion, if the trend of the Province of Modena were to be confirmed also at the national level, it is possible to predict, with due caution, a number of missed diagnoses equal to about 25.005 for women and 31.550 for man.