Final assignment

Writer: Azadeh Pirzadeh

Hypothesis: Increasing of death tolls Corona in 2020 causes the declining of life expectancy in Netherlands among the men of younger than 30 years old more than other groups.

In this assignment I used 2 datasets from CBS.

These are the links of 2 datasets:

https://www.cbs.nl/en-gb/figures/detail/71950eng

https://www.cbs.nl/en-gb/figures/detail/7052eng

First dataset

The first dataset contains the number of deaths in the population of the Netherlands by underlying cause of death (short list), sex and age-group (at time of death). Since 2013 Statistics Netherlands is using Iris software for automatic coding for cause of death. This improved the international comparison of the data. The change in coding did cause a considerable shift in the statistic. Since 2013 the (yearly) ICD-10 updates are applied.
Data available from: 1950
Status of the figures:
Figures up until and including 2020 are final.
Changes as of December 21st 2021:
- The provisional figures for 2020 have been replaced by final figures
Changes as of August 18th 2021:
- As of 2020 'COVID-19' has been added to the causes of death.

I used 3 important columns:

- **All causes of death**

Total number of deaths.
ICD-codes:
ICD-10: A00-Y89;
ICD-9: 001-E999;
ICD-8: 000-E999;
ICD-6+7: 000-E999

- **18.1 Confirmed COVID-19**

18.1 Confirmed COVID-19
ICD-codes:
ICD-10: U07.1;
ICD-6+7+8+9: --

- **Average population**

The average population in agegroup L for year t is calculated as follows:
((Population in agegroup (L) on January 1st year t)+
(Population in agegroup (L) on January 1st year t+1))/2.
The calculation of the average of children at the age of zero years differs from this definition and has been calculated as follows:
((live births in year t)+(children aged zero years on 1 January of year t+1))/2.

Second dataset

The second table represents five variants of health expectancies:
-life expectancy in perceived good health.
-life expectancy without physical limitations.
-life expectancy without chronic morbidity.
-life expectancy in good mental health.
-life expectancy without GALI-limitations
In addition, figures of 'normal' life expectancy are included, so the figures of health expectancy can be related to them.
In the table, the data on health expectancy can be split into the following characteristics:
-sex (starting from the data of 2018, the category ‘total, men + women’ is added).
-age.

Using this table one can see the developments over time of health expectancies. For example it can be seen that morbidity free life expectancy of women shortened during the eighties and nineties.
In the same period the life expectancy free of moderate and severe limitations of men increased.
Data available from: 1981
Status of the figures:
The figures in this table are definitive.
Changes as of 15 July 2021:
The figures of 2020 are added.

I used one important column:

- **Life expectancy**
  The number of years someone of a certain age is expected to live, according to a life table, assuming that the mortality chances will not change in the future.

My project consists of 12 parts, and it writes by Jupiter Notebook.I used the data of 2016 to 2020 because I think that 4 intervals (2016-2017,2017-2018,2018-2019,2019-2020) would be good to compare. In addition, there was some missing data before 2016 in life expectancy too, so I decided to compare the data from 2016.

Part1: In the first part I read the csv files and change it to data frames.

Part2: In the second part I cleaned the data frame of death (rename and chose the columns that I need), then add 7 rows to prepare the dataset to interpolate because I had missing data from 2011 to 2018.Then I estimate the data between these dates. After that I checked types of columns, transposed and deleted the extra columns. I also calculate the percentage of changing between the years in 4 intervals and finally replaced the inf column with none. So, the defeat was ready.

Part3: I cleaned the dataset of life expectancy (rename, chose the important columns), then reshape it to prepare it for concatenation, after that calculate the percentage of

changing between intervals, so the df_life was ready. Finally, I concatenate to data frames to one as df.

Part4: Made a function of QQplot to check if the data of life expectancy and number of deaths were following the normal distribution or not and you can see that they followed the normal distribution as well.

Part5: I used the t-test to check that is there any significant differences between the 2 important interval (2018-2019,2019-2020), and the result was that no, there is not any significant changes.

Part6: I plot line graph that compare the life expectancy between 2016-2020 of women and men in 3 tabs. Each tab refers to one age group (0-30,30-65,65 and more)

In this graph I wanted to visualize the changing of life expectancy in different age groups.

Part7: In this part I plotted 2 stack bars that compare the percentage of changing life expectancy in 2 important intervals (2018-2019, 2019-2020). In this graph you can see the life expectancy of women was decreased in all age groups in 2019-2020 rather than 2018-2019. Also, you can see the same for men as well. So, we can conclude that this decline would be because of death tools Corona.

Part8: in this part I plotted the bar plot that can show the differences of life expectancy in men and women in all group ages. It can also prove that the changes of life expectancy were not significantly changes.

Part9: This is 2 horizontal plots that show the differences of total death and average population (Unfortunately there is not any category for women and men in this dataset so I just can analyze it totally). In addition, you can see that the changes of different years are not surprisingly big.

Part10: I made a heatmap plot to visualize the relation between death and different age groups of females, as you can see the strong relation between number of deaths in 2019-2020 and life expectancy. Also, we can see that it affects more on elder women rather than young one.

Part11: In this part I made a heatmap for men as well. You can see the same result, while the number of deaths effects on all ages the same.

Part12: As a conclusion, in this line graph you can see the percentages of changes in all intervals for both men and women and for death number and average population. Increasing the number of deaths effects on decreasing the life expectancy of all people. So, as I see in the defeat the only thing that was not before as a reason of death was the covid diseases. In addition, we can see that the slope of line plot of men over 65-years-old is more than other. Also, I can prove that by the df_health. As you can see in column of 2019-2020(%), the value of men over 65-years-old in 2019-2020 is more than others, it means that the life expectancy in men over 65years old reduced more than others.

So, we can conclude that the death tools Covid effect on declining the life expectancy of men over 65-year-old more than other groups.