**The model**

For the selection of a model, three options were reviewed. The first option was a chain model. A chain model reviews the impact of variable A on B, and subsequently the impact of B on C.



Figure 1: Graphical display of chain model

The second option was a collider model. A collider model assumes that multiple variables have an impact on one selected variable. For instance, variable A, B and C have an impact on variable D.

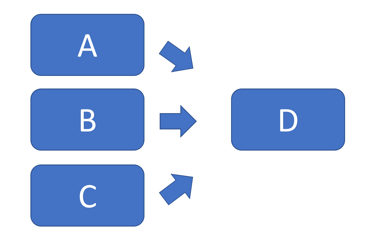


Figure 2: Graphical display of collider model

For our case we have decided to use a fork model. A fork model assumes that 1 variable has an impact on multiple other variables. Our hypothesis is that age dependency is positively correlated towards a number of variables from our selection. For this reason, a fork model is most suitable.

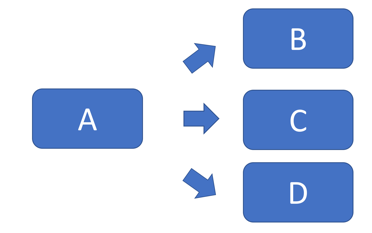


Figure 3: Graphical display of fork model

**Correlation of variables**

In figure 1, the correlation between each of the normalised variables is depicted.

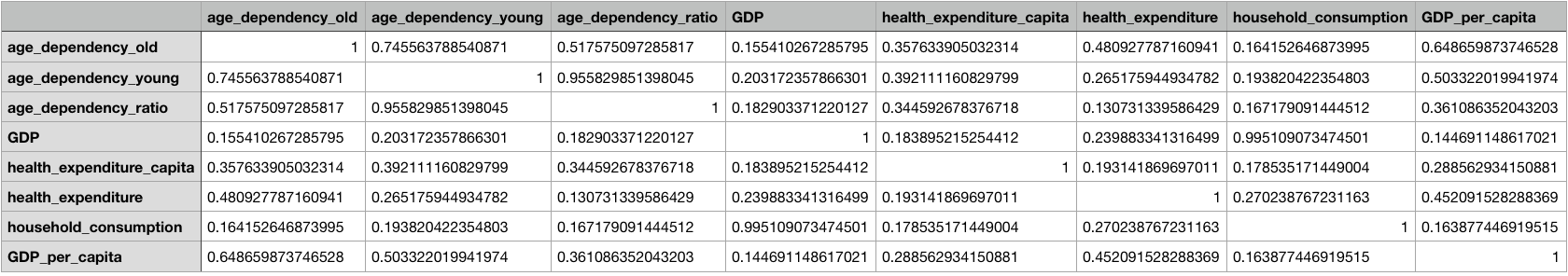
A correlation is a statistical measure for indication of the extent to which two or more variables fluctuate (together). A positive correlation indicates the extent to which the values of both variable increase or decrease in parallel. A negative correlation indicates the extent to which the value of one variable increases and the other decreases. Correlation can be useful, since it might indicate a predictive relationship.

In the figure, the square is coloured red if significant correlation is observed. Blue indicates a small correlation and yellow squares indicate that almost no correlation is observed.



Figure 4: Correlation of normalised variables

In the introduction, a hypothesis was set for the relation of each variable towards age dependency (old). Each of these hypotheses will be reviewed, based on the correlation in figure 1. During this evaluation outliers will be discussed, and ultimately a concrete fork model will be presented. In table 1, all the correlations are given numerically.



1. **Health Expenditure**

For health expenditure the following hypothesis was set in the problem understanding part:

[Hypothesis 1: Cost of healthcare is positively related to the age dependency ratio]

**Age dependency ratio:** In figure 1, it can be seen that health expenditure has no significant correlation with the age dependency ratio (0.13, yellow). Based on this it can be concluded that hypothesis 1 should be rejected.

**Age dependency young:** For age dependency young figure 1 also indicates that no significant correlation with health expenditure is observed. The correlation is a bit higher (0.26, blue), but still quite low.

**Age dependency old:** However, for age dependency old a significant correlation is observed (0.48, red).

A possible explanation for these results is that cost of healthcare is mostly increasing with the amount of elderly people. Elderly people tend to need more care and therefore in the case of a high age dependency ratio (old) the cost of healthcare is generally higher. For the overall age dependency ratio this effect is not strong enough.

1. **Health expenditure per capita**

For the health expenditure per capita the following hypothesis was set in the problem understanding part:

[Hypothesis 2: Cost of healthcare per capita is positively related to the age dependency ratio]

**Age dependency ratio:** In figure 1 and table 1, it can be seen that the health expenditure per capita shows small correlation with the age dependency ratio (0.34, blue). Based on this correlation the hypothesis set can be considered true, but the relation is not very strong.

**Age dependency young:** For the age dependency young, the figure shows a slightly stronger correlation (0.39, red).

**Age dependency old**: Age dependency old shows a correlation with health expenditure per capita that is comparable to the correlations of age dependency ratio and the age dependency young (0.35, blue)

The age dependency ratio shows small correlation with health expenditure per capita. This correlation is similar for the young and old age dependency. An important note for this conclusion is the fact that in health expenditure per capita the effect of age dependency not is included. By calculating this expenditure, it is assumed that all people pay for the costs related to health. In practice these costs are covered by the working part. In order to really test the stress that is caused by health expenses, this should be further analysed.

1. **Household consumption**

For the household consumption variable, the following hypothesis was set in the problem understanding part:

[Hypothesis 3: Household is not related to the age dependency ratio]

**Age dependency ratio:** The age dependency ratio shows a weak correlation with the household consumption (0.16, yellow).

**Age dependency young:** The correlation age dependency young to household consumption is comparable to the age dependency ratio (0.19, blue).

**Age dependency old**: The correlation of household consumption with the age dependency old is also comparable. (0.16, yellow)

Based on the number described, the hypothesis is still questionable. The observed correlation is very weak; therefore, it is uncertain whether the two are related. It is a possibility that age dependency ratio has an impact on the household consumption, but this impact is not strong enough for a stronger correlation, the other possibility is that this correlation is coincidental, or caused by other factors.

1. **GDP**

For the economic performance variable, the GDP, the following hypothesis was set in the problem understanding part:

[Hypothesis 4: Economic performance is negatively related towards the age dependency ratio]

**Age dependency ratio:** The age dependency ratio shows a weak correlation towards the GDP (0.18, yellow).

**Age dependency young:** For age dependency young the correlation is comparable (0.20, blue)

**Age dependency old:** For age dependency old the correlation is comparable as well (0.15, yellow).

The age dependency ratio is weakly correlated with the GDP of a country. Solely looking at this correlation, the hypothesis would be rejected. However, is not possible to exclude the possibility of a relation between both variables solely on these data. Reason for this is that countries might have a higher age dependency for other reasons, for instance a high level of welfare. This might cause an increase in life expectancy of the inhibitors of this specific country. Now both age dependency and GDP (assumption based on high level of welfare) are high in this situation, but a direct link is missing.

1. **GDP per capita**

For the other economic performance variable, the GDP per capita, the following hypothesis was set in the problem understanding part:

[Hypothesis 5: GDP per capita is negatively related towards age dependency ratio.]

**Age dependency ratio:** The age dependency ratio is significantly correlated with the GDP per capita (0.36, blue).

**Age dependency young:** For age dependency young the correlation is even stronger(0.50, red).

**Age dependency old:** Lastly, for the age dependency old the correlation is strongest (0.65 red).

Interpretation of these data again should be done carefully. A correlation between age dependency and GDP per capita clearly exist. However, for numerous reasons this does not have to be an indication for a causal relation. Considering the fact that age dependency old is the highest, a greater number of elderly people is expected to have a correlation with a higher GDP per capita. In order to really analyse the impact of age dependency a longer time frame should be analysed.

1. Age dependency young
2. Age dependency old

To-Do:

* Create fork diagram(s) for our case
* Map causal relations for introduction
* Convert this document / introduction to markdown file
* Add part for further research (See analysis part + state something about longer time frame)
* Include hypotheses about age dependency old and young
* Mention other correlations from figure (that not relate towards age dependency)
* Check Bayesian assumptions (for Model selection)