Solution Series 6

1) In the following there are examples given for a business problem which has to be solved. Write for each example down, a) which method(s) you would select and use or test to each other, b) which software tool(s) you probably would use, c) the two most important qualities of the model you are expecting and d) how you would test the goodness of fit or performance of the model:

   i) A retailer wants to launch some special offer and is sending out advertisements to the existing customers. The retailer wants to know among all existing customers which are responding to that given offer.

   Methods: Classification; two classes: "will respond" and "will not respond" or logistic regression

   Software: R, SAS, SPSS, KNIME, WEKA, depending, if the retailer has already some software and if not what is the budget and intention for software e.g. the retailer wants to have the model and use it regularly.

   Qualities of the model: Precision and explicit results

   Test of performance: Accuracy / error rate, ROC

   ii) A service company is building up a new help line service for their customers. The company wants to know, how much a given customer will use that new service.

   Methods: Regression; prediction of "service usage" by e.g. historical customer data

   Software: Excel, R, SPSS, it is probably a one off analysis and maybe the software will not be used anymore afterwards, thus the software has not to be so expensive and easy to use

   Qualities of the model: Robustness and explicit results

   Test of performance: R^2, R^2_{adj}, AIC, BIC

   iii) A consumer goods producing company wants to restructure their sales teams and is thinking about how to organise the different sales teams.

   Methods: Clustering; grouping the individuals based on similarities

   Software: R, Python (pandas), KNIME, WEKA, depending of course what software the company already has, but as this is most probably a one-off project and the company maybe do not have any statistical tool and you can maybe choose the one you want.

   Qualities of the model: Possibility of parameter setting and explicit results

   Test of performance: ROC but other measures are Sum of Squared Error (SSE) which measures the compactness of a cluster or comparison with a given cluster structure (Rand Index)

iv) A manufacturing company wants to improve the efficiency of its production line in one of its plants.

Methods: Simulation (eventually, combined with optimisation);

Software: Depending how the processes are modelled: Markov models; Monte Carlo simulation; System Dynamics; Discrete Event Modelling; all of them are possible depending on the detailed aspects of the production processes which have to be modelled. Hence, Matlab, R, @Risk, Vensim/PowerSim or anylogic can be used.

Qualities of the model: Possibility of parameter setting and robustness

Test of performance: Analysis of the difference between der model the status quo. One is typically first setting up the model which models the current status and calibrate it that it is sufficient close to the current results of the production line. Based on criteria which are important e.g. number of produced items during a given time period, production time, the accuracy of the model is defined. Afterwards, the model can be adjusted for optimisation purpose.

v) A trader wants to analyse if messages on social media platforms like e.g. twitter have an influence on the price of Novartis shares.

Methods: Statistical inference: testing *if* messages have an influence; logistic regression or classification if one wants to know *the direction* of the influence e.g. "+" and "-" or even "++", "+", "0", "-", "--"; or if one wants to know *how much* it influences it starts with regression but can be extended to neural networks and so on

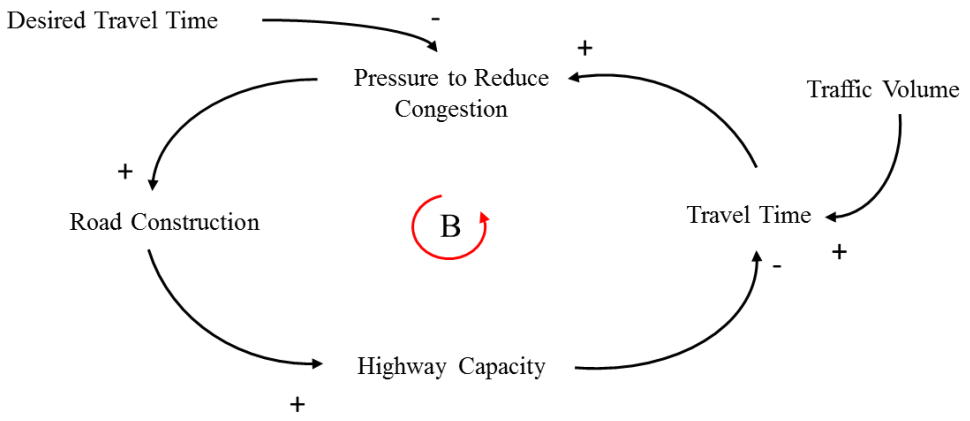
Software: Depending on the data volume it can be done with Rapidminer, Python (pandas), R or if the volume is bigger and it is done in a professional manner, then Hadoop based techniques would be needed

Qualities of the model: Diversity of types of data processed and if you want to use it as a trading strategy, then speed of the model development is important, otherwise the explicit results are important.

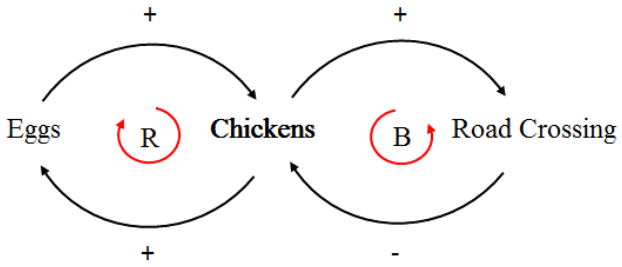
Test of performance: p-value in inference, further R^2_{adj}, AIC, accuracy / error rate, ROC (depending on the methods and the methods which are compared to each other)

2) System Dynamics: feedback and causal loop diagram for planning infrastructure: in today's world there are a lot of traffic jam and congestion. One solution is typically building more roads and increasing the capacity.

One solution – given the hints in the exercise – could look as follows:



3) System Dynamics: In the lecture we have seen the feedback and causal loop diagram of the dynamics of the eggs and chicken and the chicken and road crossing. Both diagrams can be combined for representing one system:



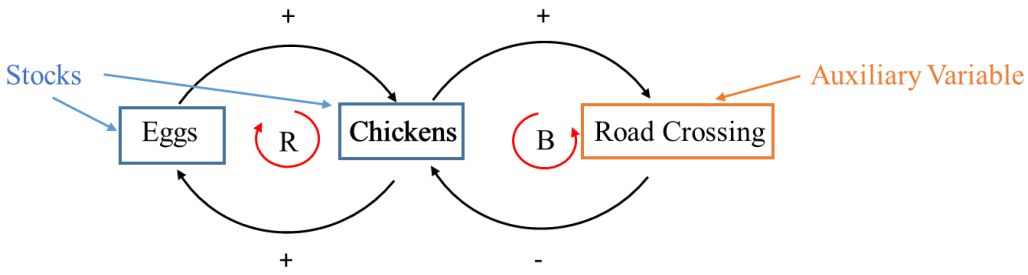
First, determine the stocks and the auxiliary variables which give the dynamics:
Stock = an inventory which contains items where the amount of items can decrease or increase

- Eggs
- Chickens

Dynamic = auxiliary variable:

- Road Crossing

Arrows: The arrows including their direction of influences i.e. "+" or "-" remain the same:



If we now translate that into a stock and flow diagram first, start with the two stocks “eggs” and “chickens” and connect them. Between the two stocks, the valve “hatching rate” is determining the growth in the chicken stock.

As every stock and flow diagram has one source and one sink (stocks outside the model), there must be a source pointing to the stock “eggs” and from the stock “chickens”

Then, the increase of the stock of the “eggs” is determined by a valve “egg laying rate” and the decrease of the stock of the “chickens” is determined by a valve “run over rate”.

Finally, “road crossing” gives a dynamics and is therefore an auxiliary variable.

Linking all together gives: