Course 42588 – week 12

Data analytics IV – choice models

Today's program

- Choice models
  - Choice sets
  - Logit models
  - Willingness to pay and elasticities
  - Example
  - Model evaluation
  - Exercise

- Work on project 3

The course plan

Week	Date	Subject/Lecture	Literature	Exercises	Teachers
1	31/1	Introduction + questions and data	AoS chap. 3	Form groups + week 1 exercise	Stefan
2	7/2	Basics on data and variables	AoS chap. 1-2 (+ OM 1)	Project 1 – start	Stefan/Guest from Genmab
3	14/2	Surveys + data types + experimental data	Paper 1 (+ OM 2-5)	Project 1 – work	Sonja / Stefan
4	21/2	Governance + causality	Paper 2 + AoS chap. 4 (+ OM 6)	Project 1 – deadline	Hjalmar / Stefan
5	28/2	More on data, e.g. real-time data, online data	Paper 3 (+ OM 7-10)	Discuss data for project 2	Guido/ Stefan
6	6/3	Visualisation	Chap. 1,5,6,7,10,23, 24,29 in Wilke + (AM 1-2)	Integrated exercises + work on project 2	Mads
7	13/3	Spatial data	Chap. 1,14 in Gimonds	Week 7 exercises + work on project 2	Mads / Guest from Niras
8	20/3	Imputation/weighting/presentation proj. 2	Paper 4	First deadline of project 2 + Week 8 exercises	Mads
9	3/4	Data analytics I	ISL ch. 3 + paper 5	Work on project 3	Stefan
10	10/4	Data analytics II	ISL ch. 6-6.1 + 7-7.4	Work on project 3	Stefan
11	17/4	Data analytics III	ISL ch. 4-4.3	Work on project 3	Stefan
12	24/4	Data analytics IV	Train 2.1-2.3 + 3.1, 3.6, 3.8-9	Work on project 3	Stefan
13	1/5	Summary and perspective	Paper 6	Project 3 – deadline	Stefan

Logistic regression for 2+ classes

- The probability formula for logistic regression generalises to a well-defined model for more than two classes (discuss 1 min. with your neighbour how).
- The book notes that this model tends not to be used that much. It mentions other models that can be used for multiclass classification, e.g. linear and quadratic discriminant analysis.
- From a machine learning perspective, multiclass logistic regression is essentially a very simple neural network. So in that sense they are used a lot.
- However, interpretability may be lost in multiclass logistic regression and especially in more advanced neural networks. So if interpretability is important, many researchers (especially in economics and social sciences) use more theoretically founded models…

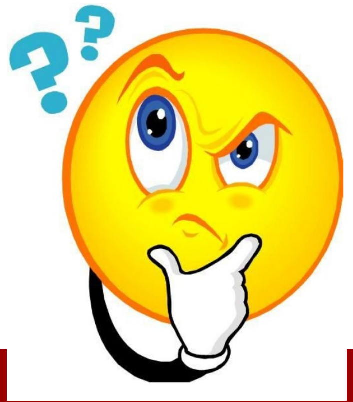
Purpose of today’s lecture

- Look at an example of an econometric model, i.e. a data science/statistical model based on theoretical considerations.
- Describe and discuss choice models
- Discuss logit models and give an example
- Compare logit models to logistic regression and binary prediction models
- Apply a logit model to analyse a choice situation
- Today’s lecture is not exam curriculum but intended to give you some perspective on classification methods!

Discrete choices - motivation

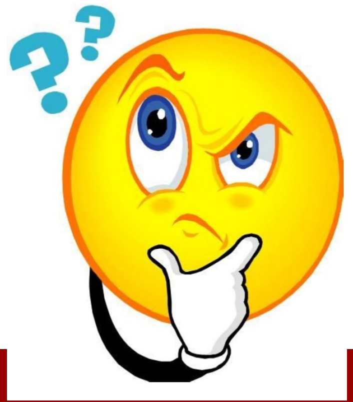
- Discrete choice models are used to model discrete variables with a finite number of outcomes.

- Examples are (discuss 1 min. with your neighbour)



Choice sets

- To use a choice model, we need to define a choice set, i.e. a set of available alternatives
- The choice set must satisfy three criteria
  - Alternatives are mutually exclusive
  - Choice set is exhaustive (i.e., all relevant alternatives are considered)
  - The number of alternatives is finite
- Think for two minutes about the choice set for commuting in Copenhagen!



Choice set definition

- Choice set size can be very large
  - Vehicle choice – there are approx. 1-2000 different car variants sold in DK each year
  - Beverage choice at Starbucks

- Choice set composition can be different
  - Individuals may share the same choice set
  - Individuals may have different choice sets, e.g. individuals without a licence or travellers of age less than 17 cannot drive a car in Denmark. Sometimes car driving is also excluded from the choice set of non-car owners.

Choice set

Universal choice set

- All potential alternatives for the population
- Restricted to relevant alternatives

Mode Choice

- Walk
- Bike
- E-scooter
- Car driver
- Car passenger
- Public transport

Choice set

Individuals choice set

- No driver's license
- No car available
- Trip length too long for walking

Mode Choice

- ~~Walk~~
- Bike
- E-scooter
- ~~Car driver~~
- Car passenger
- Public transport

Choice probabilities

- A discrete choice model gives the probabilities for each alternative conditional on explanatory variables, i.e. $P(i|x_n)$.

- The choice probabilities of the most commonly applied discrete choice model are

$$P(i|x_n) = \frac{\exp(V(x_{ni}))}{\sum_j \exp(V(x_{nj}))}$$

where $V(x_{nj})$ are functions of the explanatory variables related to alternative j. These are also know as utility functions.

- This model is know as the logit model or the multinomial logit (MNL) model. If there are only two alternatives it is know a binary logit.

Discrete choice – example

Obs ID	Alt.	Ttime (min)	Tcost (kr)	Choice
1	Car	32	40	1
1	Bike	53	10	0
1	Bus	34	29	0

- If we suppose that an individual has utility functions

$$V(x_{ni}) = \mathrm{k_i} + \beta_t * Ttime_{ni} + \beta_c * Tcost_{ni}$$

and

$$\mathrm{k_{car}} = 1, \mathrm{k_{bike}} = 2, \mathrm{k_{bus}} = 0, \beta_t = -0.2, \beta_c = -0.1$$

We can calculate the probability for choosing car, i.e. $P(i|x_n) = \frac{\exp(V(x_{ni}))}{\sum_j \exp(V(x_{nj}))}$

Discrete choice – example

- Similar calculations for the other alternatives give

Obs ID	Alt.	Ttime (min)	Tcost (kr)	Choice	Prob.
1	car	32	40	1	0.39
1	bike	53	10	0	0.32
1	bus	34	29	0	0.29

- In a scenario, we might have changes in the input, which will then change the probabilities

Obs ID	Alt.	Ttime (min)	Tcost (kr)	Prob. New	
1	1	32	31	0.61	+57%
1	2	53	10	0.20	-36%
1	3	34	29	0.18	-36%

Discrete choice models – RUM

- The discrete choice models used in mode choice and travel demand modelling are linked to what is known as random utility maximisation (RUM).

- Suppose an individual with rational preferences makes choices from a choice set, J. Then it can be shown that the alternatives in the choice set can be represented by utility functions, U_j such that the probability of a specific choice is equal to the probability of the utility being the maximum utility, i.e

$$P(i|x) = P(U_i > U_j \; for \; all \; j \; \neq i)$$

- So to calculate the probability we need to define the utility functions, U_j.

Discrete choice models – theory

- How do we derive the choice probabilities for the logit model?

- The first assumption that is normally made is to split the utility functions in a systematic part and a random part independent of the explanatory variables:

$$U_{ni}(x) = V_{ni}(x) + \varepsilon_{ni}$$

- It is common to assume that V_i is linear-in-parameters, i.e.

$$\mathrm{V}_{ni} = \beta_i + \beta_1 x_{ni1} + \cdots + \beta_k x_{nik}$$

but we can also use transformations of the x-variables, e.g.

$$\mathrm{V}_{ni} = \beta_i + \beta_1 \ln(x_{ni1}) + \cdots + \beta_k (x_{nik})^2$$

Discrete choice models – the MNL model

- If

$$U_{ni}(x) = V_{ni}(x) + \varepsilon_{ni}$$

where the ε_{ni} are IID extreme value type I (Gumbel) distributed and individuals maximise utility then it can be shown that

$$P(i|x_n) = P(U_{ni} > U_{nj}\ for\ all\ j\ \neq i) = \frac{\exp(V(x_{ni}))}{\sum_j \exp(V(x_{nj}))}$$

- So this is the reasoning behind the choice probabilities we saw earlier.

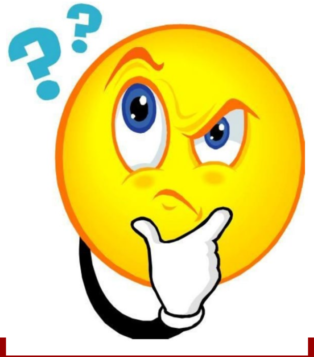
- Important: An attribute of alternative i can only enter V_{ni}!

Question – scales

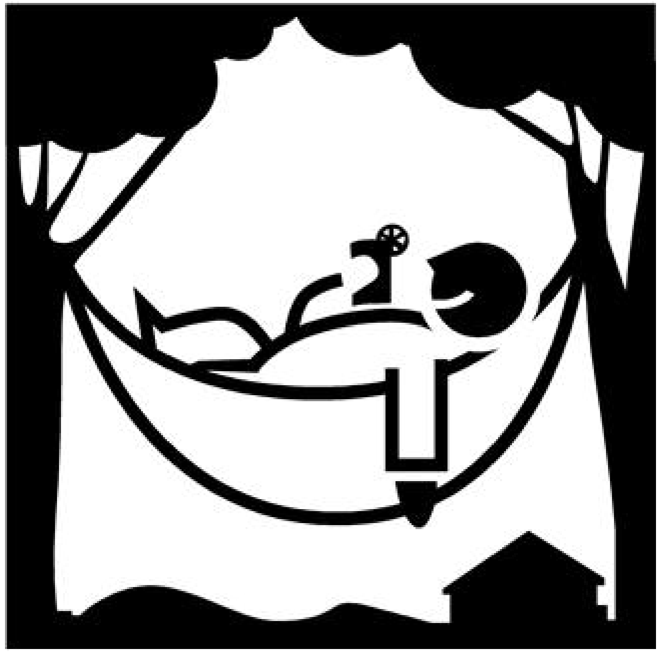
- Suppose you have an income variable measured in nine groups [0-99999], [100000-199999],…,[800000+]. How is this variable scaled?

A. *Nominally*
B. *Ordinally*
C. *Interval*
D. *Ratio*

Answer:



Break



Choice modelling using discrete choice models

- We can model the demand for a given population using probabilities.

- We see the demand for an alternative as the sum of probabilities over individuals in the population (often represented by a sample).

- If a person has probability, P_{ni}, to choose alternative i, then we get

$$N_i = \sum_{n \in Pop} P_{ni} = N * \frac{1}{N} * \sum_{n \in Pop} P_{ni} = N * \overline{P_i}$$

e.g. if the probability to choose car as travel mode is $P_{n,car}$ then we get

$$N_{car} = N * \overline{P_{car}}$$

- Since N is fixed, we often just model $\overline{P_{car}}$ instead of N_{car}.

Alternative specific constants

- We say that a model is identified if we can estimate all parameters from data.

- Due to identification, it is only possible to include K-1 alternative specific constants (ASCs). In the example, we assume $\mathrm{k}_3 = 0$.

- An unidentified model will normally have infinite many solutions. Setting $\mathrm{k}_3 = 0$ is a way to choose a specific of these many solutions that all yield the same probabilities.

- When K-1 ASCs are included, the model will reproduce the sample shares, i.e. the market shares in the sample:

$$\frac{1}{N}\sum_n P_n(i) = \frac{1}{N}\sum_n y_{in}$$

e.g. if 213 persons choose alternative i then $\sum_i P_{ni} = 213$.

MNL models - marginal utility

- The marginal utility in a discrete choice model is defined as

$$MU_{x_k}^{i} = \frac{\partial \mathrm{U}_i}{\partial x_k}$$

- So in a model that is linear in the attributes (x_k), we have

$$MU_{x_k}^{i} = \beta_k$$

- Marginal utility needs to have the correct sign (but the size is difficult to interpret)
  -travel time and cost should have negative marginal utility

- As for other models we can calculate elasticities to compare the effects of various attributes.

MNL models - elasticities

- We can calculate two kinds of elasticities for a discrete choice model

- The direct elasticity
  - This is the percentage change in the probability of an alternative as a function of a percentage change in an attribute of that alternative

- The cross elasticity
  - This is the percentage change in the probability of an alternative as a function of a percentage change in an attribute of another alternative

- Another way of formulating it is
  - **Direct elasticities:** if an attribute of alternative *i* changes by p%, how much will the demand change for alternative *i* in %?
  - **Cross elasticities:** if an attribute of alternative *i* changes by p%, how much will the demand change for alternative *j* in %?

Elasticities

- For the logit model, these become

  – The direct elasticity: $E_{nii} = \partial P_{ni}/\partial x_{ni} \cdot x_{ni} / P_{ni} = \partial V_{ni}/\partial x_{ni} \cdot x_{ni} \cdot (1-P_{ni})$

  – The cross elasticity: $E_{nij} = \partial P_{ni}/\partial x_{nj} \cdot x_{nj} / P_{ni} = -\partial V_{nj}/\partial x_{nj} \cdot x_{nj} \cdot P_{nj}$

- For the linear-in-parameters logit model, we have

$$E^{i}_{x_{ik}} = \frac{\partial \mathrm{P}_i}{\partial x_{ik}} \frac{x_{ik}}{\mathrm{P}_i} = \beta_k x_{ik}(1 - P_i) \text{ (direct elasticity)}$$

$$E^{i}_{x_{jk}} = \frac{\partial \mathrm{P}_i}{\partial x_{jk}} \frac{x_{jk}}{\mathrm{P}_i} = -\beta_k x_{jk} P_j \text{ (cross elasticity)}$$

Elasticities

- For travel time in the linear-in-parameters logit model, this gives
  - The direct elasticity: $E_{nii} = \beta_t \cdot t_{ni} \cdot (1-P_{ni})$
  - The cross elasticity: $E_{nij} = -\beta_t \cdot t_{nj} \cdot P_{nj}$
- For an individual with car travel time 20 min and mode choice probabilities equal to the shares: car 51%, train 20%, bus 29%, the elasticities are (assume $\beta_t = -0.1$)

-0.98 (direct) and 1.02 (cross)

- Note that the cross elasticity is the same for all alternative modes!

Question – elasticities!

- What is the cross elasticity of cost ($-\partial V_{nj}/\partial x_{nj} \cdot x_{nj} \cdot P_{nj}$) in the following model for mode choice if cost is in kr and time in hours:

$$U_j = \alpha_j + \beta_t t_j + \beta_c \ln(c_j) + \varepsilon_j,$$

with $\beta_t = -1.2$, $\beta_c = -0.6$, travel cost is 30 kr, and $P_{nj}=0.5$?

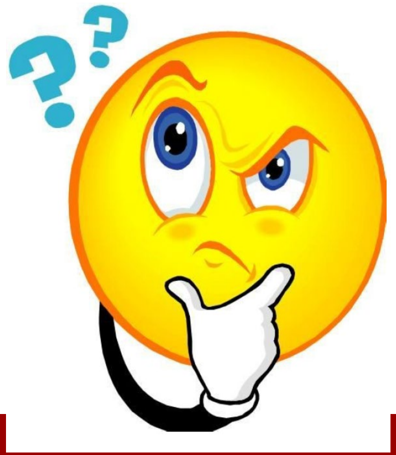
A. *-0.3*

B. *-0.3 kr/hr*

C. *0.6*

D. *0.3*

Answer:



The value of travel time

- A concept that is very important in many fields, e.g. economics, transport, and marketing, is the concept of willingness to pay (WTP).

- This is the monetary value of a marginal change in an attribute for an individual. Or the average monetary value in a population.

- Example from transport, ☺
  - A concept that is very important in transport modelling and transport planning is the value of time (VoT) or value of travel time (VTT).
  - This is the monetary value of a marginal change in travel time for an individual.
  - It is important in transport planning and evaluation of infrastructure projects because time savings are the major factor in most projects.

Discrete choice models – VoT/VTT

- The unit of VTT is money/time, e.g. DKK/hour. It is defined in a discrete choice model with utilities

$$U_i = \mathrm{k_i} + \beta_t * TT_{ni} + \beta_c * TC_{ni} + \varepsilon_{ni}$$

that include travel time (TT) and cost (TC) as

$$VTT_i = MU^i_{TT} / MU^i_{TC} = \frac{\partial \mathrm{U}_i}{\partial TT} / \frac{\partial \mathrm{U}_i}{\partial TC}$$

- In evaluation studies, it is most common to use a social VTT that is the same for everybody. The Danish official VTT for appraisal is approx. 91 DKK/hour for commuting and 373 DKK/hour for business (2020 values).

The logit model – an example

- This example is from San Francisco in the 70s when logit models were first applied for travel demand forecasting. The purpose of the study was to forecast demand for a new mode: BART.

- There are 4 modes in the before situation:
  - Drive alone (1)
  - Bus w. walk access (2)
  - Bus w. car access (3)
  - Carpooling (4)

- A logit model with linear-in-parameters utilities was estimated, see Train (2003).

The logit model – a simple example

Parameter	Estimate	t stat
Cost / wage (1-4)	-0.028	4.3
Car time (1,3,4)	-0.064	5.7
Transit in-vehicle time (2,3)	-0.026	2.9
Walk time (2,3)	-0.069	5.3
Transfer wait time (2,3)	-0.054	2.3
Number of transfers (2,3)	-0.105	0.8
Headway of first bus (2,3)	-0.032	3.2
Autos per driver with ceiling one (1)	5.00	9.7
Autos per driver with ceiling one (3)	2.33	2.7
Autos per driver with ceiling one (4)	2.38	5.3
Car alone dummy (1)	-5.26	5.9
Bus w. car assess dummy (3)	-5.49	5.3
Carpool dummy (4)	-3.84	6.36

The logit model – an example - comments

- We see that utility (hence probability) is decreasing in cost for all alternatives.
- We see that utility (hence probability) is decreasing in time attributes for relevant alternatives
- We see that more cars lead to higher probability for car choice
- All parameters are significant at the 5% level except number of transfers

- Note that the model is identified/normalised by assuming the ASC for alternative 2 (Bus w. walk access) equal to zero. Similar for the variables Autos per driver. The other attributes do not need normalisation since they vary across alternatives.

- Time aspects can be compared since they have the same unit. We see that

$$|\text{Walk}| > |\text{ car }| > |\text{ transfer wait }| > |\text{ headway }| > |\text{ transit IVT }|$$

The logit model – a simple example

- Since cost is divided by wage, VoT are reported as percentages relative to wage

	Estimate	t stat
Car time	227	3.2
Transit time	91	2.4
Walk time	243	3.1
Transfer wait time	190	2.0

Question – VTT!

- What is the Value of Travel Time in the following model for mode choice if cost is in kr and time in hours:

$$U_j = \alpha_j + \beta_t t_j + \beta_c \ln(c_j) + \varepsilon_j,$$

with $\beta_t = -1.2$, $\beta_c = -0.6$, travel cost is 30 kr, and P_{nj}=0.5?

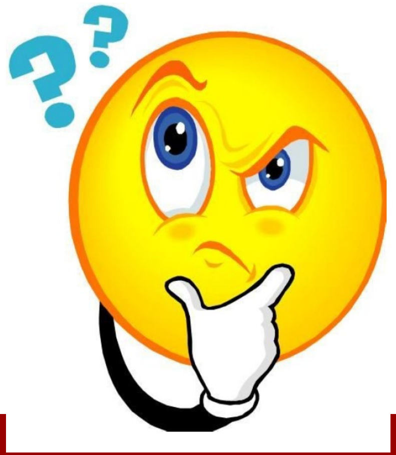
A. *90 kr/hr*

B. *60 kr/hr*

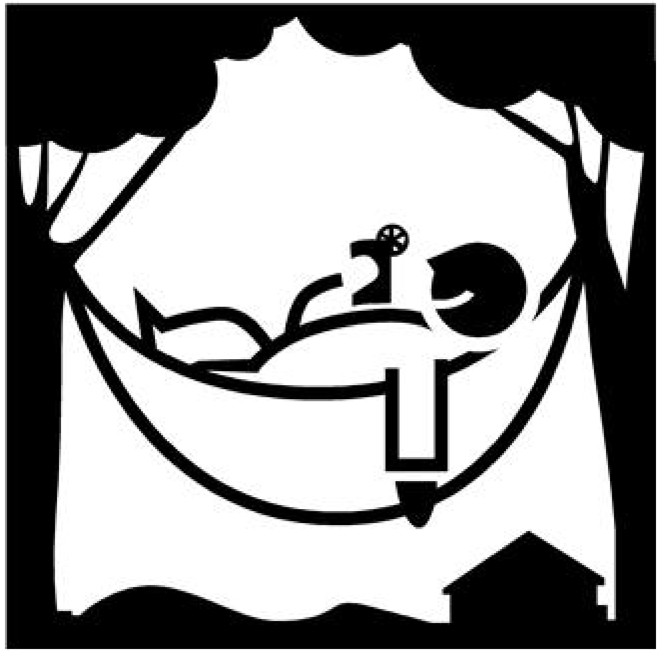
C. *30 kr/hr*

D. *2 kr/hr*

Answer:



Break



Question – t test!

- Assume that we have estimated an income paramter in a linear regression at $\hat{\beta} = 0.3$ with standard error $s = 0.09$. We would like to test if the parameter could be equal to $\beta_0 = 0$ using a t test. What expression is true?

z ~ N(0,1)	P(Z<z)
0.00	0.5
1.64	0.95
1.96	0.975
2.32	0.990
2.58	0.995

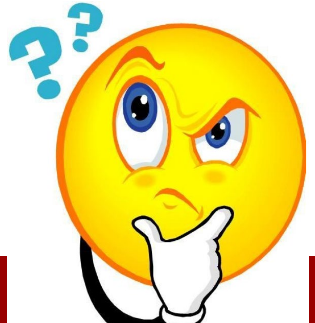
A. $\hat{\beta}$ is significantly different from 0.2 at the 10% level

B. $\hat{\beta}$ is significantly different from 0 at the 10% level but not at the 5% level.

C. $\hat{\beta}$ is not significantly different from 0 at the 10% level.

D. $\hat{\beta}$ is significantly different from 0 at the 5% level.

Answer:



Model evaluation

- Informal tests
  - Sign of parameters
  - Size of parameters
  - Ratios of parameters, e.g. VTT/VoT
  - Elasticities

- Overall goodness-of-fit measures
  - LL / AIC / BIC
  - (McFadden) rho squared and the adjusted version

- Statistical tests
  - t tests
  - LR tests

Differences 2 types of logistic regression vs. logit

- Logistic regression when seen as generalised linear models have the strength that we assume the Y variable to be Bernoulli distributed.

- Predictive logistic regression has the same probabilities but are not concerned with the distribution of error terms. These can achieve better predictive accuracy than logit.

- A main restriction of the logit model is that only variables relevant for each alternative can enter the utility function for that alternative. This restricts how well the model can fit the data. On the other hand, it improves interpretability.

- All three models can be generalised
  - Logistic regression (statistical) into e.g. generalised linear mixed-effects models
  - Logistic regression (machine learning) into e.g. neural networks
  - Logit into e.g. nested logit (and other GEV models) as well as mixed logit models

Estimating choice models

- If you estimate logistic regression, you may use
  - glm() in R
  - Statsmodels in Python
  - For both the family input should be binomial

- If you estimate choice models, you may use
  - The Apollo package in R, http://www.apollochoicemodelling.com/
  - The Biogeme panda in Python, https://biogeme.epfl.ch/

Week 12 exercise (20 min)

- Look at exercise week12 exercise 7_5
- The exercise includes a small data set with 10 observation.
- The purpose is that you set up a logit model to calculate probabilities, the value of time, and elasticities.
- You should not estimate anything so you do not need R/Python. You can do this exercise easily in Excel.

Summary of today's lecture

- Describe and discuss choice models based on random utility maximisation (RUM)
- Discuss logit (MNL) models
- Apply a logit model to analyse a choice situation
- Discuss how logit models compare to logistic regression and binary prediction models

Feedback

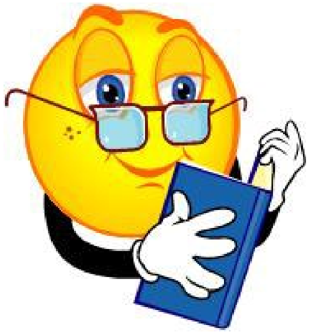
- Final questions

  1. What was the most interesting you learned during the lecture?

  2. What is your most important unanswered question based on the lecture?

- Group 8 (Jacob, Elmo, Emile, Bertram) should send their feedback to Stefan. Everyone else are very welcome to give feedback as well!

For next time



- Read for this week
  - Train chap. 2.1-2.3 + 3.1, 3.6, 3.8-9 about choice (logit) models
- Next week there will be a course summary + we will discuss a final paper.
  - Read paper 6 about how a question can be answered using various methods and how this affects the results
- Remember to evaluate your courses 17/4-7/5. Only six have completed it. Cake deadline 30/4.
- Exam plan and curriculum have been uploaded.
- Work on Project 3 (deadline 7/5).