Partial Exam SEN1221 Part 1. December 6 15:45-17:45

A electronics company considers to enter the smartphone market. To better understand the importance of attributes, such as Cost, Size, Memory storage, Camera quality and Operating System (OS) it has hired a high-end consultant to figure out how important these attributes are to consumers of different age groups and genders. Therefore, the consultant has conducted a Stated Choice experiment, in which participants faced 16 hypothetical choice tasks. The screenshot below shows one of the choice tasks. Besides the choice tasks, participants were asked about their age and gender. The data collection has just finished. In total 125, participants have completed the experiment.

	Alternative 1	Alternative 2	Alternative 3
Size [inch]	6.2	5.8	6.4
Storage [GB]	256	128	128
Camera qlt	3	1	4
OS	0	1	0
Cost	800	600	600
Choice	O	O	O

Coding scheme

The following coding scheme is used:

OS {0: Android, 1: iOS}

Camera quality {1: mediocre, 2: Good, 3: Very Good, 4: Excellent}

Age {1: Young, 2: Middle age, 3: Old}

Gender {0: Male, 1: Female}

Run this cell to create your environment locally

In [8]:
```python
# pip install .....
```

In [9]:
```python
# Import the libraries
# Do not change this code
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import biogeme.biogeme as bio
import biogeme.database as db
from biogeme import models
import biogeme.logging as blog
from biogeme.expressions import (Beta, log, exp, bioDraws, bioMultSum,MonteCarlo, Variable)
import toml
pd.set_option('display.max_columns', 500)
```

In [10]:
```python
# Set the number of draws in the .toml file to 150
# Do not change this code
with open('biogeme.toml', 'r') as file:
    tomldata = toml.load(file)

# Modify the number of draws
tomldata['MonteCarlo']['number_of_draws'] = 150

# Write the modified data back to the .toml file
with open('biogeme.toml', 'w') as file:
    toml.dump(tomldata, file)

# Create a logger to monitor the estimation progress
# if logger does not exist create it, else use it
try:
    logger
except NameError:
    logger = blog.get_screen_logger(level=blog.INFO)
```

In [20]:
```python
# Load the data in a long format
data  = pd.read_csv('data_partial_exam_long.csv', sep='\t' )
```

1 Is this a labelled or unlabelled experiment?

A. Labelled

B. Unlabelled

2 How is age distributed in the sample?

A. Approximately uniformly distributed

B. Approximately normally distributed

C. Young and old people are the most prevalent groups

D. None of the above

3 Are all 3 alternatives available to all decision makers in all observations?

A. Yes

B. No

Estimate a linear-additive RUM-MNL model [Model 1]

- Assume utility is linear and additive for the 5 attributes (hence, treat camera quality as a interval variable)
- Do not include the covariates (i.e AGE or GENDER) in your model

4 What is the final log-likelihood of this model?

A. -2197.23

B -974.54

C. -975.25

D. -1112.00

E. -972.01

F. None of the above

5 What is the correct interpretation of the rho sq of this model?

A. The rho square tells us that the data makes the model more likely than throwing a dice

B. The rho square tells us that this model is too simple to adequately explain the choice behaviour in the data

C. The rho square tells us how likely the data are

D. None of the above

6 The standard error associated with the betas for the OS and SIZE are larger than 0.05. This tells us that:

A. The OS and SIZE are not significant factors explaining cell phone choices in the population

B. There is substantial heterogeneity between people in their taste for the OS and SIZE

C. On average, people like Apple iOS more than Android OS and like larger phones better than smaller phones

D. None of the above

Estimate a new MNL model in which you interact the OS with the three age groups [Model 2].

- Use this model to infer whether there is a difference between age groups YOUNG, MIDDLE, and OLD regarding their tastes for the OS.
- Assume utility is linear and additive for all 5 attributes (hence, treat camera quality as a interval variable)
- Do not include any other covariates in the model than AGE

7 What is the final log-likelihood of the model

A. -974.55

B -1134.00

C. -950.95

D. -921.19

E. None of the above

8 Is the model with interactions statistically better than the model without interactions?
The Chi square table is supplied here

A. No

B. Yes, at 10% critical level of significance

C. Yes, at 5% critical level of significance

D. Yes, at 1% critical level of significance

9 Is there a difference in taste for the OS across the three age groups (Young, Middle, Old)?

A. Yes, the estimated betas for the OS are (significantly) different across age groups

B. No, the estimated betas for the OS are (almost) similar across all of the age groups

C. It is not possible to tell whether the estimated betas are different from each other across groups

Estimate a linear-additive PANEL Mixed Logit model [Model 3].

- Assume utility is linear and additive for all 5 attributes (hence, treat camera quality as a interval variable)
- Assume tastes for OS are normally distributed in the population: $\beta_{os}^{rnd} \sim N(\beta_{os}, \sigma_{os})$.
- For your convenience, we already prepared the data in a wide format (`data_partial_exam_wide.csv`)
- Note that the data set contains 16 choice observations per individual.

```
# Uncomment the lines below to load the data in a wide format
# df_wide = pd.read_csv('data_partial_exam_wide.csv',sep='\t')
# biodata_wide = db.Database('data_wide', df_wide)
# biodata_wide.data
```

10 What is the final log-likelihood of the Panel ML model?

Note that in the answers "+/-" means plus or minus 1 LL point

A. -967 +/- 1

B -2459 +/- 1

C. -970 +/- 1

D. -819 +/- 1

E. None of the above

11 Based on the results of the Panel ML model, what can you say about heterogeneity in tastes for the OS?

A. The fact that β_{os} is not significant tells us that people in the population don't care about the OS

B The fact that β_{os} is not significant while σ_{os} is significant tells us that only some people care about the OS

C. The fact that β_{os} is not significant while σ_{os} is significant tells us that some people prefer iOS while others prefer Android

D. The fact that β_{os} is not significant while σ_{os} is significant makes that we cannot say much about the heterogeneity of tastes for the OS

E. None of the above

12 Given the results of the three models that you have estimated so far, what is the 'best' next model to estimate?

A. A Panel Mixed Logit model which accounts for nesting effects. Thereby, we are able to uncover whether alternatives are correlated in terms of unobserved factors.

B. An MNL model in which we try to interact Gender with tastes, e.g. for size and camera quality.

C. A Panel Mixed Logit model in which we interact AGE and taste for OS. By combining the insights from Models 2 and 3 we can further refine our understanding of the importance of the OS to different age groups.

D. A fully connected MLP. Thereby, we can see how much variance is unexplained by the current models.