

Projektbeskrivelse til bachelorprojekt

Christian Heide^{a*},

^a Institut for Statskundskab, Aarhus Univeristet

* 202008359@post.au.dk

7. juli 2023

1 Introduktion

Dette er en projektbeskrivelse til mit bachelorprojekt i statskundskab med titlen *Uddannelsesvalg med bankkontoen i baghovedet? Eksperimentel evidens for en human kapital model om præferencer og adfærd ved valg af videregående uddannelse*¹.

I øjeblikket er der tale om en kodebog for mit hoveddatasæt, der findes i filen *Data.csv*. Det vil sige at dokumentet indeholder beskrivelser af variablene i datasættet med supplerende information, hvor jeg har vurderet det relevant. Ovenstående datasæt kan sammen med det medfølgende R-script bruges til at reproducere alle resultater i opgaven. Først følger dog en beskrivelse af datasættet. Tøv endelig ikke med at kontakte mig, hvis der er nogen spørgsmål eller bemærkninger i øvrigt!

2 Datasættet

Nærværende kodebog beskriver datasættet "Data.csv". Dette er et datasæt der er dannet pba. af datasættet "Data_Conjointly_RAW.csv". Datasættet er indsamlet ifm. mit bachelorprojekt, der undersøger præferencer for valg af videregående uddannelse. Indsamlingsperioden var foråret 2023,

¹<https://github.com/chr-heide/bachelorprojekt>

hvor et spørgeskema blev delt blandt gymnasieelever samt personer, der holdt sabbatår. Til indsamlingen blev survey-softwaren conjointly.com anvendt. Respondenter blev rekrutteret ved 1) gymnasielærere delte spørgeskemaet blandt deres elever, og 2) spørgeskemaet blev delt på sociale medier. I alt svarede 369 nuværende eller nylige gymnasieelever på spørgeskemaet.

To typer data blev indsamlet i spørgeskemaet, og kodebogen er opdelt efter disse. For det første blev indsamlet relevante baggrundsinformationer om respondenterne. Dette er informationer der kunne bruges til at vurdere stikprøvens repræsentativitet og til at lave relevante subgruppeanalyser.

For det andet blev indsamlet data i forbindelse med et såkaldt conjoint-surveyeksperiment. Kort beskrevet er dette et eksperiment hvor en respondent bliver præsenteret for to hypotetisk og tilfældigt konstruerede uddannelsesprofiler. De to uddannelsesprofiler varierer på forventet livstidsindkomst, arbejdsløshed, adgangskvotient, længde, åbenhed, placering og mulighed for velfærdsarbejde. Respondenten bliver herefter bedt om at vælge mellem de to uddannelsesprofiler. Hver respondent blev præsenteret for 5 af disse valgopgaver.

Formålet med eksperimentet var at estimere betydningen af forskellige karakteristika ved uddannelser for sandsynligheden for at vælge en uddannelse. For en nærmere beskrivelse af eksperimentet og teorien bag henvises til afsnit 3 i min bacheloropgave (samt litteraturen der henvises til i afsnittet), der findes på github-siden.

2.1 Struktur på datasæt

Selvom min stikprøve har 369 respondenter indeholder datasættet 3.690 observationer. Det skyldes at der er en observation for hver uddannelsesprofil en respondent blev præsenteret for. Da hver respondent blev præsenteret for 5 valgopgaver mellem 2 uddannelsesprofiler, er der 10 observationer

af hver respondent. Således er datasættet i langt format. *id* variablen der beskrives nedenfor identificere hver unikke respondent, hvorfor denne variabel indholder 369 unikke værdier.

3 Kodebog

3.1 Baggrundsvariable

id

En id-variabel der identificerer hver unikke respondent. Variablen er automatisk dannet af conjointly.

beskæftigelse

Kategorisk variabel der indikerer respondentens uddannelsesmæssige beskæftigelse.

- Værdier:

1.g

2.g

3.g

1.hf

2.hf

Sabbatår

Sabbatår

Kontinuert variabel der indikerer hvor mange sabbatår inklusive det igangværende respondenten har haft.

køn

Kategorisk variabel der indikerer respondentens køn.

- Værdier

Mand
Kvinde
Andet

storby

Kategorisk variabel der angiver om respondten bor i en af de fire store byer i Danmark (København, Aarhus, Odense og Aalborg).

- Værdier:

Uden for de fire store byer
De fire store byer

forældre_udd

Kategorisk variabel der indikerer det højeste fuldtførte uddannelsesniveau blandt respondentens forældre.

- Værdier:

Folkeskole
Gymnasial
Erhvervsuddannelse
KVU = Kort videregående uddannelse
MVU = Mellemlang videregående uddannelse
LVU = Lang videregående uddannelse
Ved ikke

karaktergennemsnit

Kontinuert variabel der angiver respondentens karaktergennemsnit fra gymnasiet.

studie_start

Dummy-variabel der angiver hvorvidt respondten på undersøgelsestidspunktet planlagde at starte på en videregående uddannelse i 2023.

- Værdier:

0 = Planlægger ikke at starte/ikke sikker

1 = Har søgt ind/planlægger at søge ind

studie_valg

String-variabel der angiver hvad respondenten planlægger at søge ind på, hvis respondten planlagde at søge ind eller allerede havde søgt ind på en uddannelse.

region

Kategorisk variabel der angiver hvilken region respondten går på gymnasie i. Er dannet pba. en gymnasie-variabel, der af anonymitetshensyn er taget ud af det publicerede datasæt.

- Værdier:

Nordjylland

Midtjylland

Syddanmark

Sjælland

Hovedstaden

højt_snit

Kategorisk variabel der angiver hvorvidt en respondents karaktergennemsnit er 9 eller under eller over 9. Er dannet pba. karaktergennemsnit-variablen.

- Værdier:

9 eller under

Over 9

3.2 Eksperimentelle variable

choice_set

Variabel der beskriver hvilke af de 5 valgopgaver i conjoint-eksperimentet, som observationen er fra.

- Variationsbredde: 1-5

profil

Variabel der beskriver hvilken af de to profiler som respondenterne kunne vælge mellem i den pågældende valgopgave observationen er fra.

- Variationsbredde: 1-2

valg

Dummy-variabel der indikerer hvorvidt den pågældende profil blev valgt

- Værdier
 - 0 = Ikke valgt
 - 1 = Valgt

livstidsindkomst

Kategorisk variabel der indikerer livstidsindkomsten for uddannelsesprofilen i mio. kr.

- Værdier:
 - 13 mio
 - 17 mio
 - 21 mio
 - 25 mio

arbejdsløshed

Kategorisk variabel der indikerer dimittendledigheden for uddannelsesprofilen.

- Værdier:

5%

10%

15%

20%

længde

Kategorisk variabel der indikerer uddannelsens længde i år.

- Værdier:

2 år

3,5 år

4 år

5 år

adgangskvotient

Kategorisk variabel der indikerer adgangskvotienten for den pågældende uddannelsesprofil.

- Værdier:

Alle optaget

Kun kvote 2

6

10

åbenhed

Kategorisk variabel der indikerer åbenheden i de jobmuligheder som uddannelsesprofilen giver

- Værdier:

Specifikke jobs

Mange jobs

velfærdsarbejde

Kategorisk variabel der indikerer om uddannelsesprofilen typisk giver mulighed for velfærdsjobs. Velfærdsjobs forstås som jobs relateret til produktion og regulering af velfærdsydelser for borgeren.

- Værdier:

Ikke mulighed for velfærdsjobs%

Mulighed for velfærdsjobs

placering

Kategorisk variabel der indikerer hvor uddannelsesprofilen kan læses.

- Værdier:

Uden for de fire store byer

I de fire store byer

target

Kategorisk variabel der indikerer hvorvidt indkomsttærsklen er opfyldt. Tærsklen er opfyldt, hvis 1) begge uddannelsesprofiler i valget har en livstidsindkomst på 21 mio. kr. eller over, eller 2) begge uddannelsesprofiler i valget har den samme livstidsindkomst.

- Værdier:

Ikke opfyldt

Opfyldt/lighed

target_variation

En alternativ tærskelvariabel der er dannet ligesom target-variablen, men hvor tilfælde hvor begge uddannelsesprofiler har den samme livstidsindkomst er ekskluderet.

- Værdier:

Ikke opfyldt

Opfyldt

arbejdsløshed_target

Kategorisk variabel med en alternativ tærskel for arbejdsløshed. Tærsklen er opfyldt i tilfælde hvor 1) begge uddannelsesprofiler har en arbejdsløshed på 10% eller under, eller 2) begge uddannelsesprofiler i valget har den samme arbejdsløshed.

- Værdier:

Ikke opfyldt

Opfyldt/lighed

åbenhed_target

Kategorisk variabel med en alternativ tærskel for åbenhed. Tærsklen er opfyldt hvis begge uddannelsesprofiler i valget har den samme grad af åbenhed.

- Værdier:

Lighed

Variation