

FORBEREDELSE TIL GRUPPEEKSAMEN I MET4 VÅR 22

I dette dokumentet vil dere finne en innledning til problemstillingen som vi skal se på ved årets gruppeeksamen i MET4. Som vedlegg til dette dokumentet vil dere finne en vitenskapelig artikkel som danner grunnlaget for problemstillingen, samt det eksakte datasettet som også blir utlevert med selve eksamensoppgavene 26. april 2022. Vi går også gjennom noen R-kommandoer som kan være nyttige for å jobbe med akkurat dette datasettet, samt en generell tilbakemelding på gruppeeksamen som ble gitt til hele kullet for et år siden.

Innledning

Samfunnsforskere har i lang tid vært opptatt av sammenhengen mellom demokratisering og tilgang til grunnskoleopplæring. På den ene siden kan vi tenke oss at demokratisk utvikling i et land gjør at flere samfunnslag, gjennom sin deltagelse i valg, krever tilgang til grunnleggende velferdsgoder som skolegang. På den andre siden kan vi tenke oss at autoritære regimer også har interesse av å øke andelen barn som går på skole siden klasserommet da kan brukes til indoktrinering, men også ganske enkelt fordi en del venstreorienterte diktaturer har basert sin legitimitet på å tilby tjenester, herunder utdanning, til fattige.

Vi kan også tenke oss at den demokratiske utviklingen som vi har sett i større eller mindre grad de siste par hundre år er en prosess som ikke henger så nøye sammen med økt grunnskoledekning.

Augustina S. Paglayan tar for seg dette problemet i artikkelen *The Non-Democratic Roots of Mass Education: Evidence from 200 Years*, som ble publisert i 2021¹. Paglayan tar for seg et rikt paneldatasett med informasjon om utdanningsdekning og demokratisering for 109 land over en periode på over 200 år. Hennes konklusjon er enkelt sagt at man ikke kan slå fast at mer demokrati fører til flere barn på skolen. Barna var i stor grad allerede på skolebenken da demokratiet ble innført.

Den aktuelle artikkelen er vedlagt sammen med et utvalg av datasettet som ble brukt². Etter å ha satt working directory til mappen der datafilen ligger kan du laste inn datasettet ved å kjøre

```
load("met4_v22.Rdata")
```

Se Tabell 1 for beskrivelse av de enkelte variablene i datasettet.

Konklusjonene til Paglayan henger i stor grad på modellen i ligning (1) i artikkelen. Dette er en såkalt “diff-in-diff”-modell som er standard i økonometrien for å estimere kausale sammenhenger i naturlige eksperimenter. I vårt tilfelle er “eksperimentet” at noen land er blitt demokratiske, mens andre land er forblitt autoritære. Demokratiseringen har også skjedd ved ulike tidspunkt og flere land har opplevd å gå fra demokratisk til ikke-demokratisk på et eller flere tidspunkt. Vi kan da se på *forskjellen i forskjellen* (“difference-in-difference”) i hvordan demokratiske lands grunnskoledekning utvikler seg før og etter demokratisering sammenlignet med ikke-demokratiske land. Denne demokratiseringseffekten svarer til parameteren δ_1 i ligning (1) i artikkelen.

¹Augustina S. Paglayan: *The Non-Democratic Roots of Mass Education: Evidence from 200 Years*. American Political Science Review (2021). Artikkelen er lagt ved.

²Hele datasettet til artikkelen er tilgjengelig på følgende adresse: <https://doi.org/10.7910/DVN/X2VJXX>. Det vedlagte datasettet er basert på filen `Education_LeeLee_Democracy_MAIN.tab`

Table 1: Variabelbeskrivelser

Variabel	Beskrivelse
<code>ccode</code>	Numerisk landkode
<code>abbrev</code>	Forkortelse av landnavn
<code>country</code>	Landnavn
<code>year</code>	Årstall for observasjon
<code>polity2</code>	En indeks publisert av 'The Polity Project' som måler graden av demokrati i landet i det aktuelle året. Skalaen går fra -10 til 10, der land med verdi over 6 regnes for å være demokatiske land
<code>year_since_suffrage</code>	Antall år <i>siden</i> det ble innført allmenn stemmerett for menn i landet. Negative tall gir dermed hvor mange år det er til stemmeretten blir innført.
<code>year_since_polity2</code>	Antall år <i>siden</i> <code>polity2</code> -variabelen passerte 6 første gang, og landet ble klassifisert som demokratisk. Negative tall gir dermed hvor mange år det er til terskelen blir passert første gang.
<code>primratio</code>	Andel av befolkning i landet som har tilgang til grunnskoleutdanning på barneskolenivå (Primær utdanning, primary education).
<code>secratio</code>	Andel av befolkning som har tilgang til utdanning på ungdomsskole og videregående nivå (Sekundær utdanning, secondary education).
<code>tertratio</code>	Andel av befolkning som har tilgang på universitetsutdanning (Tertiær utdanning, tertiary education).
<code>YearFirstPolityDemoc</code>	Årstallet da landet passerte 6 på Polity-indeksen første gang.
<code>democratic_polity</code>	Binær variabel som indikerer om landet er demokratisk i det aktuelle året, ved at <code>polity2</code> > 6. Legg merke til at indeksen for enkelte land svinger over og under grenseverdien på 6.
<code>Region</code>	Angir den geografiske regionen til landet, eventuelt om landet er klassifisert som en 'Avansert økonomi'.

Noen nyttige R-kommandoer

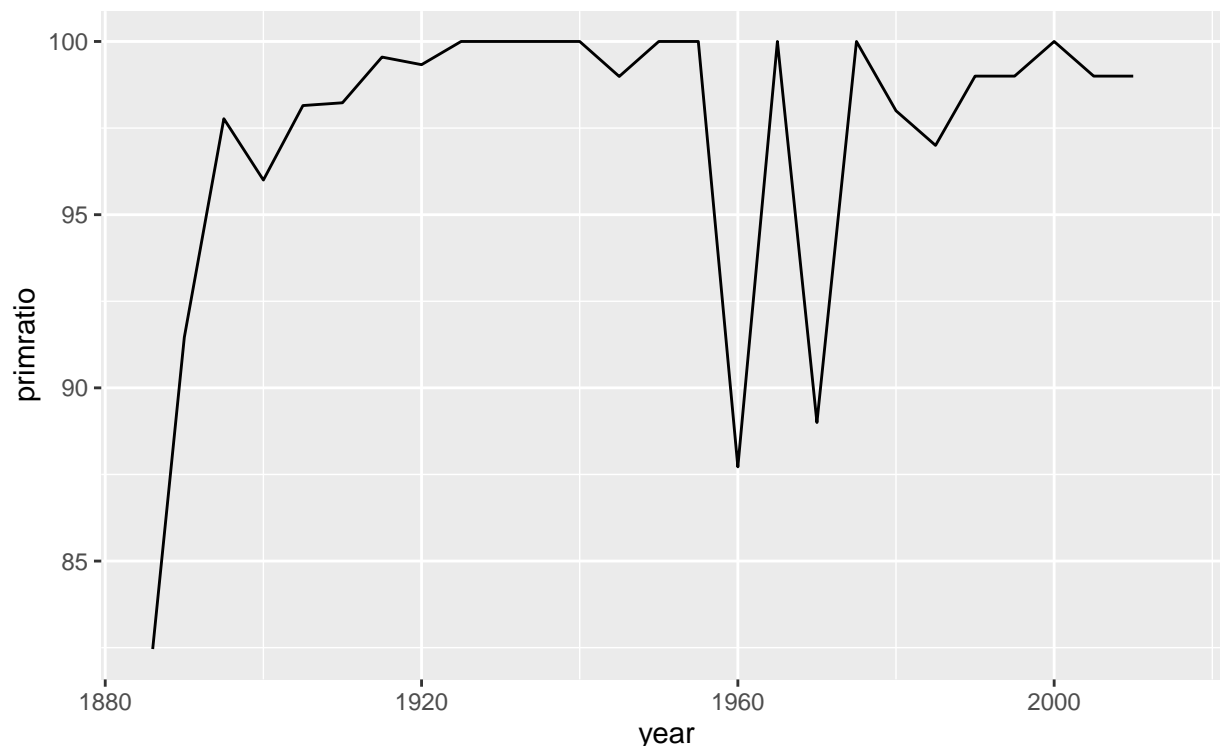
Datasettet er et klassisk paneldatasett med en rad for hver kombinasjon av år og land. Denne strukturen gjør at vi ganske sikkert får nytte av `filter()`-funksjonen som vi så på i R-introduksjonen vår. Vi får ut alle observasjonene for Norge, for eksempel, ved å filtrere på `abbrev`-variabelen:

```
dem_df %>% filter(abbrev == "NOR")
```

```
## # A tibble: 131 x 13
##   ccode abbrev country  year polity2 year_since_suffrage year_since_polity2
##   <dbl> <chr>  <chr>   <dbl>   <dbl>             <dbl>             <dbl>
## 1   385 NOR    Norway  1886     -2             -12             -12
## 2   385 NOR    Norway  1887     -2             -11             -11
## 3   385 NOR    Norway  1888     -2             -10             -10
## 4   385 NOR    Norway  1889     -2              -9              -9
## 5   385 NOR    Norway  1890     -2              -8              -8
## 6   385 NOR    Norway  1891     -2              -7              -7
## 7   385 NOR    Norway  1892     -2              -6              -6
## 8   385 NOR    Norway  1893     -2              -5              -5
## 9   385 NOR    Norway  1894     -2              -4              -4
##10   385 NOR    Norway  1895     -2              -3              -3
## # ... with 121 more rows, and 6 more variables: primratio <dbl>,
## #   secratio <dbl>, tertratio <dbl>, YearFirstPolityDemoc <dbl>,
## #   democratic_polity <lgl>, Region <chr>
```

Dette kan vi sende rett videre til `ggplot` ved hjelp av pipe-operatoren dersom vi ønsker å se utviklingen i grunnskoledekning for Norge:

```
dem_df %>%
  filter(abbrev == "NOR") %>%
  ggplot(aes(x = year, y = primratio)) +
  geom_line()
```



Når vi jobber med dette datasettet så kan vi være interessert i å regne ut gjennomsnitt av en variabel innad i grupper. For eksempel vil kanskje ha gjennomsnittlig grunnskoledekning for hver verdensdel i et gitt år. Da kan vi bruke to svært nyttige funksjoner i `dplyr`-pakken; `group_by()` og `summarise()`. Den første av disse funksjonene lager en gruppering i datasettet som alle påfølgende funksjoner i pipe-sekvensen vil respektere. `Summarise`-funksjonen er designet for å lage deskriptiv statistikk for et datasett. La oss demonstrere den raskt først. La oss si at vi ønsker å regne ut gjennomsnitt, standardavvik, min, og max for variabelen `primratio`. Vi kan selvsagt kjøre `mean(dem_df$primratio)`, `sd(dem_df$primratio)` etc. hver for seg, men vi kan også få ut en fin tabell ved hjelp av `summarise` slik som her (vi setter `na.rm = TRUE` for å ignorere manglende verdier):

```
dem_df %>%
  summarise(Gjennomsnitt = mean(primratio, na.rm = TRUE),
            Standardavvik = sd(primratio, na.rm = TRUE),
            Minimum = min(primratio, na.rm = TRUE),
            Maksimum = max(primratio, na.rm = TRUE))
```

```
## # A tibble: 1 x 4
##   Gjennomsnitt Standardavvik Minimum Maksimum
##   <dbl>         <dbl>    <dbl>    <dbl>
## 1      63.7         32.9      0      100
```

Men la oss gjøre det litt mer interessant. Vi ønsker å ta for oss kun observasjoner gjort i år 1990, og så ønsker vi *en linje med deskriptiv statistikk som over for hver region i verden*. Dette kan vi nok klare manuelt ved å lage en tabell som over for hver region i datasettet, men det er mye enklere hvis vi bare bruker `group_by()` til å fortelle at vi ønsker deskriptiv statistikk splittet opp på hver region:

```
dem_df %>%
  filter(year == 1990) %>%
  group_by(Region) %>%
  summarise(Gjennomsnitt = mean(primratio, na.rm = TRUE),
            Standardavvik = sd(primratio, na.rm = TRUE),
            Minimum = min(primratio, na.rm = TRUE),
            Maksimum = max(primratio, na.rm = TRUE))
```

```
## # A tibble: 6 x 5
##   Region                                Gjennomsnitt Standardavvik Minimum Maksimum
##   <chr>                                <dbl>         <dbl>    <dbl>    <dbl>
## 1 Advanced Economies                  95.7          5.55     75      100
## 2 Asia and the Pacific                 80.8          24.7     19.4     100
## 3 Eastern Europe                      86.5           7.05     76.4      97
## 4 Latin America and the Caribbean     86.7          15.8      26      100
## 5 Middle East and North Africa        87.5          15.7      57      100
## 6 Sub-Saharan Africa                  60.3          20.1      19      92
```

Den første funksjonen i sekvensen over er grei, den bare filtrerer ut alle observasjoner som ikke er gjort i år 1990. Den neste funksjonen *grupperer* observasjonene etter verdien av variabelen `Region`. Denne funksjonen forandrer ikke på selve innholdet i datasettet, den bare legger til metainformasjon om at alle funksjoner som kommer etterpå skal respektere gruppetilhørigheten som den definerer. Med andre ord: alt som skjer i pipe-sekvensen etter `group_by(Region)` skal skje gruppevis innen hver region. Deretter bare slenger vi på hva vi ønsker å gjøre, i dette tilfellet lage et sammendrag av `primratio`-variabelen som over, men nå med en linje per `Region`.

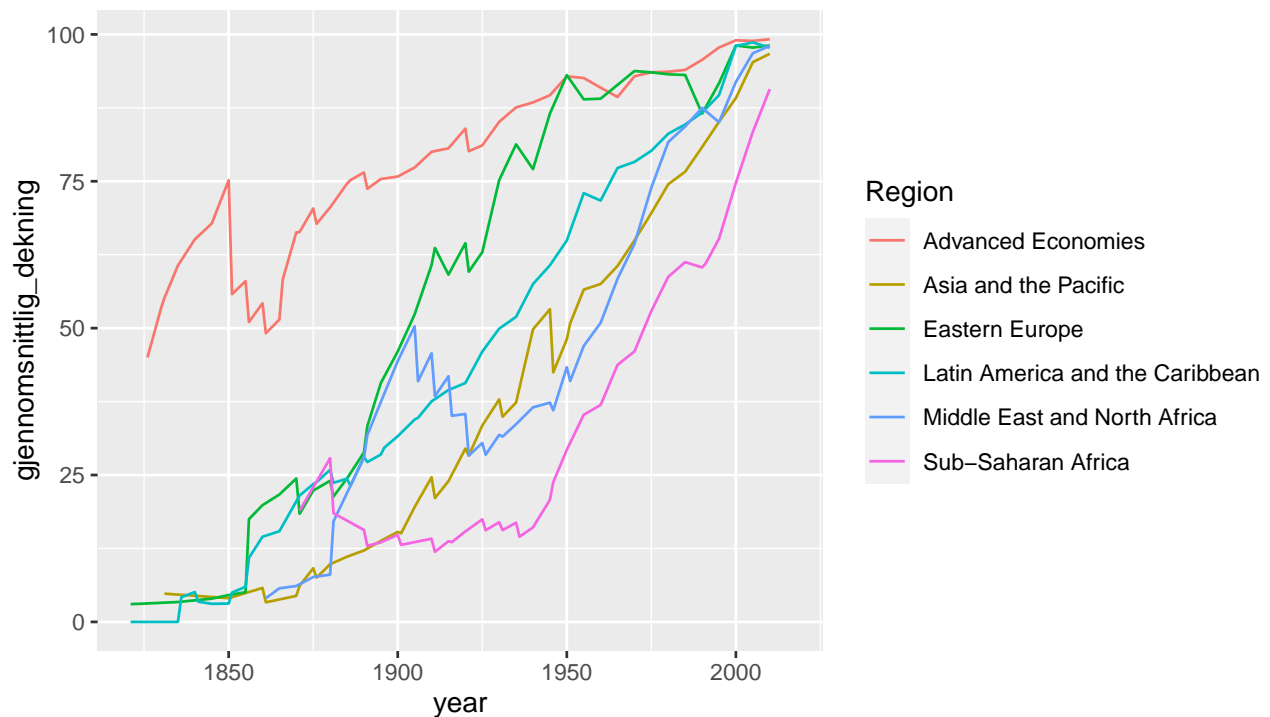
Vi kan gruppere på flere variabler også. Vi kan for eksempel se for oss et plott av *gjennomsnittlig grunnskoledekning per år per region*. Her har vi to grupperingsvariabler (`year` og `Region`), men det er ikke noe problem:

```
dem_df %>%
  group_by(Region, year) %>%
  summarise(gjennomsnittlig_dekning = mean(primratio, na.rm = TRUE))
```

```
## # A tibble: 1,071 x 3
## # Groups:   Region [6]
##   Region          year gjennomsnittlig_dekning
##   <chr>          <dbl>                <dbl>
## 1 Advanced Economies 1826                45.0
## 2 Advanced Economies 1827                47.1
## 3 Advanced Economies 1828                49.2
## 4 Advanced Economies 1829                51.3
## 5 Advanced Economies 1830                53.4
## 6 Advanced Economies 1831                55.1
## 7 Advanced Economies 1832                56.5
## 8 Advanced Economies 1833                57.9
## 9 Advanced Economies 1834                59.3
## 10 Advanced Economies 1835                60.7
## # ... with 1,061 more rows
```

Da får vi en lang data frame med et gjennomsnitt per år, per region. Denne kan sendes direkte videre til ggplot, der vi kan bruke for eksempel colour til å skille mellom regionene:

```
dem_df %>%
  group_by(Region, year) %>%
  summarise(gjennomsnittlig_dekning = mean(primratio, na.rm = TRUE)) %>%
  ggplot(aes(x = year, y = gjennomsnittlig_dekning, colour = Region)) +
  geom_line()
```



Hvis vi ønsker å gjøre klassiske hypotesetester på dette datasettet så kan det være nyttig å huske kommandoen pull for å gjøre om en data frame med en kolonne til en vektor. Grunnskoledekningen til alle demokratiske land i 1990, for eksempel, kan hentes ut som følger:

```
dem_df %>%
  filter(democratic_polity) %>%
  filter(year == 1990) %>%
  select(primratio) %>%
  pull
```

```
## [1] 98.00 98.00 26.00 75.00 100.00 90.00 93.00 71.00 76.00 87.00
## [11] 92.00 74.00 89.00 88.50 100.00 88.00 81.00 87.00 100.00 91.00
## [21] 97.00 91.00 95.00 96.00 75.00 100.00 93.00 100.00 97.00 89.00
## [31] 91.00 87.00 92.12 92.05 93.00 100.00 84.00 99.00 99.00 99.00
## [41] 96.00 52.00 92.00 100.00 100.00 76.52 37.95 100.00 98.00 100.00
```

Det kan også være aktuelt å *pare* observasjoner for å gjøre en statistisk test. Vi kan for eksempel hente ut grunnskoledekningen til alle land 5 år før de ble demokratisert første gang som følger:

```
dem_df %>%
  filter(year_since_polity2 == -5) %>%
  select(country, primratio)
```

```
## # A tibble: 78 x 2
##   country      primratio
##   <chr>         <dbl>
## 1 Canada          85.8
## 2 Haiti           44
## 3 Dominican Republic 72.2
## 4 Jamaica         100
## 5 Trinidad and Tobago 94.8
## 6 Mexico          100
## 7 Guatemala       67.7
## 8 Honduras        74.2
## 9 El Salvador     66.7
## 10 Nicaragua       76
## # ... with 68 more rows
```

Vi kan så gjøre det samme for å hente ut grunnskoledekningen fem år etter demokratiseringen ved å erstatte -5 med 5. Men finnes det en måte å sette disse to data framene sammen igjen der vi er sikker på at observasjonen fra det samme landet havner i samme rad? Ja, da kan vi bruke en funksjon som heter `full_join()`, også den i `dplyr`-pakken. La oss først lagre de to separate datasettene i hver sin variabel:

```
df1 <-
  dem_df %>%
  filter(year_since_polity2 == -5) %>%
  select(country, primratio)

df2 <-
  dem_df %>%
  filter(year_since_polity2 == 5) %>%
  select(country, primratio)
```

Vi kan så *join*e disse sammen igjen ved å be `full_join()` bruke variabelen `country`:

```
df3 <- full_join(df1, df2, by = "country")
df3

## # A tibble: 81 x 3
##   country      primratio.x primratio.y
##   <chr>         <dbl>         <dbl>
```

```
## 1 Canada                85.8      91.7
## 2 Haiti                  44        26
## 3 Dominican Republic    72.2     50.6
## 4 Jamaica               100       97.1
## 5 Trinidad and Tobago   94.8     92.4
## 6 Mexico                100       98.6
## 7 Guatemala             67.7     90.9
## 8 Honduras              74.2     89.4
## 9 El Salvador           66.7     69.4
## 10 Nicaragua            76       91.3
## # ... with 71 more rows
```

Her ser vi de to grunnskoledekningene for hvert land under hver sitt variabelnavn – `full_join()` la til `.x` og `.y` for å skille dem fra hverandre.

Til slutt kan vi nevne følgende distinksjon som kan være viktig når vi skal estimere regresjonsmodeller for dette datasettet. Legg merke til at tidsvariablene som for eksempel årstallet `year` er *numeriske* variabler i datasettet vårt. Hvis vi for eksempel ønsker å forklare variasjon i grunnskoledekning ved hjelp av tid på følgende måte:

$$Y = \beta_0 + \beta_1 t + \beta_2 X + \epsilon,$$

der X er en eller annen forklaringsvariabel og altså der β_1 representerer forventet økning i utdanning for hvert år som går, så kan vi bare bruke `year` som forklaringsvariabel i regresjonen på vanlig måte (X er bare et eksempel, vi har ingen variabler med det navnet i datasettet vårt):

```
reg1 <- lm(primratio ~ year + X, data = dem_df)
```

Dersom vi derimot ikke ønsker å anta at utviklingen i tid er lineær, kan vi heller estimere følgende modell:

$$Y = \phi_t + \beta_2 X + \epsilon,$$

der vi lar hvert år ha sin egen faste effekt ϕ_t (med andre ord, vi antar ikke at utviklingen over tid er like stor hvert år), så må vi enten bruke syntaksen i `plm`-pakker for estimering av faste effekter, eller så kan vi bare legge til variabelen `year` på samme måte som over; men da må vi huske å gi beskjed om at dette ikke lenger skal behandles som en numerisk variabel men som en kategorisk variabel der hvert år skal ha sin egen dummyvariabel. Det kan vi enkelt gjøre slik:

```
reg2 <- lm(primratio ~ as.factor(year) + X, data = dem_df)
```

Da blir derimot utskriften fra `summary()` veldig rotete siden den skriver ut estimert effekt ϕ_t for alle år i datasettet. Det kan være en idé å presentere resultatet på en annen måte, siden det gjerne bare er β_2 vi strengt tatt er interessert i.

Retteskjema

Vi gjør oppmerksom på at vi kommer til å bruke retteskjemaet som er lagt ut i fillageret på Canvas vil bli brukt til å bedømme besvarelsene.

Generell tilbakemelding etter Gruppeeksamen V21

Vi gav en generell tilbakemelding til kullet som tok gruppeeksamen i vårsemesteret 2021. Det kan være elementer der som er til hjelp også for dere inn mot eksamen. Tilbakemeldingen er gitt under:

Hei alle. Da er sensuren for den gruppebaserte hjemmeeksamen falt, og karakterene er publisert. Nivået ligger omtrent der det pleier med et snitt litt under B, som på alle måter må sies å være et godt resultat! En annen ting som er akkurat som det pleier er at ikke alle er fornøyd med karakteren sin. Vi har mottatt ganske mange forespørsler om begrunnelse, og de skal vi svare på i tur og orden, men jeg tenkte her å gi en mer generell tilbakemelding som alle kan dra nytte av i videre studier.

Forståelse av den spesifikke situasjonen:

Noe av det vi er mest fornøyd med i årets gruppeeksamen er at dere i stor grad klarte å vise forståelse for den konkrete situasjonen. Begrepet “customer analytics” er svært viktig i store deler av det private næringslivet, og det er på mange måter et åpent spørsmål hvordan vi kan dra nytte av kundedata for å forstå hvorfor kunder sier opp avtalene sine, og forhindre at det skjer. Dette handler om mer enn å bare sette opp en logistisk regresjon og se hvor godt den predikerer, det handler også om å forstå at det finnes viktige og ikke-trivielle avveininger av tilfredshet og kostnader knyttet til en kampanje rettet mot risikokunder. Videre fikk vi kanskje erfare at statistikken i dette tilfellet ikke gav oss en vidundermedisin for prediksjon, og at det ikke er lett å slå den mest naive løsningen: å bare la det være og ikke ringe til noen. Mange grupper viste god forståelse for dette gjennom nøktern analyse av de statistiske resultatene, og har blitt belønnet for det gjennom karakteren. Andre klarte i mindre grad å vise selvstendige vurderinger, og virket i større grad å gå på autopilot. Det kan være med å forklare en karakter som kanskje ikke stod til forventningene.

Kildekritikk:

Svært mange grupper presenterte følgende formulering av den logistiske regresjonsmodellen (eller en variant): $\log(y_i) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$. Dette er jo *galt*, og vi stusset lenge på hvor dette egentlig kom fra. Vi saumfarte videoer og forelesningsnotater i jakten etter synderen uten å finne en forklaring på at så mange grupper gjorde den identiske feilen. Besynderlig. Til slutt fant vi ut av det: eksempelbesvarelsen som fikk A har gjort akkurat det samme (rett før tabell 3). Ved nærmere ettersyn så viser det seg at flere grupper har lagt seg meget tett opp mot den besvarelsen, noe som kanskje er fristende gitt at eksamensoppgavene hadde store fellestrekk. Dette fremstår som svært lite sjarmerende for sensor, spesielt når man regelrett kopierer en feil. Hvis du følger deg truffet her, så vil jeg anbefale å ta med seg følgende læringspunkter i videre studier:

- Vi kommer ikke til å kjøre plagiatsaker på dette. Denne spesifikke feilen har heller ikke blitt brukt som eneste begrunnelse for å vippe noen ned fra en karakter til en annen. På samme måte som før er det totalinntrykket som til syvende og sist bestemmer karakteren.
- Det er ikke en unnskyldning at dette stod i en A-besvarelse; A betyr ikke perfekt, og i dette tilfellet kan det hende at feilen enten gikk under radaren til sensoren, eller så ble det satt på kontoen for “blingser” som vi selvsagt tillater i ellers sterke besvarelser.
- Vær kritisk til det du leser! Dobbelsjekk! Det er forskjell på kilder! Det finnes feil over alt! Dersom du skriver noe *galt* er det du som får svi for det, i alle fall dersom du ikke oppgir kilde.
- Lær deg bedre sitatskikk. En ting er å bli “inspirert”, men i min bunke var det flere besvarelser som absolutt burde referert til eksempelbesvarelsen, og som kanskje hadde måttet betale en mye høyere pris i et masterkurs for mangel på referanse. Ta dette som et vennlig råd på veien videre.

Deskriptiv statistikk:

Den deskriptive statistikken var under pari i år. Det var mange grupper som ikke egentlig forholdt seg til resten av oppgaven, som handlet om klassifisering. Da forventer vi for eksempel tydelige oversikter om hvordan variablene fordeler seg i de to klassene. Det var også muligheter til å finne noen rariteter i datasettet (i.e. negative priser) som burde bli tatt hånd om eller i det minste kommentert. En stargazer-tabell over alle variablene og et histogram eller to over noen av variablene fremstår som helt isolert fra resten av oppgaven hvis ikke er noen videre diskusjon som er relevant for det som kommer etterpå. Da har man misset poenget med deskriptiv statistikk, og må naturlig nok betale for det.

"Vi hadde alt rett, men fikk bare B (eller C)":

Dere er nå i alle fall over halvveis i bacheloren, og noen er allerede ferdige og skal videre på master. Enten vi liker det eller ikke, så må vi innse at det er på høy tid å oppdatere hvilke kriterier som gjelder for å hevde seg i toppen av karakterhierarkiet. MET4 er i grunn et godt bilde på en viktig overgang: vi jobber i større grad med *virkelige* problemer. Desverre har virkelige problemer en lei tendens til å være mye *vanskeligere*, og kanskje ikke på den måten at vi alltid trenger kompliserte modeller og lange formler. Nei, det er bare slik at den virkelige verden er mye mer nyansert enn fasiten bakerst i læreboken. Det finnes mange måter å angripe et problem på, men kanskje er det ingen måte som løser problemet fullstendig og for evig og alltid. Det er alltid noen nyanser som vi ikke klarer å ta hensyn

til, eller noen avveininger som gjør at vi kan komme et stykke på vei langs en dimensjon, men kanskje på bekostning av at et annet problem blir verre å hankses med. Med andre ord: det handler ikke bare om å regne riktig (selv om det er viktig, for all del), men om å jobbe seg frem til et svar som er så godt som mulig, som hjelper oss best mulig, og der vi samtidig har kontroll på hva vi ikke har klart å løse og hvilke konsekvenser det har.

Løsningsforslaget er ikke fullstendig. Det mangler fullstendig prosaen, argumentasjon for og mot, tolkning og diskusjon som trengs for å nå helt opp. Det inneholder noen kodelinjer som kan brukes som grunnlag for diskusjonen.

Universitets- og Høgskolerådet har generelle beskrivelser av karakterer for BØA-studiet (klikk for link) som kan være nyttige å kikke på. Blant annet ser vi der at man skal vise stor grad av selvstendighet for å få en A, og for B skal selvstendigheten og vurderingsevnen være meget god. De som lykkes godt i denne eksamensformen klarer nettopp dette, og det er klart at listen for de høyeste karakterene ligger høyere enn at man ikke har gjort direkte feil et sted. Dette er helt i tråd med Bloom's Taxonomy (klik for link) som viser at "remember" og "apply" er lavere læringsnivå enn for eksempel "analyze" og "evaluate". Jeg tror at ganske mange med stor fordel bør rekalkibrere hva som trengs for karakteruttelling i de videre studiene.

Noen leverer på et helt vilt høyt nivå

Jeg må avslutte med skryt! Det er noen grupper som leverer besvarelser som er helt eksepsjonelt gode! Vi har flere eksempler på besvarelser som analyserer problemstillingen på en nær profesjonell måte, som kommer frem til innsikter og konklusjoner basert på datasettet som er nye og genuint interressante, og som viser mye kreativitet og initiativ. Det er åpenbart at mange har brukt tiden godt gjennom semesteret og opparbeidet seg en sterk statistisk intuisjon som har kommet godt med på eksamen. Toppnivået på NHH er ulikt alt jeg har sett som ekstern sensor ved andre institusjoner her i landet.
