

i PSY2014 Eksamen 2021

PSY2014 - Kvantitativ metode

Informasjon om eksamen

- Åpen bok-eksamen 27. mai kl. 09.00 til 12.00
- Det er syv (7) faner med oppgaver, og du bør prøve å besvare alle
- Du trenger ikke å kjøre R/Rstudio under eksamen, alle nødvendige resultater blir vedlagt oppgavene
- En liste med relevante formler og en tabell med t-fordelingen er gitt på slutten av hvert vedlegg
- Du skal ikke laste opp håndtegninger
- Du trenger ikke å bruke referanser med mindre det er noe du vil referere til som er utenom pensum
- Eksamen skal være ditt eget, selvstendige arbeid, og et resultat av egen læring og arbeidsinnsats. Du kan ikke samarbeide med andre om denne eksamenen
- Du kan besvare eksamen på norsk, svensk, dansk eller engelsk
- Besvarelsen skal være anonym, ikke bruk navnet ditt i besvarelsen

Spørsmål under eksamen

- Hvis du har spørsmål under eksamen, må du sende e-post fra din UiO-mail til hjemmeeksamen@sv.uio.no.
- Merk emnefeltet med emnekoden til emnet.
- Dersom det blir gitt noe informasjon til alle kandidater under eksamen, så legges dette ut i Canvas. [Se varselinnstillinger i Canvas.](#)

Innlevering i Inspira

- [Les mer om eksamen og innlevering i Inspira.](#)
- Trykk på **Lever nå** for å levere besvarelsen.
- Etter eksamen finner du din innleverte besvarelse under **Arkiv**.

1 Ny oppgave

Oppgave 1: Visuell tracking og pupille-størrelse (60%)

I denne oppgaven skal vi analysere data fra et tenkt eksperiment. Deltagerne har som oppgave å følge en prikk på en pc-skjerm mens de blir distraheret av et antall andre bevegende prikker (distraktorer). Jo flere distraktorer det er på skjermen, jo større antas den kognitive belastningen å være. Mens oppgaven pågår måles størrelsen på pupillene til deltagerne. Pupillestørrelsen er ofte brukt som et implisitt mål på kognitiv belastning, og variabelen PUPILL inneholder den gjennomsnittlig pupillestørrelsen under oppgaven (målt i mm diameter). PUPILL vil være den avhengige variabelen i alle modellene i oppgave 1.

Følgende uavhengige variabler vil bli inkludert i regresjonsmodellene:

NDIST: Antall distraktorer (fra 0 til 15). Variabelen er sentrert på gjennomsnittet (8.4), dvs. $NDIST <- antall_distraktorer - mean(antall_distraktorer)$

STØRRELSE: Diameter på prikkene i mm.

ALDER: Deltagerens alder i år.

LANGSYN: Mål på langsynthet. Høyere verdier indikerer større grad av langsynthet.

FARGE: Fargen på de visuelle stimuliene (målet og distraktorene). 1: grå, 2: oransje, 3: grønn.

a) I modell 1 er NDIST den eneste uavhengige variabelen.

I vedlegget for modell 1 finner du R-utskrift fra funksjonene `anova()` og `confint()`. Bruk dette til å besvare følgende:

1. Hvor mange deltakere var det totalt i denne studien?
2. Hvor mye av variasjonen i pupillestørrelse kan forklares ved variasjon i antall distraktorer?
3. Hva kan du si om sammenhengen mellom NDIST og PUPILL?

Skriv ditt svar her

1. Det var 200 deltakere i denne studien. Frihetsgrader til TSS er $n-1$. Derfor blir $199 + 1 = 200$. Frihetsgradene til en statistikk er antall verdier statistikken er begrenset til å variere.

2. Forklart varians, også kalt η^2 her, regnes ut ved å ta SS_b/SS_{total} . Jeg regner derfor ut $134.80 + 344.49 = 0.3913030857$. Det vil si at antallet distraktorer forklarer ca. 39.1 prosent av variasjonen i pupillestørrelse.

3. Ettersom det er en korrelasjon på ca. 626 (kvadratroten av η^2 på 0.3913), ser man en sterk lineær sammenheng mellom NDIST og PUPILL. Det vil si at jo flere distraktorer man tar med, jo større vil pupillstørrelsen være i mm diameter.

Ut ifra modell 1 er det en svært signifikant sammenheng mellom NDIST og PUPILL, med en svært høy F-verdi på 127.29 og en veldig høy p-verdi. MS_w og MS_b (eventuelt MS_M og MSE) er forventningsrettede estimater på populasjonsvariansen, og forventede verdi på F-statistikken er derfor 1. Store verdier på F-statistikken taler imot H_0 . Nullhypotesen vil her være at dataene fra gruppene er trukket fra en normalfordeling med samme forventede verdi. Vi kan forkaste den.

Imidlertid er det ca. 60 prosent av den totale variasjonen som vi ikke kan forklare.

2 Ny oppgave

b) I modell 2 er NDIST og STØRRELSE uavhengige variabler i modellen.

1. Hva er tolkningen av de tre regresjonskoeffisientene i utskriften?
2. Hva er forventet pupillestørrelse når NDIST = 10 og STØRRELSE = 4?
3. I modell 2 blir standardfeilen til regresjonskoeffisienten for STØRRELSE (Std. Error) estimert til 0.02596. Hvordan forstår du dette tallet? Dersom du ledet dette eksperimentet, hvilke grep kunne du ta for å gjøre standardfeilen mindre ?

Skriv ditt svar her

1. Ettersom variabelen NDIST er **sentrert**, blir forventet b0 for PUPILL et direkte estimat på forventet pupillstørrelse hos en deltaker som får et gjennomsnittlig antall distraktorer (8,4). Denne verdien er svært signifikant med en svært høy p-verdi.

Jeg tolker NDIST for å være at etter som NDIST øker med 1 over gjennomsnittet, så vil gjennomsnittlig pupillstørrelse øke med 0.19448. Dersom NDIST minsker med 1 under gjennomsnittet, vil det minske 0.19448. Denne verdien er også svært signifikant.

Jeg tolker STØRRELSE her som hvor stor effekt på pupillstørrelse størrelsen på prikkene har i mm diameter. For hver millimeter denne øker, øker pupillstørrelse med 0.0548. Denne er signifikant på et alfanivå på 0.05.

Disse har en samlet forklart varians på ca. 40 prosent.

2. Jeg tolker her at NDIST = 10 betyr at NDIST er 1,6 verdi over gjennomsnittet og at det er dette jeg skal gange med. NDIST = 10 over gjennomsnittet vil være over rangen vår og ikke gi mening.

$$\text{Forventet } Y = 5.7388 + 0.19448 \times 1,6 + 0.05482 \times 4$$

$$\text{Forventet } Y = 6,269248$$

Forventet pupillestørrelse blir derfor ca. 6,27.

3. Standardfeilen til STØRRELSE er et estimat på variabiliteten (standardavviket) i stigningstallet du ville ha sett på tvers av forskjellige utvalg tatt fra den samme populasjonen. For å regne ut T-verdien, som man bruker for å estimere hvor signifikant resultatene sine er, tar man estimatet delt på standardfeilen. Jo høyere standardfeilen er, jo mindre signifikant vil resultatet være.

Det er tre forhold som påvirker standardfeilen, og som et resultat signifikans. For å minske standardfeil ville jeg øke antall observasjoner, ettersom dette gir mer statistisk styrke. Standardfeilen går ned etter som antall observasjoner øker (summen i telleren blir større). Man kan også gjøre summen i telleren mindre ved at innflytelsen av tilfeldige målefeil kan minimeres, som jeg kanskje kunne gjort ved å bruke et mer reliabelt måleinstrument eller tatt med andre variabler som kan stå for en større del av den uforklarte variasen rundt regresjonslinjen. Dette vil så gi lavere p-verdier. Det siste jeg kunne gjort var å sørge for at det er betydelig varians i STØRRELSE, ved å inkludere stor variasjon i størrelse på diameterene.

3 Ny oppgave

c) I modell 3 er NDIST og NDIST² uavhengige variabler.

1. Forklar hvorfor det er en god ide å legge til NDIST² (NDIST²), gitt plottet av residualene fra modell 1 og utskriften fra anova/AIC funksjonene som sammenlikner modell 1 og modell 3.
2. Beskriv kort sammenhengen mellom pupillestørrelse og antall distraktorer basert på resultatene fra modell 3.
3. Modellen kan brukes til å predikere pupillestørrelse når det er 30 distraktorer på skjermen? Ville du stole på dette tallet (begrunn svaret)?

Skriv ditt svar her

1. Vi sjekker rimeligheten av antagelsene i lineær regresjon ved å inspisere diagnostiske plot av residualene. Her ser man et spredningsplot. I et spredningsplot med residualer ønsker man å se ingen mønstre. Imidlertid ser man i plottet av residualene fra modell 1 at det tyder på at residualene er korrelerte. Dette vil si at antakelsen for lineær regresjon om at observasjonene er uavhengige er brutt. Ved å legge til NDIST kvadrert, kan man se om det er en ikke-lineær sammenheng her.

Anova-funksjonen ser man at under RSS, så har modell 3 mindre uforklart varians. Dette vil si at modell 3 forklarer mer varians enn modell 1. Dette kan tyde på at det er en smart å legge til NDIST².

AIC - Akaike information criterion er fra informasjonskriterier, som er en klasse statistikker som er utviklet for å optimalt balansere mellom **over og undertilpasning**, tilpasning og kompleksitet. Modellen som velges er den som har lavest verdi på informasjonskriteriet. Av AIC funksjonen ser man at modell 3 har fått lavere verdi enn modell 1. Dette tyder på at modell 3 er bedre tilpasset, og har en bedre balanse mellom å undertilpasse og overtilpasse. Trolig er modell 1 en undertilpasning av dataen.

2. Sammenhengen mellom pupillstørrelse og antall distraktorer basert på modell 3 er for det første svært signifikant. For hver gang NDIST øker med en distraktor, og øker ca. 0.2, øker det også med antallet opphøyet i andre. Dette vil si at jo flere distraktorer det blir, jo mer vil pupillestørrelse øke for hver distraktor som blir lagt til.

3. Jeg ville ikke stole på dette tallet. Man skal ikke predikere utenfor rangen til den uavhengige variabelen (**ekstrapolere**), som her er fra 0-15. I tillegg er det blitt brukt en polynomisk modell, som gjør ekstrapolering enda verre. **Polynomiske modeller har begrenset gyldighet utenfor rangen av observerte verdier av den uavhengige variabelen.**

4 Ny oppgave

d) I modell 4a legger vi til ALDER og i modell 4b legger vi også til LANGSYN.

1. Det er et problem med modell 4b. Forklar hva problemet består i, og hvilke følger det kan få for de estimerte koeffisientene til ALDER og LANGSYN.

Skriv ditt svar her

Problemet i modell 4b er **kollinearitet**. Kollinearitet er graden av lineær sammenheng mellom flere forklaringsvariabler i en multippel regresjonsmodell. Blir denne for høy (korrelasjon >0.8) kan det medføre problemer. **Blir kollinearitet for stor, blir ikke stigningskoeffisientene til å stole på grunnet stor økning i standardfeilene. Videre blir det vanskelig å vurdere hvilken av de uavhengige variablene som er av betydning.** Høye VIF-verdier (over 4 eller 5) tyder på kollinearitetsproblemer. VIF tallfester hvor mye en uavhengig variabel øker standardfeilen. Her ser man VIF verdier på 8 for både ALDER og LANGSYN. Det er derfor kollinearitetsproblemer for denne modellen.

Man ser her at det er en for stor korrelasjon mellom ALDER og LANGSYN. Når alder er lagt til alene, er den signifikant og påvirker ikke de andre variablene betydelig, men når LANGSYN er lagt til, øker ALDERs standardfeil og ALDER blir insignifikant. På statistisk grunnlag blir det her vanskelig å vite hvilken av variablene som er av betydning. I tillegg fører det til at begge blir statistisk insignifikante.

Besvart.

5 Ny oppgave

e) Den siste modellen inneholder de uavhengige variablene ALDER, STØRRELSE og FARGE.

1. Vurder påstanden «*alder er sterkere assosiert med pupillestørrelsen enn størrelsen på prikken er*».
2. Vurder påstanden «*pupillen er mindre når prikken er oransje enn når den er grå*» .
3. Hvilke variabler ville du lagt til modellen dersom du var interessert i å vurdere i hvilken grad sammenhengen mellom alder og pupillestørrelse avhenger av kognitiv belastning (antall distraktorer)?

Skriv ditt svar her

1. Påstanden «*alder er sterkere assosiert med pupillestørrelsen enn størrelsen på prikken er*» vurderer jeg som sann. Dette er ut ifra at når koeffisientene er standardisert, har ALDER en større koeffisient på ca. 0.28, i forhold til STØRRELSE sin koeffisient på ca. 0.13. Standardiserte koeffisienter gjør variablene sammenlignbare. Uten å være standardisert, er ikke koeffisientene ALDER og STØRRELSE på noen felles skala, som gjør at de vanskelig kan sammenlignes med kun koeffisientene. Når de standardisere, ved at antall enheter endring i Y per enhet økning i X, endres til antall standardavvik endring i Y når X øker tilsvarende ett standardavvik, blir disse på samme skala, og kan sammenlignes (simplifisert sett). Imidlertid er ikke STØRRELSE statistisk signifikant, så man kan ikke forkaste H_0 om at denne effekten ikke har kommet av en tilfeldighet. Derimot kan man argumentere for at p-verdien til STØRRELSE gjør den mer trolig signifikant, og at man derfor kan diskutere ut ifra den.

2. Påstanden «*pupillen er mindre når prikken er oransje enn når den er grå*» vurderer jeg for sann. Fargen grå er 1 i b0, og fargen oransje er FARGE 2, som man ser redusere verdien med ca. 0.16. Farge har blitt kodet som dummy variabler (indikator variabler). Disse tar bare verdiene 0 eller 1. Her har grå blitt kodet som 0 for både farge 2 og 3.

Man ser at oransje har en p-verdi på 0.46, som vil si at det er en 46 prosent sjanse for at denne effekten har skjedd ved en tilfeldighet. Derfor kan man ikke forkaste nullhypotesen. Farge 3 er heller ikke signifikant. Dette får meg til å vurdere om farge i seg selv i det hele tatt har en signifikant effekt. Dersom man rekoder og ser om grå har en signifikant effekt, kan man finne ut av dette. Hvis dette er tilfellet, er ikke påstanden gyldig for å vurderes ut fra dette data-settet.

3. Her er man ute etter å se en interaksjonseffekt. Med statistisk interaksjon mener vi at effekten av en variabel varierer over nivåene på en annen variabel. Jeg ville lagt til variablene NDIST og ALDER*NDIST. Et spørsmål til diskusjon her ville vært om å ta med $NDIST^2$ hadde vært nødvendig. Vi har sett at en modell med NDIST og $NDIST^2$ tilpasser seg dataen bedre. Imidlertid har jeg ikke blitt lært om man kan se på interaksjonseffekten til en variabel som er kvadrert. I så fall lurer jeg på om interaksjonseffekten hadde sett ut som dette: $NDIST*ALDER*NDIST^2$ eller om det hadde vært med $NDIST*ALDER$ og $NDIST^2$ hadde vært for seg selv.

Besvart.

6 Ny oppgave

Oppgave 2: Tetris som behandling av uønskede visuelle minner. (25%)

I denne oppgaven skal vi se på data fra en studie der Tetris (et visuelt krevende dataspill som består i å ordne geometriske figurer på bestemte måter) ble benyttet for å påvirke i hvilken grad deltagerne opplevde uønskede visuelle minner fra en film de tidligere hadde sett. Deltakerne ble delt i tre like store grupper som hver så en av tre filmer, 1: En komedie, 2: En skrekkfilm, 3: Et drama. Halvparten av deltagerne spilte så Tetris i 20 minutter samtidig som de tenkte på filmen de hadde sett. Resten av deltagerne tenkte på filmen mens de satt alene i rolige omgivelser. Neste dag blir deltagerne spurt om hvor mange ganger de har opplevd uønskede forestillingsbilder fra filmen etter at de så den.

Vedlagt finner du R-utskrift fra en variansanalyse.

1. Fyll inn de sladdede verdiene i tabellen.
2. Hvilke konklusjoner ville du trekke på bakgrunn av R-utskriften?
3. Under tabellen finner du tre figurer (A,B, og C). Hvilken av dem kan være et plot av dataene fra denne studien (begrunn svaret).

Skriv ditt svar her

1. Mean Sq = Sum Sq/Df.

$31.7/2=15.85$.

F-verdi = Mean Sq Filmsjanger:Tetris/Mean Sq Residuals.

$15.85/2.67=5.936329588$

2. Nullhypotesen er at dataen er trukket fra en normalfordeling med samme forventet verdi (gjennomsnitt). Alle gruppene i analysen viser seg å være statistisk signifikant, som gjør at jeg forkaster H_0 for alle gruppene.

Man ser at forklart varians (η^2) for filmsjanger er ca. 19,3 prosent, for tetris er det ca. 3,4 prosent og i interaksjonene mellom filmsjanger og tetris er det ca. 6 prosent. Med η^2_{sq} part, altså partial eta squared er disse verdiene høyere. Partial eta squared er variansen forklart av en variabel av den gitte variansen som er igjen etter man ekskluderer varians forklart av andre prediktorer.

Ut ifra resultatene ville jeg trukket en konklusjon om at filmsjanger har stor effekt på hvilken grad deltakerne opplevde uønskede visuelle minner etter filmen. Videre ville jeg konkludert med at å spille tetris etter å ha sett på film har en liten signifikant hovedeffekt. Deretter ville jeg fått fram at disse variablene avhenger av hverandre, og at det er en signifikant interaksjonseffekt mellom filmsjanger og å spille tetris etterpå.

3. Plot A kan være et plot av studien. Dette er fordi at man her ser en tydelig interaksjonseffekt, hvor effekten av å spille tetris etter man har sett filmen kommer an på hvilken film man har sett, her ser man en spesiell effekt når man har sett en skrekkfilm. Plot B har en mindre interaksjonseffekt, hvor det er nesten som om det kun er en hovedeffekt av filmsjanger og tetris, men en veldig liten interaksjonseffekt. I Plot C er det nesten kun en hovedeffekt. Jeg vil konkludere med at Plot A kan være et plot av dataen i denne studien.

Besvart.

7 Ny oppgave

Oppgave 3: Holdninger til firedagers arbeidsuke. (15%)

I denne oppgaven skal vi analysere holdninger til firedagers arbeidsuke blant folk i tre forskjellige sektorer, 1: Helsearbeidere, 2: Offentlige saksbehandlere, og 3: Håndverkere. Deltagerne i studien svarte på spørsmålet «Ønsker du at din arbeidsplass går over til en fire dagers arbeidsuke?»

1. Hvilke konklusjoner ville du trekke om hvorvidt ansatte i ulike yrker ønsker å gå over til firedagers arbeidsuke?

Skriv ditt svar her

H_0 her er at holdning for 4 dagers arbeidsuke og hvilken "arbeidssektor" man er i er uavhengige variabler.

Antall frihetsgrader regnes ut ved $(rader-1)(kolonner-1)$. Dette blir $2 \times 2 = 4$. Det er altså fire frihetsgrader.

X-squared verdien er en kvatifisering av relativ avstand mellom observert og forventet frekvenstabell. For å forkaste H_0 når man har fire frihetgrader, med kritisk verdi på 0,05, trenger man en χ^2 -kvadrat verdi 9.49. X-squared verdien på 23 forkaster nullhypotesen selv når man har en kritisk verdi på 0.005. Man kan altså forkaste H_0 på dette grunnlaget.

For å forstå forholdet mellom variablene, må vi se i tabellen for standardiserte residualer. Sammenligning av de observerte og forventede frekvensene lar oss teste hypotesen om uavhengighet. Forskjellen mellom f_0 og f_e , altså differansen mellom observert og forventet frekvens kalles et residual. Man standardiserer residualet for å gi det mer mening. Dette gjøres ved å dele $f_0 - f_e$ på standarderror.

Tabellen viser at Gruppe 1, 2 og 3 avviker betydelig fra hverandre. Håndtverkere blir observert å svare langt mer ja enn forventet (3.9), langt mindre kanskje enn forventet (3.5), og men avviker ikke like mye i hvor mye de sier nei som er forventet. Saksbehandlere svarer ja langt mindre enn er forventet (-4.1), langt mer kanskje enn forventet (3) og litt mer nei enn forventet. Helsearbeiderne avviker derimot ikke så mye fra det som er forventet, men svarer litt mindre nei enn forventet (-1). Alt dette tyder på at man kan forkaste H_0 . Det vil si at hvilken holdning man har til en 4 dagers arbeidsuke er ikke uavhengig av hvilken "arbeidssektor" man er i.

Det er viktig å få fram at en χ^2 -kvadrat test ikke sier noe om typen eller styrken på sammenheng her.

Besvart.