

## 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.  $NDIST <- \text{antall\_distraktorer} - \text{mean}(\text{antall\_distraktorer})$

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?  
Dette kan leses fra `anova` tabellen.  $Df(TSS) = df(SSM) + df(SSE) = 1 + 198 = 199$ . Videre er  $df(TSS) = n - 1$ , og det følger av det at  $n = 200$ .
2. Hvor mye av variasjonen i pupillestørrelse kan forklares ved variasjon i antall distraktorer?  
Dette kan regnes ut av ANOVA tabellen.  
 $R^2 = 1 - (SSE/TSS) = 1 - 209.69 / (209.69 + 134.80) = 0.39130 = 39.1\%$
3. Hva kan du si om sammenhengen mellom NDIST og PUPILL?  
Konfidensintervallet viser at vi kan forkaste  $H_0$  ( $H_0: b_1 = 0$  i populasjonen vi har trukket dataene fra) på et 0.05 nivå.
  - $R^2$  kan gjere nevnes i denne sammenheng (NDIST kan forklare 39.1% av variansen i dataene).

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

1. Hva er tolkningen av de tre regresjonskoeffisientene i utskriften?  
Dette er en multipl regressjonsanalyse, så det er viktig at det kommer frem at koeffisientene for de to uavhengige variablene er partielle regresjonskoeffisienter, og må tolkes som *forventet endring i den avhengige variabelen når den uavhengige variabelen endres en enhet, men den andre holdes konstant*.
  - $\hat{b}_0$ : Forventet pupillestørrelse når  $NDIST = 0$  (ved gjennomsnittlig antall distraktorer, 8.4), og  $STØRRELSE = 0$ . (Strengt tatt kan størrelse ikke være 0).

- $\hat{b}_1$ : Forventet forskjell i pupillestørrelse mellom to eksperimentelle betingelser som skiller seg med én distraktor.
  - $\hat{b}_2$ : Forskjell i forventet pupillestørrelse mellom to betingelser som skiller seg med én enhets størrelse (1mm i diameter).
2. Hva er forventet pupillestørrelse når NDIST = 10 og STØRRELSE = 4?  
 Forventet pupillstørrelse =  $5.73880 + 0.19448 \cdot 10 + 0.05482 \cdot 4 = 7.90288 \approx 7.9$ .
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?  
 Det bør fremkomme av svaret at standardfeilen er standardavviket i samplingfordelingen til regresjonskoeffisienten. Det vi ser i utskriften er et estimat på det standardavviket.  
 Særlig tre forhold vil kunne påvirke standardfeilen:
1. Utvalgsstørrelsen (mindre SE ved større utvalg)
  2. Variansen til STØRRELSE (større varians vil gi mindre SE)
  3. Varians rundt regresjonslinjen. Denne er en funksjon av målefeil i pupillstørrelsen og innflytelsen av andre umålte variabler. Større varians vil gi større SE. Den kan minimeres ved å få et mer reliabelt mål på pupillstørrelse, eller inkludere andre uavhengige variabler som kan redusere den uforklarte variansen.

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

1. Forklar hvorfor det er en god ide å legge til NDIST2 ( $NDIST^2$ ), gitt plottet av residualene fra modell 1 og utskriften fra anova/AIC funksjonene som sammenlikner modell 1 og modell 3.  
 Vi ønsker ikke å se noen form for struktur i plottet av residualene. Det er en rimelig klar kvadratisk form på residualene i plottet.  
 Anova tabellen viser at det er en signifikant bedring i tilpasningen når vi legger til en kvadratisk effekt av NDIST. AIC-verdien (AIC er et informasjonskriterie som balanserer modelltilpasning mot modell parsimoni, og slik skal hjelpe oss unngå overtilpasning) er minst for modell 3, noe som favoriserer modell 3. Alle indikasjoner peker derfor i retning av at modell 3 er bedre enn modell 1.
2. Beskriv sammenhengen mellom pupillestørrelse og antall distraktorer basert på resultatene fra modell 3.  
 Her er det viktig at det fremkommer at modellen antyder at forholdet mellom NDIST og PUPILL *ikke er helt lineært*, men kan bedre beskrives bedre som en annengradsfunksjon av NDIST. Størrelsen på pupillene øker relativt lite med få distraktorer, men sammenhengen øker i styrke ettersom antall distraktorer øker.
3. Modellen kan brukes til å predikere pupillestørrelse når det er 30 distraktorer på skjermen? Ville du stole på dette tallet (begrunn svaret)?  
 Vi må være forsiktig med å bruke modellen til å predikere verdier langt utenfor den rangen vi har observasjoner (ekstrapolasjon). Maks antall distraktorer i eksperimentet er 15, så 30 ligger langt utenfor, og det er ingen garanti at modellen er gyldig for alle verdier av NDIST. Dette er særlig problematisk ved polynomregresjon, da modellen tilsier at kurven fortsetter å krumme oppover, slik at pupillstørrelsen blir i økende grad ekstrem for store verdier på NSIDT.

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

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.

Problemet er kolinearitet. VIF verdiene antyder at ALDER og LANGSYN er høyt korrelerte, og dette fører til en voldsom økning av standardfeilen. I modell 4a er ALDER sterkt signifikant, mens i modell 4b er verken ALDER eller LANGSYN signifikant. Det mest rimelige å gjøre her er å ta en av variablene ut av modellen.

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».  
Viktig her at det vises til de standardiserte koeffisientene. Den standardiserte koeffisienten for ALDER er større enn koeffisienten for STØRRELSE, noe som støtter påstanden (selv om vi ikke har signifikanttestet forskjellen).
2. Vurder påstanden «pupillen er mindre når prikken er oransje enn når den er grå» .  
Grå er referansekategori. Koeffisienten  $\text{factor(FARGE)}^2$  er derfor hvordan estimatet for forventet forskjell fra å gå fra grå til oransje. Siden den estimerte verdien er negativ (-0.158779) er det en viss støtte for påstanden, men p-verdien er ikke mindre enn 0.05, så strengt tatt er det ikke rimelig å hevde at påstanden er sann.
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)?  
Her ønsker jeg i utgangspunktet at kandidaten skal forstå at vi leter etter en interaksjon. I så tilfelle er det nødvendig å legge til variablene NDIST og  $\text{NDIST} \cdot \text{ALDER}$ . Vi vet riktignok at det er en ikke lineær sammenheng mellom NDIST og PUPILL, så det er ikke urimelig å legge til  $\text{NDIST}^2$  og  $(\text{NDIST}^2) \cdot \text{ALDER}$ , men jeg ville ikke gi trekk for ikke å ha tenkt på det.

Noen kan forstå oppgaven dithen at vi ønsker å kontrollere for NDIST. Dette kan gi noe uttelling, men bør ikke være et fullstendig svar på denne oppgaven.

## 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 deltagere 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.

```
> summary(ANOVA1<-aov(Tanker~Filmsjanger*Tetris, data=DAT))
              Df Sum Sq Mean Sq F value    Pr(>F)
Filmsjanger    2  103.8    51.91   19.416 3.42e-08 ***
Tetris          1   18.2    18.17    6.797 0.01009 *
Filmsjanger:Tetris  2   31.7    15.83    5.920 0.00338 **
Residuals     144  385.0     2.67
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- $\text{Mean Sq} = \text{SS} / \text{df} = 31.7/2=15.85$  (forskjellen til tabellen skyldes at ikke alle desimalene i SS er oppgitt).
- $\text{F-value} = \text{MSM}/\text{MSE} = 15.83 / 2.67 = 5.928839$

2. Hvilke konklusjoner ville du trekke på bakgrunn av R-utskriften?

Vi ser en signifikant hovedeffekt av Filmsjanger, en signifikant hovedeffekt av Tetris, og en signifikant interaksjon mellom Filmsjanger og Tetris.  
Eta\_sq burde trekkes frem og tolkes.

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).

Det er kun figur A der det er klar evidens for en interaksjon, dette er derfor et plott av dataene fra studien. Det bør forklares hvordan man ser at det er en interaksjon i figuren.

### 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?

```
> chisq.test(TAB)

Pearson's Chi-squared test

data:  TAB
X-squared = 23.131, df = 4, p-value = 0.0001192
```

Ut ifra tabellen kan vi se at den kritiske verdien ved  $\text{df}=4$  er 9.49. Vi kan forkaste  $H_0$  med god margin, og konkludere med at det er svært lite trolig at uttrykt ønske om firedagers uke er uavhengig av yrke/sector.

For å gå mer detaljert inn i sammenhengen mellom de to variablene, er det rimelig å kommentere de standardierte residualene. Residualer med en absoluttverdi større enn 2 informerer oss om hvilke celler i tabellen som særlig avviker fra forventet frekvens under  $H_0$ . Det er altså vesentlig færre saksbehandlere som svarer ja på at de ønsker å gå over til en fire dagers arbeidsuke, men vesentlig flere håndverkere.