

Oppgave 1a: Sosial angst hos ungdommer

I denne oppgaven skal vi utforske ulike faktorer som kan påvirke nivået av selvrapportert sosial angst blant ungdom i alderen 13 til 18 år.

'SosialAngst' vil være den avhengige variabelen i alle analysene i oppgave 1. Høyere skåre indikerer mer selvrapportert sosial angst.

I deloppgave 1a vil vi se på to regresjonsmodeller som du finner utskrift fra i vedlegget. Vi inkluderer følgende uavhengige variabler:

- TimerDataspill: Antall timer en ungdom bruker på å spille dataspill per uke.
- AntallSosken: Antall søsken ungdommen har.
- Sovn: En selvrapportert vurdering av søvnkvaliteten til ungdommen.
- KoffeinBrus: Antall glass med koffeinholdig brus som ungdommen drikker per uke.

1. I modell_1 inkluderer vi de to uavhengige variablene TimerDataspill og AntallSosken.
 - Regresjonskoeffisienten for TimerDataspill blir estimert til 0.3792. Forklar hvordan du tolker dette tallet.

Regresjonskoeffisienten (vi har også brukt begrepet partiell regresjonskoeffisient) for TimerDataspill er 0.3792 i modell_1. Dette betyr at for hver ekstra time en ungdom bruker på å spille dataspill per uke, forventes det en økning på 0.3792 i skåren for selvrapportert sosial angst, *gitt at alle andre variabler i modellen holdes konstante (eventuelt, ikke endrer seg)*. Kandidaten kan alternativt vise til et regresjonsplan, der stigningstallet mens man beveger seg parallelt med TimerDataspill akse er 0.3792.

- Bruk utskriften for modell_1 til å finne forventet nivå av sosial angst for en ungdom som spiller 8 timer dataspill i uken og har to søsken.

For å finne forventet nivå av sosial angst for en ungdom som spiller 8 timer dataspill i uken og har to søsken, kan vi bruke regresjonskoeffisientene fra modell_1. Vi setter inn verdiene i regresjonsligningen:

$$\text{SosialAngst} = b_0 + \text{TimerDataspill } b_1 + \text{AntallSosken } b_2$$

$$\text{SosialAngst} = 14.8813 + 8 * 0.3792 - 2 * 0.5050$$

$$\text{SosialAngst} = 14.8813 + 3.0336 - 1.0100$$

$$\text{SosialAngst} = 16.9049 \approx 16.9$$

Så forventet nivå av sosial angst for en ungdom som spiller 8 timer dataspill i uken og har to søsken er ca. 16.9.

2. I modell_2 legger vi også til variablene Sovn og KoffeinBrus. Se resultatene i vedlegget og vurder om de ulike statistikkene gir støtte til modell_1 eller modell_2.

Her er det mange statistikker som bør trekkes frem for å vurdere om statistikkene gir støtte til modell_1 eller modell_2.

- **R2 og justert R2:** R2, er en statistikk som kan tolkes som andelen av variansen for den avhengige variabelen som kan forklares av de uavhengige variablene i modellen. I modell_1 er R-squared 0.08521, noe som betyr at omtrent 8.5% av variansen i SosialAngst kan forklares av TimerDataspill og AntallSosken. I modell_2 er R-squared 0.0922, noe som betyr at omtrent 9.2% av variansen i SosialAngst kan forklares av alle fire uavhengige variablene. Justert R-squared tar hensyn til antall prediktorer i modellen, og justerer R-squared nedover hvis antall prediktorer ikke bidrar til en bedre modell. I modell_1 er justert R-squared 0.07592, og i modell_2 er det 0.07358. Dette indikerer at inkludering av de to ekstra variablene i modell_2 ikke nødvendigvis forbedrer modellen betydelig, tvert imot synker justert R2 fra modell1 til modell2. Dette skjer fordi justert R2 tar hensyn til utvalgsstørrelse og antall parametere. I og med at vi legger til to parametere som forklarer lite, synker justert R2. Kort oppsummert, R2 og særlig justert R2 indikerer at modell_2 utgjør en svært liten forbedring over modell_1.

- **AIC:** Akaike's informasjonskriterium (AIC) er en statistikk for å sammenligne modeller for et gitt sett med data. AIC tar hensyn til modellens kompleksitet, og straffer modeller som er for komplekse. Lavere AIC-verdier indikerer bedre modeller. I dette tilfellet er AIC for modell_1 1050.666, og for modell_2 er det 1053.131. Dette indikerer at modell_1 kan være en bedre modell, siden den har en lavere AIC-verdi. (ting som kan nevnes, men som ikke forventes kjent er at AIC kan favorisere en modell der parametrene ikke er signifikante

- **ANOVA:** ANOVA (Analysis of Variance) er en statistisk metode som brukes til å sammenligne to eller flere (vanligvis nøyttede) modeller. Resultatene hr viser at p-verdien er 0.4733, noe som er større enn 0.05, og dermed ikke statistisk signifikant. Dette indikerer at det ikke er en signifikant forskjell mellom modell_1 og modell_2. Dette peker igjen på at vi ikke oppnår en vesentlig forbedring ved å legge til de to variablene Sovn og Koffeinbrus.

- **lm.beta koeffisienter:** lm.beta koeffisienter er standardiserte regresjonskoeffisienter. De gir en indikasjon på styrken og retningen av forholdet mellom hver uavhengig variabel og den avhengige variabelen. Siden koeffisientene indikerer hvor mange standardavvik den avhengige variabelen endrer seg for en endring på ett standardavvik i den uavhengige, kan vi bedre sammenlikne koeffisienter i multiple regresjonsmodeller. I modell_2 er lm.beta koeffisientene for TimerDataspill og AntallSosken henholdsvis 0.21464310 og -0.20874820, noe som indikerer at det er en positiv sammenheng mellom TimerDataspill og SosialAngst, og en negativ sammenheng mellom AntallSosken og SosialAngst. Vi kan ikke direkte sammenlikne modellene ved å se på beta koeffisientene, men det at de estimerte verdiene er så mye mindre for Sovn og KoffeinBrus, er nok en indikasjon på at de er av liten betydning, og følgelig at det er lite støtte for den økte kompleksiteten til modell_2.

- **Konfidensintervaller:** Vi kan se på konfidensintervallene for de to ekstra uavhengige variablene i modell_2 at ingen av dem er statistisk signifikante på et 0.05 nivå (begge intervallene rommer verdien 0). Dette indikerer igjen at de to ekstra uavhengige variablene ikke gir mye til modellen, og at man burde velge den enklere modell_1.

- **Øvrige p-verdier:** P-verdiene gir oss sannsynligheten for å se de observerte dataene gitt at H_0 er sann. Dette gir ingen informasjon utover intervallene, men viser at kun TimerDataspill og AntallSosken er signifikante. Fordi den er maskert må vi selv regne ut t-verdien for Sovn ($-0.14097 / 0.11633 = -1.21181$). I et utvalg av denne størrelsen ($n=200$) kan vi bruke normalapproximasjon for de kritiske verdiene, så Sovn har derfor en p-verdi >0.05 . KoffeinBrus har en t-verdi på -0.395 og vil derfor ha en p-verdi $>>0.05$ (antatt kjent, fordi i et utvalg av denne størrelsen kan vi approksimere de kritiske verdiene for en t-fordeling, med de fra en standard normalfordeling, som vi vet er ± 2 (eller mer presist 1.96)).

Oppgave 1b: Sosial angst hos ungdommer

I deloppgave 1b skal vi undersøke hvordan tiden brukt på sosiale medier og foreldrenes utdanningsnivå er assosiert med sosial angst blant ungdom. Du vil finne utskrifter fra en regresjonsmodell (modell_3) i det vedlagte pdf-dokumentet.

De uavhengige variablene i modell 3 er:

- **SosialeMedier:** Antall minutter en ungdom i snitt bruker på sosiale medier per dag.
- **ForeldreUtdanning:** En binær (dummy) variabel som indikerer om minst én av foreldrene har høyere utdanning (1) eller ikke (0).
- **SosialeMedier*ForeldreUtdanning.**

1. Hvordan tolker du sammenhengen mellom sosial angst, og variablene sosiale medier og foreldrenes utdanning, basert på disse resultatene?

I modell_3 er det en interaksjon mellom SosialeMedier og ForeldreUtdanning. Dette betyr at effekten av SosialeMedier på SosialAngst avhenger av nivået på ForeldreUtdanning, og omvendt.

Regresjonskoeffisienten for SosialeMedier er 0.055270 (kan gjerne runnes av, f.eks. til 0.01). Dette betyr at for hver ekstra minutt en ungdom bruker på sosiale medier per dag, forventes det en økning på 0.055270 i skåren for selvrapportert sosial angst, gitt at alle andre variabler i modellen holdes konstante. Denne koeffisienten har en svært lav p-verdi. Verdien av koeffisienten er lav, men det er fordi den viser til bruk av sosial media i enheter på minutter. Det er altså grunn til å konkludere med at det er en hovedeffekt av SosialeMedier, altså økt bruk er assosiert med mer sosial angst også uavhengig av foreldres utdannelse.

Regresjonskoeffisienten for ForeldreUtdanning er -2.009363. Dette betyr at ungdommer hvis minst én av foreldrene har høyere utdanning, forventes å ha en 2.009363 lavere skåre for selvrapportert sosial angst enn de hvis ingen av foreldrene har høyere utdanning, gitt at alle andre variabler i modellen holdes konstante. Men denne effekten er moderert av

SosialeMedier. Denne koeffisienten har en p-verdi på 0.1528, så vi kan ikke forkaste nullhypotesen om at foreldres utdanning alene er uavhengig av barnas sosiale angst.

Regresjonskoeffisienten for interaksjonstermen SosialeMedier:ForeldreUtdanning er -0.026994. Dette betyr at effekten av SosialeMedier på SosialAngst avtar med 0.026994 for hver ekstra minutt brukt på sosiale medier per dag for ungdommer hvis minst én av foreldrene har høyere utdanning, sammenlignet med de hvis ingen av foreldrene har høyere utdanning. Det vesentlige i oppgaven er at kandidaten forstår at resultatene antyder at det er en interaksjon mellom de to variablene SosialeMedier og ForeldreUtdanning, og kan beskrive dette forholdet. Tallene antyder altså at høyere utdanning hos foreldre er assosiert med en lavere negativ effekt av sosiale medier.

2. I vedlegget finner du også et plot av residualene fra modell 3 mot variabelen SosialeMedier. Hvordan tenker du det er rimelig å justere modellen i lys av disse residualene?

Plottet av residualene viser en svakt negativ kvadratisk sammenheng over nivåer av sosiale medier. Dette kan tyde på at det er en ikke-lineær sammenheng mellom SosialeMedier og SosialAngst som ikke er fanget opp av den nåværende modellen. Det vil derfor bli oppstå en korrelasjon mellom residualer med lik X-verdi, og det er en forutsetning for korrekt inferens i regresjonsanalyser at disse residualene er uavhengige. En mulig justering kan derfor være å inkludere en kvadratisk term for SosialeMedier i modellen (dvs. $SosialeMedier^2$) for å fange opp denne ikke-lineære sammenheng. En måte å spesifisere denne modellen i R ville være `modell_3b <- lm(SosialAngst ~ SosialeMedier*ForeldreUtdanning + I(SosialeMedier^2), data = df)`.

Kandidaten må gjerne forklare hvilke antagelser vi gjør oss i lineære regresjonsmodeller mer generelt, eller beskrive hensiktsmessige innflytelsesstatistikker.

Oppgave 1c: Sosial angst hos ungdommer relatert til skoletype

Vi fortsetter med 'SosialAngst' som avhengig variabel.

1. I 'modell_4' er 'Skoletype' den eneste uavhengige variabelen. 'Skoletype' er en kategorisk variabel som indikerer hvilken type skole deltageren går på. Variabelen tar en av følgende mulige verdier; 'Offentlig', 'Steinerskole' og 'Internasjonal'.
 - Hvordan tolker du resultatene fra 'modell_4'?

I modell_4 er Skoletype den eneste uavhengige variabelen. Skoletype er en *kategorisk variabel* med tre nivåer: Offentlig (referansekategori), Steinerskole, og Internasjonal.

Regresjonskoeffisienten for SkoletypeInternasjonal er -0.5728, men p-verdien er 0.3355, som er over den vanlige signifikansgrensen på 0.05. Dette indikerer at det ikke er en signifikant forskjell i nivået av sosial angst mellom ungdommer som går på internasjonal skole og de som går på offentlig skole, når vi ikke kontrollerer for andre variabler.

Regresjonskoeffisienten for SkoletypeSteinerskole er 1.1049, og p-verdien er 0.0484, som er under signifikansgrensen på 0.05. Dette indikerer at det er en signifikant forskjell i nivået av sosial angst mellom ungdommer som går på Steinerskole og de som går på offentlig skole, med de på Steinerskole rapporterer høyere nivåer av sosial angst, når vi ikke kontrollerer for andre variabler.

Kandidaten kan også med fordel trekke frem og tolke *F-verdien* (4.354). Denne lar oss teste nullhypotesen om at begge koeffisientene over er lik null i populasjonen. Dette har vi allerede forkastet over, men man kunne diskutere at det er mulig at ingen av koeffisientene enkeltvis er signifikante, men at det totalt sett kan være en signifikant samlet forskjell mellom gruppene.

- Hva kan du konkludere om forskjellen mellom sosial angst hos de som går på Steinerskolen og de som går på internasjonal skole?

For å forstå forskjellen i sosial angst mellom studenter som går på en internasjonal skole og studenter som går på en Steinerskole, må vi derfor ta hensyn til koeffisientene for begge disse kategoriene i forhold til referansekategorien. *Dette kan ikke leses direkte fra tabellen uten denne sammenligningen.*

For å finne forskjellen i forventet sosial angst mellom studenter som går på en internasjonal skole og studenter som går på en Steinerskole, kan vi trekke koeffisienten for "SkoletypeInternasjonal" fra koeffisienten for "SkoletypeSteinerskole". Det vil si, forskjellen er $1.1049 - (-0.5728) = 1.6777$. Dette betyr at studenter som går på en Steinerskole har en forventet sosial angst som er 1.6777 enheter høyere enn studenter som går på en internasjonal skole, gitt at alle andre variabler holdes konstante. Men husk at p-verdien for "SkoletypeInternasjonal" er større enn 0.05, noe som betyr at denne forskjellen ikke er statistisk signifikant.

2. I modell_5 inkluderer vi en ekstra uavhengig variabel, 'MobbingBarneskole'. Dette er en variabel som måler hvor mobbet en deltaker opplevde seg på barneskolen. Hvordan forstår du endringene i resultater fra modell_4 til modell_5?

Når vi inkluderer MobbingBarneskole i modellen, endres regresjonskoeffisientene for SkoletypeInternasjonal og SkoletypeSteinerskole, og deres tilhørende p-verdier øker (til henholdsvis 0.1859 og 0.0806). Dette indikerer at når vi kontrollerer for opplevd mobbing på barneskolen, er det ikke lenger en signifikant forskjell i nivået av sosial angst mellom ungdommer som går på Steinerskole og de som går på offentlig skole, og forskjellen mellom de som går på internasjonal skole og de som går på offentlig skole forblir ikke-signifikant.

Regresjonskoeffisienten for MobbingBarneskole er 2.1057, og p-verdien er svært lav ($<2e-16$), noe som indikerer at det er en sterk og signifikant positiv sammenheng mellom opplevd mobbing på barneskolen og nivået av sosial angst.

Samlet sett tyder disse resultatene på at opplevd mobbing på barneskolen er en sterkere prediktor for sosial angst enn skoletype. Det er også en styrke i besvarelsen at kandidaten

forklarer at MobbingBarneskole er en *konfunderende variabel* i forholdet mellom skoletype og sosial angst, og forklarer hva dette betyr.

Oppgave 2: Emosjonsreguleringsteknikker og emosjonell intensitet

Emosjonsregulering refererer til prosessen hvor individer påvirker, opprettholder eller endrer opplevelsen og uttrykket av sine egne følelser. Denne studien fokuserer på tre emosjonsreguleringsteknikker:

Kognitiv omstrukturering: Endre måten man tenker på en emosjonelt utfordrende situasjon for å redusere den negative emosjonelle opplevelsen (for eksempel ved å finne positive aspekter ved en situasjon).

Oppmerksomhetsforskyvning: Flytte oppmerksomheten bort fra den emosjonelt utfordrende situasjonen eller tankene og rette fokus mot noe nøytralt eller positivt.

Suppresjon: Aktivt undertrykke eller ignorere den negative emosjonelle opplevelsen i stedet for å konfrontere den.

I denne oppgaven skal vi undersøke sammenhengen mellom bruken av disse emosjonsreguleringsteknikkene og opplevd emosjonell intensitet. Deltakerne i studien ble bedt om å rapportere hvor ofte de brukte de nevnte teknikkene. Emosjonell intensitet ble målt og kategorisert i tre nivåer: lav, middels og høy. Basert på R-utskriften som er vedlagt, drøft hvilke konklusjoner du kan trekke om sammenhengen mellom bruken av ulike emosjonsreguleringsteknikker og emosjonell intensitet.

Basert på R-utskriften, ser det ut til at det er en signifikant sammenheng mellom bruken av ulike emosjonsreguleringsteknikker og emosjonell intensitet. Dette er indikert av Kji-kvadrat testen, der kji-kvadrat statistikken har en verdi på 52.649. Antall frihetsgrader (df) og p-verdien er ikke gitt i utskriften, men krysstabellen (emosjonsregulering x intensitet) er 3 x 3, og vi vet at df da er gitt av $df=(C-1)(R-2)=(3-1)(3-1)=4$.

Fra tabellen kan vi da lese kritisk verdi (kan gjerne defineres i oppgaven) til å bli 9.49. Fra dette vet vi at p-verdien må være mindre enn 0.05, og at vi kan forkaste hypotesen om at emosjonsreguleringsstrategi og emosjonell intensitet er uavhengige. Det er altså signifikante forskjeller i opplevd intensitet på tvers av de ulike emosjonsreguleringsstrategiene.

Her kan stolpediagrammet gjerne kommenteres. Dette viser den samme informasjonen som er gitt i krysstabellen, der omstrukturering er forbundet med den laveste intensiteten, mens suppresjon, en noe mindre forekommende strategi, har den relativt sett høyeste andelen deltagere som angir høy intensitet.

For å forstå sammenhengen mer detaljert, kan vi altså se på residualene og standardiserte residualene fra chi-kvadrat-testen. Residualene gir oss en indikasjon på avviket mellom de observerte og forventede frekvensene for hver celle i tabellen. Standardiserte residualer er residualene delt på deres standardavvik, og gir en indikasjon på størrelsen på avviket i standardavviksenheter. Siden disse er angitt i standardavviksenheter, er vi særlig oppmerksom på verdier mindre enn -2 eller større enn 2. Igjen ser vi at det er signifikant flere observasjoner i cellen «lav intensitet» og «Kognitiv omstrukturering» enn forventet under H_0 , og signifikant færre i cellen «høy intensitet» og «Kognitiv omstrukturering». Samtidig er det vesentlig færre i cellen «Suppresjon» og «Lav», og flere i «Suppresjon» og «Høy». Det er også nevneverdig at ingen av de andre cellene avviker vesentlig fra forventningen under H_0 .

Det er også bakgrunnsstoff som kan trekkes inn ved besvarelsen av denne oppgaven. For eksempel rasjonale bak χ^2 -kvadrat statistikken og antagelser for valid tolkning av χ^2 -kvadrat verdien. Disse funnene tyder på at det er en sammenheng mellom bruken av ulike emosjonsreguleringsteknikker og emosjonell intensitet, men det er også en styrke å understreke at resultatene er ikke nødvendigvis indikerer kausale forhold. Det kan være andre faktorer som påvirker både valg av emosjonsreguleringsteknikk og emosjonell intensitet.

Oppgave 3: Fag, undervisningsmetoder og elevdeltagelse

Elevdeltagelse er en viktig indikator for studenters engasjement og læring i skolen. I denne oppgaven skal vi utforske hvordan fag (FAG) og undervisningsmetoder (METODE) kan påvirke elevdeltagelsen. Vi ser på deltagelse på tvers av tre fagområder: matematikk, samfunnsfag og språk.

I tillegg har vi data fra to undervisningsmetoder; tradisjonell og nettbasert undervisning.

Tradisjonell undervisning innebærer at læreren underviser elevene i et klasserom, mens nettbasert undervisning skjer gjennom digitale plattformer og teknologier.

I vedlegget finner du et par figurer, samt resultater fra to variansanalyser og post-hoc analyser.

1. Tolk og diskuter resultatene fra enveis variansanalysen over faktoren FAG.

En enveis ANOVA-analyse over faktoren FAG avdekker en betydelig forskjell i elevengasjementet mellom de tre fagene (F -verdi = 19.5, p -verdi < 0.001). Dette betyr at elevengasjementet varierer basert på faget som undervises. Det vil være fornuftig å konsultere box-plottet i denne sammenhengen, og forklare hvordan det skal tolkes – hva strekene og punktene representerer, og hvorfor det er rimelig å anta signifikante forskjeller mellom fagene basert på dette.

Fra box-plottet kan vi også observere noen særegenheter. Det er fire uteliggere i samfunnsfag-gruppen som potensielt kan påvirke resultatene. Videre er det en viss skjevhet i fordelingen i denne gruppen, som muligens ikke har stor innflytelse, men som likevel bryter med forutsetningene for analysen.

ANOVA er en omnibus-test, og en signifikant verdi tillater oss å forkaste nullhypotesen (at dataene er hentet fra en populasjon der gjennomsnittet er det samme for hvert fag). For å forstå

sammenhengen mellom engasjement og fag bedre, vil det være hensiktsmessig å utføre parvise post-hoc-sammenligninger etter en signifikant ANOVA. Verdien av slike analyser er at de gir mer detaljert informasjon om forskjellene mellom fagene. Uten korreksjon for antall tester, ville vi ha konkludert med at det er en signifikant forskjell i engasjement mellom samfunnsfag og språk. Dersom vi derimot justerer for antall tester ved hjelp av en Bonferroni-korreksjon, er forskjellen ikke lenger signifikant. Behovet for slike korreksjoner er viktig og bør diskuteres.

Til slutt, selv etter å ha tatt hensyn til Bonferroni-korreksjonen, er det en signifikant forskjell i engasjement mellom parvise sammenligninger av matematikk-samfunnsfag og matematikk-språk.

2. Vedlegget inneholder også utskrift fra en toveis variansanalyse over FAG*METODE.

- [Tolk og diskuter resultatene fra toveis variansanalysen. Sammenlikn resultatene med de fra enveis variansanalysen.](#)

En toveis ANOVA-analyse av FAG*METODE avdekker en betydelig hovedeffekt for både FAG (F-verdi = 22.294, p-verdi < 0.001) og METODE (F-verdi = 12.911, p-verdi = 0.000484). I tillegg observeres det en signifikant interaksjonseffekt mellom FAG og METODE (F-verdi = 3.411, p-verdi = 0.036414). Dette indikerer at elevengasjementet er påvirket av både fag og undervisningsmetode, og at undervisningsmetodens effekt på elevengasjementet varierer mellom de ulike fagene.

Sammenlignet med enveis ANOVA, gir toveis ANOVA et mer detaljert bilde av hvordan både fag og undervisningsmetode påvirker elevengasjementet, og på hvilken måte disse to faktorene interagerer. Spesielt merker vi at F-verdien for fag er høyere i toveis analysen. Dette skyldes at vi i denne analysen beregner F-verdien som forholdet mellom mean square error på mean-square error/residuals i en modell der effekten av metode er kontrollert for. Den reduserte nevneren resulterer i en større F-verdi. En slik økt effektstørrelse kan være en av fordelene ved å inkludere flere faktorer i en ANOVA-analyse.

- [Forklar hvilket av de tre linjediagrammene \(versjon A, B eller C\) som stemmer best overens med utskriften.](#)

Det riktige svaret er «Versjon C». Alle figurene samsvarer generelt med box-plottet fra forrige side i vedlegget. Imidlertid viser Versjon A en inkonsistens med resultatene fra toveis ANOVA. Ut fra dette plottet ville vi ikke forvente en hovedeffekt av undervisningsmetode, ettersom forskjellene mellom tradisjonell og nettbasert undervisning jevnes ut over fagene. Derfor ville vi, basert på Versjon A, forvente en hovedeffekt av fag og en interaksjonseffekt, men ikke en signifikant effekt av undervisningsmetode.

I Versjon B ser vi en tydelig hovedeffekt av fag, samt en svak, men sannsynligvis signifikant forskjell mellom undervisningsmetodene. Men ut fra Versjon B ville vi ikke forvente å se en interaksjonseffekt, ettersom linjene er nesten helt parallelle. Dette står i motsetning til den signifikante interaksjonseffekten som fremkommer i ANOVA-resultatene.