Sensorveiledning PSY2014 V24

# Oppgave 1a:

**a) Modell 1: KLIMAANGST som eneste uavhengige variabel**

**Spørsmål 1: Hvordan ville du oppsummert resultatene fra modell 1 basert på denne utskriften?**

Momenter som bør adresseres:

1. Modellens overordnede tilpasning:
   * R-squared: 0.0894, som indikerer at omtrent 8,94% av variasjonen i fremtidsplaner forklares av klimaangst.
2. Tolkning av koeffisientene:
   * Intercept: 91.2386 (p < 2e-16), som representerer den forventede skåren på fremtidsplaner når klimaangst er null.
   * Klimaangst: -0.7465 (p = 0.000202), som indikerer at for hver enhets økning i klimaangst, reduseres fremtidsplaner med omtrent 0.75 enheter.
   * Begge koeffisientene er statistisk signifikante, som indikert av p-verdiene.

Andre momenter som kan påpekes:

* + Residual standard error er 6.92, som antyder variabiliteten i fremtidsplaner rundt regresjonslinjen.
  + F-statistikk: 14.53, med en p-verdi < 0.0002019, som indikerer at modellen er statistisk signifikant.

**Spørsmål 2: Hva er forventet skåre på FREMTIDSPLANER for en ungdom med klimaangst på 10?**

* Bruk regresjonsligningen:
* For Klimaangst = 10: Fremtidsplaner =91.2386−0.7465×10=83.7736

**Spørsmål 3: I utskriften finner du et spredningsplott med to observasjoner merket A (rødt) og B (blått). Hvordan tror du disse to observasjonene påvirker resultatene? Tenker du at det er rimelig å fjerne en eller begge fra datasettet?**

Momenter som bør adresseres:

1. Punkt A:
   * Leverage: 0.009 (lav)
   * Cook's Distance: 0.011 (lav)
   * Residual: 10.843 (høy)
   * Tolkning: Punkt A har en høy residual men lav leverage og Cook's distance, som indikerer at det er en utligger med hensyn til den avhengige variabelen, men påvirker ikke regresjonskoeffisientene vesentlig.
2. Punkt B:
   * Leverage: 0.191 (høy)
   * Cook's Distance: 0.171 (høy)
   * Residual: 7.544 (høy)
   * Tolkning: Punkt B har høy leverage og Cook's distance, som indikerer at det er både en utligger og et influensielt punkt. Dette punktet kan potensielt forvrenge regresjonsresultatene.
3. Anbefalinger:
   * Punkt B bør vurderes for fjerning eller videre undersøkelse på grunn av dets høye leverage og innflytelse på modellen. Punkt A, til tross for høy residual, er mindre innflytelsesrik og trenger ikke fjernes.
   * Studentene skal begrunne beslutningen sin basert på leverage, Cook's distance og residualverdier.

# Oppgave 1b:

**b) Modell 2: Inkludering av EKSTROVERSJON og SKOLEKARAKTERER som uavhengige variabler**

**Spørsmål: Forklar hvorfor koeffisienten til klimaangst endrer seg når ekstroversjon og skolekarakterer blir lagt til modellen.**

**Momenter som bør adresseres:**

Det er forventet at studenten kan gjenkjenne en konfunderende sammenheng,. Utfordringen i oppgave også i å lese statistikkene slik at man kan begrunne hvilken av variablene (skolekarakterer eller ekstroversjon) som fører til endringer i estimatet av klimaangst, og hvorfor.

1. **Endring i koeffisienten til klimaangst:**
   * **Modell 1:** Koeffisienten til klimaangst er -0.7465, med en standardfeil på 0.1958 og en t-verdi på -3.812 (p = 0.000202).
   * **Modell 2:** Koeffisienten til klimaangst endres til -0.9382, med en standardfeil på 0.1978 og en t-verdi på -4.743 (p = 4.97e-06).

* Når ekstroversjon og skolekarakterer legges til modellen, endrer koeffisienten til klimaangst seg. Dette kan skyldes at skolekarakterer er en konfunderende variabel som forklarer noe av variansen i fremtidsplaner, som opprinnelig ble tilskrevet klimaangst.

**Forklaring av konfunderende variabler:**

* **Skolekarakterer:** Har en positiv koeffisient på 3.4033 (p = 0.000892), noe som tyder på at høyere skolekarakterer er assosiert med bedre fremtidsplaner.
* Partielle korrelasjoner viser at skolekarakterer er signifikant korrelert med fremtidsplaner (r = 0.270, p = 0.000892). Matrisescatterplottet viser også en signifikant korrelasjon mellom skoleprestasjoner og både fremtidsplaner og klimaangst.
* **Ekstroversjon:** Har en koeffisient på 0.2165 (p = 0.669605), som ikke er signifikant. Dette indikerer at ekstroversjon ikke har en vesentlig effekt på fremtidsplaner og ikke bidrar til endringen i koeffisienten til klimaangst.

1. **Partielle korrelasjoner:**
   * Partielle korrelasjoner viser at når vi kontrollerer for klimaangst og ekstroversjon, er skolekarakterer signifikant korrelert med fremtidsplaner (r = 0.270, p = 0.000892).
   * Ekstroversjon er ikke signifikant korrelert med fremtidsplaner (r = 0.035, p = 0.669605), noe som indikerer at denne variabelen ikke bidrar vesentlig til modellen.
2. **Modelltilpasning:**
   * **Modell 1:** R-squared er 0.0894, som indikerer at 8.94% av variasjonen i fremtidsplaner forklares av klimaangst.
   * **Modell 2:** R-squared øker til 0.1583, som indikerer at 15.83% av variasjonen i fremtidsplaner forklares når ekstroversjon og skolekarakterer inkluderes. Dette viser at modellen forbedres ved å inkludere flere relevante variabler.

Momenter som kan trekkes frem:

* **Residualstandardfeil:**
  + **Modell 1:** Residual standard error er 6.92.
  + **Modell 2:** Residual standard error reduseres til 6.699, noe som indikerer en forbedring i modellens prediksjonsnøyaktighet.
* Studentene skal vise forståelse for hvordan og hvorfor koeffisientene til klimaangst endrer seg når nye variabler legges til modellen.
* Studentene skal helst vise forståelse av hva partielle korrelasjoner er, og hvordan de hjelper til med å forstå forholdet mellom variablene når man kontrollerer for andre variabler.
* Studentene bør diskutere forbedringen i modelltilpasning (R-squared) og hva dette innebærer for modellens evne til å forklare variasjonen i fremtidsplaner.
* Forvent en grundig analyse av hvorfor skolekarakterer er en viktig konfunderende variabel og hvordan dette påvirker resultatene.
* Studentene bør også forklare hvorfor det ikke er ekstroversjon som står for endringen i koeffisienten til klimaangst, ved å referere til de statistiske resultatene som viser at ekstroversjon ikke er signifikant korrelert med fremtidsplaner.

# Oppgave 1c:

**c) I modell 3 blir de uavhengige variablene SES og sosial\_stotte lagt til.**

1. **Noen av tallene i R-utskriften fra denne modellen er sladdet. Hvilke av følgende uavhengige regresjonskoeffisienter i modell 3  [klimaangst, ekstroversjon, skulekarakterar, sesLav, sesMiddels, sosial\_stotte]**
2. Signifikans av  koeffisienter:
   * Klimaangst: Koeffisienten er -0.7833 med en p-verdi på 1.77e-06, noe som betyr at den er statistisk signifikant ved et 0,05-nivå.
   * Skolekarakterer: Koeffisienten er 3.3722 med en p-verdi på 4.25e-05, noe som betyr at den er statistisk signifikant ved et 0,05-nivå.
   * SES Lav: Vi kan regne ut t-verdien. -10.8271/1.1392  = -9.504  . Gitt utvalgets størrelse kan vi konkludere (z-approksimasjon) at koeffisienten er meget signifikant.
   * SES Middels: Øvre grense for konfidensintervallet er under 0, hvilket indikerer at denne er signifikant ved et 0,05-nivå.
   * Sosial støtte: Koeffisienten er 0.1428 med en t-verdi på 1.342, noe som indikerer at den ikke er signifikant ved et 0,05-nivå.
   * Ekstroversjon: Konfidensintervallet for ekstroversjon (-0.6026 til 0.9763) inkluderer null, noe som indikerer at denne variabelen ikke er signifikant ved et 0,05-nivå.

**Uavhening av signifikans, hvilken av variablene "ekstroversjon" og "sosial\_stotte" ville du argumentere for at er sterkest assosiert med fremtidsplaner?**

Det viktige her er at kandidaten forstår at man må se til de standardiserte koeffisientene, og gjerne beskriver hvordan de skal tolkes.

* Ekstroversjon**:** Har en standardisert koeffisient på 0.028, som ikke er signifikant og har en lav korrelasjon med fremtidsplaner.
* Sosialstøtte**:** Har en standardisert koeffisient på 0.082, som også ikke er signifikant, men er noe høyere enn ekstroversjon.
* Basert på standardiserte koeffisienter og korrelasjoner, kan det argumenteres for at sosial støtte har en sterkere, om enn ikke signifikant, assosiasjon med fremtidsplaner enn ekstroversjon.

**Hvordan ville du oppsummert forholdet mellom sosioøkonomisk status (ses) og fremtidsplaner?**

Studenten bør vise at de forstår hvordan man tolker en kategorisk prediktor.

* SES Lav: Koeffisienten er -10.8271, som er betydelig og negativ, noe som tyder på at ungdommer med lav SES har dårligere fremtidsplaner sammenlignet med de med høy SES.
* SES Middels: Koeffisienten er -5.3213, som også er negativ og signifikant, men ikke så stor som SES Lav, noe som indikerer en moderat negativ effekt på fremtidsplaner.
* Konklusjon: Generelt viser modell 3 at ungdommer med lav og middels SES har dårligere fremtidsplaner sammenlignet med de med høy SES. Dette er viktig for å forstå hvordan sosioøkonomiske faktorer kan påvirke ungdommers fremtidsutsikter.

# Oppgave 1d:

**Momenter som bør adresseres:**

1. **Forholdet mellom klimaangst og alder:**

Studenten burde gjennkjenne modellen som en polynom regresjon, og tolke den korrekt.

* **Modell 4a:**
  + Koeffisienten for alder er 0.97203, noe som indikerer en positiv lineær sammenheng mellom alder og klimaangst.
  + Dette betyr at for hver ettårig økning i alder, øker klimaangst med 0.972 enheter.
  + Modell 4a har en R-squared på 0.4663, som indikerer at 46.63% av variansen i klimaangst kan forklares av alder alene.
* **Modell 4b:**
  + Koeffisienten for alder er 8.55301, mens koeffisienten for alder^2 er -0.25130. Dette indikerer en kurvelineær sammenheng mellom alder og klimaangst.
  + Den positive koeffisienten for alder og den negative koeffisienten for alder^2 antyder at klimaangst øker med alderen, men denne økningen avtar etter hvert (konkav kurve).
  + Modell 4b har en R-squared på 0.5937, noe som betyr at 59.37% av variansen i klimaangst kan forklares av både alder og alder^2. Dette er en forbedring sammenlignet med modell 4a.
* **Sammenlikning av modeller:**
  + Den lavere AIC-verdien for modell 4b (653.0342) sammenlignet med modell 4a (691.9248) indikerer at modell 4b har en bedre modelltilpasning.
  + ANOVA-resultatene viser en signifikant forbedring i modelltilpasning når alder^2 legges til modellen (F = 46.067, p = 2.619e-10).

1. **Diagnostiske plot:**
   * **Spredningsplotter for modell 4a og 4b:**
     + Modell 4a: Det skal være mulig å spore en svak kvadratisk funksjon i residualene. Dette antyder at en slik struktur ikke er fanget opp av modellen, og at det kan være hensiktsmessig å legge inn et kvadratisk ledd.
     + Modell 4b: Spredningsplottet viser en mer tilfeldig fordeling av residualer langs aldersskalaen, noe som indikerer en bedre modelltilpasning. I dette residualplottet er det lite/ingen struktur.
   * **Diagnostiske plot for modell 4b:**
     + Studenten må helst vise (gjerne svært kort) at de forstår hvordan man leser histogrammer av residualene, og Qq-plots.
       - **Histogram av residualer:** Histogrammet bør vise en tilnærmet normalfordeling av residualene, noe som indikerer at modellens antagelser om normalfordelte residualer er oppfylt.
       - **Q-Q plot:** Q-Q plottet bør vise at residualene følger en rett linje, noe som ytterligere bekrefter at residualene er normalfordelte.
       - **Residualer vs. alder:** Plottet bør vise en tilfeldig fordeling av residualene rundt null, noe som indikerer at modellens antagelser om homoskedastisitet er oppfylt.

**Generelle notater:**

* Studentene skal vise evne til å tolke og sammenlikne koeffisientene for alder og alder^2.
* Studentene bør demonstrere forståelse for hvordan modelltilpasning kan forbedres ved å legge til en ekstra variabel (alder^2).
* Forvent at studentene kan tolke diagnostiske plot og vurdere modellens antagelser basert på disse plottene.
* Det er viktig at studentene kan forklare hvorfor modell 4b gir en bedre tilpasning enn modell 4a, basert på både statistiske målinger (R-squared, AIC, F-test) og visuelle inspeksjoner (diagnostiske plot).

# Oppgave 2:

**a) Forklar kort begrepene "innen-gruppe varians" og "mellom-gruppe varians", og hvordan forholdet mellom dem kan brukes til å vurdere nullhypotesen i en variansanalyse.**

Momenter som bør adresseres:

* **Innen-gruppe varians:** Variasjonen i dataene innenfor hver gruppe. Det måler hvor mye verdiene innenfor en gruppe avviker fra gruppens gjennomsnitt.
* **Mellom-gruppe varians:** Variasjonen i dataene mellom gruppene. Det måler hvor mye gruppens gjennomsnitt avviker fra det totale gjennomsnittet for alle gruppene.
* **Forholdet mellom dem:** Brukes i ANOVA til å vurdere om det er en signifikant forskjell mellom gruppene. En høy mellom-gruppe varians sammenlignet med innen-gruppe varians indikerer at gruppene er forskjellige fra hverandre.
* **Nullhypotesen:** Nullhypotesen i ANOVA er at det ikke er noen forskjell mellom gruppenes gjennomsnitt. Hvis mellom-gruppe varians er signifikant høyere enn innen-gruppe varians, avviser vi nullhypotesen.

**b) I vedlegget finner du utskrift fra en enveis variansanalyse over faktoren MUSIKKTYPE. Fyll inn de manglende verdiene i den vedlagte tabellen, og konkluder med hensyn på resultatene.**

Df (degrees of freedom) for musikk: 2  
Sum Sq (Sum of Squares) for musikk: 1708.4  
Mean Sq (Mean Square) for musikk: 854.2  
F value: 37.26  
Pr(>F): 2.04e-12

Post-hoc testene burde diskuteres.

* **Bonferroni-justerte p-verdier:**
  + Ingen musikk vs Electronic dance music: p < 0.001
  + Ingen musikk vs Hip-hop: p < 0.001
  + Electronic dance music vs Hip-hop: p = 1 (ikke signifikant)

**Konklusjon:**

* Post-hoc testene viser at treningsprestasjonene med "Ingen musikk" er signifikant forskjellige fra både "Electronic dance music" og "Hip-hop".
* Det er ingen signifikant forskjell mellom "Electronic dance music" og "Hip-hop", verken med ujusterte eller Bonferroni-justerte p-verdier.

**c) Forklar med egne ord hva vi mener med en interaksjonseffekt.**

* **Interaksjonseffekt:** Når effekten av en uavhengig variabel på en avhengig variabel avhenger av nivået til en annen uavhengig variabel.
* **Eksempel:** I denne konteksten kan det være at effekten av musikktype på treningsprestasjon varierer avhengig av treningsformen (utholdenhet eller styrketrening).

**d) I neste omgang blir en toveis anova gjennomført, med faktorene MUSIKKTYPE og TRENINGSFORM. Konkluder med hensyn på resultatene fra analysen, og forklar hvorfor F-verdien til faktoren MUSIKKTYPE har endret seg fra enveis-analysen.**

* **Musikktype:** F (2, 84) = 42.640, p < 0.001. Det er en signifikant hovedeffekt av musikktype på treningsprestasjon.
* **Treningsform:** F (1, 84) = 3.758, p ≈ 0.056. Det er ingen signifikant hovedeffekt av treningsform på treningsprestasjon. (men den er i grenseland)
* **Interaksjonseffekt:** F (2, 84) = 5.898, p < 0.01. Det er en signifikant interaksjonseffekt mellom musikktype og treningsform.

Studenten burde vise til plottet, og forklare hvordan de kan se en interaksjonseffekt.

**Forklaring på endringen i F-verdi for musikktype:**

* **Enveis ANOVA:** Når kun musikktype er inkludert, reflekterer F-verdien forskjellene mellom musikktypene uten å kontrollere for treningsform.
* **Toveis ANOVA:** Når treningsform også inkluderes, kontrolleres effekten av treningsform, og interaksjonseffekten mellom musikk og trening tas også med i betraktningen. Dette kan føre til en økning i F-verdien for musikktype hvis effekten av musikk blir mer tydelig når variansen knyttet til treningsform også vurderes.
* **Endring i F-verdi for musikktype:** F-verdien for musikktype har økt fra 37.26 i enveis ANOVA til 42.64 i toveis ANOVA. Dette kan skyldes at inkluderingen av treningsform i modellen forklarer mer av variasjonen i treningsprestasjoner, noe som gir en mer nøyaktig estimering av effekten av musikktype.

# Oppgave 3:

**Momenter som bør adresseres:**

* **Kji-kvadratverdi:** 12.198
* **Frihetsgrader:** 4
* **P-verdi:** 0.0158
* **Statistisk signifikans:** Resultatet er statistisk signifikant på et 0.05-nivå (p < 0.05), noe som betyr at vi kan avvise nullhypotesen og konkludere at det er en signifikant sammenheng mellom arbeidsmiljø og innovasjonsevne.

**Momenter som bør adresseres:**

* **Kji-kvadratverdi:** 12.198
* **Frihetsgrader:** 4
* **P-verdi:** 0.0158
* **Statistisk signifikans:** Resultatet er statistisk signifikant på et 0.05-nivå (p < 0.05), noe som betyr at vi kan avvise nullhypotesen og konkludere at det er en signifikant sammenheng mellom arbeidsmiljø og innovasjonsevne.

**Spørsmål 2: Hva vil du konkludere om sammenhengen mellom arbeidsmiljø og innovasjonsevne basert på resultatene?**

**Momenter som bør adresseres:**

* **Sammenheng mellom arbeidsmiljø og innovasjonsevne:** Resultatene viser at det er en signifikant sammenheng mellom arbeidsmiljø og innovasjonsevne.
* **Residualer og standardiserte residualer:** Residualene og de standardiserte residualene gir innsikt i hvilke kategorier som avviker mest fra det forventede antallet under nullhypotesen.
  + Clean desk: Har flere ansatte med høy innovasjonsevne enn forventet, og færre med lav innovasjonsevne enn forventet.
  + Fast plass i åpent kontorlandskap: Har færre ansatte med høy innovasjonsevne enn forventet, og flere med middels innovasjonsevne enn forventet.
  + Individuelle kontorer: Har omtrent det forventede antallet i alle kategorier, men litt flere med lav innovasjonsevne.

**Konklusjon:**

* Basert på resultatene kan vi konkludere at det er en signifikant sammenheng mellom arbeidsmiljø og innovasjonsevne.
* Ansatte i "Clean desk" miljøet ser ut til å ha høyere innovasjonsevne sammenlignet med de andre miljøene, mens ansatte i "Fast plass i åpent kontorlandskap" ser ut til å ha lavere innovasjonsevne.
* Residualene indikerer at "Clean desk" miljøet er mer gunstig for høy innovasjonsevne, mens "Fast plass i åpent kontorlandskap" kan være mindre gunstig.