Kladd

1

I denne studien vil nullhypotesen være at språkforståelse ikke er assosiert med popularitet. Den alternative hypotesen er dermed at det faktisk er en sammenheng mellom disse to variablene.

For å oppsummere sammenhengen mellom POPULARITET og SPRÅK, så er det verdt å merke seg visse verdier:

- Den uavhengige variabelen SRPÅK har en signifikant p-verdi med et alpha-nivå på 0.001. Vanligvis setter man alpha-nivå til 0.05, så denne verdien sier at sannsynligheten for at verdien vår faller innenfor 2 standardavvik på begge sider av gjennomsnittet, altså innenfor 95% av verdiene er ganske usannsynlig. Vi ser også at SPRÅK er signifikant ved hjelp av utskriften av konfidensintervallet, og at dette ikke inneholder 0.

- Videre kan vi se på p-verdien til hele modellen, som også er signifikant.

- For å se hvor mye av variansen modellen er i stand til å forklare ser man på 'Multiple R-squared'. I dette tilfellet har vi bare én uavhengig variabel, så det holder å se på 'Multiple R-squared'. Ergo, vi trenger ikke å se på 'Adjusted R-squared'. Denne regnes ut ved: r^2 = 1 - SSE/TSS. SSE finner vi i anova-utskriften og denne er 619.26. Vi må finne TSS ved hjelp av å ta SSM+SSE: 182.396+619.26=801.656. Dermed setter vi inn alle verdiene:

r^2= 619.26/801.656=0.2275 , runder dette tallet av til 0.23.

Vi ser at selv om p-verdiene i modellen vår er signifikante, så er modellen bare i stand til å forklare 23% av variansen. Dette er ikke stort. Høye R-squared-verdier tilsier at modellen er i stand til å forklare mer enn lave R-squared-verdier, og vi ønsker oss ideelt høye verdier.

Den forventede skåren på POPULARITET for et barn som har skåren 10 på variabelen SPRÅK finner man ved å putte inn verdien i regresjonslina vår:

y^hat = b^hat 0 + b^hat 1

y^hat = 10.0395 + 1.051 (10)

y^hat = 20.5495 , runder opp til 20.55.

2

Ved å kontrollere for SES i modell 2a, ser vi at SPRÅK går ned i signifikans og estimat (stigningstall). SPRÅK er dog fortsatt signifikant i modell 2a.

Ved å kontrollere for INTROVERSJON i modell 2b, ser vi at SPRÅK holder lik signifikans som i modell 1 og nesten likt estimat som i modell 1.

Det som skjer i modell 2a er at SES undertrykker SPRÅK. SES fungerer altså som er suppressor-variabel, og fortrenger assosiasjonen mellom POPULARITET og SPRÅK-variabelen.

3

1. Residualene fra modell 3 bryter med antagelsene i en lineær regresjonsanalyse på spesielt ett punkt. Ved å observere plottet av residualene, så kan vi se at modell 3 har et ikke-lineært forhold mellom x og y. Det ser ut til at dataen passer en annengradsfunksjon (et kvadratisk uttrykk) bedre siden vi kan se at den synker, kommer til et bunnpunkt, og stiger igjen.

Det er derfor mer hensiktsmessig å kjøre modell 4 fordi modell 3 bryter med antagelsene i en lineær regresjonsanalyse, mens modell 4 ikke gjør det. Vi ser at residualene i modell 4:

- er normalfordelt

- ikke har noen uteliggere med stor innflytelse

-  ikke er systematisk organisert, man ser ingen tegn til mønster i dataen

- residualene er ikke korrelerte, de klumper seg ikke sammen på visse steder

2. Forventet nivå på tilfredshet for en person som rapporterer en stress-skåre på 10 er gitt ved:

y^hat = b^hat 0 + b^hat 1 + b^hat 2

\*Runder av verdiene fra utskriften til to desimaler\*

y^hat= 5.24 + 0.29x  + 0.15x\*2

er det samme som:

y^hat= 5.24 + 0.29x  + 0.15x\*2

y^hat= 5.24 + 0.29 (10) + 0.15 (10)^2

y^hat = 8,14 + 15

y^hat = 23.14

forholdet mellom stress og tilfredsstillhet på arbeidsplassen er uttrykket av STRESS og STRESS kvadrert. Dette betyr at forholdet ikke er fullstendig lineært, men beskrevet bedre som en annengradsfunksjon av STRESS, altså STRESS kvadrert. Slik med annengradsfunksjoner, så er det en større økning jo flere enheter på den avhengige variabelen. Ergo, tilfredshet øker ganske lite i starten, men så øker den ettersom stress øker, og sammenhengen mellom de to forsterkes.