Sensorveiledning
Eksamen kvantitativ metode V04, SOS4020

a) Utvalgsfordelingen er tilnærmet klokkeformet, men med en hale mot høyere verdier – dvs. svakt høyreskjev.
Et konfidensintervall angir statistiske usikkerhetsmarginer og er et intervall rundt det observerte parameterestimatet (punktestimatet). Intervallet har den egenskap at det er en nærmere spesifisert sannsynlighet for at populasjonsverdien skal ligge innenfor intervallet. Siden antallet observasjoner er stort, er sannsynlighetsfordelingen en normalfordeling og intervallet blir

Øvre grense 2.77 + 1.96*0.055 = 2.88
Nedre grense 2.77 + 1.96*0.055 = 2.66

b) Kji-kvadratobservatoren brukes (her) til å teste nullhypotesen om at det ikke er sammenheng mellom informantens svar på de to spørsmålene, dvs. at de to variablene er ukorrelert. Observatoren er et mål på diskrepansen mellom den "tabellinnmaten" (cellefrekvensene) en ville forvente dersom nullhypotesen er riktig, gitt marginalfrekvensene, og den "tabellinnmaten" en faktisk observerer. Når diskrepansen er stor, er det grunn til å forkaste nullhypotesen.
Logikken kan eksempelvis utlegges som følger: Totalt oppgir en andel på 718/1420 = 0.5056 at de aldri savner å ha en god venn. Hvis nullhyptesen er riktig forventer en at samme andel gir dette svaret for alle grader av ensomhet. Siden 470 svarer at de aldri føler seg ensom, burde den forventede frekvensen i cellen "aldri/aldri" derfor være 470*0.5056 = 237.6.
De forventede cellefrekvensene kan også beregnes som produktet av kolonne- og radsummene for cellen, dividert med det totale antallet observasjoner. Eksempelvis vil det forventede antallet i cellen "aldri, aldri" være 718*470/1420 = 237.6.
Antall frihetsgrader = (antall rekker – 1)*(antall kolonner – 1) = (4–1)*(4–1) = 9
Det er 5 prosent sjanse for at tilfeldigheter skal gi en diskrepans som er større enn 16.9. Vi forkaster derfor nullhypotesen på 5 prosentnivået dersom den observerte kji-kvadratobervatoren overstiger denne kritiske verdien.
I det aktuelle tilfellet forkaster vi derfor nullhypotsen og trekker den konklusjon at det er sammenheng mellom svarene på de to spørsmålene.

c) Standardfeilen til korrelasjonskoeffisienten blir

$$SE(\hat{r}) = \frac{1}{\sqrt{1420-1}} = 0.0265$$

Korrelasjonskoeffisienten er et mål på lineær sammenheng mellom variablene (slik disse er kodet). Nullhypotesen sier at det ikke er noen lineær sammenheng, mens den alternative hypotesen sier at det er en slik sammenheng (tosidig test):

$H_0 : r = 0,$ $H_A : r \neq 0$

Vi bruker en t-test, med testobservator

$$t = \frac{\hat{r}}{SE(\hat{r})} = \frac{0.47}{0.0265} = 17.7$$

Siden antallet observasjoner er stort, kan vi bruke normalfordelingen. Kritisk t-verdi blir 1.96. Siden den observerte t-verdien langt overskrider dette, forkaster vi nullhypotesen.

d) Cronbachs alpha-koeffisient er et mål på generaliserbarhet. Koeffisienten angir i hvilken grad den indeksen en har konstruert vil samsvare med en tilsvarende indeks basert på litt andre spørsmål, trukket fra det samme universet av spørsmål. Koeffisientens størrelse vil være avhengig av hvor godt samsvar det er (dvs. hvor høy korrelasjon det er) mellom svarene på de enkelte spørsmålene som inngår i indeksen, samt hvor mange spørsmål indeksen er basert på.
Koeffisienten er i dette tilfellet 0.71, og det er ikke veldig høyt. Man kan med andre ord bare forvente et moderat sterkt samsvar med andre mål for det samme fenomenet.
Jo flere spørsmål en indeks er basert på, desto større blir alpha. Man skulle med andre ord forvente høyere alpha ved flere spørsmål.

e) Jentene rapporterer en noe høyere grad av sosial isolasjon enn guttene. Forskjellen er på 0.63 indekspoeng. (Det er også en liten forskjell i spredningen hos de to kjønn – standardavviket er litt større blant jenter.)
Standardfeilen til differansen mellom de to gjennomsnittene er

$$SE(\hat{m}_1 - \hat{m}_2) = \sqrt{(0.114)^2 + (0.105)^2} = 0.155$$

Derved får vi

$$t = \frac{\hat{m}_1 - \hat{m}_2}{SE(\hat{m}_1 - \hat{m}_2)} = \frac{0.63}{0.155} = 4.06$$

Siden antallet observasjoner er stort, kan vi bruke normalfordelingen. Vi forkaster med andre ord nullhypotesen om at det ikke er noen kjønnsforskjell, siden t-verdien er langt større enn 1.96.

f) For jentene er gjennomsnittlig indeksverdi blant 13-åringene (alder = 0) lik 3.38 (=konstantleddet), og graden av isolasjon synker med økende alder. (Nedgangen er statistisk signifikant på 5 prosentnivået, idet t = –0.120/0.057 = 2.10.) For guttene er gjennomsnittet blant 13-åringene 2.28, dvs. 1.1 indekspoeng lavere, og graden av isolasjon øker med alder. (Økningen er imidlertid ikke statistisk signifikant, t = 0.096/0.057 = 1.68.)
Standardfeilen til forskjellen mellom de to regresjonskoeffisientene for alder er

$$SE(\hat{b}_1 - \hat{b}_2) = \sqrt{(0.057)^2 + (0.057)^2} = 0.081$$

Testobservatoren blir derfor

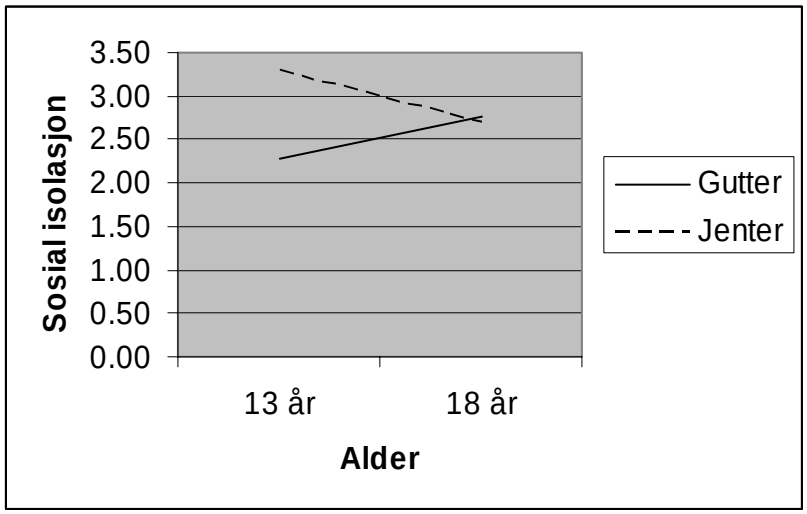
$$t = \frac{\hat{b}_1 - \hat{b}_2}{SE(\hat{b}_1 - \hat{b}_2)} = \frac{0.096 - (-0.120)}{0.081} = 2.67$$

og vi må forkaste hypotesen om at de to regresjonskoeffisientene er like.
Resultatet forteller oss med andre ord at det er store kjønnsforskjeller i selvrapportert ensomhet blant de yngste, men kjønnsforskjellen blir mindre med økende alder. Det med andre ord samspill mellom kjønn og alder.

g) Hensiktet med produktleddet er å fange opp samspillet. Koeffisienten for kjønn er nå ikke lenger et generelt mål for kjønnsforskjell, siden kjønnsforskjellen varierer med alder. Nevnte koeffisient er kjønnsforskjellen blant dem som har verdien null på den andre variabelen som inngår i samspillsleddet, dvs. blant 13-åringene. På samme måte er koeffisienten for alder aldereffekten blant de med null på kjønnsvariabelen, dvs. gutter. Som en ser er denne den samme som i den kjønnsspesifikke modellen for gutter (modell #2) – slik en skulle vente. Alderseffekten blant jenter blir lik summen av alderkoeffisienten og samspillskoeffisienten: 0.096 + (–0.215) = –0.119, som er (nesten) det samme som i modell #1. For å beregne kjønnsforskjellen i ulike aldersgrupper må en ta kjønnsvariabelens koeffisient og legge til samspillsleddet etter å ha multiplisert sistnevnte med alder.

h) Vi får følgende resultat:

	13 år	18 år
Gutter:	2.28	2.76
Jenter:	3.285	2.69



Vi ser med andre ord at kjønnsforskjellene bare er tilstede blant de yngre. Blant 18-åringene er det praktisk talt ingen forskjell mellom gutter og jenter med hensyn til selvrapportert isolasjon.

i) En dummyvariabel er en dikotom variabel, vanligvis kodet 0 og 1, som representerer en kvalitativ variabel. Sammenhengen mellom sosial isolasjon og variabelen "bor med begge foreldre" er signifikant ($t = -0.259/0.106 = 2.44$) på 5 prosentnivået. Resultatet viser at de som bor med begge foreldre rapporterer mindre sosial isolasjon enn de som ikke gjør det, kontrollert for kjønn og alder. Resultatet kan muligens tyde på at skilsmissebarn mv. føler sterkere sosial isolasjon, eksempelvis pga. mindre kontakt med den fraværende av foreldrene, og/eller fordi barn av aleneforeldre har større tilbøyelighet til å isolere seg i forhold til jevnaldrene.

j) Resultatet viser at de med god foreldrekontakt rapporterer vesentlig mindre sosial isolasjon enn de med god foreldrekontakt, alt annet konstant. Forskjellen er klart signifikant ($t = -0.541/0.066 = 8.20$). Videre er koeffisienten for "bor med begge foreldre" nær null, og ikke lenger signifikant ($t = -0.023/0.108 = 0.21$). Det betyr at det ikke spiller noen rolle for isolasjonen om man bor med én eller begge foreldre, gitt graden av foreldrekontakt. Det er med andre ord foreldrekontakten som sådan som er avgjørende, ikke om man bor med én eller begge. Sammenhengen mellom isolasjon og hvem man bor sammen med i modell #4 var med andre ord spuriøs. Denne spuriøse korrelasjonen må være forårsaket av en korrelasjon mellom foreldrekontakt og hvem man bor sammen med. Den innsiktsfulle kandidat vil også innse at denne korrelasjonen må være positiv og at de som bor sammen med begge foreldre har bedre foreldrekontakt enn de som bor sammen med bare én av dem.

k) Variabelen foreldrekontakt har fire verdier, 0, 1, 2 og 3. En av disse må velges som referansekategori, og det konstrueres én dummyvariabel for hver av de tre gjenværende verdiene. Vi må således konstruere tre dummyvariabler. Dersom sammenhengen mellom variabelen foreldrekontakt og sosial isolasjon ikke er helt lineær, vil denne fremgangsmåten være å foretrekke. Den metoden som er benyttet i modell #5 forutsetter linearitet.