

Skoleeksamen i SOS1120 - Kvantitativ metode

Tirsdag 30. mai 2017 (4 timer)

Hjelpemidler

- Ordbok
- Alle typer kalkulatorer
- Lærebok (det er mulig mulig å ha med en annen, tilsvarende pensumbok, som erstatning for den på pensumlisten).
- Utskrift av online ressurser til læreboka: Formelark og kapittelet om multippel regresjon
- Det er lov med notater og post-it-lapper i boka

Eksamensoppgaven

Denne eksamenen består av to deler. Den første delen er flervalgsoppgaver (multiple choice), mens den andre delen består av ordinære oppgaver. Alle oppgaver skal besvares. Enkelte steder er fagtermer også angitt på engelsk slik det brukes i pensumboka. Dette gjøres første gang ordet brukes.

Del 1 besvares på oppgavearket ved å krysse av for ett svaralternativ. Det skal ikke legges ved noen utregninger i denne delen.

Del 2 besvares på eget ark. For alle oppgaver der det er utregninger skal utregningen vises. Konklusjoner og begrunnelser skrives med fullstendige setninger.

Sensur, begrunnelse og klage

Sensur for eksamen faller 20. juni kl. 14.00. Sensuren publiseres i StudentWeb ca. kl. 15.00 samme dag. Sensuren regnes som mottatt av studentene når den publiseres i StudentWeb.

Vi minner om at kandidater som ønsker begrunnelse må be om dette **senest 1 uke etter at sensur er falt**. Frist for å klage på karakteren er tre uker etter at sensur er falt. Informasjon om rutiner for begrunnelser og klager ved ISS finnes på emnesiden.

Oppgavesettet består av **10 sider** inkludert denne.

Levering av besvarelsen

I del 1 krysser du direkte av på oppgavearket. **Del 1 skal leveres inn** som en del av din eksamensbesvarelse. På del 1 brukes det ikke gjennomslagsark, så du leverer kun ett eksemplar av del 1. **Husk å notere kandidatnummeret ditt på del 1.**

På del 2 skal du levere både originalen og kopien av besvarelsen. Kladd skal ikke leveres.

NB! Skriv hardt nok til at kopien blir leselig. Tusj penn kan ikke brukes.

Husk å notere deg kandidatnummeret ditt.

Lykke til!

Kandidatnummer_____

Del 1: Flervalgsoppgaver

Del 1 besvares på oppgavearket ved å krysse av for ett svaralternativ. Det skal ikke legges ved noen utregninger i denne delen.

Oppgave 1.1. En student noterer hvor mange timer hun studerer hver dag i løpet av ni dager og får følgende resultat:

6.2 5.0 4.3 7.4 5.8 7.2 8.4 1.2 6.3

Hva er gjennomsnitt og median for denne perioden?

- a. gjennomsnitt = 6.15 og median = 6.20
- b. gjennomsnitt = 5.76 og median = 6.20
- c. gjennomsnitt = 6.20 og median = 5.76
- d. gjennomsnitt = 5.76 og median = 6.15
- e. gjennomsnitt = 6.15 og median = 5.76

Oppgave 1.2. En student noterer hvor mange timer hun studerer hver dag i løpet av ni dager og får følgende resultat:

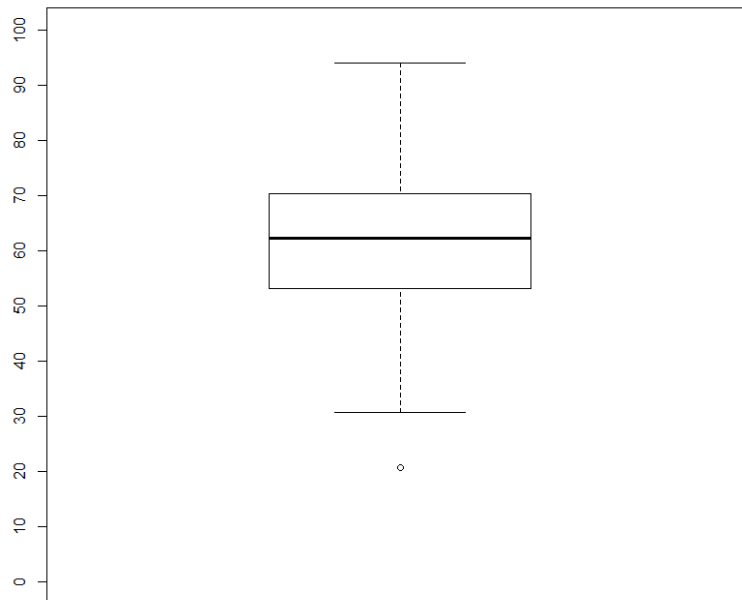
6.2 5.0 4.3 7.4 5.8 7.2 8.4 1.2 6.3

Hva betyr det at standardavviket er 2.11? (Engelsk: "standard deviation")

- a. Gjennomsnittet ligger 2.11 fra null
- b. p-verdien er 0.0174
- c. Datapunktene ligger i gjennomsnitt 2.11 timer fra gjennomsnittet
- d. 95% av dagene jobber hun mellom $\bar{x} \pm 1.96 \cdot 2.11$ timer
- e. At dataene er normalfordelt

Kandidatnummer_____

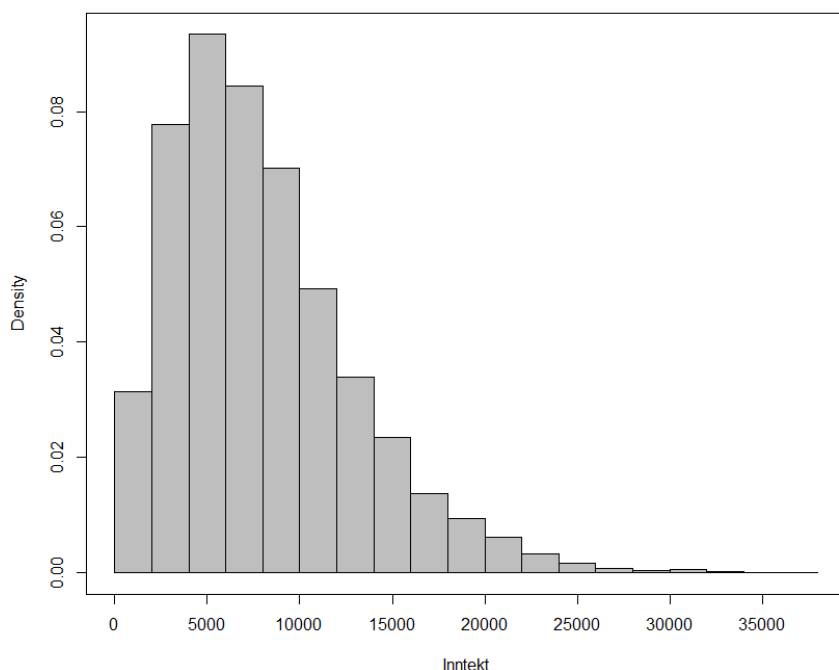
Oppgave 1.3. På en eksamen i kvantitative metoder er det mulig å få opptil 100 poeng.
Boksplotet nedenfor viser fordelingen for alle som tok eksamen.



Omtrent 25% av studentene fikk mindre enn...

- a. 30 poeng
- b. 53 poeng
- c. 62 poeng
- d. 70 poeng
- e. 59 poeng

Oppgave 1.4. Histogrammet nedenfor viser inntektsfordelingen i et utvalg.



Hvilket av følgende utsagn er riktige:

- a. gjennomsnittet er lavere enn medianen
- b. gjennomsnittet er lik 2 ganger standardavviket
- c. gjennomsnittet er høyere enn medianen
- d. gjennomsnittet er lik medianen
- e. man ikke si noe om gjennomsnittet fra figuren

Oppgave 1.5. Bruk standard normalfordelingen. (engelsk: "normal distribution"). Hvilken sannsynlighet tilsvare $Z > -1.62$? Med andre ord: hvor stor andel av fordelingen ligger over -1.62 ?

- a. 0.0044
- b. 0.0526
- c. 0.9474
- d. 0.9956
- e. 62%

Oppgave 1.6. Bruk standard normalfordelingen. Hvilken sannsynlighet tilsvare $Z < 1.62$? Med andre ord: hvor stor andel av fordelingen ligger under 1.62 ?

- a. 0.0044
- b. 0.0526
- c. 0.9474
- d. 0.9956
- e. 62%

Kadidatnummer_____

Oppgave 1.7. Bruk standard normalfordelingen. Hvilken Z-verdi gir at andelen med lavere verdi enn Z er 0.975?

- a. 1.96
- b. 1.68
- c. -1.96
- d. 2.00
- e. 0.025

Oppgave 1.8. Gjennomsnittlig samlet inntekt i Norge, var i 2014 kr 340 330 med standardavvik 655 803. Hvis man trekker et tilfeldig utvalg personer vil målingen ha en usikkerhet uttrykt ved standardfeilen. (Engelsk: "standard error"). Hvis man trekker et utvalg på 200 personer, hva er da standardfeilen?

- a. 3 279
- b. 46 372
- c. 1.93
- d. 655 803
- e. 1701.5

Oppgave 1.9. Gjennomsnittlig samlet inntekt i Norge, var i 2002 kr 205 862 med standardavvik 998 847. Hvis man trekker et tilfeldig utvalg på 200 personer, hva er da sannsynligheten for at gjennomsnittet \bar{x} i dette utvalget er høyere enn kr. 500.000? Vi kan skrive dette som: $P(\bar{x} > 500\ 000)$?

- a. 0.500
- b. 0.050
- c. 0.010
- d. 0.001
- e. praktisk talt null

Oppgave 1.10. Hva betyr det når vi snakker om samplingfordelingen til et gjennomsnitt? (Engelsk: samplingfordelingen = "Sampling distribution", utvalg = "sample")

- a. Fordelingen i utvalget
- b. Vår beste gjetning på fordelingen i populasjonen ut fra utvalget
- c. Ved et enkelt tilfeldig utvalg fra en normal populasjon
- d. Fordelingen av alle gjennomsnitt det er mulig å få hvis man trakk utvalget veldig mange ganger
- e. Den fordelingen vi bruker for å regne ut standardfeilen

Kandidatnummer_____

Oppgave 1.11. Hva er standardfeil? (Engelsk: "standard error")

- a. Et mål på hvor mye vi har bommet på den sanne verdien
- b. Et mål på standardavviket i populasjonen
- c. Vår beste gjetning på standardavviket i samplingfordelingen
- d. Det kvadrerte avviket fra gjennomsnittet
- e. Vår beste gjetning på gjennomsnittet i samplingfordelingen

Oppgave 1.12. Se for deg at du intervjuer et tilfeldig utvalg på 10 arbeidstakere og spør hvor lang reisevei de har til jobben (f.eks. målt i antall kilometer). For disse 10 kan du regne gjennomsnittlig reisevei. Se så for deg at du gjentar denne prosedyren veldig mange ganger med nye tilfeldige utvalg, og for hver gang noterer deg gjennomsnittet i hvert av utvalgene. Hvis du så lager et histogram av disse gjennomsnittene, så representerer dette histogrammet...

- a. en eventuell skjevhet i utvalgsmetoden
- b. den sanne reiseveien i populasjonen
- c. et enkelt tilfeldig utvalg
- d. samplingfordelingen til gjennomsnittet
- e. "de store talls lov"

Oppgave 1.13. Når man regner et 95% konfidensintervall sier man ofte at man er «95% sikker». Hva mener man egentlig med dette?

- a. 95% av observasjonene ligger innenfor dette intervallet
- b. Sannsynligheten for at man tar feil er 0.05
- c. Intervallet er regnet ut med en metode som fanger den sanne verdien 95% av gangene
- d. At man bruker normalfordelingen
- e. At intervallet er ± 1.96 ganger standardfeilen

Del 2: Utrekningsoppgaver

Del 2 besvares på eget ark. For alle oppgaver der det er utregninger skal utregningen vises. Konklusjoner og begrunnelser skrives med fullstendige setninger.

Oppgave 1.14. Tabellen viser andel menn 40-45 år som vurderer egen helse som god eller mindre god. Utvalget er et tilfeldig utvalg menn i Oslo 2000-2001 i alderen 40-45 år. (Kilde: Humbro undersøkelsen).

	God	Mindre god	(N)
Ytre vest	641	71	712
Indre vest	358	54	412
Ytre øst	756	226	982
Indre øst	205	190	395

- Formuler en nullhypotese og alternativ hypotese om sammenhengen mellom bosted og helse.
- Hvis nullhypotesen er sann, hvor mange ville du da forvente i tabellcellen for god helse i ytre øst? (Dette er første trinn du må gjøre for å regne kji-kvadrattesten). Vis utregningen.
- Hvis nullhypotesen er sann, hvor mange ville du da forvente i tabellcellen for mindre god helse i ytre vest? Vis utregningen.
- Hvor mange frihetsgrader har kji-kvadrattesten for denne tabellen?
- Kjikvadrattesten gir $\chi^2 = 238.3$. Hva er konklusjonen på testen?

Oppgave 1.15.

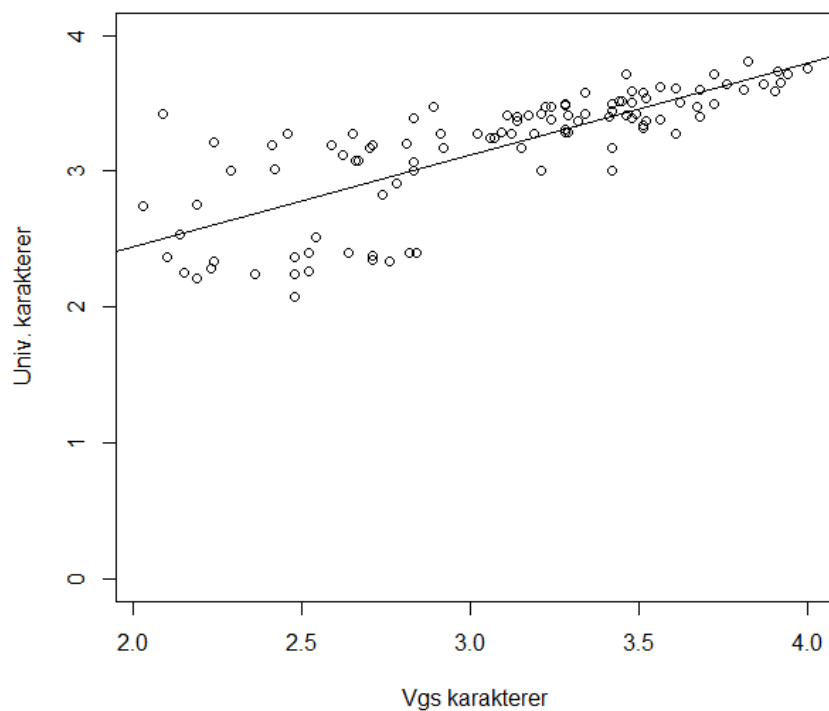
I en levekårsundersøkelse blant innsatte i norske fengsler (N= 264) ble det oppgitt at 38 prosent hadde vokst opp i en familie der noen hadde narkotika eller alkoholproblemer.

- Hva er standardfeilen til dette estimatet? Vis utregningen.
- Regn ut et 95% konfidensintervall for dette estimatet og gi en tolkning av resultatet.

Oppgave 1.16. I en amerikansk studie ble det trukket et tilfeldig utvalg på 105 ungdommer som så ble fulgt til de avsluttet høyere utdanning. Det ble undersøkt sammenhengen mellom gjennomsnittlig karakterer i videregående skole med gjennomsnittlig karakter etter endt høyere utdanning ved hjelp av lineær regresjon. Begge karakterskalaer går fra 0 til 4, der 4 er best.

(Oppgave 1.16 fortsetter).

Plottet nedenfor viser dataene og tilpasset regresjonslinje, $y = \alpha + \beta \cdot x$. I regresjonslinja er Y-variabelen karakter ved høyere utdanning, og x er karakter fra videregående skole. Regresjonslinja er estimert til å være: $y = 1.097 + 0.675 \cdot x$. Standardfeilen for β er 0.053.



- Ved å se på plottet, er en lineær regresjon en rimelig metode for å beskrive sammenhengen? Begrunn svaret.
- Beskriv med egne ord hva estimatet for β betyr.
- Formuler en nullhypotese og en alternativ hypotese for β .
- Gjør en t-test for nullhypotesen og konkluder med valgt signifikansnivå. (Engelsk: "significance level"). Vis utregningen.
- Regn ut et 99 prosent konfidensintervall for β . (Engelsk: "confidence interval"). Vis utregningen og si hva intervallet betyr med egne ord.
- Hva er den predikerte karakteren fra høyere utdanning for en person som har 3.4 fra videregående skole?

(Oppgave 1.16 fortsetter).

- g. R^2 for denne regresjonsmodellen er 0.675. Gi en tolkning av R^2 . I hvilke sammenhenger er R^2 nyttig?
- h. Forklar med egne ord hva en residual er. Hvilken nytte kan vi ha av å se nærmere på residualene?

Oppgave 1.17. I et tilfeldig utvalg personer fra norske befolkning var det to grupper på henholdsvis 30 og 35 år gamle. Tabellen gir gjennomsnittlig inntekt, standardavvik og utvalgsstørrelse for begge grupper.

	30-åringer	35-åringer
Gjennomsnitt	408 806	479 008
Standardavvik	159 512	219 689
N	3 500	2 500

- a. Regn ut standardfeilen til gjennomsnittet for 30-åringer.
- b. Gi et 90% konfidensintervall for gjennomsnittlig inntekt for gruppen av 30-åringer. Gi en tolkning av hva dette konfidensintervallet betyr.
- c. Regn ut differansen for gruppene og standardfeilen til denne differansen.
- d. Regn ut et 95% konfidensintervall for differansen mellom gruppene. Si med egne ord hva dette intervallet betyr.
- e. Er forskjellen statistisk signifikant på signifikansnivå 0.01? Gjør en t-test og konkluder.

Oppgave 1.18. I en britisk studie av alkoholkonsum (Britton et al 2015, *BMC Medicine*, 13:47) ble det estimert hvordan alkoholkonsum endret seg med alderen.

Dataene er fra et tilfeldig utvalg (N=1485) av befolkningen bosatt på vestkysten av Skottland, og det var omtrent like mange av hvert kjønn. I aldersgruppen 50 til 80 år, ble det konsumert i gjennomsnitt 0.187 færre alkoholenheter i uka for hvert år eldre man ble. Dette estimatet har standardfeil 0.051. Menn på 50 år drakk i gjennomsnitt 15 enheter i uka. 50-årige kvinner drakk i gjennomsnitt 3.5 enheter i uka. (En alkoholenhet tilsvarer 33cl øl eller et lite glass vin).

Ta utgangspunkt i at variabelen for alder er omkodet slik at den tar verdien 0 ved 50 år og øker med 1 for hvert år. Med andre ord: variabelen *alder* har verdier fra 0 til 30. Variabelen *kjønn* er en indikatorvariabel med verdiene 0 for menn og 1 for kvinner.

(Oppgave 1.18 fortsetter).

- a. Skriv opp formelen for en multippel lineær regresjonsmodell som beskriver hvordan alkoholkonsumet for menn og kvinner varierer med alder i denne aldersgruppen. Angi regresjonsparameterne med tall.
- b. Det er noe usikkerhet i hvor mye alkoholkonsumet øker med alderen. Gi et 95 prosent konfidensintervall for dette estimatet. Forklar med egne ord hva dette intervallet betyr når det gjelder alkoholkonsum.
- c. Gjør en hypotesetest av den estimerte regresjonsparameteren. Sett opp hypoteser og gjennomfør testen, og konkluder.
- d. Det viser seg at kvinner ikke reduserer alkoholbruket like mye med alderen som menn. For hvert år eldre de blir reduserer de i gjennomsnitt alkoholforbruket med 0.025 enheter i uka. Skriv nå en regresjonsmodell som viser endringer for menn og kvinner som viser denne forskjellen.