NTNU

KANDIDAT

10083

PRØVE

SØK1004 1 Statistikk for økonomer

Emnekode	SØK1004
Vurderingsform	Skriftlig eksamen
Starttid	26.11.2024 14:00
Sluttid	26.11.2024 18:00
Sensurfrist	17.12.2024 22:59
PDF opprettet	16.12.2024 16:30

Seksjon 1

Oppgave	Tittel	Oppgavetype
ℹ	Front page	Informasjon eller ressurser
1	1	Langsvar
2	2	Langsvar
3	3	Langsvar
4	4	Langsvar
5	5	Langsvar
6	6	Langsvar

1 1

Spørsmål 1 (16p)

Forklar kort følgende begrep

1. Frekvensfordeling
2. Korrelasjonskoeffisient
3. Simpson’s paradoks
4. Ordinal variabel
5. Sentralgrenseteorem
6. «Matched sample design»
7. Trimmet gjennomsnitt
8. Statistisk inferens

Skriv ditt svar her

1. **Frekvensfordeling**

I noen tilfeller deler man data inn i "klasser". Man tar utgangspunkt i et sett med innsamlet data, finner "passende" klasser der man deretter plasserer de individuelle dataverdiene. Eksempelvis dersom det er foretatt en undersøkelse der deltagerne skal gi et tall mellom 1 og 100 kan de ulike klassene være 1-10, 11-20, 21-30 osv. opp til 100. Viktig når det kommer til hvordan disse klassene er formet, er at ingen klasser overlapper. Slik at ingen verdier kan passe inn i flere klasser. Alle tall som blir samlet inn i undersøkelsen som er fra og med 1 til og med 10, vil plasseres i den første klassen. Frekvensfordelingen forteller hvor mange tall som befinner seg i hver enkelt klasse. Relativ frekvensfordeling finner man ved å ta frekvensen til hver enkelt klasse og dele på antallet deltagere. Prosent frekvensfordeling finner man ved å gange den relative frekvensfordelingen med 100.

2. **Korrelasjonskoeffisient**

Korrelasjonskoeffisienten forteller oss noe om sammenhengen mellom to variable. Hvorvidt den er positiv eller negativ, og hvorvidt sammenhengen er sterk eller svak. En korrelasjonskoeffisient lik 1 sier man er perfekt positiv, mens en korrelasjonskoeffisient lik -1 sier man er perfekt negativ. Dersom korrelasjonskoeffisienten er lik 0, er det ingen sammenheng mellom variablene.

Korrelasjonskoeffisienten er altså alltid et tall mellom -1 og 1.

3. **Simpson's paradoks**

4. **Ordinal variabel**

En ordinal variabel er en variabel som kan være både numerisk eller ikke-numerisk, og til forskjell fra en nominal variabel, kan en ordinal variabel rangeres.

5. **Sentralgrenseteoremet**

Enhver fordeling kan tilnærmes som normalfordelt når størrelsen på utvalget blir stor. En sier ofte at dersom n blir høyere enn 30 kan man bruke en normalfordeling.

6. **"Matched sample design"**

"Matched sample design" går ut på hvordan utvalget "velges". Og at for å få best resultater, resultater som kan fortelle noe om hele populasjonen, må utvalget være representativt for

populasjonen. Matched sample design handler da om at utvalget og populasjonen i så stor grad som mulig "matcher".

7. **Trimmet gjennomsnitt**
Gjennomsnittet man kommer frem til etter man har "trimmet" utvalget for uteliggere, altså unaturlig store eller unaturlig små verdier, eller andre statistiske "feil".

8. **Statistisk inferens**
Resultater av en statistisk analyse. Som regel blir utvalg valgt med en hensikt om at resultatene fra undersøkelsen kan anvendes til hele populasjonen utvalget er tatt fra. Den statistiske inferensen er de endelige resultatene fra en undersøkelse eller en analyse, som forhåpentligvis kan anvendes og fortelle noe om hele populasjonen.

Ord: 403

Knytte håndtegninger til denne oppgaven?
Bruk følgende kode:

6055935

2 2

Spørsmål 2 (24p)

En iskremkiosk i sentrale Norge er åpen i månedene 3-10. I tabellen nedenfor kan du se totalt salg for hver uke og antall dager over 20 grader.

Måned	Antall dager over 20 grader (x)	Salg i NOK (tusen kroner) (y)
3	2	2
4	12	9,4
5	13	9
6	18	16,9
7	18	22,2
8	16	18,3
9	15	14,8
10	14	14,9

a) Beregn gjennomsnitt, median, varians og standardavvik for månedlig salg ved iskrembaren (y). Kommenter resultatene (6p)

b) Vil du si at det er uteliggere i salg i NOK? Forklar hvorfor eller hvorfor ikke. (4p)

c) Beregn kovariansen mellom antall dager over 20 grader og salg i NOK. Hva er tolkningen av kovariansen? (7p)

d) Beregn korrelasjonskoeffisienten mellom antall dager over 20 grader og salg i NOK (7p). Bruk $S_x = 5,13$. Tolk resultatene.

Skriv ditt svar her

Hoveddel av oppgaven besvares på ark.

Begrunner valg av formler. Ettersom dataene er samlet fra dager med over 20 grader, og ikke *alle* dagene i månedene 3-10, kan man si at det er tatt ut et utvalg. Derfor bruker jeg formler for utvalgs-gjennomsnitt, utvalgs-varians osv.

a)
Median: median kan være et mer hensiktsmessig mål på "midt"-verdien i en samling data, dersom det eksempelvis foreligger uteliggere. I deloppgave c) konkluderer vi med at det ikke er noen uteliggere i utvalgsdataene, så utregning av gjennomsnitt er mer presist og passer best. Median = 14.85 er ikke veldig langt unna gjennomsnittet = 13.4375 som også er en indikasjon på at det ikke er noen uteliggere i dataverdiene.

c) Korrelasjonskoeffisienten lik 0.918 forteller oss at det er en sterk positiv sammenheng mellom antall dager i løpet av månedene med over 20 grader (x), og omsetning den måneden (y). Altså jo flere dager med over 20 grader, jo større omsetning har iskremkiosken den måneden.

Ord: 159

Knytte håndtegninger til denne oppgaven?
Bruk følgende kode:

2 5 7 8 5 2 9

Håndtegning 1 av 3

Fill out Question Code and Test Information on every sheet. Fyll inn oppgavekode og emneinformasjon på alle skisseark.

Question Code Oppgavekode	Date Dato	Subject code Emnekode	Candidate ID KandidatID	Question no Oppgavenr	Page number Sidetall	Number of pages Antall ark
2 5 7 8 5 2 9	26.11.24	SØK1004	10083	2	1	9

Writing area Skriveområde

Spørsmål 2

Antagelser / begrunnelser på PC.

a) Gjennomsnitt:

$$\frac{2 + 9{,}4 + 9 + 16{,}9 + 22{,}2 + 18{,}3 + 14{,}8 + 14{,}9}{8}$$

$$= \frac{107.5}{8} = 13{,}4375$$

$$\bar{y} = 13{,}4375$$

Median:

2, 9, 9.4, 14.8, 14.9, 16.9, 18.3, 22.2

De to verdiene i midten: 14.8 og 14.9

Median: $\frac{14.8 + 14.9}{2} = 14.85$

Varians:

$$s_Y^2 = \frac{1}{8-1}\Big((2 - 13{,}4375)^2 + (9.4 - 13.4375)^2 + (9 - 13.4375)^2$$

$$+ (16.9 - 13{,}4375)^2 + (22.2 - 13.4375)^2 + (18.3 - 13.4375)^2$$

$$+ (14.8 - 13.4375)^2 + (14.9 - 13.4375)^2\Big)$$

$$\Rightarrow \frac{1}{7}(130{,}8 + 16.3 + 19.69 + 11{,}99 + 76.78 + 23{,}64$$

$$+ 1{,}86 + 2{,}14) \Rightarrow \frac{1}{7}(283.2) = 40{,}46$$

Standardavvik:

$$s_y = \sqrt{40{,}46} = 6.36$$

Form Number: INSPERA-02-160919-1 Provider: CADA Sistemas Informáticos www.cada.es

Håndtegning 2 av 3

Fill out Question Code and Test Information on every sheet. Fyll inn oppgavekode og emneinformasjon på alle skisseark.

Question Code Oppgavekode	Date Dato	Subject code Emnekode	Candidate ID KandidatID	Question no Oppgavenr	Page number Sidetall	Number of pages Antall ark
2578529	26.11.24	SØK1004	10083	2	2	9

Writing area Skriveområde

$\bar{y}$ = 13,4375 : Gjennomsnittlig salg på dagene over 20 grader, er på 13 437,5 kr

Median: 14.85 : Utdypning skrevet på PC

Varians = 40,46, forteller noe om hvordan dataene varierer i forhold til hverandre, men tallverdien man får ved å regne ut varians forteller ikke så mye.

Standardavvik: forteller oss den gjennomsnittlige avstande fra gjennomsnittet. Her: lik 6.36

b) Det er ikke uteliggere i salg av NOK. Dette kan man bruke den empiriske regel til å finne ut av. Den hevder blant annet at ca alle dataverdier er innen for 3 standardavvik fra gjennomsnittet. Det tilsvarer z-score mindre eller større enn 3,

$z \leq |3|$

Vi kan regne ut dette for høyeste og laveste salg i NOK:

$$z = \frac{2 - 13{,}4375}{6.36} = -1{,}79$$

$$z = \frac{22.2 - 13{,}4375}{6.36} = 1{,}377$$

Ingen av disse salgene ~~er~~ har z-verdi høyere/lavere enn |3|.

Form Number: INSPERA-02-160919-1 Provided by: CADA Sistemas Informáticos www.cada.es

Håndtegning 3 av 3

Fill out Question Code and Test Information on every sheet. Fyll inn oppgavekode og emneinformasjon på alle skisseark.

Question Code Oppgavekode	Date Dato	Subject code Emnekode	Candidate ID KandidatID	Question no Oppgavenr	Page number Sidetall	Number of pages Antall ark
2 5 7 8 5 2 9	26.11.24	SØK1004	10083	2	3	9

Writing area Skriveområde

c) Kovarians : $s_{xy} = \frac{\sum_{i=1}^{8} (x_i - \bar{x})(y_i - \bar{y})}{n-1}$

Må finne $\bar{x}$, bruker samme formel som for x i deloppgave a).

$\bar{x} = 13,5$

$s_{xy} = \frac{1}{7}(131,53 + 6.05 + 2.22 + 15.58 + 39.43 + 12.15 + 2.04 + 0.73)$

$s_{xy} = \frac{1}{7} \cdot 209.73 = \underline{29.96}$

En kovarians på 29.96 forteller oss at det er en positiv sammenheng mellom de to variablene, men ikke mye mer.

d) Korrelasjonskoeffisient:

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

$$= \frac{29.96}{5.13 \cdot 6.36} = \underline{0,918}$$

Korrelasjonskoeffisient lik 0,918, forteller oss at det er en sterk positiv sammenheng mellom x og y.

Form Number:INSPERA-02-160919-1 Provider:© CIDA Sistemas Informáticos www.cida.es

3 **3**

Spørsmål 3 (20p)

Vurder den følgende hypotesen:

H_0: $\mu \leq 12$

H_a: $\mu > 12$

Anta at utvalgsgjennomsnittet er x=14, utvalgsstandardavviket er s=4,27 og utvalgsstørrelsen er n=25

a) Beregn t-verdien (test statistic) (5p)

b) Hva forteller t-fordelingen deg om p-verdien? (5p)

c) Til α=0.05, hva er konklusjonen din? (5p)

d) Hva er forkastingsregelen ved å bruke kritisk verdi? (5p)

Skriv ditt svar her

Ord: 0

Knytte håndtegninger til denne oppgaven?
Bruk følgende kode:

5 0 5 9 0 5 4

Håndtegning 1 av 1

Fill out Question Code and Test Information on every sheet. Fyll inn oppgavekode og emneinformasjon på alle skisseark.

Question Code Oppgavekode	Date Dato	Subject code Emnekode	Candidate ID KandidatID	Question no Oppgavenr	Page number Sidetall	Number of pages Antall ark
5059054	26.11.24	SØK1004	10083	3	4	9

Writing area Skriveområde

Oppgave 3

$H_0: \mu \leq 12$ $H_A: \mu > 12$

$\bar{x} = 14$ $s = 4.27$ $n = 25$

a) t-verdi:

$$t = \frac{14 - 12}{\frac{4.27}{\sqrt{25}}} = 2.34$$

b) $df = n - 1 = 25 - 1 = 24$ (Finner i tabell)

Ved 24 frihetsgrader vil p-verdien være mellom ~~0,05 o~~ 0,025 og 0.01.

c) Med $\alpha = 0,05$ betyr det at p-verdien er lavere enn signifikansnivået, og vi kan derfor forkaste null-hypotesen.

d) Forkastingsregel ved å bruke kritisk verdi:

dersom $t_\alpha <$ teststatistikk kan nullhypotesen forkastes.

$t_{0.05}(24) = 1.711$ og $1.711 < 2.34$

Form Number: INSPERA-02-160919-1 provided by: CADA Sistemas Informáticos www.cada.es

4 4

Spørsmål 4 (18p)

Kredittkortgjelden for et utvalg av 60 husholdninger i USA hadde gjennomsnittlig gjeld på $10,912 med et utvalgsstandardavvik på $4,832

a) Beregn et 90 % konfidensintervall for gjennomsnittlig kredittkortgjeld for populasjonen av husholdninger i USA. Anta at populasjonen er normalfordelt. (8p)

b) Beregn et 95 % konfidensintervall for gjennomsnittlig kredittkortgjeld for populasjonen av husholdninger i USA. Anta at populasjonen er normalfordelt (6p)

c) Diskuter forskjellene mellom konfidensintervallene i a og b. Hvorfor er de ulike? (4p)

Skriv ditt svar her

Ord: 0

Knytte håndtegninger til denne oppgaven?
Bruk følgende kode:

5 6 0 2 2 8 4

Håndtegning 1 av 2

Fill out Question Code and Test Information on every sheet. Fyll inn oppgavekode og emneinformasjon på alle skisseark.

Question Code Oppgavekode	Date Dato	Subject code Emnekode	Candidate ID KandidatID	Question no Oppgavenr	Page number Sidetall	Number of pages Antall ark
560284	26.11.24	SØK1004	10083	4	5	9

Writing area Skriveområde

Spørsmål 4

$n = 60 \qquad \bar{x} = 10912 \qquad s = 4832$

a) 90% konfidensintervall

$$\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \qquad \alpha = 0.1$$

$df = n - 1 = 60 - 1 = 59$

$t_{\frac{\alpha}{2}} = t_{0.05}(59) = 1.671$

$\frac{s}{\sqrt{n}} = \frac{4832}{\sqrt{60}} = 623.8$

$\Rightarrow 10912 \pm 1{,}671 \cdot 623.8 = 10912 \pm 1042.37$

$KI = [9869.63 ; 11954.37]$

b) 95% konfidensintervall, $\alpha = 0.05$

$t_{\frac{\alpha}{2}} = t_{0.025}(59) = 2.001$

$\frac{s}{\sqrt{n}} = 623.8$

$10912 \pm 2.001 \cdot 623.8 \Rightarrow 10912 \pm 1248.22$

$KI = [9663.78 ; 12160.22]$

Form Number: INSPERA-02-160919-1 Provider: CADA Sistemas Informáticos www.cada.es

Håndtegning 2 av 2

Fill out Question Code and Test Information on every sheet. Fyll inn oppgavekode og emneinformasjon på alle skisseark.

Question Code Oppgavekode	Date Dato	Subject code Emnekode	Candidate ID KandidatID	Question no Oppgavenr	Page number Sidetall	Number of pages Antall ark
560284	26.11.24	SØK1004	10083	4	6	9

Writing area Skriveområde

c) Konfidensintervallet i oppg. 4a) forteller oss at vi med 90% sikkerhet kan si at populasjonsgjennomsnittet for kredittkortgjeld i husholdninger i USA er mellom $9869.63 og $11954.37.

Konfidensintervallet i oppgave 4b) forteller oss at ~~gjennomsnittlig kredittkort~~ vi med ~~90~~ 95% sikkerhet kan si at gjennomsnittlig kreditt kortgjeld for populasjonen av husholdninger i USA er mellom $9663.78 og $12160.22.

Konfidensintervallet i deloppgave b) er større. Dette er fordi konfidensnivået er høyere, 95% > 90%. Ettersom alt annet likt, må intervallet bli større for at man med høyere/større sikkerhet kan si at populasjonsgjennomsnittet ligger innenfor intervallet. Derfor er konfidensintervallene ulike.

Form Number:INSPERA-02-160919-1 Provider:CADA Sistemas Informáticos www.cada.es

5 **5**

Spørsmål 5 (12p)

En tidligere undersøkelse har undersøkt kostnadene på å leie biler i USA og fant et gjennomsnitt på $55 per dag for å leie en middels stor bil med et standardavvik på $8.88.

Anta at prosjektdirektøren ønsker å estimere populasjonens gjennomsnittlig daglige leiebilkostnader slik at det er 90 % sannsynlig at «sampling error» er $2 eller mindre.

Hvor stort utvalg trengs for å møte disse kravene?

Skriv ditt svar her

Ord: 0

Knytte håndtegninger til denne oppgaven?
Bruk følgende kode:

6 1 6 1 2 3 8

Håndtegning 1 av 2

Fill out Question Code and Test Information on every sheet. Fyll inn oppgavekode og emneinformasjon på alle skisseark.

Question Code Oppgavekode	Date Dato	Subject code Emnekode	Candidate ID KandidatID	Question no Oppgavenr	Page number Sidetall	Number of pages Antall ark
6 1 6 1 2 3 8	26.11.24	SØK1004	10083	5	7	9

Writing area Skriveområde

Spørsmål 5

$$\underbrace{\bar{X} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}}_{E}$$

Ettersom σ er kjent, lik \$8.88, benytter jeg meg av denne formelen.

E = feilmargin

Ettersom oppgaven spør om hvor stort utvalg som trengs for å møte kravene gitt av oppgaveteksten, kan man skrive om formelen, og løse for n.

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow n = \frac{(z_{\alpha/2})^2 \cdot \sigma^2}{E^2}$$

Oppgaveteksten "90% sannsynlig" $\Rightarrow \alpha = 0.1$

$$z_{\alpha/2} = z_{0,05} = 1.645$$

$E \leq 2$

$$n = \frac{(1.645)^2 \cdot (8,88)^2}{2^2} = \frac{2,7 \cdot 78.8544}{4}$$

$n = 53,22 \approx 54.$

Tolkning på nytt ark →

~~Utvalget må være minst 54 for at prosjekt-direktøren kan estimere populasjonens gjennom-snittlige leiekostnader slik at det er 90%~~

Form Number:INSPERA-02-160919-1 Provider:© CADA Sistemas Informáticos www.cada.es

Håndtegning 2 av 2

Fill out Question Code and Test Information on every sheet. Fyll inn oppgavekode og emneinformasjon på alle skisseark.

Question Code Oppgavekode	Date Dato	Subject code Emnekode	Candidate ID KandidatID	Question no Oppgavenr	Page number Sidetall	Number of pages Antall ark
6161238	26.11.24	SØK1004	10083	5	8	9

Writing area Skriveområde

Utvalget må være 54 slik at det er 90% sannsynlig at "sampling error" er $2 eller mindre

Form Number:INSPERA-02-160919-1 Provider:CADA Sistemas Informáticos www.cada.es

6 6

Spørsmål 6 (10p)
Et utvalg på 16 enheter har et standardavvik på 9,5. Test følgende hypotese med å bruke
α = 0.05.
H_0: $\sigma^2 \leq 50$
H_1: $\sigma^2 > 50$
Hva er konklusjonen din? Bruk både p-verditilnærmingen og kritisk verdi-tilnærmingen.

Skriv ditt svar her

Ord: 0

Knytte håndtegninger til denne oppgaven?
Bruk følgende kode:

4 2 3 8 4 9 8

Håndtegning 1 av 1

Fill out Question Code and Test Information on every sheet. Fyll inn oppgavekode og emneinformasjon på alle skisseark.

Question Code Oppgavekode	Date Dato	Subject code Emnekode	Candidate ID KandidatID	Question no Oppgavenr	Page number Sidetall	Number of pages Antall ark
4 2 3 8 4 9 8	26.11.24	SØK1004	10083	6	9	9

Writing area Skriveområde

Spørsmål 6

$H_0: \sigma^2 \leq 50$ $\alpha = 0.05$

$H_A: \sigma^2 > 50$

$n = 16$ $s = 9.5$

Ettersom hypotesene omhandler varians benytter jeg meg av kjikvadrattest.

Da er det hensiktsmessig å ~~benytte seg å~~ finne frihetsgrader $df = n-1 = 16-1 = 15$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \Rightarrow \frac{(16-1)\cdot 9.5^2}{50} = \frac{15 \cdot 90.25}{50}$$

$$= 27.075$$

Finner i tabell at $0.025 \leq$ p-verdi < 0.05.

Ettersom $\alpha = 0.05$ og p-verdi < 0.05 kan vi forkaste nullhypotesen om populasjonsvarians mindre eller lik 50.

Kommer til samme konklusjon ved kritisk verdi-tilnærmingen:

$$\chi^2_{0.05}(15) = 24.996$$

Kritisk verdi < teststatistikk $\Rightarrow$ kan forkaste H_0.

Form Number: INSPERA-02-160919-1 provided by: CADA Sistemas Informáticos www.cada.es