

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i STK1000 — Innføring i anvendt statistikk.

Eksamensdag: Torsdag 9. oktober 2008.

Tid for eksamen: 15:00–17:00.

Oppgavesettet er på 5 sider.

Vedlegg: Ingen

Tillatte hjelpemidler: Lærebok, ordliste for STK1000, godkjent kalkulator.

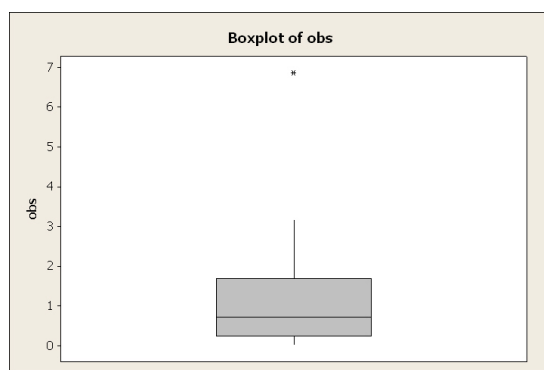
Kontroller at oppgavesettet er komplett før du begynner å besvare spørsmålene.

Husk å fylle inn kandidatnummer under.

Kandidatnr: _____

Alle 20 oppgaver teller likt. For hver oppgave skal du merke av for bare ett svaralternativ. Du får ett poeng for hvert riktige svar, maksimum 20 poeng. Dersom du svarer feil eller lar være å krysse av på et spørsmål, får du null poeng. Du blir altså ikke “straffet” med minuspoeng for å svare feil. *Lykke til!*

Oppgave 1. Figuren viser et boks-plott for 30 observasjoner.



Oppsummerende mål for det samme datasettet er gitt ved

Descriptive Statistics: obs

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
obs	30	0	1.185	0.258	1.416	0.004	0.241	0.722	1.681	6.876

(Fortsettes på side 2.)

De tre horisontale linjene i boks-plottet representerer

- a) minimum, median, maksimum
- b) minimum, gjennomsnitt, maksimum
- c) 1. kvartil, median, 2. kvartil
- d) 1. kvartil, median, 3. kvartil

Oppgave 2. Fra opplysningene i oppgave 1 kan vi konkludere med at fordelingen til observasjonene

- a) er normal, gitt at vi har mer enn 20 observasjoner
- b) har en hale mot høyre (høyreskjev)
- c) har en hale mot venstre (venstreskjev)
- d) er tilnærmet symmetrisk

Oppgave 3. Interkvartil-bredden (IQR) for datasettet i oppgave 1 er

- a) 1.44
- b) 6.872
- c) 0.481
- d) 1.681

Oppgave 4. Potensielle uteliggere (outliers) i datasettet fra oppgave 1 vil være observasjoner som er større enn

- a) 2.16
- b) 3.17
- c) 3.84
- d) 3.51

Oppgave 5. Dagtemperaturen i Ålesund i juli er tilnærmet normalfordelt med forventning 15 grader og standardavvik 5 grader. Hva er sannsynligheten for at dagtemperaturen en tilfeldig valgt dag i juli er lavere enn 10 grader?

- a) 0.0228
- b) 0.2072
- c) 0.1587
- d) 0.8413

Oppgave 6. Vi antar at høyden til kvinner i Norge er normalfordelt med forventning 170 cm og standardavvik 6.5 cm. Hvor høy må en kvinne minst være for å være blant de 5% høyeste (kvinnene) i befolkningen?

- a) 185.1
- b) 183.2
- c) 181.4
- d) 180.7

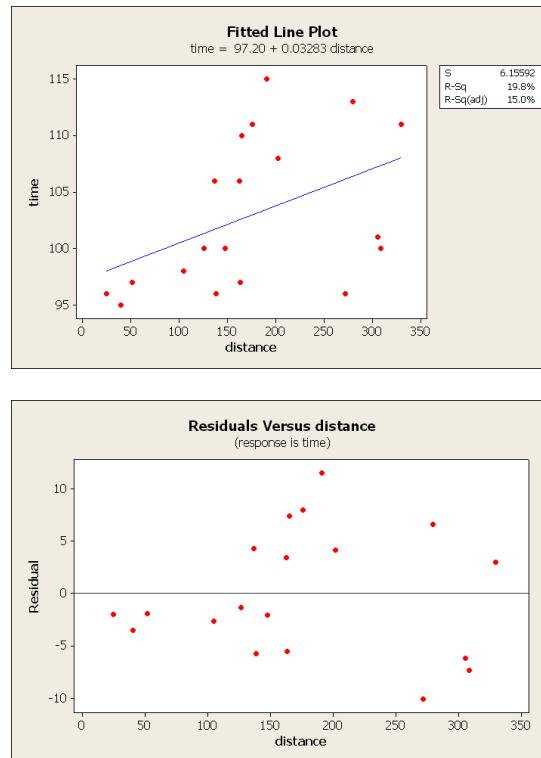
Oppgave 7. For å teste en persons reaksjonsevne med datamus, utføres følgende forsøk: I det man klikker innenfor en sirkel på skjermen, dukker en ny sirkel opp, og tiden det tar til man klikker i den nye sirkelen måles. Vi har data for 19 slike reaksjonstider i millisekunder ('time' i utskriften) for samme person, med tilhørende avstand ('distance') fra kursoren til senteret i den nye sirkelen i det denne dukker opp på skjermen. Resultatet fra en minste kvadraters regresjonsanalyse finnes nedenfor og på neste side. Analysen er tema for oppgave 7, 8, 9 og 10.

Regression Analysis: time versus distance

The regression equation is
time = 97.20 + 0.03283 distance

S = 6.15592 R-Sq = 19.8% R-Sq(adj) = 15.0%

(Fortsettes på side 3.)



Hva er riktig utsagn i forhold til opplysningene gitt ovenfor?

- a) Reaksjonstiden er forklaringsvariabel og avstanden responsvariabel
- b) Avstanden er forklaringsvariabel og reaksjonstiden er responsvariabel
- c) Avstanden øker med 0.03282 når tiden øker ett millisekund
- d) Forklaringsvariabelen øker med 0.03282 når responsen øker med ett millisekund

Oppgave 8. Korrelasjonen mellom avstand og reaksjonstid er

- a) 0.415 b) 0.150 c) 0.112 d) 0.445

Oppgave 9. R-Sq=19.8% i utskriften betyr

- a) 19.8 % av reaksjonstidene avhenger av avstanden
- b) 19.8 % av variasjonen i reaksjonstid forklares av den lineære sammenhengen med avstand
- c) 19.8 % av variasjonen i reaksjonstid forklares av prediksjon
- d) 19.8 % av variasjonen fra person til person kan forklares med avstanden

Oppgave 10. Den første målingen gav at personen hadde en reaksjonstid på 115 millisekunder ved en avstand på 190.7 enheter. Residualet til dette datapunktet er

- a) 9.34 b) 5.02 c) 11.54 d) 10.39

(Fortsettes på side 4.)

Oppgave 11. En observator for parameteren p er forventningsrett (unbiased) dersom

- a) observatoren har liten varians
 b) utvalgsfordelingen er tilnærmet normal
 c) observatoren har forventning 0
 d) observatoren har forventning p

Oppgave 12. Hvis A og B er to disjunkte hendelser er den betingede sannsynligheten $P(A|B)$ lik:

- a) $P(A)$ b) $\frac{P(A)}{P(B)}$ c) $\frac{P(B)}{P(A)}$ d) 0 (null)

Oppgave 13. En medisinsk test har sannsynlighet 0.85 for å avsløre om en pasient har en bestemt sykdom. Sannsynligheten for at testen viser at pasienten er syk når pasienten egentlig er frisk, er 0.03. Man vet at 1% av befolkningen har sykdommen. Da er sannsynligheten for at pasienten faktisk har sykdommen når testen viser det

- a) 0.338 b) 0.255 c) 0.223 d) 0.201

Oppgave 14. La X være en diskret tilfeldig variabel med sannsynlighetsfordeling

x_i	0	1	2	3	4	5	6	7
$P(X = x_i)$	0.05	0.20	0.28	0.25	0.14	0.05	0.02	0.01

Hva er $P(1 < X < 4)$?

- a) 0.53 b) 0.73 c) 0.87 d) 0.67

Oppgave 15. Den diskrete tilfeldige variabelen X har sannsynlighetsfordeling

x_i	0	1	2	3	4
$P(X = x_i)$	0.2	0.2	0.3	0.2	0.1

Hva er forventningen μ_X og standardavviket σ_X ?

- a) $\mu_X = 1.8$ og $\sigma_X = 3.25$ b) $\mu_X = 1.8$ og $\sigma_X = 1.25$
 c) $\mu_X = 2.0$ og $\sigma_X = 4.82$ d) $\mu_X = 1.8$ og $\sigma_X = 1.56$

Oppgave 16. To tilfeldige variable X og Y har forventning $\mu_X = \mu_Y = 0$ og standardavvik $\sigma_X = \sigma_Y = 1$. Korrelasjonen mellom X og Y er 0.5. Da har $3X - Y$ varians

- a) 7 b) 3 c) 5 d) 9

Oppgave 17. I en midtveiseksamen som denne, med 20 spørsmål med 4 svaralternativer på hver, er forventet antall riktige svar dersom man krysser av helt tilfeldig (uavhengig fra spørsmål til spørsmål)

- a) 4 b) 5 c) 8 d) 10

(Fortsettes på side 5.)

Oppgave 18. Standardavviket til antall riktige svar i situasjonen ovenfor vil være

- a) 2.94 b) 3.75 c) 1.94 d) 1.71

Oppgave 19. Ferske tall viser at 15% av alle som nå blir fedre, tar ut pappaperm i mer enn de obligatoriske seks ukene. I et utvalg på 200 nybakte fedre vil fordelingen til antallet som tar ut mer enn seks uker pappaperm være tilnærmet

- a) $N(15,5.05)$
 b) $N(30,6.05)$
 c) $N(15,0.025)$
 d) $N(30,5.05)$

Oppgave 20. I oppgave 5 fikk vi vite at høyden til kvinner i Norge følger en $N(170,6.5)$ -fordeling. Sannsynligheten for at gjennomsnittshøyden for 9 tilfeldig valgte kvinner i Norge er mindre enn 168 cm er

- a) 0.821 b) 0.378 c) 0.622 d) 0.179

Det var det!