

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

UNDERVEISEKSAMEN I: STK1000 – INNFORING I ANVENDT STATISTIKK.
EKSAMENSDAG: FREDAG 13.10.2006.
TID FOR EKSAMEN: KL. 09.00–11.00.
TILLATTE HJELPEMIDLER: LÆREBOK: MOORE & MCCABE “INTRODUCTION TO THE PRACTICE OF STATISTICS”. ORDLISTE FOR BRUK I STK1000, KALKULATOR.

OPPGAVESETTET ER PÅ 5 SIDER.

FOR HVERT SPØRSMÅL SKAL EN MERKE AV FOR BARE ETT SVARALTERNATIV. SE GJENNOM ALLE SVARALTERNATIVENE FØR DU KRYSSER AV.

KANDIDATNR. _____

Oppgave 1. Et stilk-og-blad-plott (stem and leaf) for et bestemt datasett ser slik ut:

1 6
2 2489
3 01123678
4 058
5 018
6 11
7 9

Datasettet består av heltall, der den minste observasjonen er 16 og den største er 79. Hvilket utsagn om fordelingen til dette datasettet er riktig?

- a) Fordelingen er tydelig symmetrisk
- b) Fordelingen er venstreskjev, siden de fleste observasjonene er små
- c) Fordelingen har en svak skjevhet mot høyre

Oppgave 2. Gjennomsnittet for datasettet i oppgaven ovenfor

- a) er mindre enn medianen, siden fordelingen er venstreskjev
- b) er 36.5
- c) er 40.0
- d) kan ikke beregnes fra et slikt plott

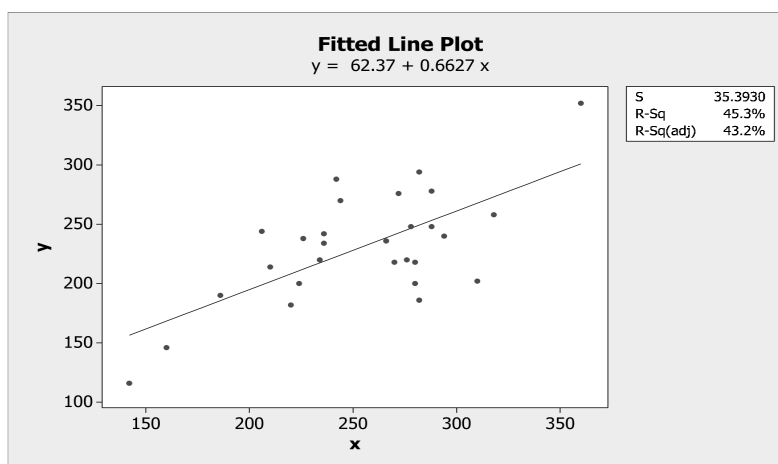
Oppgave 3. Høyt glukosenivå i blodet er tegn på sykdommen diabetes, en sykdom som blir stadig mer utbredt i den vestlige verden. Et tilfeldig utvalg på 180 individer fikk målt glukosenivået i blodet. Fem-talls-oppsummering av resultatene var

52 79 91 119 220

Hvor mange i utvalget hadde glukosenivå over 119?

a) 25 b) 135 c) 45

Oppgave 4. Et sett med 28 observasjonspar (x_i, y_i) er plottet mot hverandre sammen med minste kvadraters regresjonslinje i figuren nedenfor.



Nedenfor ser du også en utskrift (fra Minitab) med resultater knyttet til plottet ovenfor.

The regression equation is

$$y = 62.4 + 0.663 x$$

Predictor	Coef	SE Coef	T	P
Constant	62.37	36.86	1.69	0.103
x	0.6627	0.1428	4.64	0.000

S = 35.3930 R-Sq = 45.3% R-Sq(adj) = 43.2%

Korrelasjonen mellom x_i -ene og y_i -ene er

a) 0.453 b) 0.432 c) 0.673 d) 0.569

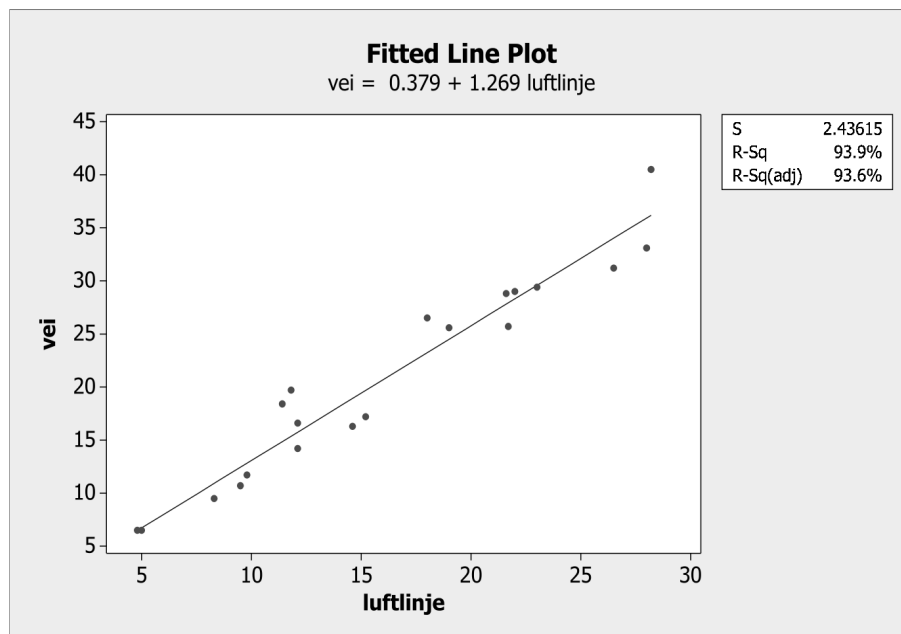
Oppgave 5. Punktet lengst til høyre er en innflytelsesrik observasjon. Hvilket utsagn er riktig?

- a) Uten dette punktet, hadde korrelasjonen mellom x og y vært sterkere
- b) Uten dette punktet, hadde korrelasjonen mellom x og y vært svakere
- c) Ett enkelt punkt kan ikke påvirke korrelasjonen

Oppgave 6. Avstanden mellom to steder på et kart er enkel å måle i luftlinje, dvs. den rett linjen mellom de to stedene. Hvordan relaterer denne avstanden seg til faktisk veiaavstand? Målinger av kartavstand og veiaavstand (i miles) mellom 20 par av steder i England er plottet nedenfor. Noen tall fra Minitab-analysen er også vedlagt.

The regression equation is
 $\text{vei} = 0.379 + 1.269 \text{ luftlinje}$

S = 2.43615 R-Sq = 93.9% R-Sq(adj) = 93.6%



Linjen i plottet er minste kvadraters regresjonslinje med veilengde som respons og luftlinje-avstand som forklaringsvariabel. Stigningstallet til regresjonslinjen er

- a) 1.269. Det betyr at for hver ekstra miles i luftlinje-avstand kan man forvente ca. 1.27 miles mer å kjøre.
- b) 0.379. Det betyr at for hver ekstra miles i luftlinje-avstand kan man forvente ca. 0.38 miles ekstra å kjøre.
- c) 1.269. Det betyr at for hver ekstra miles å kjøre kan man forvente ca. 1.27 miles mindre i luftavstand.

Oppgave 7. Verdien $R\text{-Sq} = 93.9\%$ i utskriften betyr at

- a) 93.9% av de 20 veilengdene følger en klar lineær sammenheng med avstandene i luftlinje.
- b) 93.9% av variasjonen i veilengdene kan forklares ved variasjonen i luftlinje-avstandene.
- c) 93.9% av variasjonen i veilengdene kan forklares ved den lineære sammenhengen mellom luftlinje- og vei-avstand.

Oppgave 8. Hvis et par av steder har et stort positivt residual, betyr det at

- a) veien mellom de to stedene er usedvanlig kronglete (svinger/bakker)
- b) veien mellom de to stedene er lenger enn mellom de fleste andre steder
- c) veien mellom de to stedene er usedvanlig rett og flat

Oppgave 9. Modellen som er tilpasset lengde-dataene har en svakhet, som består i at

- a) vi ikke har sjekket om dataene er normalfordelte med et normalfordelingsplott
- b) vi vet at linjen må gå gjennom origo og burde ha tilpasset en annen modell
- c) vi vet at linjen må gå gjennom origo og burde ha brukt $\text{vei} = 1.269$ luftlinje

Oppgave 10. Et nordisk gallup-firma baserer sine undersøkelser på telefonintervju med 1000 tilfeldig valgte voksne i hvert av de nordiske landene. I Sverige er populasjonen omtrent dobbelt så stor som i Norge. Betyr det at variabiliteten i de norske resultatene er mindre enn for de svenske?

- a) Ja.
- b) Ja, fordi andelen av populasjonen som er intervjuet er dobbelt så stor i Norge.
- c) Nei, det er bare størrelsen på utvalget som påvirker variabiliteten.
- d) Nei, den er mindre i Sverige.

Oppgave 11. Hvis hendelsene A og B er disjunkte, er de også

- a) uavhengige
- b) avhengige
- c) umulige hendelser

Oppgave 12. En diagnosemetode har sannsynlighet 0.99 for å avsløre at en person har en bestemt sykdom. Samtidig er sannsynligheten for å feildiagnostisere en frisk person 0.006. Man vet at 1% av populasjonen har sykdommen. Hva er sannsynligheten for at en person faktisk er syk, når diagnosen viser det?

- a) 0.650
- b) 0.625
- c) 0.375
- d) 0.350

Oppgave 13. For en kontinuerlig tilfeldig variabel X som er uniformt fordelt på intervallet $[0,2]$, hvilket utsagn er riktig?

- a) Forventningen er større enn medianen
- b) Sannsynligheten for at X er mellom 0.2 og 0.5 er lik 0.3
- c) Sannsynligheten for at X er større enn 1.5 er lik 0.25.

Oppgave 14. En diskret tilfeldig variabel X med sannsynlighetsfordeling

Verdi av X	0	1	2	3
Sannsynlighet	0.1	0.4	0.3	0.2

har standardavvik lik

- a) 0.840 b) 0.931 c) 0.917 d) 0.810

Oppgave 15. La X og Y være to uavhengige stokastiske variable. Begge er standard normalfordelte, dvs. har forventning 0 og standardavvik 1. Differansen $X-Y$ har standardavvik

- a) 2 b) $\sqrt{2}$ c) 0 d) $1/\sqrt{2}$

Oppgave 16. Sannsynligheten for at X definert i oppgaven over ligger mellom 0 og 1 er

- a) 0.8413 b) 0.3413 c) 0.1587

Oppgave 17. I Norge er det straffbart å kjøre bil dersom alkoholkonsentrasjonen i blodet overstiger 0.2 promille. La X være resultatet av en promille-måling. Hvis en person som testes har akkurat 0.2 i promille, kan vi anta at målingene av promillen er uavhengige og normalfordelte med forventning 0.2 promille og standardavvik 0.025 promille. Sannsynligheten for at en måling blir større enn 0.22 er

- a) 0.2119 b) 0.7881 c) 0.5669

Oppgave 18. Hvis man foretar 12 uavhengige målinger hos personen med 0.2 i promille, er forventet antall ganger man får en måling som er høyere enn 0.22

- a) 6.240 b) 4.096 c) 2.543

Oppgave 19. Det er vanlig å bruke gjennomsnittet av to promille-målinger for å redusere usikkerheten. Forventningen til gjennomsnittet av to målinger er

- a) 0.2 promille b) 0.1 promille c) 0.0

Oppgave 20. Standardavviket til gjennomsnittet av to målinger er ca.

- a) 0.0125 promille b) 0.05 promille c) 0.0353 promille
 d) 0.0177 promille

SLUTT