

UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

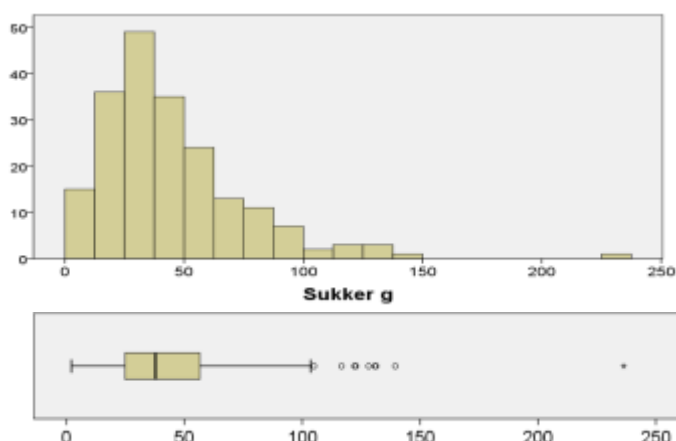
Eksamen i: STK1000 – Innføring i anvendt statistikk
Eksamensdag: Fredag 28. oktober 2016
Tid for eksamen: 14.00 – 16.00
Oppgavesettet er på 6 sider
Tillatte hjelpemidler: Godkjent kalkulator,
Lærebok (alle utgaver)
Ordliste for STK1000
Vedlegg: Svarark

Kontroller at oppgavesettet er komplett før du begynner å besvare spørsmålene. Svarene føres på eget svarark.

Alle 20 oppgaver teller likt. For hver oppgave skal du merke av for bare ett svaralternativ. Du får ett poeng for hvert riktige svar, maksimalt 20 poeng. Dersom du svarer feil eller lar være å krysse av på en oppgave, får du null poeng. Du blir altså ikke straffet med minuspoeng for å svare feil. Legg merke til at i spørsmålene er noen av svaralternativene avrundet.

Lykke til!

Oppgave 1



Her ser du et histogram og et tilhørende boksplott for det daglige sukkerinntaket for 200 gravide kvinner på Rikshospitalet. Hvilken påstand er feil:

- A Fordelingen er skjev.
- B Fordelingen har flere outliere eller ekstremverdier.
- C Gjennomsnittet og standardavviket er gode oppsummeringstall for denne fordelingen.
- D Fordelingen er entoppet eller unimodal.

Oppgave 2

Her ser du deskriptiv statistikk for de samme dataene. Hvilken påstand er feil:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	sd
2.1000	24.825	37.800	44.737	56.575	236.400	30.5165

- A Ca 50 av kvinnene i undersøkelsen spiser mindre enn 25 gram sukker daglig.
- B 50% av kvinnene i undersøkelsen spiser mellom 25 og 45 gram sukker daglig.
- C Minst 50 av kvinnene spiser mer enn 50 gram sukker daglig.
- D Ca 95% av kvinnene spiser mellom 0 og 105.7 gram sukker daglig.

Oppgave 3

Hva er variansen for observasjonene (daglig sukkerinntak for gravide kvinner) i oppgave 2?

- A 931.3 B 61.0 C 31.75 D 234.3

Oppgave 4

Hvis du får oppgitt at på et spørreskjema om livskvalitet, der minimumsscore er 0, og maksimumsscore er 100, så er gjennomsnittet 88.4, og standardavviket 12.3, hvordan ser du for deg at fordelingen er da:

- A Den er symmetrisk om gjennomsnittet
- B Det er umulig å vite
- C Den er skjev med høyrehale
- D Den er skjev med venstrehale

Oppgave 5

Hvilken påstand er riktig:

- A I en normalfordeling er 99% av verdiene mellom $\mu - 2.81 \cdot \sigma$ og $\mu + 2.81 \cdot \sigma$
- B I en normalfordeling er 80% av verdiene mellom $\mu - 1.28 \cdot \sigma$ og $\mu + 1.28 \cdot \sigma$
- C I en normalfordeling er 95% av verdiene mellom $\mu - 1.645 \cdot \sigma$ og $\mu + 1.645 \cdot \sigma$
- D I en normalfordeling er 97.5% av verdiene mellom $\mu - 1.96 \cdot \sigma$ og $\mu + 1.96 \cdot \sigma$

Oppgave 6

Temperaturmålinger hos en frisk person kan antas å være fordelt som $N(37.0, 0.3)$. Da er sannsynligheten for at en måling av en frisk person viser mer enn 37.5

- A 0.0764 B 0.0344 C 0.0475 D 0.9525

Oppgave 7

Her ser du vekstkurver (som viser sammenhengen mellom vekt og alder), basert på Vekststudien i Bergen (BGS), SYSBARN-undersøkelsen, og WHO's internasjonale vekstkurver for gutter og jenter i alderen 0 - 5 år.

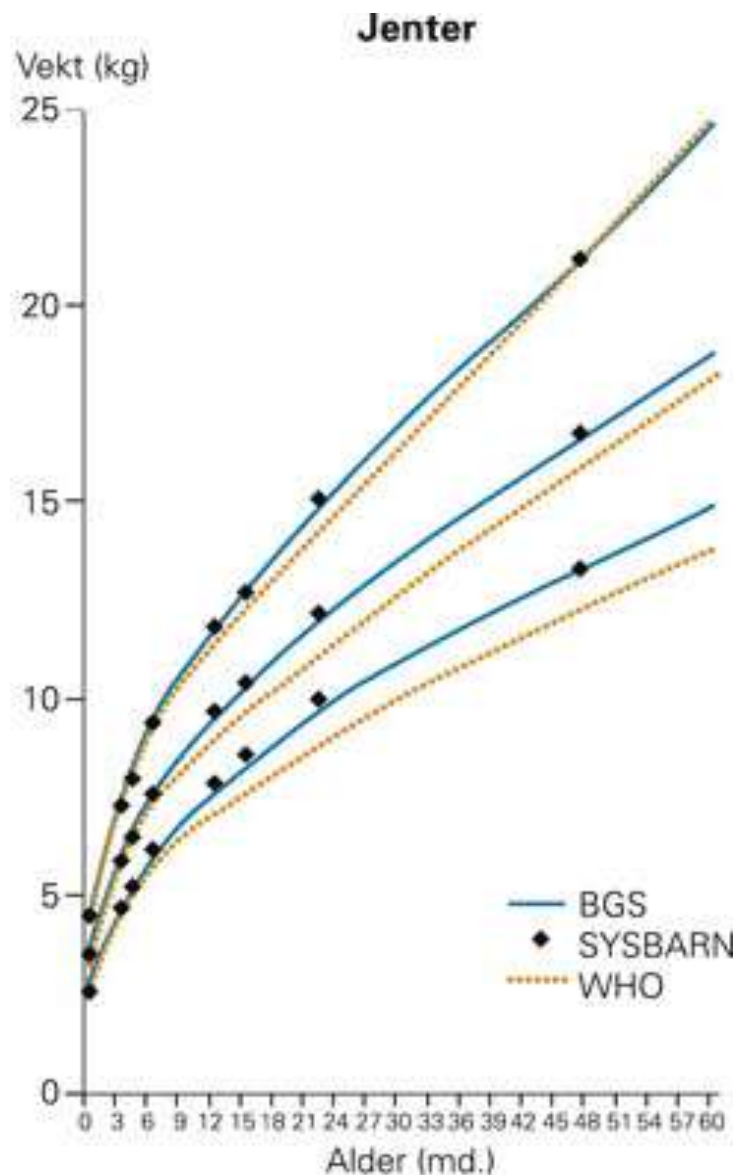
Linjene viser 2.5-, 50- og 97.5-prosentilene (percentilene).

Figuren er hentet fra Tidsskriftet for Den norske legeforening 2009;129: 281-6

Ta utgangspunkt i de nye, norske vekstkurvene fra BGS, altså de heltrukne linjene.

Hvilken påstand stemmer ikke:

- A 95% konfidensintervall for forventet vekt i kg for en 4-årig jente er [13, 21.1]
- B Over 50% av 4-årige jenter veier mer enn 15 kg
- C 95% av 4-årige jenter veier mellom 13 og 21.1 kg
- D Vekstfordelingen for 4-åringer er relativt symmetrisk, men litt høyreskjev.



Oppgave 8

I en studie av barns motoriske ferdigheter, laget amerikanske forskere et spørreskjema som ga en totalscore. Denne scoren ble undersøkt i et utvalg av amerikanske barn, og beregnet slik at den fulgte en $N(50,10)$ -fordeling. Hvilken påstand er feil:

- A Et barn som scorer 30 har en z-score på -2.

- B Sannsynligheten for at et barn scorer lavere enn -2 er 2.3%
- C Gjennomsnittet av to barns scorer vil være fordelt som $N(50, 7.1)$
- D Sannsynligheten for at to uavhengige barn scorer lavere enn -2, er 4.6%

Oppgave 9

Hvis vi samler nok data, altså at utvalgsstørrelsen n blir stor nok, så vil tre av disse påstandene være sanne. Hvilken av dem skjer ikke?

- A Gjennomsnittet vil ligne mer og mer på forventningsverdien i populasjonen.
- B Histogrammet over dataene vil bli normalfordelt.
- C Fordelingen til gjennomsnittet blir normalfordelt.
- D Histogrammet over dataene vil ligne på populasjonsfordelingen.

Oppgave 10

Hvis vi har to tilfeldige utvalg, ett på $n=25$, som gir et gjennomsnitt \bar{x}_{25} og et standardavvik sd_{25} , og ett på $n=100$, som gir et gjennomsnitt \bar{x}_{100} og et standardavvik sd_{100} , og histogrammene for begge utvalgene ser rimelig symmetriske ut, så vil tre av påstandene være riktige. Hvilken påstand er feil?

- A Percentilene i de to utvalgene er like, altså i samme størrelsesorden.
- B Estimeringsusikkerheten (standardavvikene) til de to gjennomsnittene er ganske like, altså i samme størrelsesorden.
- C Gjennomsnittene i de to utvalgene er ganske like, altså i samme størrelsesorden.
- D Standardavvikene i de to utvalgene er ganske like, altså i samme størrelsesorden.

Oppgave 11

En prosess kalles tilfeldig hvis

- A Individuelle utfall er usikre og de er umulig å predikere
- B Man har ingen idé om hva som kommer til å skje
- C Utfallene man er interessert fordeler seg i to grupper med 50% i hver gruppe
- D Individuelle utfall er usikre men skjer i et forutsigbart mønster over tid

Oppgave 12

Randomisering i forbindelse med utvelgelse av individer er viktig for å

- A skape rettferdighet mellom individer
- B redusere skjevhet (bias)
- C få liten variabilitet
- D få en symmetrisk og entoppet (unimodal) fordeling av responsvariabelen

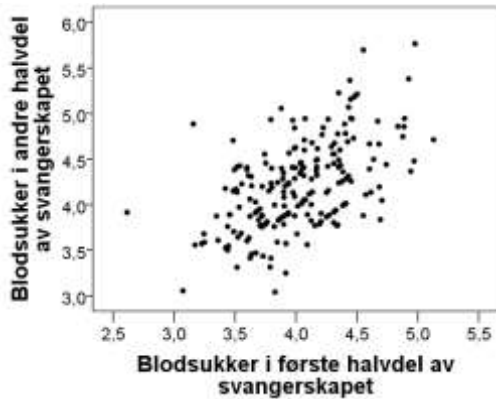
Oppgave 13

Betrakt et spill der kostnaden ved å spille er 20 NOK. Hvis du vinner et spill, mottar du 100 NOK mens hvis du taper, mottar du ingenting. Sannsynligheten for å vinne et enkelt spill er

0.1, og de ulike spillene har uavhengige utfall. Anta du spiller tre spill på rad og du er interessert i den tilfeldige variabelen X som er hvor mye du har tjent etter tre spill.

- A Utfallsrommet er $\{0; 100; 200; 300\}$ og sannsynligheten for $X = 300$ er 0.729
- B Utfallsrommet er $\{-20; 80; 160; 240\}$ og sannsynligheten for $X = 160$ er 0.243
- C Utfallsrommet er $\{-60; 40; 140; 240\}$ og sannsynligheten for $X = 240$ er 0.001
- D Utfallsrommet er $\{-60; 40; 140; 240\}$ og sannsynligheten for $X = 40$ er 0.027

Oppgave 14



Den empiriske korrelasjonen mellom blodsukker målt tidlig og sent i svangerskapet (vist i figuren over) er ett tallene oppgitt under. Angi hvilket.

- A 0.54
- B 0.36
- C -0.62
- D 0.82

Oppgave 15

En diskret tilfeldig variabel X antar verdiene 1 til 4, med sannsynligheter gitt i følgende tabell

x	1	2	3	4	Sum
$P(X=x)$	0.4	0.3	0.2	0.1	1

Da er forventningen til X , μ_X , lik

- A 1.8
- B 1.9
- C 2.0
- D 2.1

Oppgave 16

For X gitt i forrige oppgave blir standardavviket lik

- A 1.1
- B 1.0
- C 5.0
- D 1.2

Oppgave 17

Variabelen X i de to forrige oppgavene viser fordelingen av gjester til middag, i en barnefamilie med tre barn. La variabelen Y være det totale antallet personer rundt middagsbordet. Da er $Y = X + 5$. Hvilken påstand er feil:

- A $E(Y) = \mu_Y = 7$, hvis man runder av til en desimal.
- B X og Y er korrelerte.
- C Standardavviket til Y er lik standardavviket til X
- D Standardavviket til Y er større enn standardavviket til X

Oppgave 18

La A og B være to disjunkte hendelser, hver med positiv sannsynlighet. Da er

- A $P(A|B)=P(A)P(B)$
- B $P(A|B)=P(A)$
- C $P(A|B)=0$
- D $P(A|B)=1$

Oppgave 19

La A og B være to uavhengige hendelser, hver med positiv sannsynlighet. Da er

- A $P(A|B)=P(A)P(B)$
- B $P(A|B)=P(A)$
- C $P(A|B)=0$
- D $P(A|B)=1$

Oppgave 20

En kvinne tar en mammografiundersøkelse. Sannsynligheten for at mammografiet viser tegn på brystkreft hvis kvinnen har brystkreft er 0.95. Sannsynligheten er 0.034 for at mammografiet viser tegn på kreft selv om kvinnen egentlig ikke har kreft. Vi vet også at sannsynligheten for at en tilfeldig kvinne har brystkreft er 0.005. Anta at mammografiet viser tegn på brystkreft. Hva er da sannsynligheten for at kvinnen virkelig har brystkreft?

- A 0.27
- B 0.16
- C 0.13
- D 0.10

SLUTT