

UNIVERSITETET I OSLO

Matematisk Institutt

MIDTVEISEKSAMEN I: **STK 1000: Innføring i anvendt statistikk**
TID FOR EKSAMEN: **Onsdag 9. oktober 2013, 11:00–13:00**
HJELPEMIDLER: **Lærebok, ordliste for STK1000, godkjent kalkulator**

Oppgavesettet er på 5 sider og består av 20 oppgaver. I tillegg kommer eget svarark, der dine svar skal føres.

Hver av de 20 oppgaver teller likt. For hver oppgave får du fire svaralternativer, hvorav kun ett er korrekt. Du skal for hver oppgave merke av for bare ett svaralternativ. Du får ett poeng for hvert riktige svar, og følgelig maksimalt 20 poeng. Dersom du svarer feil, eller lar være å krysse av på et spørsmål, eller setter flere kryss enn ett, får du null poeng for vedkommende punkt. Du blir altså ikke «straffet» med minuspoeng for å svare feil. Lykke til!

Oppgave 1. Et bestemt datasett med 24 observasjoner er som følger, ordnet i stigende rekkefølge:

0.02 0.04 0.11 0.13 0.14 0.15 0.21 0.35
0.37 0.38 0.44 0.52 0.58 0.76 0.92 1.08
1.18 1.38 1.40 1.45 1.53 1.81 1.85 3.55

For dette datasettet gir MINITAB (blant annet) følgende:

Variable	Count	Mean	StDev	Minimum	Q1	Median	Q2	Maximum
C1	24	0.8479	0.8251	0.02	0.18	0.55	1.39	3.55

Fem-talls-oppsummeringen for dette datasettet er

A 0.11 0.18 0.55 1.81 3.55
B 0.8479 0.8251 0.18 0.55 1.39
C 0.02 0.04 0.55 1.85 3.55
D 0.02 0.18 0.55 1.39 3.55

Oppgave 2. For datasettet over, hvor mange potensielle outliers er det, ved $1.5 \times \text{IQR}$ -regelen?

A 0 **B** 1 **C** 2 **D** 3

Oppgave 3. En sannsynlighetsfordeling med utfallsom $\{0, 1, 2, 3\}$ har sannsynligheter gitt i denne tabellen:

0	1	2	3
0.5	0.3	p	p

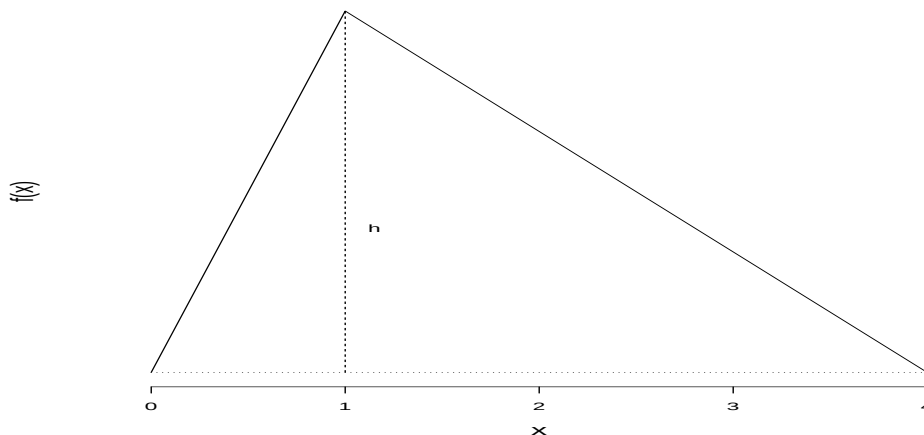
Her er p altså det samme tallet for hvert av utfallene '2' og '3'. Da må forventningsverdien (the mean value) for X være

A 0.80 **B** 0.90 **C** 1.00 **D** 1.10

Oppgave 4. Anta at X, Y, Z er tre uavhengige stokastiske variable, hver med den sannsynlighetsfordelingen som er gitt i foregende oppgave. Sannsynligheten for at hver av disse tre er minst lik 2 er da:

- A** 0.001 **B** 0.004 **C** 0.006 **D** 0.008

Oppgave 5. En sannsynlighetstetthet (probability density curve) $f(x)$ har formen angitt i figur 1. Den stiger altså som en rett linje fra $x = 0$ til $x = 1$, og avtar deretter som en annen rett linje fra $x = 1$ til $x = 4$. Høyden ved punktet 1 er lik h . NB: Figuren er laget slik at skalaen på y-aksen ikke er lik skalaen på x-aksen, så man kan altså ikke lese ut størrelsen av h bare ved å måle.



Figur 1: Sannsynlighetstettheten (density curve) $f(x)$ beskrevet i oppgave 5.

Høyden h må være lik

- A** 0.5 **B** 1.0 **C** 2.0 **D** 4.0

Oppgave 6. Hvis X er en stokastisk variabel med sannsynlighetstettheten (density curve) fra oppgaven over, hva er da $P(X \geq 1)$?

- A** $\frac{1}{4}$ **B** $\frac{1}{2}$ **C** $\frac{3}{4}$ **D** $\frac{5}{6}$

Oppgave 7. Et bestemt eksperiment har utfallsrom $\{1, 2, 3, 4, 5, 6\}$, med sannsynligheter henholdsvis $\frac{1}{21}, \frac{2}{21}, \frac{3}{21}, \frac{4}{21}, \frac{5}{21}, \frac{6}{21}$. Vi definerer begivenhetene (events)

$$A = \{2, 4, 6\}, \quad B = \{1, 3, 5\}, \quad C = \{1, 6\}.$$

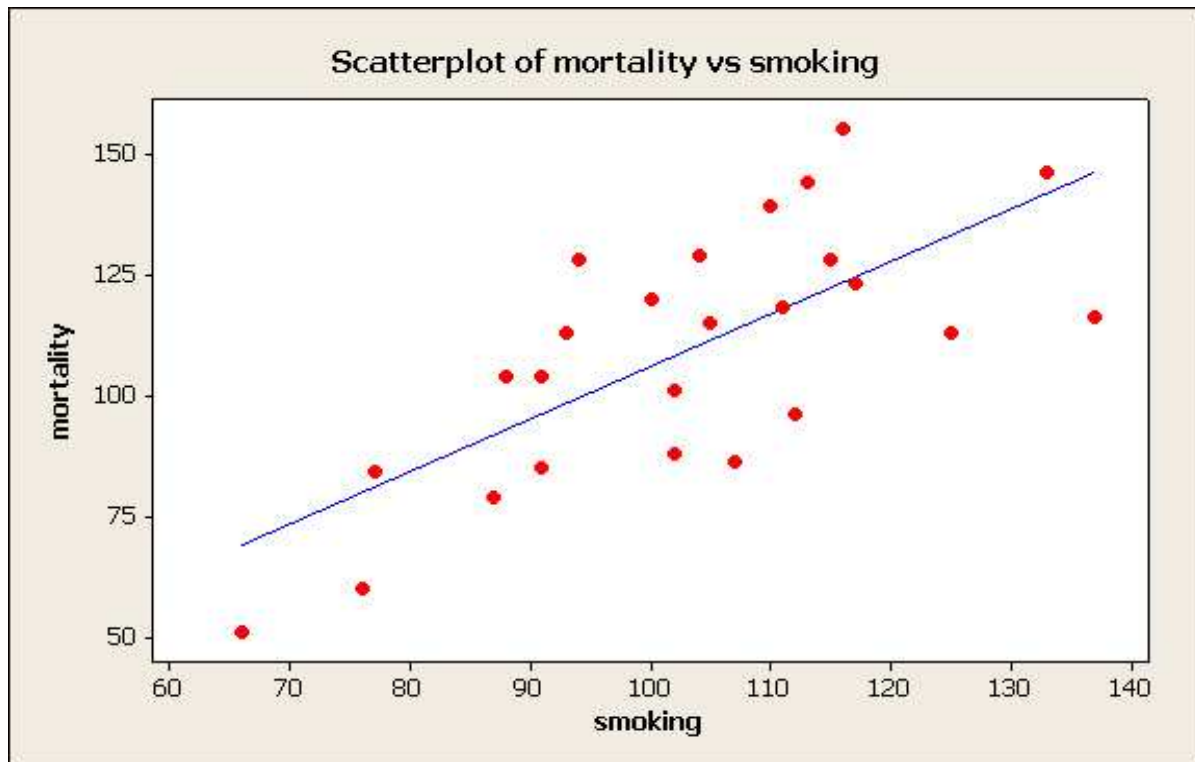
Da er

- A** A og B uavhengige; **B** $P(A) = P(B)$;
C A og B disjunkte; **D** $P(A \cap C) = \frac{1}{6}$.

Oppgave 8. For utfallsrommet og sannsynlighetene over, gjelder

- A** $P(A)P(B) = P(C)$; **B** $P(A)P(B) > P(C)$;
C $P(C^c) = \frac{3}{4}$; **D** $P(B \cup C) = \frac{5}{7}$.

Oppgave 9. En forskningsgruppe leverte i 1978 en rapport til den engelske regjering, der ulike sammenhenger mellom røyking og sykdommer ble vurdert. Ett av gruppens funn er oppsummert i figuren under. Her er avmerket punkter (x, y) for 25 ulike yrkeskategorier, der x (smoking) er røykevaner for denne gruppen og y (mortality) dødelighetstendensen for gruppen. Smoking-score og mortality-score er beregnet via diverse tidligere undersøkelser, og skalaene det er tale om er laget slik at verdien 100 representerer normalen i befolkningen (tilsvarende er $x > 100$ en indikasjon på mer røyking enn for gjennomsnittet og $y > 100$ et signal om høyere dødelighet enn for gjennomsnittet).



Korrelasjonskoeffisienten for disse (x, y) data er

- A 0.113 B 0.248 C 0.716 D 0.950

Oppgave 10. En MINITAB-kjøring for datasettet over gir blant annet det følgende, der C1 er smoking og C2 er mortality:

The regression equation is

$$C2 = -2,9 + 1,09 C1$$

Predictor	Coef	SE Coef	T	P
Constant	-2,89	23,03	-0,13	0,901
C1	1,0875	0,2209	4,92	0,000

Yrkesgruppen som både røyker minst og lever lengst, i denne undersøkelsen, med $(x, y) = (66, 51)$, er den som arbeider med tekniske fag, samt kunstnere. Anta at en yrkesgruppe røyker enda litt mindre enn disse, med en smoking-score på 60. Den forventede mortality-score for en slik gruppe er da

- A 48.5 B 52.7 C 58.2 D 62.4

Oppgave 11. Yrkesgruppen som i undersøkelsen beskrevet i oppgavene 9 og 10 røyker aller mest, med smoking-score på 137, er gruve- og havnearbeidere, og de har mortality-score på 116. Residualen, for denne av de 25 gruppene, blir da

A -30.10 **B** -11.15 **C** 12.88 **D** 30.10

Oppgave 12. Vi ser på to stokastiske variable X og Y , begge med utfallsrom $\{1, 2, 3, 4\}$, og med sannsynlighetsfordelinger gitt ved

$$\begin{aligned} P(X = 1) &= 0.1, & P(X = 2) &= 0.2, & P(X = 3) &= 0.3, & P(X = 4) &= 0.4, \\ P(Y = 1) &= 0.3, & P(Y = 2) &= 0.3, & P(Y = 3) &= 0.3, & P(Y = 4) &= 0.1. \end{aligned}$$

Forventningsverdien til variabelen $Z = (X + Y)/2$ er da

A 2.3 **B** 2.4 **C** 2.5 **D** 2.6

Oppgave 13. I situasjonen over gjelder det følgende, for variansen til variabelen $Z = (X + Y)/2$:

- A** den er 0.98;
- B** den er 1.10;
- C** den er 0.49;
- D** vi har ikke tilstrekkelig informasjon til å bestemme den.

Oppgave 14. Vi besøker et bestemt samfunn av kvinner og menn. Anta (for denne oppgavens vedkommende) at menn utgjør 40% av denne populasjonen og snakker sant 70% av tiden, mens kvinner snakker sant 90% av tiden. Sannsynligheten for at en vilkårlig valgt person snakker sant er da:

A 0.78 **B** 0.80 **C** 0.82 **D** 0.85

Oppgave 15. For dette samfunnet beskrevet over, anta du får vite at en person faktisk snakker sant. Sannsynligheten for at vedkommende er en kvinne er da:

A 0.600 **B** 0.614 **C** 0.659 **D** 0.700

Oppgave 16. Anta at den stokastiske variabel X har en $N(3.33, 2.22)$ -fordeling (altså en normalfordeling med forventning $\mu = 3.33$ og standardavvik $\sigma = 2.22$), og at Y , uavhengig av X , har en $N(0, 1)$ -fordeling. Da vil den stokastiske variabelen $Z = X + 2Y$ også være normalfordelt, med forventning og standardavvik

A (3.33, 2.99) **B** (3.33, 4.22) **C** (3.33, 8.92) **D** (6.66, 5.44)

Oppgave 17. I situasjonen over, se også på den stokastiske variabelen $W = X - 2Y$. Denne er også normalfordelt, med forventning og standardavvik

A (3.33, 2.99) **B** (3.33, 0.22) **C** (3.33, 2.92) **D** (6.66, 5.44)

Oppgave 18. Forskere fra et universitet i South Carolina undersøkte i 1989 ulike helseparametre for nyfødte barn av mødre som jevnlig brukte kokain. Blant annet fant de at fødselsvektene for slike barn stemmer godt overens med en normalfordeling, si $N(\mu, \sigma)$, der $\mu = 2.733$ kg og $\sigma = 0.599$ kg. Sannsynligheten for at et nyfødt barn, av en kokainbrukende mor, skal veie mer enn 3.0 kg, er da:

A 0.229 **B** 0.328 **C** 0.421 **D** 0.599

Oppgave 19. At en observator (statistic) for en parameter p er forventningsrett (unbiased) betyr

A at den har liten varians;

B at dens forventningsverdi er p ;

C at dataene den er basert på er representative;

D at dens sannsynlighetsfordeling er eksakt eller tilnærmet en normalfordeling.

Oppgave 20. Man er interessert i å estimere en viss lengdeparameter μ , som f.eks. kan måles på centimeter-skalaen. For å anslå μ kan man utføre et eksperiment, som gir en måling X , med forventning $EX = \mu$ og standardavvik $\text{stdev } X = \sigma = 10$ cm. Siden ett enkelt eksperiment ikke er tilstrekkelig utføres flere uavhengige slike, med resultater $X_1, X_2, \dots, X_n$, der hver observasjon altså har forventning μ og standardavvik σ . Som estimator for μ anvendes så gjennomsnittet av observasjonene (datapunktene),

$$\bar{x} = (X_1 + X_2 + \dots + X_n)/n.$$

Det er et krav til den rapport som skal publiseres at estimatet $\bar{x}$ er så presist at dets standardavvik er på høyst 0.50 cm. Hvor mange observasjoner er det da nødvendig å skaffe seg?

A 200 **B** 400 **C** 500 **D** 1000