

# STK 1000 midtveiseksamen H23

2023-10-12

## Forside

STK 1000 – Innføring i anvendt statistikk

Torsdag 12. oktober 2023 kl. 15.00 -17.00 (2 timer)

20 flervalgsoppgaver: Hvert riktig svar gir 1 poeng, det gis 0 poeng for galt svar og 0 poeng for ikke besvart oppgave. Maksimal poengsum for hele oppgavesettet er 20 poeng.

Tillatte hjelpemidler: Godkjent kalkulator, lærebok og “ordliste for STK1000”. En lenke til ordlisten er gitt under oppgavelinjen.

Læreboka er: Moore McCabe Craig Introduction to the practice of statistics 9th edition, 2017.

Tidligere utgaver av læreboka “Moore McCabe Craig, Introduction to the practice of statistics” er alternativt tillatt hjelpemiddel (studenten får ha med en versjon av læreboka, i tillegg til kalkulator og ordliste).

## Oppgave 1

Antall hjort drept av bil eller tog i jaktårene 2015/16, 2016/17, 2017/18, 2018/19, 2019/20, 2020/21 og 2021/22 var henholdsvis 639, 873, 1223, 1038, 1025, 983 og 1084. Hva er medianen til disse tallene?

- A. 1025
- B. 981
- C. 1011
- D. 931

## Oppgave 2

Antall hjort drept av bil eller tog i jaktårene 2015/16, 2016/17, 2017/18, 2018/19, 2019/20, 2020/21 og 2021/22 var henholdsvis 639, 873, 1223, 1038, 1025, 983 og 1084. Hva er interkvartil avstand (IQR) til disse tallene?

- A. 55
- B. 133
- C. 211
- D. 584

## Oppgave 3

Hvilken fordeling passer best til å beskrive antall hjort drept av bil eller tog i løpet av et jaktår?

- A. Normalfordeling
- B. Eksponentialfordeling
- C. Binomisk fordeling
- D. Poissonfordeling

## Oppgave 4

Vi tilpasser en lineær modell for sammenhengen mellom  $X$  og  $Y$  og får  $Y = 30 - 0.9x$ . I tillegg har vi at standardavvikene er gitt ved  $s_x = 2$  og  $s_y = 9$ . Hva blir korrelasjonen  $r$  i dette tilfelle?

- A. 0.3
- B. 0.2
- C. -0.2
- D. -0.3

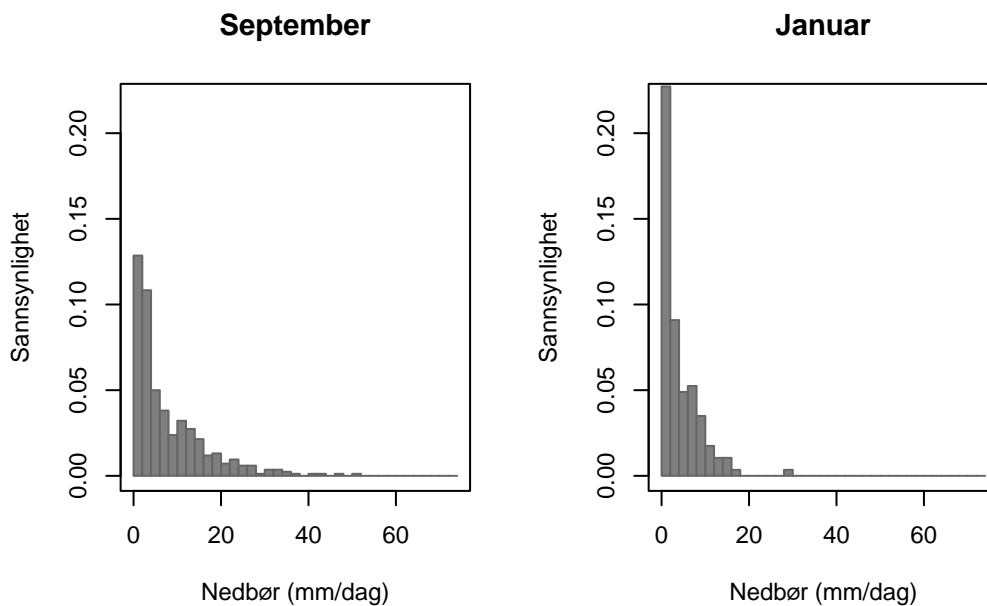
## Oppgave 5

Korrelasjonen mellom  $X$  og  $Y$  er gitt ved  $r = 0.4$ . Hva er korrelasjonen mellom  $2X$  og  $(-Y)$ ?

- A. 0.4
- B. -0.4
- C. 0.8
- D. -0.8

## Oppgave 6

Vi ser på døgnnedbør fra værstasjonen Blindern i Oslo, målt fra 1980 til i dag. Vi kaller en dag "våt" hvis det har regnet mer enn 0.5 mm den dagen. 34% av dager i september er våte dager og 33% av dager i januar er våte dager. Nedenfor vises histogrammer for nedbørsmengde på våte dager for hver av de månedene.



Hvilken av følgende påstander er feil?

- A. Det er omtrent like mange våte dager i september og januar
- B. Det er omtrent like mange våte dager i september og januar hvor nedbøren er mellom 10 og 20 mm per dag
- C. Det regner i gjennomsnitt mer på en våt dag i september enn (det regner eller snør) på en våt dag i januar

D. Fordelingen for september har større median enn fordelingen for januar

### Oppgave 7

Femtalls-oppsummering for nedbør på våte dager i september fra forrige oppgave er gitt ved

[1] 0.50 2.00 4.35 11.60 50.10

I perioden fra 1980 til i dag er det registrert en måling på 1220 våte septemberdager. På hvor mange av disse dagene har det regnet mer en 11.6 mm?

- A. 407
- B. 610
- C. 220
- D. 305

### Oppgave 8

Vi har 1220 målinger på døgnet nedbør på våte dage i september. Disse målingene har et gjennomsnitt på 7.9 mm og standardavvik på 8.4 mm. Vi trekker 100 tilfeldige våte septemberdager og måler gjennomsnittet av døgnet nedbør på disse dagene. Hvor stor er sannsynligheten for at gjennomsnittet av disse 100 målingene er større enn 9 mm?

- A. 0.10
- B. 0.45
- C. 0.13
- D. 0.87

## Oppgave 9

En forsker studerer effekten av et nytt medikament for å redusere høyt blodtrykk. Forskeren har grunn til å tro at medikamentet vil ha en annen effekt på kvinner enn på menn. På grunn av dette bør forskeren bruke

- A. et eksperiment med blokkdesign
- B. et eksperiment med matchede par
- C. et stratifisert eksperiment
- D. et stratifisert tilfeldig utvalg

## Oppgave 10

Karl Pearson (1857-1936) har samlet inn data på høyden av far og sønn for 1078 far/sønn par. Høydene er målt i centimeter. Vi estimerer en lineær modell til å predikere høyden på sønnen gitt høyden på faren:

Call:

```
lm(formula = son.height ~ father.height)
```

Residuals:

Min	1Q	Median	3Q	Max
-22.5480	-3.8466	-0.0201	4.1364	22.7799

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	86.07198	4.65418	18.49	<2e-16 ***
father.height	0.51409	0.02705	19.01	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.189 on 1076 degrees of freedom

Multiple R-squared: 0.2513, Adjusted R-squared: 0.2506

F-statistic: 361.2 on 1 and 1076 DF, p-value: < 2.2e-16

I følge regresjonslinjen, hva er predikert høyde på en sønn hvis far er 170 centimeter?

- A. 175
- B. 169.8

- C. 170
- D. 173.5

### Oppgave 11

I datasettet til Karl Pearson i forrige oppgave, så er gjennomsnittshøyden til de 1078 farene 171.9 cm med standardavvik 7.0 cm mens sønnene har gjennomsnittshøyde på 174.4 cm med standardavvik 7.1 cm. Vi trekker tilfeldige 100 menn fra fedregruppen og tilfeldige 100 menn fra sønnegruppen. Hva er forventet forskjell på gjennomsnittshøyden i de to gruppene?

- A. 2.5 cm
- B. 0.3 cm
- C. 0.1 cm
- D. 2 cm

### Oppgave 12

Hvilken av de følgende R kommandoer gir ikke det samme svar som de tre andre?

- A. `pnorm(0.67, lower.tail=FALSE)`
- B. `pnorm(0.67, lower.tail=TRUE)`
- C. `pnorm(-0.67)`
- D. `1 - pnorm(0.67)`

### Oppgave 13

P og Q vurderer å søke jobb. Sannsynligheten for at P søker jobben er  $\frac{1}{4}$ , sannsynligheten for at P søker jobben gitt at Q søker jobben er  $\frac{1}{2}$ , og sannsynligheten for at Q søker jobben gitt at P søker jobben er  $\frac{1}{3}$ . Hva er sannsynligheten for at Q søker jobben?

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

### Oppgave 14

Et eksperiment har lik sjanse for å ha et negativt og et positivt utfall. Hva er sannsynligheten for å oppnå maksimalt et negativt utfall i fem forsøk?

- A.  $\frac{6}{32}$
- B.  $\frac{3}{32}$
- C.  $\frac{2}{32}$
- D.  $\frac{1}{32}$

### Oppgave 15

En diskret tilfeldig variabel  $X$  har sannsynlighetsfordeling

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

Hva er  $P(X \leq 2)$ ?

- A. 0.5
- B. 0.3
- C. 0.6
- D. 0.4

### Oppgave 16

Vi ser igjen på en diskret tilfeldig variabel  $X$  med sannsynlighetsfordeling

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

Hva er forventningen  $\mu$  og variansen  $\sigma^2$  for denne sannsynlighetsfordelingen?

- A.  $\mu = 2$  og  $\sigma^2 = 1.1$
- B.  $\mu = 2.1$  og  $\sigma^2 = 1.1$
- C.  $\mu = 2.1$  og  $\sigma^2 = 1.29$

D.  $\mu = 1.5$  og  $\sigma^2 = 3.05$

### Oppgave 17

I en sannsynlighetsmodell der  $S$  er utfallsrommet,  $A$  og  $B$  er disjunkte hendelser og,  $P(A) > 0$  og  $P(B) > 0$ , er en av påstandene feil:

- A.  $P(S) = 1$
- B.  $P(A^C) = 1 - P(A)$
- C.  $P(A \text{ eller } B) = P(A) + P(B)$
- D.  $P(A \text{ og } B) = P(A) \times P(B)$

### Oppgave 18

For uavhengige og identisk fordelte tilfeldige variabler  $X_1, \dots, X_n$  med forventning  $\mu$  og standardavvik  $\sigma$  vil gjennomsnittet  $\bar{x}$  være tilnærmet normalfordelt når antall observasjoner  $n$  er stort. Dette resultatet kalles

- A. sentralgrenseteoremet
- B. et randomisert eksperiment
- C. Bayes regel
- D. store talls lov

### Oppgave 19

I Norge fødes det i snitt 103 gutter per 100 jenter. En norsk by har to sykehus, en er større med 45 fødsler per dag i gjennomsnitt og den andre er mindre med 15 fødsler per dag i gjennomsnitt. Hvilket av de to sykehusene har større sjanse for å ha en dag hvor 60% eller mer av de nyfødte er en gutt?

- A. Begge sykehus har omtrent lik sjanse
- B. Det større sykehuset
- C. Det mindre sykehuset
- D. Det er ikke nok informasjon i oppgaven til å svare på spørsmålet



## Oppgave 20

En foreleser lager en eksamen med 20 flervalgsoppgaver hvor hver oppgave har fire svarmuligheter. Hvilken av følgende R kommandoer kan brukes til å trekke, på en tilfeldig måte, hvilken svarmulighet skal være riktig for hver oppgave?

- A. `rpois(20, lambda=2.5)`
- B. `rbinom(20, size=4, prob=0.25)`
- C. `sample(c(1:4), size=20, replace=TRUE)`
- D. `sample(c(1:4), size=20, replace=FALSE)`