

i Forside

Institutt for matematiske fag

Eksamensoppgave i ISTA1001, ISTA1002, ISTA1003, ISTG1001, ISTG1002, ISTG1003, ISTT1001, ISTT1002, ISTT1003, VB6200 Statistikk

Eksamensdato: 11. august 2023

Eksamenstid (fra-til): 09:00 - 12:00 (3 timer)

Hjelpemiddelkode/Tillatte hjelpemidler: C

Godkjent kalkulator

Digitale ressurser under eksamen: Formelark og fem tabeller (binomisk kumulativfordeling, Poisson kumulativfordeling, normal kumulativfordeling, normalfordeling kritisk verdi og t-fordeling kritisk verdi) er lagt ved eksamen som pdf-filer

Faglig kontakt under eksamen:

Thea Bjørnland (41123849)

Faglig kontakt møter i eksamenslokalet: NEI

ANNEN INFORMASJON:

Henvend deg til en eksamensvakt hvis du ønsker å kontakte faglig kontaktperson under eksamen. Noter spørsmålet ditt på forhånd. Faglig kontaktperson skal kun kontaktes dersom det er direkte feil eller mangler i oppgavesettet.

Vekting av oppgavene: er oppgitt for hver oppgave. Det gis ikke minus-poeng for gale eller manglede svar.

Lagring: Besvarelsen din i Inspira Assessment lagres automatisk hvert 15. sekund.

Varslinger: Hvis det oppstår behov for å gi beskjeder til kandidatene underveis i eksamen (f.eks. ved klare mangler eller feil i oppgavesettet), vil dette bli gjort via varslinger i Inspira. Et varsel vil dukke opp som en dialogboks på skjermen i Inspira. Du kan finne igjen varselet ved å klikke på bjella øverst i høyre hjørne på skjermen.

Trekk fra eksamen: Blir du syk under eksamen, eller av andre grunner ønsker å levere blankt/trekke deg, gå til «hamburgermenyen» i øvre høyre hjørne og velg «Lever blankt». Dette kan ikke angres selv om prøven fremdeles er åpen. Tilkall så eksamensvakt og følg instruksene fra eksamensvakten.

Tilgang til besvarelse: Du finner besvarelsen din i Arkiv etter at sluttida for eksamen er passert.

1 Oppgave 1: Observasjoner fra normalfordeling

La X_1 , X_2 og X_3 være uavhengige normalfordelte stokastiske variabler slik at $X_i \sim N(\mu, \sigma)$ for $i = 1, 2, 3$.

Vi har gjort følgende tre observasjoner:

$x_1 = 6.6$, $x_2 = 3.8$ og $x_3 = 5.9$

Hva blir gjennomsnittet av observasjonene?

Velg ett alternativ:

- ☐ 6.89
- ☐ 4.31
- ☐ 7.25
- ☐ 5.43
- ☐ 2.98
- ☐ 3.62

Hva blir empirisk standardavvik?

Velg ett alternativ

- ☐ 3.89
- ☐ 5.36
- ☐ 2.37
- ☐ 1.46
- ☐ 4.27
- ☐ 0.62

Hvilket er et 95% konfidensintervall for μ ?

Velg ett alternativ

- ☐ [1.21, 8.34]
- ☐ [4.21, 6.62]
- ☐ [2.45, 7.69]
- ☐ [0.93, 8.76]
- ☐ [3.49, 9.67]
- ☐ [1.81, 9.05]

Maks poeng: 3

2 Oppgave 2: Simultanfordeling

Tabellen viser simultanfordelingen til to stokastiske variabler X og Y , $P(X = x, Y = y)$. Vi kan for eksempel lese ut at $P(X = 3, Y = 1) = 0.3$.

x / y	$y = 1$	$y = 2$
$x = 1$	0.2	0.1
$x = 2$	0.1	0.2
$x = 3$	0.3	0.1

Fra denne simultanfordelingen kan man for eksempel regne ut marginalfordelingen til variablen X som blir

$P(X = 1) = 0.3$, $P(X = 2) = 0.3$, og $P(X=3) = 0.4$.

a) Hva blir marginalfordelingen til Y ?

Velg ett alternativ:

- ☐ $P(Y=1) = 0.6$, $P(Y=2) = 0.4$
- ☐ $P(Y=1) = 0.2$, $P(Y=2) = 0.8$
- ☐ Det er umulig å regne ut fra tabellen
- ☐ $P(Y=1) = 0.1$, $P(Y=2) = 0.9$
- ☐ $P(Y=1) = 0.7$, $P(Y=2) = 0.3$
- ☐ $P(Y=1) = 0.5$, $P(Y=2) = 0.5$

b) Hva blir $E(X)$?

Velg ett alternativ

- ☐ 1.9
- ☐ 1.3
- ☐ 3.2
- ☐ 0.4
- ☐ 2.8
- ☐ 2.1

c) Hva blir $P(Y = 2 \mid X < 3)$?

Velg ett alternativ

☐ 0.2

☐ 0.4

☐ 0.1

☐ 0.3

☐ 0.6

☐ 0.5

Maks poeng: 3

3 Oppgave 3: Utregning av sannsynligheter med Python

a) I Python har vi skrevet følgende kode:

```
from scipy import stats  
print(stats.binom.pmf(4,8,0.4))
```

Hvilket tall skrives ut (avrundet til to desimaler)?

Velg ett alternativ

- ☐ 0.62
- ☐ 0.17
- ☐ 0.23
- ☐ 0.48
- ☐ 0.34
- ☐ 0.51

b) I Python har vi skrevet følgende kode:

```
from scipy import stats
```

La Z være en standard normalfordelt stokastisk variabel. Hvilken kommando vil gi oss sannsynligheten $P(Z \leq 2)$?

Velg ett alternativ:

- ☐ stats.norm.cdf(2, 1, 0)
- ☐ stats.norm.cdf(-2, 0, 1)
- ☐ stats.norm.pdf(2, 0, 1)
- ☐ stats.norm.pdf(2, 1, 0)
- ☐ stats.norm.pdf(2, 0, 1)
- ☐ stats.norm.cdf(2, 0, 1)

Maks poeng: 2

4 Oppgave 4: Drikkevann

I et renseanlegg for drikkevann gjennomføres det daglige kontroller for bakterien *E. coli*. Dersom konsentrasjonen av *E. coli* i drikkevannet overstiger 1 cfu/100ml sendes det varsel til innbyggerne om at vannet må kokes før bruk.

Varslingssystemet er forbundet med noe usikkerhet. Selv om drikkevannet *ikke* er forurensset med *E. coli* er det likevel en sannsynlighet $p = 0.001$ for at et varsel sendes ut til innbyggerne. Dersom vannet faktisk er forurensset med *E. coli* vil systemet avdekke dette med sannsynlighet 0.98.

a) Dersom vannet over en periode på 40 dager *ikke* er forurensset, hva er sannsynligheten for at det likevel sendes ut minst ett varsel til innbyggerne? De daglige kontrollene kan antas å være uavhengige forsøk med sannsynlighet for suksess $p = 0.001$.

Velg ett alternativ:

- ☐ 0.039
- ☐ 0.058
- ☐ 0.001
- ☐ 0.047
- ☐ 0.062
- ☐ 0.027

b) Anta nå at vannet faktisk er forurensset med *E. coli*. Hva er sannsynligheten for at et varsel *ikke* blir sendt ut den først dagen med forurensset vann?

Velg ett alternativ

- ☐ 0.020
- ☐ 0.001
- ☐ 0.040
- ☐ 0.032
- ☐ 0.027
- ☐ 0.005

Eierne av renseanlegget ønsker å teste en ny metode for deteksjon av E. coli. Metoden skal i følge leverandøren fungere på følgende måte: Dersom den faktiske konsentrasjonen av E. coli i vann er μ cfu/100ml så vil prøvene som tas være normalfordelte med forventning μ cfu/100ml og standardavvik $\sigma = 0.01$ cfu/100ml.

La X være en stokastisk variabel som representerer målinger gjort med den nye metoden på en kontrollprøve der konsentrasjonen av E. coli er kjent og lik 1 cfu/100ml. Vi har altså at $X \sim N(1, 0.01)$.

c) Hva er sannsynligheten for at en slik prøve gir et måleresultat mellom 0.98 og 1.02 cfu/100ml?

Velg ett alternativ

- ☐ 0.893
- ☐ 0.581
- ☐ 0.954
- ☐ 0.469
- ☐ 0.673
- ☐ 0.781

Eierne av renseanlegget mistenker at den nye metoden har en systematisk feil slik at målingene som gjøres i snitt viser for lave konsentrasjoner. De vil derfor gjennomføre en venstresidig Z-test ved signifikansnivå $\alpha = 0.05$ for å teste

$H_0: \mu = 1$ cfu/100ml mot $H_1: \mu < 1$ cfu/100ml

basert på fem uavhengige målinger av en kontrollprøve med kjent konsentrasjon av E. coli lik 1 cfu/100ml:

$x_1=0.964$, $x_2=1.001$, $x_3=0.948$, $x_4=0.954$, $x_5=1.021$

d) Basert på disse fem observasjonene, hvilken konklusjon er korrekt?

Velg ett alternativ

- ☐ Basert på observasjonen $z = 1.782$ kan vi ikke forkaste H_0 til fordel for H_1 ved signifikansnivå 0.05
- ☐ Basert på observasjonen $z = -1.568$ kan vi ikke forkaste H_0 til fordel for H_1 ved signifikansnivå 0.05
- ☐ Basert på observasjonen $z = -5.009$ kan vi forkaste H_0 til fordel for H_1 ved signifikansnivå 0.05
- ☐ Basert på observasjonen $z = 4.018$ kan vi forkaste H_0 til fordel for H_1 ved signifikansnivå 0.05
- ☐ Basert på observasjonen $z = 3.824$ kan vi forkaste H_0 til fordel for H_1 ved signifikansnivå 0.05
- ☐ Basert på observasjonen $z = -3.012$ kan vi ikke forkaste H_0 til fordel for H_1 ved signifikansnivå 0.05

Maks poeng: 4

5 Oppgave 5: Elektrisk tungtransport

Ved en ladestasjon for elektrisk tungtransport ankommer el-lastebiler ifølge en Poissonprosess med rate $\lambda = 3$ el-lastebiler per time. El-lastebiler som ankommer ladestasjonen blir alltid stående for å lade i en time og ladestasjonen har plass til at tre el-lastebiler lader samtidig.

a) Hva er sannsynligheten for at det i løpet av en time ankommer nøyaktig to el-lastebiler?

Velg ett alternativ

- ☐ 0.49
- ☐ 0.63
- ☐ 0.56
- ☐ 0.14
- ☐ 0.22
- ☐ 0.35

b) Hva er sannsynligheten for at det i løpet av en time ankommer mer enn tre el-lastebiler?

Velg ett alternativ

- ☐ 0.16
- ☐ 0.48
- ☐ 0.35
- ☐ 0.23
- ☐ 0.56
- ☐ 0.60

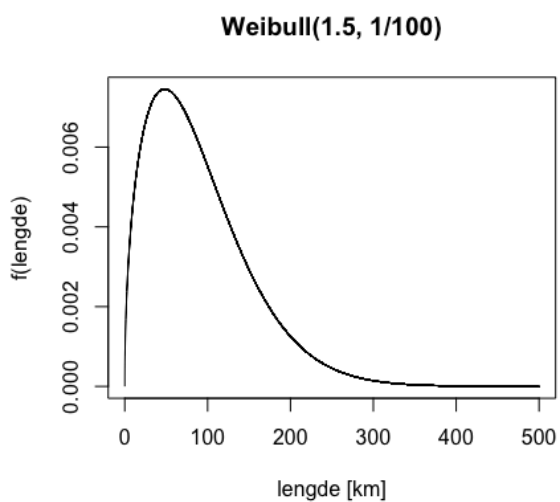
c) Klokka 14.05 har det nettopp ankommet en el-lastebil til ladestasjonen. Hva er sannsynligheten for at den neste el-lastebilen ankommer i løpet av 30 minutter?

Velg ett alternativ

- ☐ 0.78
- ☐ 0.81
- ☐ 0.53
- ☐ 0.65
- ☐ 0.97
- ☐ 0.47

Et transportfirma kjører el-lastebiler med en rekkevidde på 250 km. Firmaet vet av erfaring at på et tilfeldig valgt oppdrag er kjørelengden (distanse i km fra firmaets sentral og til leveringsstedet) Weibullfordelt med parametere $\alpha = 1.5$ og $\lambda = 1/100$.

Sannsynlighetstetthetsfunksjonen til denne weibullfordelingen er illustrert i figuren under.



d) Hva er sannsynligheten for at et tilfeldig valgt oppdrag har en kjørelengde på mer enn 200 km?

Velg ett alternativ

- ☐ 0.56
- ☐ 0.28
- ☐ 0.40
- ☐ 0.06
- ☐ 0.39
- ☐ 0.13

e) Hva er sannsynligheten for at et oppdrag kan fullføres (det vil si: kjøre fra firmaets sentral og til leveringsstedet - og *tilbake igjen*) uten lading? Du kan anta at el-lastebilen er fulladet når oppdraget starter.

Velg ett alternativ:

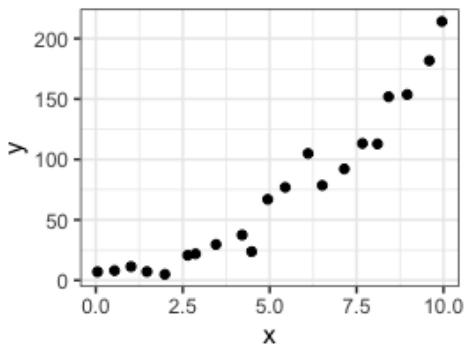
- ☐ 0.41
- ☐ 0.64
- ☐ 0.52
- ☐ 0.75
- ☐ 0.93
- ☐ 0.87

Maks poeng: 5

6 Oppgave 6: Enkel lineær regresjon

Vi skal studere sammenhengen mellom x og y ved hjelp av en enkel lineær regresjonsmodell. Målet vårt er å predikere en verdi for y når $x = 5$.

Figuren viser et kryssplott av 21 parvise observasjoner $(x_1, y_1), \dots, (x_{21}, y_{21})$.

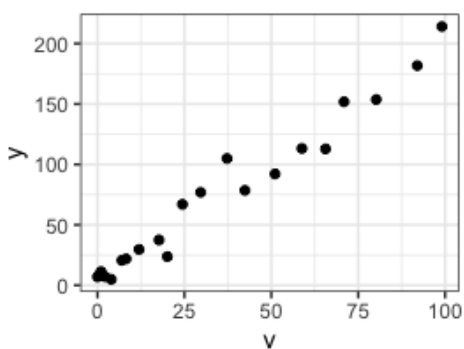


a) Hvilken antagelse fra lineær regresjon ser ut til å være brutt?

Velg ett alternativ

- ☐ Det er en lineær sammenheng mellom y og x
- ☐ Variasjonen i y er den samme uansett verdi til x
- ☐ Observasjonsparene er uavhengige
- ☐ Feilleddene er normalfordelte

Hvis vi i stedet plotter y mot $v = x^2$ får vi en figur der antagelsene for lineær regresjon ser ut til å være oppfylt.



For datasettet bestående av de parvise observasjonene $(y_1, v_1), \dots, (y_{21}, v_{21})$, der $v_i = x_i^2$ har vi følgende resultater:

$$\bar{y} = 72.3 \text{ og } \bar{v} = 34.5$$

$$\sum_{i=1}^{21} (v_i - \bar{v})(y_i - \bar{y}) = 40472.6$$

$$\sum_{i=1}^{21} (v_i - \bar{v})^2 = 20824.3 \text{ og}$$

$$\sum_{i=1}^{21} (y_i - \bar{y})^2 = 81401.5$$

b) Hva blir det estimerte stigningstallet for den lineære regresjonsmodellen $Y_i = \beta_0 + \beta_1 v_i + e_i$?

Velg ett alternativ

- ☐ 0.32
- ☐ 2.58
- ☐ 3.26
- ☐ 1.94
- ☐ 0.89
- ☐ 1.03

c) Tilbake til utgangspunktet vårt, hva blir predikert y når $x = 5$? Bruk modellen fra oppgave b) for å predikere y .

Velg ett alternativ

- ☐ 79.4
- ☐ 62.7
- ☐ 42.2
- ☐ 31.8
- ☐ 53.9
- ☐ 28.2

Maks poeng: 3