i Forside

Institutt for matematiske fag

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

Eksamensdato: 11. august 2023

Eksamenstid (frå-til): 09:00 - 12:00 (3 timar)

Hjelpemiddelkode/Tillatne hjelpemiddel: C
Godkjend kalkulator

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

Fagleg kontakt under eksamen:
Thea Bjørnland (41123849)

Fagleg kontakt kjem til eksamenslokalet: NEI

ANNA INFORMASJON:

Vend deg til ei eksamensvakt om du ønskjer å kontakte fagleg kontaktperson under eksamen. Noter spørsmålet ditt på førehand. Fagleg kontaktperson skal berre kontaktast dersom det er direkte feil eller manglar i oppgåvesettet.

Vekting av oppgåvene: er gjeve for kvar oppgåve. Det blir ikkje gjeve minus-poeng for feil svar på ei oppgåve.

Lagring: Svara dine i Inspera Assessment blir lagra automatisk kvart 15. sekund.

Varslinger: Dersom det oppstår behov for å gje beskjedar til kandidatane medan eksamen er i gang (til dømes ved klare manglar i oppgavesettet), vil dette bli gjort via varslingar i Inspera. Eit varsel vil dukke opp som ein dialogboks på skjermen i Inspera. Du kan finne att varselet ved å klikke på bjølla i øvre høgre hjørne på skjermen.

Trekk frå eksamen: Blir du sjuk under eksamen, eller av andre grunnar ønskjer å levere blankt/trekkje deg, gå til «hamburgarmenyen» i øvre høgre hjørne og vel «Lever blankt». Dette kan ikkje angrast sjølv om prøven framleis er open. Tilkall så eksamensvakt og følg instruksane frå eksamensvakta.

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 vere uavhengige normalfordelte stokastiske variablar slik at $X_i \sim N(\mu, \sigma)$ for i = 1, 2, 3.

Vi har gjort følgande tre observasjonar:
$x_1 = 6.6$, $x_2 = 3.8$ og $x_3 = 5.9$

Kva blir gjennomsnittet av observasjonane?
Vel eitt alternativ

- 7.25
- 2.98
- 6.89
- 3.62
- 4.31
- 5.43 ✔

Kva blir empirisk standardavvik?
Vel eitt alternativ

- 0.62
- 4.27
- 5.36
- 3.89
- 1.46 ✔
- 2.37

Kva er eit 95% konfidensintervall for μ?

Vel eitt alternativ

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

Maks poeng: 3

2 Oppgave 2: Simultanfordeling

Tabellen synar simultanfordelinga til to stokastiske variablar X og Y, P(X = x, Y = y). Vi kan til dømes 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

Frå denne simultanfordelinga kan ein til dømes rekne ut marginalfordelinga til variablen X som blir P(X = 1) = 0.3, P(X = 2) = 0.3, og P(X=3) = 0.4.

a) Kva blir marginalfordelinga til Y?
Vel eitt alternativ

- P(Y=1) = 0.7, P(Y=2) = 0.3
- P(Y=1) = 0.6, P(Y=2) = 0.4 ✔
- P(Y=1) = 0.5, P(Y=2) = 0.5
- P(Y=1) = 0.1, P(Y=2) = 0.9
- P(Y=1) = 0.2, P(Y=2) = 0.8
- Det er umuleg å rekne ut frå tabellen

b) Kva blir E(X)?
Vel eitt alternativ

- 2.1 ✔
- 2.8
- 1.9
- 3.2
- 1.3
- 0.4

c) Kva blir P(Y = 2 | X < 3)?

Vel eitt alternativ

- 0.2
- 0.6
- 0.5 ✔
- 0.1
- 0.4
- 0.3

Maks poeng: 3

3 Oppgave 3: Utregning av sannsynligheter med Python

a) I Python har vi skrive følgande kode:

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

Kva tal blir skrive ut (avrunda til to desimalar)?
Vel eitt alternativ

- 0.62
- 0.48
- 0.17
- 0.51
- 0.34
- 0.23 ✔

b) I Python har vi skrive følgende kode:

```
from scipy import stats
```

La Z vere ein standard normalfordelt stokastisk variabel. Kva kommando vil gje oss sannsynet $P(Z \leq 2)$?
Vel eitt alternativ

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

Maks poeng: 2

4 Oppgave 4: Drikkevann

I eit renseanlegg for drikkevatn gjerast det daglege kontrollar for bakterien E. coli. Di som konsentrasjonen av E. coli i drikkevatnet overstig 1 cfu/100ml sendast det varsel til innbyggerne om at vatnet må kokast før bruk.

Varslingssystemet er forbunde med noko usikkerheit. Sjøl om drikkevatnet *ikkje* er forureina med E. coli er det likevel eit sannsyn p = 0.001 for at eit varsel sendast ut til innbyggarane. Si som vatnet faktisk er forureina med E. coli vil systemet avdekke dette med eit sannsyn på 0.98.

a) Di som vatnet over ein periode på 40 dagar *ikkje* er forureina, kva er sannsynet for at det likevel sendast ut minst eit varsel til innbyggarane? Dei daglege kontrollane kan antakast å vere uavhengige forsøk med sannsyn for suksess p = 0.001.

Vel eitt alternativ

- 0.001
- 0.027
- 0.047
- 0.039 ✔
- 0.062
- 0.058

b) Anta no at vatnet faktisk er forureina med E. coli. Kva er sannsynligheten for at eit varsel *ikkje* blir sendt ut den først dagen med forureina vann?

Vel eitt alternativ

- 0.020 ✔
- 0.040
- 0.032
- 0.001
- 0.027
- 0.005

Eigarana av renseanlegget ynskjer å teste ein ny metode for deteksjon av E. coli. Metoden skal i følge leverandøren fungere på følgande måte: De som den faktiske konsentrasjonen av E. coli i vatn er μ cfu/100ml så vil prøvene som takast vere normalfordelte med forventning μ cfu/100ml og standardavvik σ = 0.01 cfu/100ml.

La X vere ein stokastisk variabel som representerar målingar gjort med den nye metoden på ei kontrollprøve der konsentrasjonen av E. coli er kjend og lik 1 cfu/100ml. Vi har då at X ~ N(1, 0.01).

c) Kva er sannsynet for at ei slik prøve gjer eit måleresultat mellom 0.98 og 1.02 cfu/100ml?
Vel eitt alternativ

- 0.469
- 0.954 ✔
- 0.581
- 0.893
- 0.781
- 0.673

Eigarane av renseanlegget mistenkjer at den nye metoden har ein systematisk feil slik at målingane som gjerast i snitt viser for låge konsentrasjonar. Dei vil difor gjere ein venstresidig Z-test ved signifikansnivå α = 0.05 for å teste

H_0: μ = 1 cfu/100ml mot H1: μ < 1 cfu/100ml

basert på fem uavhengige målingar av ei kontrollprøve med kjend 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 desse fem observasjonane, kva konklusjon er korrekt?

Vel eitt alternativ

- Basert på observasjonen z = 3.824 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 = -1.568 kan vi ikkje forkaste H_0 til fordel for H_1 ved signifikansnivå 0.05
- Basert på observasjonen z = -3.012 kan vi ikkje forkaste H_0 til fordel for H_1 ved signifikansnivå 0.05
- Basert på observasjonen z = 1.782 kan vi ikkje 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 ✔

Maks poeng: 4

5 Oppgave 5: Elektrisk tungtransport

Ved ein ladestasjon for elektrisk tungtransport kjem el-lastebiler i følge ein Poissonprosess med rate $\lambda = 3$ el-lastebilar per time. El-lastebilar som kjem til ladestasjonen blir alltid ståande for å lade i ein time og ladestasjonen har plass til at tre el-lastebiler ladar samstundes.

a) Kva er sannsynet for at det i løpe av ein time kjem nøyaktig to el-lastebilar?
Vel eitt alternativ

- 0.63
- 0.49
- 0.22 ✔
- 0.56
- 0.35
- 0.14

b) Kva er sannsynet for at det i løpet av ein time kjem meir enn tre el-lastebilar?
Vel eitt alternativ

- 0.23
- 0.35 ✔
- 0.60
- 0.56
- 0.16
- 0.48

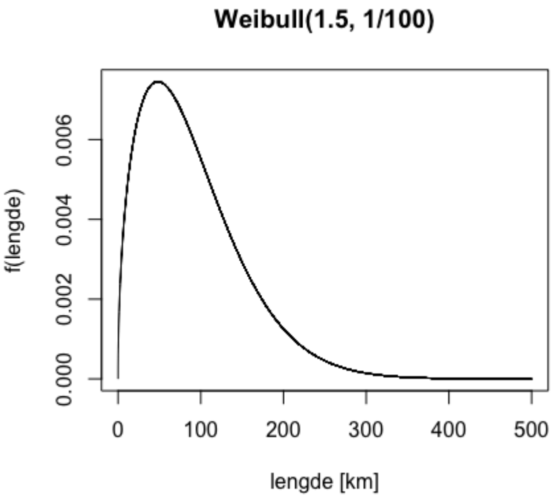
c) Klokka 14.05 har det nettopp kome ein el-lastebil til ladestasjonen. Kva er sannsynet for at den neste el-lastebilen kjem i løpet av 30 minutt?

Vel eitt alternativ

- 0.53
- 0.65
- 0.97
- 0.47
- 0.78 ✔
- 0.81

Eit transportfirma køyrer el-lastebilar med ei rekkevidde på 250 km. Firmaet veit av erfaring at på eit tilfeldig vald oppdrag er køyrelangda (distansa i km frå firmaet sin sentral og til leveringsstaden) Weibullfordelt med parametere $\alpha = 1.5$ og $\lambda = 1/100$.

Sannsynstettleiksfunksjonen til denne weibullfordelingen er illustrert i figuren under.



d) Kva er sannsynet for at eit tilfeldig vald oppdrag har ei kjørelengde på meir enn 200 km?

Vel eitt alternativ

- 0.28
- 0.39
- 0.13
- 0.40
- 0.06 ✔
- 0.56

e) Kva er sannsynet for at eit oppdrag kan fullførast (det vil sei: køyre frå firmaet sin sentral og til leveringsstaden - *og tilbake igjen*) uten lading? Du kan anta at el-lastebilen er fullada når oppdraget startar.

Vel eitt alternativ

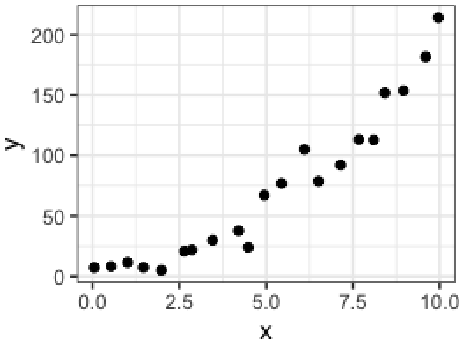
- 0.64
- 0.41
- 0.75 ✔
- 0.87
- 0.52
- 0.93

Maks poeng: 5

6 Oppgave 6: Enkel lineær regresjon

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

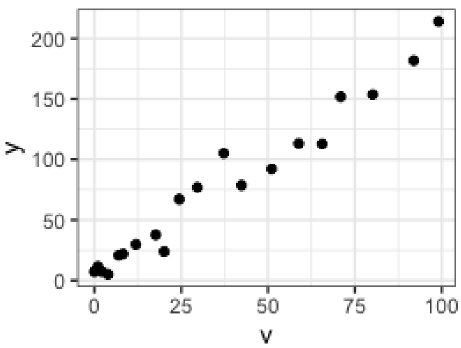
Figuren syner eit kryssplott for 21 parvise observasjonar (x_1,y_1), ..., (x_{21},y_{21}).



a) Kva antaking frå lineær regresjon ser ut til å vere brutt?
Vel eitt alternativ

- ◯ Feilledda er normalfordelte
- ◯ Det er ein lineær samanheng mellom y og x ✔
- ◯ Observasjonspara er uavhengige
- ◯ Variasjonen i y er den same uansett verdi til x

Om vi i steden plottar y mot v = x^2 får vi ein figur der antakelsane for lineær regresjon ser ut til å vere oppfyld.



For datasettet beståande av dei parvise observasjonane (y_1, v_1), ..., (y_{21}, v_{21}), der $v_i = x_i^2$ har vi følgande resultat:

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

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

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

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

b) Kva blir det estimerte stigningstalet for den lineær regresjonmodellen $Y_i = \beta_0 + \beta_1 v_i + e_i$?

Vel eitt alternativ

- 0.32
- 2.58
- 3.26
- 1.94 ✔
- 1.03
- 0.89

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

Vel eitt alternativ

- 42.2
- 62.7
- 79.4
- 31.8
- 28.2
- 53.9 ✔

Maks poeng: 3