

i Forside

Institutt for matematiske fag

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

Eksamensdato: 20.12.2023

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

Hjelpemiddelkode/Tillatte hjelpemidler:

Hjelpemiddelkode C. Godkjent kalkulator. Ingen håndskrevne eller trykte hjelpemidler, 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), Stine Marie Berge (93478689)

Faglig kontakt møter i eksamenslokalet: NEI

ANNEN INFORMASJON:

Skaff deg overblikk over oppgavesettet før du begynner på besvarelsen din.

Les oppgavene nøye, gjør dine egne antagelser og presiser i besvarelsen hvilke forutsetninger du har lagt til grunn i tolkning/avgrensing av oppgaven. Faglig kontaktperson kontaktes kun dersom det er direkte feil eller mangler i oppgavesettet. Henvend deg til en eksamensvakt hvis du mistenker feil og mangler. Noter spørsmålet ditt på forhånd.

Ingen håndtegninger: Denne eksamenen tillater ikke bruk av håndtegninger. Har du likevel fått utdelt skanne-ark, er dette en feil. **Arkene vil ikke bli akseptert for innlevering, og de vil derfor heller ikke sendes til sensur.**

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

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

Trekk fra/avbrutt eksamen: Blir du syk under eksamen, eller av andre grunner ønsker å levere blankt/avbryte eksamen, gå til "hamburgermenyen" i øvre høyre hjørne og velg «Lever blankt». Dette kan ikke angres selv om prøven fremdeles er åpen.

Tilgang til besvarelse: Etter eksamen finner du besvarelsen din i arkivet i Inspira. Merk at det kan ta én virkedag før eventuelle håndtegninger vil være tilgjengelige i arkivet.

1 Oppgave 1

Oppgave 1

a) En stokastisk variabel X kan ta verdiene 1, 2 og 3 og sannsynlighetsfordelingen er gitt i tabellen:

x	1	2	3
$P(X=x)$	0.5	0.1	0.4

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

Velg ett alternativ

☐ 0.3

☐ 0.1

☐ 0.6

☐ 0.4

☐ 0.2

☐ 0.5

b) La X representere en stokastisk variabel med $E(X) = 5$ og $\text{Var}(X) = 1$ og la Y representere en stokastisk variabel med $E(Y) = 2$ og $\text{Var}(Y) = 2$. Videre er $\text{Cov}(X, Y) = 0.5$.

Hva er $\text{Var}(X-Y)$?

Velg ett alternativ

☐ 2

☐ 0

☐ -1

☐ 0.5

☐ 1.5

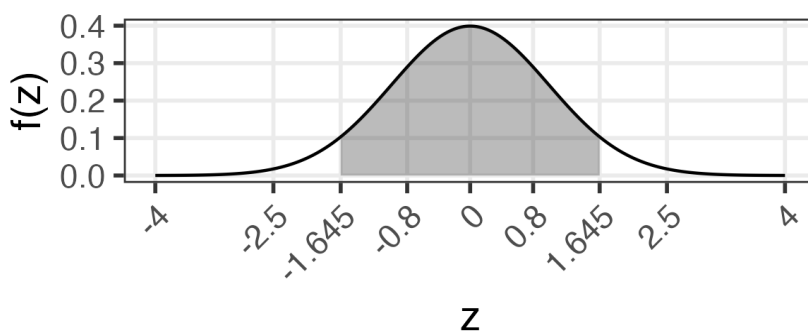
☐ 3

c) For en stokastisk variabel $T \sim \text{Weibull}(\alpha = 0.4, \lambda = 2)$, hva er $P(T \leq 1)$?

Velg ett alternativ

- ☐ 0.58
- ☐ 0.36
- ☐ 0.41
- ☐ 0.89
- ☐ 0.73
- ☐ 0.62

d) I figuren under ser vi sannsynlighetstettheten til en standard normalfordelt stokastisk variabel Z . Hva er arealet av det fargede (grå) området?



Velg ett alternativ

- ☐ 0.85
- ☐ 0.99
- ☐ 0.90
- ☐ 0.75
- ☐ 0.80
- ☐ 0.95

e) En ivrig hobbyfisker er på jakt etter den beste fiskeklassen i Mjøsa. Antall fisk som fanges per time kan modelleres med en Poissonprosess. Fiskeren har funnet seg en ny fiskeplass og ønsker å estimere fangstraten (suksessraten) λ . På 2 timer fanger fiskeren 12 fisk. Hva blir et tilnærmet 95% konfidensintervall for λ ?

Velg ett alternativ

- ☐ [3.45, 8.55]
- ☐ [2.61, 9.39]
- ☐ [0.38, 10.62]
- ☐ [5.21, 7.79]
- ☐ [4.78, 7.22]
- ☐ [1.98, 8.02]

f) Vi har et tilfeldig utvalg $X_1, X_2, \dots, X_{16}$ der $X_i \sim N(\mu, \sigma)$ med $\sigma = 4$, for $i = 1, \dots, 16$.

En hypotesetest for $H_0: \mu = 2$ mot $H_1: \mu < 2$ forkastes dersom snittet av de 16 observasjonene er mindre enn den kritiske verdien 0.35. Hva blir teststyrken dersom H_1 er sann og $\mu = 1.5$?

Velg ett alternativ

- ☐ 0.8621
- ☐ 0.9177
- ☐ 0.1251
- ☐ 0.4602
- ☐ 0.0026
- ☐ 0.6664

g) I Python har vi et datasett med to kolonner som heter v og w. Når vi kjører følgende Python-kode for lineær regresjon, hvilken regresjonsmodell estimerer vi?

```
import statsmodels.formula.api as smf
smf.ols('w ~ v',data).fit()
```

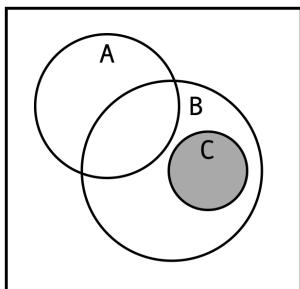
Velg ett alternativ:

- ☐ $v = \beta_1 w + e$
- ☐ $w = \beta_1 v + e$
- ☐ $y = \beta_0 v + \beta_1 w + e$
- ☐ $w = \beta_0 + \beta_1 v + e$
- ☐ $v = \beta_0 + \beta_1 w + e$
- ☐ $y = \beta_0 w + \beta_1 v + e$

Maks poeng: 7

2 Oppgave 2

Oppgave 2: Hendelser og sannsynlighet



a) Det grå området i venndiagrammet viser hendelsen C, hvilket av uttrykkene under er også en riktig beskrivelse av det grå området?

Velg ett alternativ

- ☐ $(B \cap C)^c$
- ☐ $A \cap C$
- ☐ $B \cup C$
- ☐ C^c
- ☐ $B \cap C$
- ☐ $(A \cap B)^c$

b) Dersom $P(A) = 0.4$, $P(B) = 0.3$ og $P(A \cap B) = 0.1$, hva er $P((A \cup B)^c)$?

Velg ett alternativ

- ☐ 0.7
- ☐ 0.8
- ☐ 0.3
- ☐ 0.4
- ☐ 0.5
- ☐ 0.6

Du er trener for et barnefotball-lag som består av 9 barn. Når barna spiller kamp kan laget kun ha 7 spillere på banen.

c) På hvor mange forskjellige måter kan du velge 7 barn som skal starte kampen? (Merk at det eneste som har noe å si for tellingen er om barnet er på eller av banen)

Velg ett alternativ:

- ☐ 56
- ☐ 48
- ☐ 24
- ☐ 72
- ☐ 64
- ☐ 36

d) To av barna på laget har blitt uvenner og nekter å spille sammen. Du skal nå trekke tilfeldig ett av de mulige lag-oppsettene på 7 barn, hva blir sannsynligheten for at du ikke kan bruke lag-oppsettet?

Velg ett alternativ

- ☐ 0.58
- ☐ 0.21
- ☐ 0.36
- ☐ 0.63
- ☐ 0.42
- ☐ 0.74

Maks poeng: 4

3 Oppgave 3

Oppgave 3: Aluminiumsplater

En maskin produserer aluminiumsplater som ifølge produktspesifikasjonen skal veie 100 gram. Det blir akseptert et lite avvik fra denne vekten, men dersom vekten avviker mer enn ± 1 gram fra denne verdien, blir aluminiumsplaten vurdert til å være defekt.

Anta at vekten på aluminiumsplatene som produseres er normalfordelt med forventning 100 gram og standardavvik 0.8 gram.

Ti plater pakkes i en pappeske. Pappesken veier 50 gram. Platenes vekt er uavhengige.

a) Hva er forventet vekt av pappesken med ti tilfeldig valgte plater?

Velg ett alternativ:

- ☐ 1000
- ☐ 200
- ☐ 100
- ☐ 2050
- ☐ 1050
- ☐ 150

b) Hva er standardddevviket?

Velg ett alternativ

- ☐ 4.76
- ☐ 5.41
- ☐ 1.92
- ☐ 6.81
- ☐ 2.53
- ☐ 3.87

c) I en eske med ti plater, hva er sannsynligheten for at minst 2 plater er defekte?

Velg ett alternativ

- ☐ 0.65
- ☐ 0.53
- ☐ 0.35
- ☐ 0.48
- ☐ 0.73
- ☐ 0.24

Det har kommet klager fra kunder som fører til at en ansatt ved bedriften mistenker at vekten på aluminiumsplatene har en snittvekt på mer enn 100 gram. Standardavviket antas å være 0.8 gram som tidligere.

Ved å veie 25 tilfeldig valgte plater skal bedriften gjennomføre en ensidig z-test for å teste om forventet vekt av en plate er større enn 100 gram, altså $H_0: \mu = 100$ gram mot $H_1: \mu > 100$ gram. De vil benytte signifikansnivå 0.001.

d) Et tilfeldig utvalg på 25 plater veies og samlet vekt blir målt til 2512 gram. Hva blir konklusjonen av hypotesetesten?

Velg ett alternativ

- ☐ Basert på observasjonen $z = 1.8$ kan vi ikke forkaste H_0 til fordel for H_1 ved signifikansnivå 0.001
- ☐ Basert på observasjonen $z = 3.0$ kan vi forkaste H_0 til fordel for H_1 ved signifikansnivå 0.001
- ☐ Basert på observasjonen $z = 1.8$ kan vi forkaste H_0 til fordel for H_1 ved signifikansnivå 0.001
- ☐ Basert på observasjonen $z = 3.0$ kan vi ikke forkaste H_0 til fordel for H_1 ved signifikansnivå 0.001
- ☐ Basert på observasjonen $z = -5.5$ kan vi forkaste H_0 til fordel for H_1 ved signifikansnivå 0.001
- ☐ Basert på observasjonen $z = -5.5$ kan vi ikke forkaste H_0 til fordel for H_1 ved signifikansnivå 0.001

4 Oppgave 4

Oppgave 4: Kobbertråd

I en tynn kobbertråd finnes det svakheter i materialet tilfeldig fordelt langs tråden. Antall svakheter på en tråd av lengde s millimeter kan antas å være Poissonfordelt med forventning λs der $\lambda = 0.2$ svakheter/millimeter.

a) Dersom du undersøker en 3 mm lang kobbertråd, hvor mange svakheter forventer du å finne? Det vil si, la X være antall svakheter på en kobbertråd av lengde 3 mm og regn ut forventningsverdien til X .

Velg ett alternativ:

- ☐ 0.5
- ☐ 1.6
- ☐ 1.5
- ☐ 2.6
- ☐ 2.5
- ☐ 0.6

b) Hva er sannsynligheten for å finne 2 eller færre svakheter på den 3 mm lange tråden, $P(X \leq 2)$?

Velg ett alternativ

- ☐ 0.56
- ☐ 0.41
- ☐ 0.85
- ☐ 0.98
- ☐ 0.67
- ☐ 0.73

Du trenger en bit med kobbertråd som ikke har svakheter. Du starter i enden av tråden og bestemmer deg for å kutte tråden så snart du finner den første svakheten. La den stokastiske variabelen S representere lengden til den feilfrie biten med tråd.

c) Hva er forventningsverdien til S ?

Velg ett alternativ

- ☐ 4 mm
- ☐ 3 mm
- ☐ 2 mm
- ☐ 5 mm
- ☐ 6 mm
- ☐ 7 mm

d) Du har studert 2 mm med tråd og har ikke funnet en svakhet. Hva er sannsynligheten for at du finner den første svakheten på den neste millimeteren? Det vil si beregn $P(S \leq 3 \mid S \geq 2)$.

Velg ett alternativ

- ☐ 0.73
- ☐ 0.64
- ☐ 0.35
- ☐ 0.21
- ☐ 0.87
- ☐ 0.18

Maks poeng: 4

5 Oppgave 5

Oppgave 5: Lineær regresjon

Vi har et datasett bestående av fem parvise observasjoner av høyde (i cm) og skostørrelse. Datasettet er lastet inn i Python og et skjermbilde vises her:

	skostr	hoyde
0	40	170
1	39	165
2	45	193
3	42	175
4	38	161

En lineær regresjonsmodell med høyde som responsvariabel og skostørrelse som forklaringsvariabel (kovariat) har blitt tilpasset. Modellen kan representeres slik:

$$\text{hoyde} = \beta_0 + \beta_1 \text{skostr} + \varepsilon, \varepsilon \sim N(0, \sigma)$$

Deler av modellsammendraget fra Python vises her:

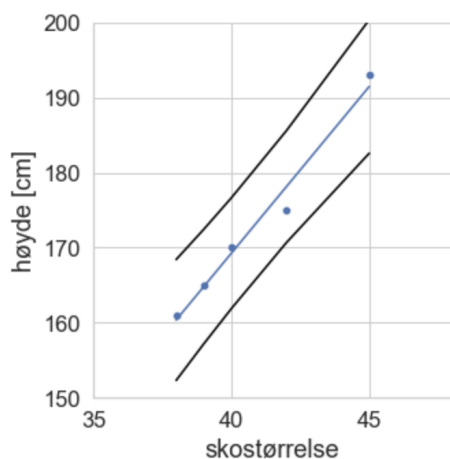
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-8.4156	15.446	-0.545	0.624	-57.573	40.742
skostr	4.4416	0.378	11.754	0.001	3.239	5.644

a) Hvilken hypotesetest utføres i linja som starter med "skostr"?

Velg ett alternativ

- ☐ $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 > 0$
- ☐ $H_0 : \beta_0 = 1$ mot $H_1 : \beta_0 \neq 1$
- ☐ $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$
- ☐ $H_0 : \beta_0 = 0$ mot $H_1 : \beta_0 \neq 0$
- ☐ $H_0 : \beta_1 = 1$ mot $H_1 : \beta_1 \neq 1$
- ☐ $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 < 0$

I figuren under ser vi et plott av de fem observasjonene, den tilpassede regresjonslinja og et 95% prediksjonsintervall.



b) Dersom du møter en person med skostørrelse 43, hva er din beste gjetning (i følge modellen) på personens høyde?

Velg ett alternativ

- ☐ 185.7 cm
- ☐ 182.6 cm
- ☐ 178.7 cm
- ☐ 180.8 cm
- ☐ 179.2 cm
- ☐ 183.8 cm

c) Hva er et 95% prediksjonsintervall på høyden til en tilfeldig valgt person med skostørrelse 40?

Velg ett alternativ:

- ☐ [162 cm, 177 cm]
- ☐ [166 cm, 174 cm]
- ☐ [156 cm, 184 cm]
- ☐ [168 cm, 172 cm]
- ☐ [158 cm, 182 cm]
- ☐ [160 cm, 175 cm]

