

i Eksamensoppgave i ISTA1001, ISTA1002, ISTA1003, ISTG1001, ISTG1002, ISTG1003, ISTT1001, ISTT1002, ISTT1003, VB6044 Statistikk

Dato: 19. desember 2024

Tid: 09:00 - 12:00

Faglig kontakt: Thea Bjørnland

Møter i eksamenslokalet: NEI

Hjelpemiddelkode/Tillatte hjelpemidler: Hjelpemiddelkode C. Godkjent kalkulator. Ingen håndskrevne eller trykte hjelpemidler da 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.

ANNEN INFORMASJON:

Faglig kontaktperson skal kun kontaktes 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.

FAGSPESIFIKK INFORMASJON

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:

Det er 22 deloppgaver og alle deloppgaver teller likt. Det gis ikke minus-poeng for gale eller manglede svar.

Varslinger:

Eventuelle beskjeder under eksamen (f.eks. ved feil i oppgavesettet), sendes ut 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:

Dersom du 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.

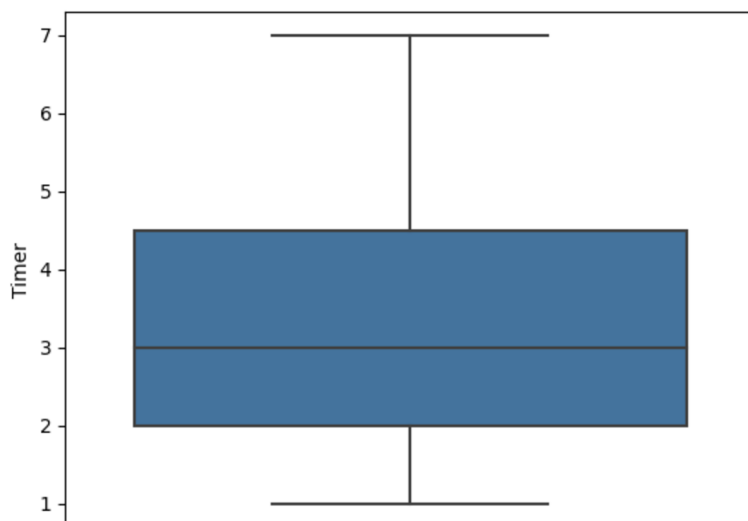
Tilgang til besvarelse:

Etter eksamen finner du besvarelsen din under tidligere prøver i Inspira. Merk at det kan ta én virkedag før eventuelle håndtegninger vil være tilgjengelige under «tidligere prøver».

1 Oppgave 1: Beskrivende statistikk

I en undersøkelse har 10 studenter rapportert tidsbruk, x_i (i timer), på en øving. Svarene er gitt i tabellen og illustrert i boksplottet.

Student i	1	2	3	4	5	6	7	8	9	10
Timer x_i	2	2	3	7	6	3	3	2	1	5



Det oppgis at $\sum_{i=1}^{10} x_i = 34$ og at $\sum_{i=1}^{10} (x_i - \bar{x})^2 = 34.4$.

a) Hva blir gjennomsnitt $\bar{x}$ og median m_x i tidsbruk for de ti studentene?

Velg ett alternativ

- ☐ $\bar{x} = 4.7$ timer og $m_x = 3$ timer
- ☐ $\bar{x} = 3.4$ timer og $m_x = 3$ timer
- ☐ $\bar{x} = 3.4$ timer og $m_x = 3.5$ timer
- ☐ $\bar{x} = 2.9$ timer og $m_x = 3$ timer
- ☐ $\bar{x} = 2.9$ timer og $m_x = 3.5$ timer
- ☐ $\bar{x} = 4.7$ timer og $m_x = 3.5$ timer

b) Hva blir det empiriske standardavviket?

Velg ett alternativ

- ☐ 1.32
- ☐ 1.96
- ☐ 1.53
- ☐ 2.08
- ☐ 0.87
- ☐ 2.55

Maks poeng: 2

2 Oppgave 2: Hendelser og sannsynlighet

En nettbutikk har fått inn en bestilling og skal sende en vare til en kunde. Nettbutikken har en garanti på at pakker kommer frem innen 3 dager.

Nettbutikken benytter to ulike budtjenester, PostlBox og NordPost. Dersom PostlBox er ledige er det denne tjenesten som velges. Av erfaring vet nettbutikken at PostlBox er ledige 60% av tiden, og at det med PostlBox er 99% sikkert at pakken kommer frem i tide. Dersom PostlBox ikke er ledige benyttes NordPost, men da er det bare 80% sjanse for at pakken kommer frem i tide.

a) Hva er sannsynligheten for at nettbutikken må bruke NordPost og at pakken ikke kommer frem i tide?

Velg ett alternativ

- ☐ 0.532
- ☐ 0.340
- ☐ 0.400
- ☐ 0.211
- ☐ 0.080
- ☐ 0.165

b) Hva er sannsynligheten for at varen fra en vilkårlig bestilling kommer frem i tide?

Velg ett alternativ

- ☐ 0.914
- ☐ 0.873
- ☐ 0.407
- ☐ 0.645
- ☐ 0.512
- ☐ 0.784

c) I en periode er det særdeles travelt i nettbutikken. Det er da 55% sjanse for at butikken bruker PostlBox, 40% sjanse for NordPost, og 5% sjanse for at pakken ikke blir sendt i det hele tatt. Sannsynlighetene for at en tilfeldig pakke kommer frem i den travle perioden er uendret: 99% for PostlBox, og 80% for NordPost.

En kunde har ikke mottatt bestillingen sin innen 3 dager, hva er sannsynligheten for at butikken sendte pakken med NordPost?

Velg ett alternativ

- ☐ 0.772
- ☐ 0.370
- ☐ 0.263
- ☐ 0.415
- ☐ 0.678
- ☐ 0.590

Maks poeng: 3

3 Oppgave 3: Diskrete sannsynlighetsfordelinger

a) La X være binomisk fordelt med parametere $n = 6$ og $p = 0.7$. Hva er $P(X=3)$?

Velg ett alternativ

- ☐ 0.451
- ☐ 0.590
- ☐ 0.185
- ☐ 0.038
- ☐ 0.219
- ☐ 0.377

b) La X være binomisk fordelt med parametere $n = 6$ og $p = 0.7$. Hva er $P(X \geq 3)$?

Velg ett alternativ

- ☐ 0.681
- ☐ 0.570
- ☐ 0.863
- ☐ 0.429
- ☐ 0.728
- ☐ 0.930

c) La X være binomisk fordelt med parametere $n = 6$ og $p = 0.7$. Hva er $E(2X+3)$?

Velg ett alternativ

- ☐ 7.9
- ☐ 10.7
- ☐ 9.2
- ☐ 12.8
- ☐ 11.4
- ☐ 8.5

d) La Y være Poissonfordelt med forventning $E(Y) = 4.5$. Hva er $P(Y > 5 \mid Y > 4)$?

Velg ett alternativ

- ☐ 0.756
- ☐ 0.129
- ☐ 0.372
- ☐ 0.593
- ☐ 0.481
- ☐ 0.635

e) La Y være Poissonfordelt med $E(Y) = 4.5$. Hva er $\text{Var}(2Y+3)$?

Velg ett alternativ

- ☐ 14.5
- ☐ 16.2
- ☐ 18.0
- ☐ 10.3
- ☐ 12.0
- ☐ 8.5

Maks poeng: 5

4 Oppgave 4: Normalfordeling

La V være normalfordelt med forventning $\mu_V = 2$ og standardavvik $\sigma_V = 2$.

a) Hva er $P(V \leq 2)$?

Velg ett alternativ:

☐ 0.50

☐ 0.40

☐ 0.60

☐ 0.45

☐ 0.35

☐ 0.55

b) Hva er $P(0 \leq V \leq 4)$?

Velg ett alternativ

☐ 0.78

☐ 0.68

☐ 0.51

☐ 0.96

☐ 0.49

☐ 0.32

c) Hvilket tall k oppfyller $P(V \geq k) = 0.10$?

Velg ett alternativ

- ☐ 6.59
- ☐ 4.56
- ☐ 2.76
- ☐ 5.27
- ☐ 3.98
- ☐ 1.30

d) Vi har fem uavhengige observasjoner $x_1, x_2, \dots, x_5$ fra en normalfordeling med ukjent forventning μ og kjent standardavvik $\sigma = 1$. Vi ønsker å undersøke $H_0 : \mu = 0$ mot $H_1 : \mu \leq 0$. Vi har observert $\bar{x} = -0.48$. Hva blir testens p-verdi, sannsynligheten for at snittet er lavere enn -0.48, under antagelsen om at null-hypotesen er sann?

Velg ett alternativ

- ☐ 0.14
- ☐ 0.18
- ☐ 0.01
- ☐ 0.03
- ☐ 0.36
- ☐ 0.46

Maks poeng: 4

5 Oppgave 5: Lyslenker og levetid

Anna har kjøpt en gammel lyslenke til juletreet sitt på Finn.no. Alle lyspærene fungerer og Anna vet at levetiden til hver av lyspærene er eksponentialfordelt med forventning 125 timer.

a) Hva er sannsynligheten for at en tilfeldig valgt pære fungerer i mer enn 100 timer?

Velg ett alternativ

- ☐ 0.860
- ☐ 0.629
- ☐ 0.381
- ☐ 0.712
- ☐ 0.449
- ☐ 0.532

b) Lyslenken består av lyspærer som er koblet i serie, det vil si at lenken fungerer bare dersom alle lyspærene fungerer. Levetidene til de n lyspærene er uavhengige. Hvilket uttrykk representerer sannsynligheten for at hele lyslenken fungerer i mer enn 100 timer?

Velg ett alternativ

- ☐ $(e^{-0.8})^n$
- ☐ $1 - n \cdot e^{-0.8}$
- ☐ $n \cdot e^{-0.8}$
- ☐ $1 - (1 - e^{-0.8})^n$
- ☐ $(1 - e^{-0.8})^n$
- ☐ $1 - (e^{-0.8})^n$

Axel har kjøpt nye trådløse julelys til sitt juletre. Han antar at levetidene til disse lysene er uavhengige og eksponentialfordelte med ukjent forventet levetid $1/\lambda$. Axel skrur på alle de $n = 30$ lysene samtidig og skal observere dem i 200 timer.

c) La X betegne antall lys som slutter å fungere i løpet av 200 timer. Vi lar antall 'suksesser' representere antall lys som slutter å fungere. Da vil X være binomisk fordelt med parametere $n = 30$ og suksess-sannsynlighet p gitt ved

Velg ett alternativ

- ☐ $p = \lambda/200$
- ☐ $p = 1 - e^{-200\lambda}$
- ☐ $p = e^{-200\lambda}$
- ☐ $p = 1 - e^{-\lambda/200}$
- ☐ $p = 200\lambda$
- ☐ $p = e^{-\lambda/200}$

d) Etter 200 timer observerer Axel at 8 av 30 lys har sluttet å fungere. Hva blir et (tilnærmet) 95% konfidensintervall for suksess-sannsynligheten p ?

Velg ett alternativ

- ☐ [0.252, 0.740]
- ☐ [0.093, 0.386]
- ☐ [0.108, 0.425]
- ☐ [0.231, 0.539]
- ☐ [0.051, 0.606]
- ☐ [0.171, 0.428]

e) Anton har kjøpt $m = 20$ trådløse lys av akkurat samme type som Axel. Han lar Y representere antall 'suksesser' (lys som går i stykker) i løpet av 200 timer. Dermed vil Y være binomisk fordelt med samme suksess-sannsynlighet p som hos Axel. Videre er det rimelig å anta at X og Y er uavhengige stokastiske variabler. Dersom vi skal kombinere Axel og Antons observasjoner til én estimator for p , hvilken av disse estimatorene vil vi foretrekke, og hvorfor? Bruk $n = 30$ og $m = 20$.

$$\hat{p}_1 = \frac{X+Y}{n+m}$$

$$\hat{p}_2 = 0.5 \left(\frac{X}{n} + \frac{Y}{m} \right)$$

Velg ett alternativ

- ☐ Vi foretrekker $\hat{p}_1$ siden $\hat{p}_2$ ikke er forventningsrett
- ☐ Vi foretrekker $\hat{p}_2$ siden $\hat{p}_1$ ikke er forventningsrett
- ☐ Estimatorene er like gode
- ☐ Vi foretrekker $\hat{p}_1$ siden begge estimatorene er forventningsrette men variansen til $\hat{p}_1$ er lavere enn variansen til $\hat{p}_2$
- ☐ Vi foretrekker $\hat{p}_2$ siden begge estimatorene er forventningsrette men variansen til $\hat{p}_2$ er lavere enn variansen til $\hat{p}_1$
- ☐ Ingen av estimatorene er forventningsrette, men vi foretrekker $\hat{p}_1$ siden variansen til $\hat{p}_1$ er lavere enn variansen til $\hat{p}_2$

Maks poeng: 5

6 Oppgave 6: Enkel lineær regresjon

Vi har $n = 9$ parvise observasjoner $(x_i, y_i), i = 1, \dots, n$. I dataene er $x_1 = x_2 = x_3 = x_4 = 3$ og $x_5 = x_6 = x_7 = x_8 = x_9 = 5$ slik at $\bar{x} = 37/9$.

Videre er $\bar{y} = 3.45$ og empirisk varians av y -verdiene er 0.748.

Vi har tilpasset en enkel lineær regresjonsmodell

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

der vi antar $\varepsilon_i \sim N(0, \sigma)$ og uavhengige feilledd $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_9$.

Modellsammendraget fra Python er gitt under.

OLS Regression Results						
Dep. Variable:	y	R-squared:	0.579			
Model:	OLS	Adj. R-squared:	0.519			
Method:	Least Squares	F-statistic:	9.645			
Date:	Wed, 27 Nov 2024	Prob (F-statistic):	0.0172			
Time:	08:22:52	Log-Likelihood:	-7.0384			
No. Observations:	9	AIC:	18.08			
Df Residuals:	7	BIC:	18.47			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	1.7782	0.851	2.090	0.075	-0.234	3.790
x	0.6248	0.201	3.106	0.017	0.149	1.100
Omnibus:	1.733	Durbin-Watson:	2.394			
Prob(Omnibus):	0.420	Jarque-Bera (JB):	0.765			
Skew:	0.141	Prob(JB):	0.682			
Kurtosis:	1.599	Cond. No.	19.0			

a) Basert på modellsammendraget, hvilken påstand er sann?

Velg ett alternativ

- ☐ Et 95% konfidensintervall for skjæringspunktet er [0.851, 2.090].
- ☐ Standardavviket i x -verdiene er 0.6248.
- ☐ Et 95% konfidensintervall for stigningstallet er [0.149, 1.100].
- ☐ Formelen for den estimerte regresjonslinja er $\hat{y} = 0.6248 + 1.7782x$
- ☐ I en to-sidig hypotesetest for stigningstallet er p -verdien lik 0.075.
- ☐ Ved signifikansnivå 0.05 har vi ikke grunn til å forkaste nullhypotesen om at stigningstallet er 0.

b) Fra modellsammendraget kan vi lese at $R^2 = 0.579$, og vi har oppgitt at empirisk varians i y-verdiene er 0.748. Hva blir estimatet på variansen σ^2 til støyleddene ε_i i regresjonsmodellen? Hint: Estimatet for variansen i y-verdiene kan uttrykkes ved $SST/(n-1)$, mens estimatet for variansen til støyleddene kan uttrykkes ved $SSE/(n-2)$.

Velg ett alternativ

- ☐ 0.36
- ☐ 2.96
- ☐ 1.20
- ☐ 2.21
- ☐ 0.78
- ☐ 1.92

c) Vi ønsker å teste $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$ med en to-sidig t-test og med signifikansnivå $\alpha = 0.01$. Testobservatoren er

$$T = \hat{\beta}_1 / SE(\hat{\beta}_1)$$

Hva blir riktig forkastningsregel?

Velg ett alternativ

- ☐ Vi forkaster null-hypotesen dersom vi observerer $T > 1.960$
- ☐ Vi forkaster null-hypotesen dersom vi observerer $T < -3.499$ eller $T > 3.499$
- ☐ Vi forkaster null-hypotesen dersom vi observerer $T < -2.776$ eller $T > 2.776$
- ☐ Vi forkaster null-hypotesen dersom vi observerer $T < -1.645$
- ☐ Vi forkaster null-hypotesen dersom vi observerer $T < -2.718$ eller $T > 2.718$
- ☐ Vi forkaster null-hypotesen dersom vi observerer $T < -4.541$ eller $T > 4.541$

Maks poeng: 3