Eksamen 5. juni 2018
Eksamenstid 4 timar

IR201812 Statistikk og Simulering
Nynorsk

Tillatne hjelpemiddel er kalkulator og læreboka i statistikk. Eigne notat i boka er lov.
Ta med **all mellomrekning** som trengst for å grunngje svaret.

Oppgåve 1 .... *(14%)*

I eit vekentleg minilotteri er der 10% sannsyn for å vinne 50 kr., 10% for 150 kr., og 5% for 300 kr. Loddet kostar 40 kr. (Eit lodd kan berre vinna éin av dei tre premiane, og lodd som ikkje vinn ein av desse, vinn ingenting.) Lat den stokastiske variabelen X_j vera summen ein deltakar vinn i lotteriet i veke j.

(a) Rekn ut forventingsverdien for X_j.

Solution: $\mu = 0{,}1 \cdot 50 + 0{,}1 \cdot 150 + 0{,}05 \cdot 300 = 35$

(b) Rekn ut standardavviket for X_j.

Solution:

$$\begin{aligned}\sigma^2 &= 0{,}1 \cdot (50-35)^2 + 0{,}1 \cdot (150-35)^2 + 0{,}05 \cdot (300+35)^2 + 0{,}7 \cdot (0-35)^2 \\ &= 0{,}1 \cdot 15^2 + 0{,}1 \cdot 115^2 + 0{,}05 \cdot 265^2 + 0{,}7 \cdot 35^2 \qquad (1) \\ &= 5775\end{aligned}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{5775} \approx 75{,}99 \qquad (2)$$

Jan deltek 50 veker på rad i lotteriet. Lat den stokastiske variabelen Y vera den totale summen han vinn, minus den totale summen han betaler for lodda.

(c) Skriv Y som ein funksjon av X_j.

Solution:

$$Y = \sum_{i=1}^{50}(X_i - 40) = \left(\sum_{i=1}^{50} X_i\right) - 2000$$

(d) Rekn ut forventingsverdien for Y.

Solution:

$$E(Y) = \sum_{i=1}^{50}(E(X_i) - 40) = 50 \cdot (-5) = -250$$

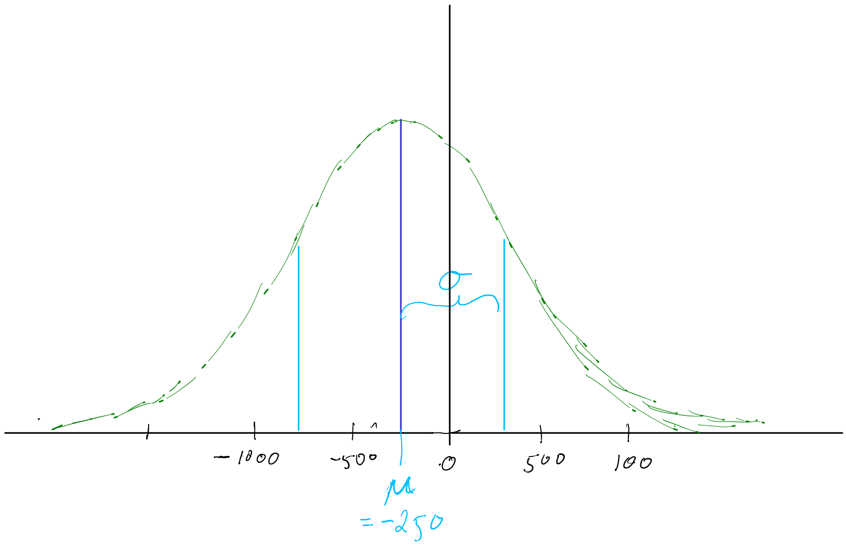
(e) Rekn ut standardavviket for Y.

Solution:

$$\mathsf{var}(Y) = \sum_{i=1}^{50} \mathsf{var}(X_i) = 50 \cdot 5775$$

$$\mathsf{StdDev}(Y) = \sqrt{\mathsf{var}(Y)} = \sqrt{50 \cdot 5775} = 573{,}35$$

(f) Teikn ei skisse av sannsynsfordelinga (PDF) for Y, der du òg indikerer forventingsverdien og standardavviket.



Solution:

(g) Rekn ut sannsynligheten for at Jan har tjent mer enn han har betalt etter å ha delteke 50 gongar i lotteriet. (Det vil seia: Rekn ut $P(Y > 0)$).

Solution: Y er normalfordelt med $\mu = -250$ og $\sigma = 534{,}50$. Dermed er Z standardnormalfordelt dersom me definerer

$$Z = \frac{Y + 250}{534{,}50}$$

Det gjev

$$P(Y > 0) = P\big(Z > \frac{250}{534{,}50}\big) = P(Z > 0{,}47) \approx 0{,}319$$

Oppgåve 2 . . . . *(4%)*

Bruk sannsynstabellane i boka og finn fylgjande:

(a) $P(Z > 2{,}0)$ der Z er standardnormalfordelt.

Solution: $P(Z > 2{,}0) = 0{,}0228$

(b) Den kritiske verdien t slik at $P(T < t) = 0{,}05$ der T har Students t-fordeling med fem fridomsgradar.

Solution:

$$t \approx -2{,}015$$

Oppgåve 3 . . . . *(6%)*

Ein stokastisk variabel X er binomialfordelt med n forsøk og punktsannsyn $\pi = 0{,}2$. Svar på fylgjande spørsmål, og forklar korleis du kjem fram til svaret.

(a) Lat $n = 6$, og finn $P(X \leq 2)$.

Solution:

$$\begin{aligned} P(X \leq 2) &= P(X=0) + P(X=1) + P(X=2) \\ &= \binom{6}{0} \cdot 0{,}2^0 \cdot 0{,}8^6 + \binom{6}{1} \cdot 0{,}2^1 \cdot 0{,}8^5 + \binom{6}{2} \cdot 0{,}2^2 \cdot 0{,}8^4 \\ &= 0{,}9011 \end{aligned} \tag{3}$$

(b) Lat $n = 180$, og finn $P(X \leq 40)$.

Solution: Fordi n er stor, kan me bruka normalfordelinga (sentralgrensesatsen). Me har $\mu = n\pi = 36$ og $\sigma^2 = n\pi(1-\pi) = 28.8$.
Me normaliserer med formelen

$$Z = \frac{X - \mu}{\sigma}$$

Dette gjev

$$P(X \leq 60) = P(Z \leq \frac{40-36}{\sqrt{28.8}}) = P(Z \leq 0{,}7454) \approx 0{,}772$$

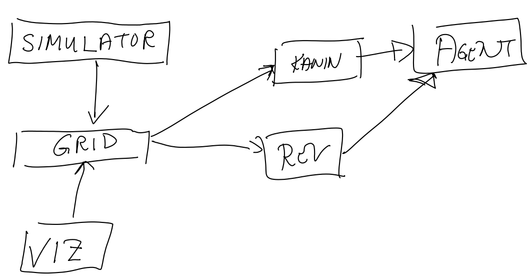
Oppgåve 4 . . . . *(12%)*

Sjå for deg ein agentbasert modell for eit økosystem med rev og kanin (rov- og byttedyr), der agentane flytter seg over eit rutenett eller raster (*grid*), og ein *enkel* simulatorimplementasjon av denne modellen.

(a) Forklar hovudprinsippet for agentbasert simulering.

Solution: Grunnlaget for agentbasert simulering er agentbasert modellering, der kvart einskild individ (agent) er modellert individuelt. I agentbasert simulering vert kvart individ implementert som ein modul (objekt i OOP), og vha. polymorfi kan simulatormotoren halda greie på alle agentane utan å kjenna korleis dei er modellert eller implementert.

(b) Teikn ei skisse til klassediagram for ein enkel simulator for rovdyr/byttedyr-problemet.



Solution:

(c) Forklar kort funksjon og formål for kvar av klassene.

Solution: Agent-klassen er ei abstrakt klasse. Alle agentane er ansvarlege for sin eigen oppførsel.

Grid definerer landskapet og held styr på kvar agentane er. Eit grid-objekt kan òg fortelja agentane kva dei ser.

Viz representerer all den grafiske visualiseringa av simuleringa.

Simulator har ansvaret for å instantiera modellen og styrer klokka for å fortelja agentane når dei skal handla.

(d) Kva klasser i klassediagrammet ditt er agentar?

Solution: Rev og kanin er agentar, og klassene arvar frå ei felles agentklasse.

(e) Gje to døme på at ulike agenttypar i modellen din har (eller kan ha) ulike eigenskapar.

Solution: Reven vil leita etter kaninen, medan kaninen vil unngå reven.

Reven kan svelta ihjel, medan kaninen berre døyr når han vert eten.

Oppgåve 5 . . . . *(10%)*

Sjå no på ein agentbasert simulering av ein varemarknad, der me ynskjer å studera samanhengen mellom pris og omsetnad. Bedriftene sel svært liknande produkt, men dei kan gjera strategiske val som prissetting, reklame, osv. Kundane kan velja kor mykje dei vil kjøpa og av kven. Bedriftene er nøydde til å tena pengar, ellers går dei konkurs og forsvinn frå modellen.

(a) Samanlikn problemet med rovdyr/byttedyr-modellen i forrige oppgåve. Kva element kan du gjenbruka, og kva må skrivast om?

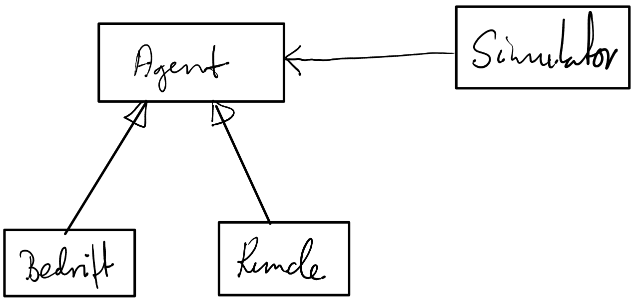
Solution: Me bruker agentbaserte modellar i båe fall, og dermed kan me gjenbruka sjølve simulatormotoren og den abstrakte agentklassa.

Den største skilnaden er at rov- og byttedyra lever i ein todimensjonal landskap som regulerer kva agentar som kan samhandla (les eta) kven. Marknadsproblemet kan ta ulike variantar, men slik det er skildra er det naturleg å la alle personar handla frå alle produsentar.

Dei einskilde agentane (underklassene av agent) må utviklast spesielt for kvart problem.

(b) Teikn klassediagram for marknadssimulatoren, og forklar kva dei viktigaste klassene gjer.

Solution:



(c) Forklar kort korleis du vil modellera og implementera kva agentane gjer kvar runde, for kvar agenttype. Modellen kan godt vera enkel.

Solution: Bedriftene må sette pris og reklamebudsjett kvar runde. Når kundane har handla, må dei rekna ut forteneste og kapital. Dersom de har negativ kapital går de konkurs og forsvinn.

Kundene må sjå kva bedrifter som sel til kva pris. Det er naturleg å tenkja at kundane sannsynglegvis handlar hjå same bedrift som forrige runde, dvs. at dei må hugsa kven dei har handla hos før, men at der er eit visst sannsyn for å byta bedrift, særleg dersom nokon er billigare, eller er vort godt kjende gjennom reklame.

Her er mange variantar, og detaljert forståing av bedriftsøkonomiske samanhengar er ikkje pensum. Målet er å visa døme på korleis enkle og lettfattelege mekanismar kan modellerast.

Oppgåve 6 . . . . *(15%)*

Billys burger lagar burgerane for hand. Dei har fått ein del klagar på at burgerane ikkje alltid har rett storleik, og dette må dei undersøkja. Det er viktig både at vekta er rett i gjennomsnitt, og at standardavviket er so lite som råd. Me skal difor estimere både forventingsverdien og standardavviket.

Billy veg seks burgerar og finn fylgjande vekter i gram:

$$160, 170, 180, 180, 185, 190$$

(a) Finn eit (punkt)estimat for forventa vekt (forventingsverdien).

Solution: $\bar{x} = \frac{1}{6}(160 + 170 + 180 + 180 + 185 + 190) = 177{,}5$

(b) Rekn ut (eit estimat for) standardfeilen for estimatoren som du bruker.

> **Solution:** Standardfeilen er standardavviket åt estimatoren $\bar{X}$.
> Me rekner fyrst ut s for utvalet.
>
> $$s^2 = \frac{1}{5}((-17{,}5)^2 + (-7{,}5)^2 + 2{,}5^2 + 2{,}5^2 + 7{,}5^2 + 12{,}5^2) = 117{,}5$$
>
> og $s = \sqrt{117{,}5} \approx 10{,}840$.
> Me estimerer standardfeilen som
>
> $$\hat{\sigma}_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{s}{\sqrt{6}} \approx 4{,}56$$

(c) Finn eit 95% konfidensinterval for forventa vekt. Presiser kva føresetnader du må gjera om sannsynsfordelinga for vekta.

> **Solution:** Utvalet er for lite til å bruka sentralgrensesatsen. Me må difor forutsetja at variabelen er normalfordelt, slik at me kan bruka t-fordelinga med fem fridomsgradar.
> Eit 95% konfidensintervall er da gjeve som
>
> $$\bar{x} - t_{0{,}025}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{0{,}025}\frac{s}{\sqrt{n}}$$
>
> eller
>
> $$177{,}5 - 2{,}571 \cdot 4{,}56 < \mu < 177{,}5 + 2{,}571 \cdot 4{,}56$$
>
> eller
>
> $$165{,}78 < \mu < 189{,}22$$

(d) Finn eit (punkt)estimat for standardavviket til vekta på ein burger.

> **Solution:** Som estimator for standardavviket bruker me utvalsstandardavviket $s = 11{,}18$.

(e) Forklar korleis du kan bruka *bootstrap* for å estimera standardfeilen på estimatoren for standardavviket. (Du vel sjølv om du vil forklara vha. programkode, pseudokode, reknedøme eller anna.)

> **Solution:** Standardfeilen er standardavviket til utvalsstandardavviket. Dersom me kan generera eit utval med fleire observasjonar av utvalsstandardavviket, kan me rekna ut utvalsstandardavviket for dette utvalet (av utvalsstandardavvik) og bruuka det som estimator.
>
> Bootstrap simulerer nye utval ved å ta utval av utvalet. Frå det opprinnelege utvalet (gjeve i oppgåve) trekk me N tilfeldige *bootstrap*-utval, kvar på $n = 6$ element, *med tilbakelegging.* Me finn utvalsstandardavviket for kvart *bootstrap*-utval, og får N observasjonar $s_1, \ldots, s_N$. Me reknar so ut utvalsstandardavviket s' for $s_1, \ldots, s_N$. Det er estimatet vårt for standardfeilen.
>
> Det er kritisk at (1) kandidaten presiserer at trekninga er med tilbakelegging, og at (2) det er tydleg at me reknar utvalsstandardavvik av utvalsstandardavvik.

Oppgåve 7 .... *(15%)*

Me studerer samanhengen mellom reklame og omsetnad for ei bedrift. Lat X vera summen ho bruker på reklame og Y omsetnaden, begge delar i tusen kroner. Me observerer følgende:

X	Y
10	68
45	210
60	250
72	305
90	400
100	440

$$\sum_i x_i = 377$$

$$\sum_i y_i = 1673$$

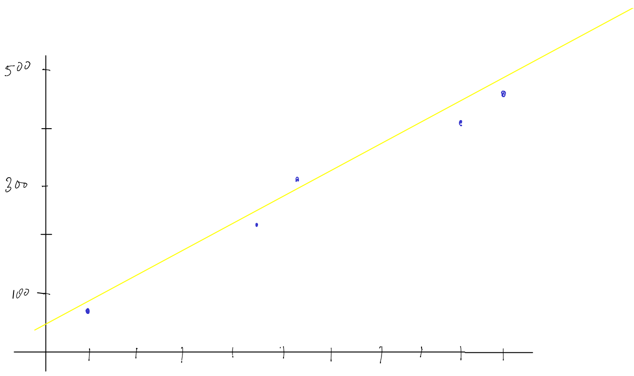
$$\sum_i x_i^2 = 29\,009$$

$$\sum_i y_i^2 = 557\,849$$

$$\sum_i x_i y_i = 127\,090$$

(a) Skisser datapunkta i eit plott.

Solution:



(b) Er der korrelasjon mellom X og Y? Forklar svaret.

Solution: Det ser slik ut fra plottet, siden datapunktene ligger rundt en linje med stigningstall ulik 0 eller ∞. Men vi må huske at dette bare er et utvalg fra populasjonen, så vi kan ikke si med sikkerhet at X og Y er korrelerte.

(c) Kva kan me seia om avhenget mellom X og Y? Forklar svaret.

Solution: Siden X og Y er korrelerte, sier vi at det er en avhengighet mellom disse variablene. Men det betyr ikke automatisk at det finnes en *logisk* avhengighet mellom X og Y. F. eks. kan både X og Y være styrt av en tredje variabel.

(d) Bruk minste kvadratsum-metoden for å finna den beste rette lina for å beskriva datasettet.

Solution:

$$\begin{aligned} b =& \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \\ =& \frac{6 \cdot 127090 - 377 \cdot 1673}{6 \cdot 29009 - 377^2} = 4{,}129 \\ a =& \bar{y} - b \cdot \bar{x} = \frac{1673}{6} - 4{,}129 \cdot \frac{377}{6} = 19{,}39 \end{aligned} \tag{4}$$

Vi finner derfor regresjonslinja $y = 19{,}39 + 4{,}129 \cdot x$

(e) Rekn ut korrelasjonskoeffisienten mellom X og Y.

Solution:

$$\begin{aligned}S_{xx} &= \sum x^2 - n\bar{x}^2 = 29\,009 - 6 \cdot (\frac{377}{6})^2 = 5320{,}8333 \\ S_{yy} &= \sum y^2 - n\bar{y}^2 = 557\,849 - 6 \cdot (\frac{1673}{6})^2 = 91\,360{,}8333 \\ S_{xy} &= \sum xy - n\bar{x}\bar{y} = 127\,090 - 6 \cdot \frac{377}{6} \cdot \frac{1673}{6} = 21\,969{,}833\,33 \\ r &= \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{21\,969{,}833\,33}{\sqrt{5320{,}8333 \cdot 91\,360{,}8333}} = 0{,}9965\end{aligned} \tag{5}$$

Oppgåve 8 .... *(12%)*

Ein partikkel flyttar seg på eit 2D raster. Sannsynsfordelinga etter eitt steg er gjeve som

		40%		
	20%	*start*	20%	
		20%		

(a) Rekn ut sannsynsfordelingen for posisjonen etter to steg.

Solution:

		16%		
	16%		16%	
4%		24%		4%
	8%		8%	
		4%		

(b) Rekn ut sannsynsfordelingen for posisjonen etter tre steg.

Solution:

			6,4%			
		9,6%		9,6%		
	4,8%		19,2%		4,8%	
0,8%		12%		12%		0,8%
	2,4%		9,6%		2,4%	
		2,4%		2,4%		
			0,8%			

Posisjonen (X_i, Y_i) etter i steg er gjeve ved to stokastiske variablar X_i og Y_i. No skal me berre sjå på rørsla langs Y-aksa.

(c) Finn forventingsverdien $E(Y_{100})$.

Solution: Sannsynsfordelinga for Y_1 er gjeve ved at $P(Y_1 = +1) = 0{,}4$, $P(Y_1 = -1) = 0{,}2$ og $P(Y_1 = 0) = 0{,}4$, og dermed har me $E(Y_1) = 0{,}2$.
Sidan Y_{100} er summen av 100 identisk fordelte steg, får me $E(Y_{100}) = 100 \cdot 0{,}2 = 20$.

(d) Finn variansen åt Y_{100}.

Solution:

$$\begin{aligned}\mathsf{var}(Y_1) &= 0{,}4 \cdot (1-0{,}2)^2 + 0{,}4 \cdot (-1-0{,}2)^2 + 0{,}2 \cdot (0-0{,}2)^2 \\ &= 0{,}4 \cdot (0{,}8)^2 + 0{,}4 \cdot (-1{,}2)^2 + 0{,}2 \cdot (-0{,}2)^2 \\ &= 0{,}84\end{aligned} \tag{6}$$

Sidan Y_{100} er summen av 100 uavhengige, identisk fordelte steg, får me $\mathsf{var}(Y_{100}) = 100 \cdot 0{,}84 = 84$.

Oppgåve 9 .... *(12%)*

Når me simulerer stokastiske prosessar bruker me slumptal. Ein klassisk slumptalsgenerator er den lineære kongruensgeneratoren:

$$x_i = a \cdot x_{i-1} + c \pmod{m}.$$

(a) Forklar kva me meiner med perioden til generatoren. På kva måte er perioden vesentleg for bruken av generatoren i praksis?

> **Solution:** Perioden er det minste talet på slumptal du kan trekkja før generatoren tek til å gjenta den same sekvensen.

(b) Finn den minste perioden til generatoren over når $a = 2$, $c = 3$ og $m = 18$.

> **Solution:** Med 1 som frø får me fylgja $1, 5, 13, 11, 7, 17, 1$, med periode 6. Med 3, har me $3, 9, 3$, og periode 2. Med 15 har me $15, 15$, og den minste perioden er dermed 1. Det rekk å visa at 15 har periode 1, sidan 0 ikkje er mogleg.

(c) Slumptalsgeneratoren krev eit *frø* (*seed*). Kva meiner me med eit frø og korleis vert det brukt for å gje oss ein serie slumptal frå generatoren over?

> **Solution:** Frøet er starttilstanden for generatoren, dvs. x_0 i dømet over. Kvart slumptal er ein funksjon av føregåande, og ein treng altso eit nullte tal for å kunne starta generatoren.

(d) Korleis ville du ha vald frø i eit praktisk system, og kvifor?

> **Solution:** Ein vanleg teknikk er å bruka dei minst signifikante bitsa i klokketida. Det er enkelt å få til i praksis, relativt portabelt mellom platformar, og det er relativt uforutsigbart (tilfeldig).