Statistikk og statistisk programmering, våren 2022

Obligatorisk oppgave 2

Innleveringsfrist: Fredag 11. februar 2022, kl. 23.59

Leveres som én pdf-fil på Canvas.

Alle oppgavene unntatt den siste skal i utgangspunktet gjøres uten Python.
Den siste oppgaven (hvor dere *skal* bruke Python) må gjøres for å få godkjent obligen.

Oppgave 1

La den stokastiske variabelen X være antall seksere ved kast med tre terninger.

a) Finn verdimengden til X.

b) Finn sannsynlighetsfordelingen (punktsannsynligheten) til X.

c) Finn forventningsverdien til X.

d) Forklar hva forventningsverdien til X uttrykker.

e) Finn standardavviket til X.

Oppgave 2

La X være antall systemsvikt kommende uke i et nytt datasystem. Sannsynlighetsfordelingen (punktsannsynlighetene) er gitt i følgende tabell:

x	0	1	2	3	4
$P(X = x)$	0.45	0.31	0.14	0.07	0.03

a) Tegn et sannsynlighetshistogram for denne fordelingen.

b) Finn forventningsverdi.

c) Finn varians og standardavvik.

d) Tegn den kumulative fordelingsfunksjonen $F(x)$ for X.

e) Finn $P(X > 2)$.

f) $P(X < 2)$.

Anta nå at bedriften som leverer dette datasystemet må betale et gebyr basert på hvor mange ganger i løpet av en uke systemet svikter. Dette gebyret beregnes som

$$g(X) = 2000X^2 + 5000X + 1000$$

g) Hva er forventningsverdien til gebyret?

Oppgave 3

En stokastisk variabel har følgende sannsynlighetsfordeling:

x	1	2	3	4	5	6	7
$P(X = x)$	0.08	0.15	0.31	0.24	0.11	0.07	k

a) Hva må k være for at dette skal være en sannsynlighetsfordeling?

b) Finn $P(X < 2)$.

c) Finn $P(X \geq 5)$.

d) Finn $P(2 \leq X < 7)$.

e) Finn forventningsverdien til X.

f) Finn variansen og standardavviket til X.

Oppgave 4

La X_1, X_2 og X_3 være uavhengige stokastiske variabler med lik sannsynlighetsfordeling med forventningsverdi $\mu = 2$ og varians $\sigma^2 = 9$.

a) Beregn forventning og varians til de stokastiske variablene
$Y = X_1 + X_2 + X_3$ og
$Z = X_1 + X_1 + X_1$.
Sammenlign svarene for Y og Z.

b) Beregn forventning og varians til $W = 3X_1 - 4X_2$.

c) Beregn forventningsverdien til $V = X_1 \cdot X_2 \cdot X_3$

Oppgave 5
Følgende tabell viser simultanfordelingen til X og Y.

x	y			
	0	1	2	$P(X = x)$
0	0.1	0.2	0.1	0.4
1	0.3	0.0	0.3	0.6
$P(Y = y)$	0.4	0.2	0.4	

a) Beregn $E(X)$ og $E(Y)$.

b) Finn $E(X \cdot Y)$.

c) Finn kovariansen og korrelasjonen.

d) Undersøk om variablene er uavhengige.

Oppgave 6
Denne oppgaven krever grunnleggende kunnskap om integraler. De som ikke har hatt integrasjon trenger ikke å gjøre denne oppgaven.

Noen personer kaster mynter mot en vegg, og det gjelder å komme nærmest mulig veggen for å vinne. La X være avstanden i meter fra veggen til der Christians mynt lander. Følgende sannsynlighetstetthet er en god modell:

$$f(x) = \begin{cases} 3(1-x)^2 & 0 < x \le 1 \\ 0 & ellers \end{cases}$$

a) Beregn sannsynligheten $P(0.5 < X < 0.8)$.

b) Beregn hvor nær veggen Christian kan forvente at mynten lander.

c) Finn standardavviket til X.

Oppgave 7
Et firma lager en elektronisk komponent. Erfaring tilsier at 20 % av komponentene er defekte.

a) Vi plukker tilfeldig ut n komponenter. La X betegne antall defekte komponenter i utvalget. Begrunn at X er binomisk fordelt med parametere n og $p = 0.2$.

b) En kunde ber om å få tilsendt så mange komponenter at han kan *forvente* at minst åtte av dem er feilfrie. Hvor mange komponenter må firmaet sende?

c) Hvor stor er da sannsynligheten for at firmaet sender minst åtte feilfrie komponenter?

Oppgave 8

Denne oppgaven må være gjort for å få godkjent obligen.

Lag et Python-program for en binomisk forsøksrekke.

Programmet skal be brukeren taste inn antall forsøk (n) og sannsynligheten for hendelsen i hvert forsøk (p).

Programmet skal så plotte punktsannsynligheten og fordelingsfunksjonen.

Tips:

i) Man kan bruke `input` for å lese inn data fra brukeren.

ii) For å få Python til å beregne sannsynligheter i en binomisk fordeling, kan man bruke
`scipy.stats.binom.pmf`
(pmf er "probability mass function", altså punktsannsynligheten)
og
`scipy.stats.binom.cdf`
(cdf er "cumulative distribution function", altså fordelingsfunksjonen $F(x)$)

iii) For å beregne «alle på en gang», altså sannsynlighetene for $x = 0, 1, 2, \ldots, n$, kan man lage seg et array av disse tallene ved å bruke
`numpy.arange(n+1)`
Merk at man med «arange» mener et verdiområde («a range»), og ikke å «ordne», altså ikke «arrange» som man kan bli forledet til å tro. Merk også at `arange(n+1)` betyr en tallfølge på $n + 1$ tall fra og med 0. Følgelig er det tallene fra og med 0 til og med n, altså ikke til og med $n + 1$.

iv) For å plotte kan man bruke
`matplotlib.pyplot.plot`