NTNU
Kunnskap for en bedre verden

Institutt for lærerutdanning

SENSURVEILEDNING til eksamensoppgave i MGLU2503 Matematikk 2 (5-10) emne 2

Eksamensdato: 26. november 2019

Eksamenstid (fra-til): 09.00-15.00

Hjelpemiddelkode/Tillatte hjelpemidler: Bestemt, enkel kalkulator tillatt**.**

Annen informasjon: Alle oppgavene skal besvares og svarene begrunnes. Karakteren blir satt etter vurdering med prosentvurderingsmetoden og en helhetsvurdering av besvarelsen.

Oppgave 1
Du får oppgitt gjennomsnittet og variasjonsbredden til et sett med datapunkter. Gjennomsnittet er 5 og variasjonsbredden er 8.

a) Hva kan du si om datasettet på bakgrunn av disse tallene?
Det er to vanlige tolkninger av gjennomsnitt for diskrete data: gjennomsnittet som balansepunktet til datasettet, og gjennomsnittet som lik fordeling. Forskjellen mellom største og minste verdi i datasettet er 8.

b) Hva kan du i tillegg si om datasettet hvis du også får oppgitt medianen?
Medianen er et mål på midtpunktet i datasettet. Halvparten av datapunktene i datasettet er større og halvparten er mindre enn medianen. Hvis forskjellen mellom median og gjennomsnitt er liten indikerer det at datasettet er symmetrisk fordelt, mens hvis forskjellen er stor indikerer det at datasettet er skjevfordelt.

c) Hva kan du fortsatt ikke vite om datasettet hvis du også får oppgitt medianen?
Vi vet ikke noe om fordelingen av datapunktene (formen på dataene). Variasjonsbredden påvirkes av ekstremverdier (outliers). Vi vet ikke om de fleste verdiene ligger tett inntil gjennomsnittet eller langt fra. Noen muligheter kan være at dataene er normalfordelt rundt gjennomsnittet på 5 med høy eller lav varians, eller uniformt fordelt rundt gjennomsnittet på 5, eller ha en bimodal fordeling med mange datapunkter rundt 0 og 8.

d) Gi konkrete eksempler på minst tre ulike datasett (med dataverdier) som illustrerer egenskaper ved ulike sentral- og variasjonsmål. Forklar valgene dine.
Her kan man velge datapunkter som for eksempel illustrerer stor og liten varians, symmetrisk fordeling og skjevfordeling, unimodal og bimodal fordeling, m.m. Datasettene trenger ikke ha egenskapene i oppgaveteksten. Et diagram over fordelingene kan godkjennes på lik linje med tallverdier.

Eksempler:
X1 = {0,3,5,7,7,7,9,12,13}. Symmetrisk. Gjennomsnitt = median.
X2 = {0,7,7,7,7,7,7,7,13}. Symmetrisk. Gjennomsnitt = median. Samme variasjonsbredde som X1, men lavere varians.
X3 = {5,5,5,5,5,18}. Skjevfordelt. Gjennomsnitt = 7, median = 5.
X4 = {0,0,0,0,0,0,0,5,6,7,8,9,13,13,13,13,13,13,13}. Symmetrisk. Gjennomsnitt = median. Samme variasjonsbredde som X1, men høy varians.

Oppgave 2
Gjør rede for de ulike trinnene i datadetektivens syklus. Velg ut fire sentrale begreper fra statistikken og beskriv hvordan elevene kan tilegne seg kunnskap om disse gjennom å arbeide med datadetektivens syklus. Skriv maksimalt én side.

Datadetektivens fem trinn:
i. problem,
ii. plan,
iii. data,
iv. analyse,
v. konklusjon

Fire sentrale begreper:
Her er det mange begreper å velge mellom. En god besvarelse knytter begrepene opp mot datadetektivens syklus og beskriver et undervisningsopplegg eller en aktivitet hvor det vises konkret hvordan elevene kan få erfaring med og utvikle forståelse av begrepene.

Et eksempel kan være å skrive om hvordan innsamling av data fra en egenkomponert spørreundersøkelse kan føre til at man samler inn både kategoriske og numeriske data, og at man får inn svar som er feil, eller oppgitt i ulike måleenheter (f.eks. noen svar i cm og noen i m). Her kan man beskrive begreper som rensing av data, ulike datatyper, og begreper knyttet til å analysere ulike typer data.

Ikke utfyllende liste over relevante begreper:
Strategisk tenking innen statistikk.
Ulike typer data: kategoriske data og tallmateriale. Diskrete, kontinuerlige, og grupperte data.
Lokalisering, innsamling og rensing av data.
Sentralmål og variasjonsmål
Fordelinger av (formen til) data
Ulike visuelle fremstillinger av data
Hyppighet og frekvens
Årsakssammenheng, korrelasjon, regresjon

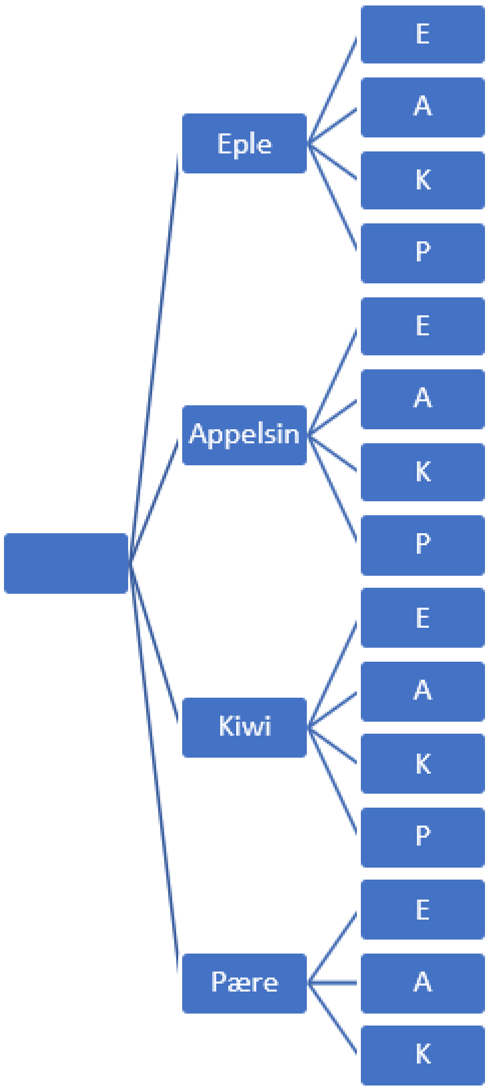
Oppgave 3
I ei fruktkurv ligger det 3 epler, 4 appelsiner, 2 kiwi og 1 pære.

a) Du ønsker å trekke deg en frukt. Skriv opp utfallsrommet.
U = {eple, appelsin, kiwi, pære}. En god besvarelse skriver kort og presist om hva et utfallsrom er og hvorfor {eple, appelsin, kiwi, pære} er utfallsrommet.

b) Lag en modell som viser hva som kan skje dersom du skal trekke to frukter.
For eksempel:
1)

1. trekk	Eple	OK	OK	OK	OK
	Appelsin	OK	OK	OK	OK
	Kiwi	OK	OK	OK	OK
	Pære	OK	OK	OK	UMULIG
		Eple	Appelsin	Kiwi	Pære
		2. trekk			

2) 1. trekk 2. trekk



c) Rikke ønsker å trekke seg frukt. Hendelse A er at Rikke trekker en sitrusfrukt (appelsin eller kiwi) og hendelse B er at Rikke trekker det hun anser som en god frukt til bruk i smoothie (eple, kiwi og pære).
Forklar hva det følgende betyr i konteksten som er gitt:

i) P(B|A)
Sannsynligheten for at du har trukket en god smoothiefrukt dersom du vet at den frukten du har trukket er en sitrusfrukt.

ii) P(A|B)
Sannsynligheten for at du har trukket en sitrusfrukt dersom du vet at den frukten du har trukket er en god smoothiefrukt.

iii) P(BUA)
Sannsynligheten for at frukten du har trukket er en god smoothiefrukt eller en sitrusfrukt eller en frukt som er begge delene.

iv) P(B∩A)
Sannsynligheten for at frukten du har trukket er både en god smoothiefrukt og en sitrusfrukt.

d) Finn de fire sannsynlighetene i)-iv) og forklar hva disse sannsynlighetene betyr. Ville du brukt nøyaktig denne oppgaven i undervisning? Hvorfor/hvorfor ikke?

i) $P(B|A) = P(B \cap A)/P(A) = (2/10) / (6/10) = 2/6 = 1/3$
1 av 3 sitrusfrukter er en god smoothiefrukt.

ii) $P(A|B) = P(B \cap A)/P(B) = (2/10) / (6/10) = 2/6 = 1/3$
1 av 3 smoothiefrukter er en sitrusfrukt.

iii) $P(BUA) = P(A)+P(B)-P(A \cap B) = 6/10+6/10-2/10 = 10/10 = 1$
Det at sannsynligheten for unionen av to hendelser er 1 betyr at utfallene i disse hendelsene er det samme som hele utfallsrommet. Det er umulig å trekke en frukt som verken er en sitrus- eller en god smoothiefrukt.

iv) $P(B \cap A) = 2/10 = 1/5$
1 av 5 frukter er både sitrus- og god smoothiefrukt.

I og med at $P(B|A) = P(A|B)$ og $P(AUB) = 1$ er dette en dårlig egna oppgave for å utforske og forstå hva som ligger i begrepene betinget sannsynlighet og union. Generelt er ikke $P(B|A) = P(A|B)$ og $P(AUB) = 1$ i sannsynlighetsoppgaver. For å bruke samme kontekst kunne en endra på antallet frukter slik at $P(B|A)$ ikke blir lik $P(A|B)$ og $P(AUB)$ ikke blir lik 1.
Oppgaven kan bli brukt som et eksempel på at $P(B|A)$ noen ganger kan være lik $P(A|B)$ og at $P(AUB)$ kan være lik $P(U) = 1$, men det er i så fall viktig at det kommer fram at dette ikke gjelder generelt.

Oppgave 4
Vi ser fortsatt på samme fruktkurv som i oppgave 3.

a) Magne skal trekke seg fire frukter til fruktsalat. Han deler opp og legger fruktene han trekker i fruktsalatskåla umiddelbart etter hver av fruktene er valgt. Hvor mange måter kan trekkinga gjøres på dersom

i) rekkefølgen han trekker fruktene i ikke betyr noe
Mulige kombinasjoner:
3 epler: EEEA, EEEK, EEEP
2 epler: EEAA, EEAK, EEAP, EEKK, EEKP
1 eple: EAAA, EAAK, EAAP, EAKK, EAKP, EKKP
0 epler: AAAA, AAAK, AAAP, AAKK, AAKP, AKKP
Totalt 20 mulige kombinasjoner.

ii) rekkefølgen han trekker fruktene i betyr noe
Hvis 4 like: 1 kombinasjon
Hvis 3+1 like: 4 kombinasjoner
Hvis 2+1+1 like: 12 kombinasjoner
Hvis 2+2 like: 6 kombinasjoner
Hvis 4 ulike: 24 kombinasjoner
3 epler: EEEA, EEEK, EEEP
(totalt $3 \cdot 4 = 12$ kombinasjoner med tre epler)
2 epler: EEAA, EEAK, EEAP, EEKK, EEKP
(totalt $2 \cdot 6 + 3 \cdot 12 = 48$ kombinasjoner med to epler)

1 eple: EAAA, EAAK, EAAP, EAKK, EAKP, EKKP
(totalt 1·4+4·12+1·24 = 76 kombinasjoner med 1 eple)
0 epler: AAAA, AAAK, AAAP, AAKK, AAKP, AKKP
(totalt 1·1+2·4+1·6+2·12 = 39 kombinasjoner med 0 epler)

Totalt blir dette 12+48+76+39 = 175 mulige kombinasjoner.

b) Andrine ønsker å spise alle ti fruktene i fruktkurva i løpet av en dag. Hun velger seg en frukt klokka 10 og deretter en ny frukt hver av de 9 påfølgende timene. På hvor mange forskjellige måter kan de ti fruktene velges?
En kan tenke at en vil plassere en av frukttypene ut på x av 10 plasser først, deretter neste frukttype på y av 10-x plasser. Deretter den tredje frukta på z av 10-x-y plasser. Siste frukttype kan da bare plasseres på én måte på de resterende 10-x-y-z plassene. Antallet mulige plasseringer av x frukter på 10 plasser er 10 over x:

$$\binom{10}{x} = \frac{10!}{(10-x)!\,x!}$$

Totalt blir antall mulige plasseringer av frukter på disse ti timene 10C3 ·7C4·3C2·1C1 = 10C4·6C3·3C2·1C1 = 10C1·9C4·5C2·3C3 = 12600

c) Dersom Andrine hadde ei større fruktkurv med 10 av hver av fruktene og fortsatt skulle velge seg 10 frukter i løpet av en dag: På hvor mange måter kunne hun valgt fruktene nå?
Her har hun nok frukt til at en kan tenke på dette som et ordna utvalg med tilbakelegging (selv om det egentlig er uten tilbakelegging). Antall kombinasjoner blir dermed 4^{10}= 1 048 576.

Oppgave 5

I Stokastik-boka og i artikkelen til Shaughnessy (2007) blir det gitt fem generelle anbefalinger til hvordan en bør gjennomføre undervisning i sannsynlighet. En av anbefalingene er at en skal lage koblinger mellom statistikk og sannsynlighet. Forklar hvordan dette kan gjøres ved å beskrive et konkret undervisningsopplegg basert på en hensiktsmessig kontekst.

En av anbefalingene Shaughnessy (2007) er at en skal lage koblinger mellom statistikk og sannsynlighet. Ut fra Shaughnessys forklaring "The concept of statistical variation in a probability distribution is closely connected to the concept of sample space in a probability task," finner vi at forbindelse mellom de to stokastiske begrepene er gjennom statistisk variasjon og stikkprøve av sannsynlighet. En kan beskrive koblingen ved eksempel som viser stikkprøve av sannsynligheter og variasjon fra data i et kontekstspesifisert eksempel (kast mynt n ganger, eller terningkast m ganger, ...).
En god besvarelse beskriver et hensiktsmessig undervisningsopplegg og forklarer hvorfor det er en hensiktsmessig kontekst/et hensiktsmessig opplegg som er beskrevet.

Oppgave 6

En middels stor skog inneholder 640 elger, hvorav 140 som er merket. En ivrig elgjeger feller til sammen åtte elger i jakta. La Y være antall merkede elger i jegerens fangst.

a) Anta at Y er hypergeometrisk fordelt. Finn $P(Y=3)$ og $P(Y \leq 3)$.

N = Antall elger
S= Antall merkede elger i skogen
n = Antall forsøk, altså det antallet elger elgjegeren feller
Y = Antall merkede elger i jegerens fangst

Y er hypergeometrisk fordelt (fordi vi har to grupper: Merkede og umerkede elger som trekkes ut tilfeldig uten tilbakelegging). Av N=64 elger, er vi interessert i antall merkede elger, Y, av n=8 forsøk.

Sannsynlighet for Y=3 er

$$P(Y=3)=\frac{\binom{140}{3}\binom{640-140}{8-3}}{\binom{640}{8}}=\frac{\binom{140}{3}\binom{500}{5}}{\binom{640}{8}} \approx 0{,}1710$$

Ønsker å finne $P(Y \leq 3)=P(Y=0)+P(Y=1)+P(Y=2)+P(Y=3)$
Hvor

$$P(Y=2)=\frac{\binom{140}{2}\binom{640-140}{8-2}}{\binom{640}{8}}=\frac{\binom{140}{2}\binom{500}{6}}{\binom{640}{8}} \approx 0{,}3067$$

$$P(Y=1)=\frac{\binom{140}{1}\binom{640-140}{8-1}}{\binom{640}{8}}=\frac{\binom{140}{1}\binom{500}{7}}{\binom{640}{8}} \approx 0{,}3114$$ og

$$P(Y=0)=\frac{\binom{140}{0}\binom{640-140}{8-0}}{\binom{640}{8}}=\frac{\binom{140}{0}\binom{500}{8}}{\binom{640}{8}} \approx 0{,}1371.$$

Derfor er $P(Y \leq 3) \approx 0{,}171+0{,}3067+0{,}3114+0{,}1371=0{,}9262 \approx 0{,}926$

I en annen, kjempestor skog er det 6400 elger, hvorav 1400 er merket. Alle jegerne i denne skogen feller til sammen 800 elger. La X være antall merkede elger blant de felte.

b) Finn P(X ≤ 200) og P(160 ≤ X ≤ 185).

X er tilnærmet binomisk med $n = 800$ og $p = \frac{14}{64}$.

X er også tilnærmet normalfordelt med $\mu = np = 175$ og $\sigma^2 = np(1-p) = 136.7$.

$$P(X \leq 200) = P\left(\frac{X-\mu}{\sigma} \leq \frac{200-\mu}{\sigma}\right) = P\left(Z \leq \frac{200-175}{\sqrt{136.7}}\right)$$

$$= P\left(Z \leq \frac{25}{\sqrt{136.7}}\right) \approx P(Z \leq 2{,}14) = 0.9838$$

$$P(160 \leq X \leq 185) = P\left(\frac{160-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{185-\mu}{\sigma}\right)$$

$$= P\left(\frac{160-175}{\sqrt{136.7}} \leq Z \leq \frac{185-175}{\sqrt{136.7}}\right) = P\left(\frac{-15}{\sqrt{136.7}} \leq Z \leq \frac{10}{\sqrt{136.7}}\right)$$

$$\approx P(-1{,}28 \leq Z \leq 0{,}86) = P(Z \leq 0{,}86) - (1 - P(Z \leq 1{,}28))$$

$$= 0{,}8051 - 0{,}1003 \approx 0{,}71$$

Oppgave 7

Gjennomsnittlig høyde for alle norske vernepliktige menn er 180 cm. Unge menn fra Finnmark er tilsynelatende lavere. I 2011 var det 218 vernepliktige fra Finnmark, med gjennomsnittshøyde 177,5 cm (kilde: SSB). Anta at kroppshøyden til en tilfeldig ung mann fra Finnmark, er normalfordelt med forventning μ og standardavvik $\sigma = 7$cm. Anta også at det er tilfeldig hvem som utfører verneplikten. Er det grunnlag for å hevde at unge menn fra Finnmark er lavere enn landsgjennomsnittet? Utfør hypotesetest med signifikansnivå 0,01.

Gjennomfør en hypotesetest av gjennomsnitt

H_0: Finnmark normalt høye $(\mu \geq 180)$, og

H_1: Finnmark lavere $(\mu \leq 180)$,

Med $\mu_0 = 180$, $n = 218$, $\bar{X} = 177.5$, $\sigma = 7$ og $\alpha = 0{,}01 = 1\,\%$.

Dette er hypotesetesting av μ når standardavviket er kjent. Så vi beregner

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{177{,}5 - 180}{7/\sqrt{218}} = -5{,}273$$

Da forkaster vi H_0 siden $\alpha = 0{,}01$ er stor.

(Fra tabellene i vedlegg 3/4 ser en at $Z = -5{,}273$ gir en verdi som er mindre enn $0{,}01$. $Z = -2{,}33$ er grensa.)