Institutt for lærerutdanning

SENSURVEILEDNING til eksamensoppgave i LGU52015 Matematikk 2 (5-10) emne 2

Eksamensdato: 10. desember 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 (20%)

Du får oppgitt gjennomsnittet og variasjonsbredden til et sett med datapunkter. Gjennomsnittet er 7 og variasjonsbredden er 13.

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 13.

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 7 med høy eller lav varians, eller uniformt fordelt rundt gjennomsnittet på 7, eller ha en bimodal fordeling med mange datapunkter rundt 0 og 13.

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 (13%)

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 (15%)

En vanlig kortstokk består av fire sorter med 13 kort i hver sort. Hver sort består av kort med verdi fra og med 2 til og med 10, knekt, dame, konge og ess. I tillegg er det to jokere som en kan velge til å være hvilken som helst verdi og sort. De to jokerne er helt like.

I kortspillet Rome blandes to slike kortstokker sammen, slik at bunken av kort som brukes totalt har 108 kort. Anta at kortstokken er godt stokket, og at du skal trekke ett kort.

a) Forklar hva som menes med utfall, utfallsrom og hendelse.

   Notasjon: s=spar, Kn=knekt, D=dame, K=konge, E=ess

   Utfall: En konkret ting som kan skje i et (stokastisk) forsøk. I dette tilfellet kan en for eksempel trekke en knekt i spar. Dette er et mulig utfall av forsøket "trekke ett kort".

   Hendelse: En kombinasjon av ett eller flere utfall i et forsøk. For eksempel kan en hendelse ved forsøket "trekke ett kort" være å trekke en spar eller å trekke et bildekort.

   Utfallsrom: Mengden av alle mulige utfall av et forsøk. I dette tilfellet består utfallsrommet av alle de mulige kortene en kan trekke (E, 2, 3, …, 10, Kn, D, K i alle de fire sortene i tillegg til joker).

   En god besvarelse forklarer de tre begrepene både generelt og i kortstokk-konteksten.

b) I Rome kan det være lurt å samle på kort i én av sortene. Jokere kan være en valgfri sort. Vi ser på hendelsen A: "En trekker en spar eller en joker". Skriv opp utfallene som gir denne hendelsen. Hvor mange utfall er det i hendelse A og hva er sannsynligheten for hendelse A? Forklar.

   Det er to mulige måter å skrive opp A:
   A = {sE, s2, s3, s4, s5, s6, s7, s8, s8, s9, s10, sKn, sD, sK, joker}
   A = {spar, joker}

   Avhengig av hvordan en definerer utfall og hendelse A (se over) har en tre mulige antall utfall i A: 2 (spar og joker), 14 (alle de 13 spar-verdiene og joker) og 30 (to av hver av de 13 spar-verdiene og fire jokere). En god besvarelse må forklare hvorfor/hvordan en har kommet fram til antallet en har skrevet ned, og gjerne også kommentere at det kan være flere korrekte svar.

$$P(A) = 30/108 = 15/54 = 5/18 \approx 0{,}278$$

c) I Rome kan det også være lurt å samle på små kort. En annen mulig hendelse, *B*, er at en trekker et kort med verdi større eller lik 2 og mindre eller lik 7 (eller en joker). Hva er sannsynligheten for hendelse *B*?

Det er 52 av 108 kort som gir en verdi mellom 2 og 7 eller er en joker (6·4·2 + 4).

$$P(B) = 52/108 = 13/27 \approx 0{,}481$$

d) Systematiser opplysningene om hendelsene *A* og *B* i en krysstabell.

Det er to mulige krysstabeller:

	A	ikke A	SUM
B	16	36	52
ikke B	14	42	56
SUM	30	78	108

	P(A)	P(ikke A)	SUM
P(B)	16/108	36/108	52/108
P(ikke B)	14/108	42/108	56/108
SUM	30/108	78/108	1

En god besvarelse setter opp en korrekt tabell og forklarer hvordan en kom fram til verdiene i tabellen.

e) Finn P(B∩A), P(BUA) og P(A|B). Forklar hva de tre ulike sannsynlighetene sier noe om.

P(B∩A) = 16/108 = 4/27 = 0,148

Dette er sannsynligheten for å trekke et kort som både er spar/joker og mellom 2-7 (eller joker).

P(BUA) = (52+30-16)/108 = (82-16)/108 = 66/108 = 11/18 = 0,611

Dette er sannsynligheten for å trekke et kort som enten er spar/joker, har verdi mellom 2-7 (eller joker) eller er en spar mellom 2 og 7 (eller joker).

P(A|B) = 16/52 = 4/13 = 0,308

Dette er sannsynligheten for at kortet du har trukket er spar/joker gitt at du vet at kortet har verdi mellom 2 og 7 (eller er joker).

En god besvarelse må forklare hvorfor en har brukt de tallene en har brukt i utregninga og forklare hva sannsynlighetene betyr i konteksten (og gjerne generelt også).

Oppgave 4 (16%)

Vi ser nå på en enkelt, vanlig kortstokk (som beskrevet i oppgave 3) uten jokere.

a) Magne skal trekke seg fire kort. Han legger kortene han har trukket i en egen bunke. Hvor mange forskjellige bunker kan Magne trekke dersom

i. rekkefølgen han trekker kortene i ikke betyr noe?

Dette er et uordna utvalg uten tilbakelegging, som gir at antall mulige kombinasjoner er 52C4 = 52!/((52-4)! 4!) = 270 725.
En kan også tenke at det er 52 muligheter på første trekk, 51 på andre, 50 på tredje og 49 på fjerde, og at antall mulige måter å plassere fire ting etter hverandre på er 4·3·2·1 = 24.Da blir det også 52·51·50·49 / 24 = 270 725.

En god besvarelse må forklare hvorfor den valgte fremgangsmåten er korrekt.

ii. rekkefølgen han trekker kortene i betyr noe?

Dette er et ordna utvalg uten tilbakelegging, noe som gir at antall mulige kombinasjoner er 52P4 = 52!/(52-4)! = 6 497 400. Her kan en også tenke at det er 52 valg på første trekk, 51 på andre, 50 på tredje og 49 på fjerde.

En god besvarelse må forklare hvorfor den valgte fremgangsmåten er korrekt.

b) Andrine ønsker å legge alle kortene i kortstokken i ei lang rekke. Hvor mange ulike rekker kan hun lage?

Her kan to svar godtas så lenge det er forklart hvordan en har tenkt. En god besvarelse har med begge svarene med forklaring.

1) 52!
   Dersom det er viktig hvilken side en begynner fra når en ser på rekkefølgen på kortene.
2) 52!/2
   Dersom det kommer på det samme hvilken side av rekka en ser kortene fra blir det halvparten så mange kombinasjoner.

   Det er ikke hensiktsmessig å regne ut noen av de involverte tallene. Svaret bør skrives med fakultetnotasjon.

c) Dersom Andrine har 10 kort fra en kortstokk: Hvor mange ulike kombinasjoner av sorter kan hun ha?

Dette kan en se på som et uordna utvalg med tilbakelegging (selv om det egentlig er uten tilbakelegging), der en har fire utfall for hvert trekk (spar, kløver, hjerter, ruter) siden en har flere enn ti av hver sort. En får da

$$\binom{10+4-1}{10} = \binom{13}{10} = \frac{13!}{(13-10)!\,10!} = 286$$

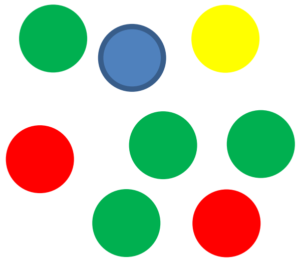
Systematisk telling godtas også som svar. Det må da komme tydelig fram hvordan en er sikker på at en har fått med alle kombinasjonene uten å telle noen flere ganger. I den (kanskje) mest oversiktlige tellemåten kan en utnytte at trekanttallene (1-3-6-10-15-21-28-36-45-55-66) dukker opp. Svaret er summen av de ti første trekanttallene. En ser da på tilfellene med f.eks. 10 spar, 9 spar, 8 spar, osv. ned til 0 spar, og finner antall mulige kombinasjoner av de resterende sortene.

Oppgave 5 (12%)

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 introdusere sannsynlighet gjennom data, og en av måtene å gjøre dette på er å jobbe med statistikk før en jobber med sannsynlighet. Forklar bakgrunnen for denne anbefalingen, og beskriv et konkret undervisningsopplegg basert på en hensiktsmessig kontekst som følger denne anbefalingen.

Blant de fem generelle anbefalinger fra Shaughnessy (2007) til hvordan en bør gjennomføre undervisning i sannsynlighet, er en at ***en skal introdusere sannsynlighet gjennom data***. Tidligere forskning anbefalte at sannsynlighet burde bli introdusert ut fra eksperimentering, eksempelvis med kast med en mynt eller terning, hvor resultatet er tilfeldig. Ut fra slike eksperiment generes data som en kan bruke til å beregne sannsynligheter. I Shaughnessy (2007) er anbefalingen at læreren begynner fra et statistikk-datamateriale og lager spørsmål til dette for å finne sannsynligheter.

For eksempel:
Læreren kan presentere de følgende brikkene:



Da kan elevene se/telle hvor mange grønne, blå og gule brikker det er.
Etter de kan stille spørsmål som: Hvis jeg har trukket en rød brikke, hva er så sannsynligheten for at jeg trekker enda en rød brikke. Dette kan være en innledning til å jobbe med sannsynlighet.

Et annet eksempel er at en bruker fritidsaktivitetene til barna i klassen (eller andre ting elevene har kjennskap til). Da kan en f.eks. få en situasjon der 10 barn driver med fotball, 7 håndball, 3 karate, 6 musikk og 2 innebandy i en klasse på 18 (noen barn har flere enn 1 aktivitet, noen har ingen, osv.). Da kan en (etter å ha samlet og sortert dataene) f.eks. se på sannsynligheten for at en tilfeldig valgt elev har samme interesse som en bestemt elev osv. Da jobber en med sannsynlighet ut fra noe som både har grunnlag i statistikk og noe som elevene har en tilknytning til/kan sette seg inn i.

En god besvarelse beskriver et hensiktsmessig undervisningsopplegg og forklarer hvorfor det er en hensiktsmessig kontekst/et hensiktsmessig opplegg som er beskrevet.

Oppgave 6 (12%)

En liten skog inneholder 64 elger, hvorav 14 som er merket. En ivrig elgjeger feller til sammen ti elger i jakta. La Y være antall merkede elger i jegerens fangst.

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

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

Vi skal anta at Y er binomisk fordelt (selv om dette egentlig ikke stemmer helt overens med konteksten fordi vi har to grupper elger (merkede og umerkede) som trekkes ut tilfeldig UTEN tilbakelegging). Vi må her anta at jakta skjer med tilbakelegging for å få lik sannsynlighet i alle forsøk, f.eks. ved at det kun brukes bedøvelse for å sjekke helsetilstanden til elgene. I en god besvarelse bør det kommenteres at det enten 1) hadde vært mer naturlig å anta at Y er hypergeometrisk fordelt, eller 2) hvilke antakelser en må gjøre for at Y skal være binomisk fordelt.

Fra N=64 elger, er vi interessert i antall merkede elger, Y, i fangsten der sannsynligheten for merkede elger i jegerens fangst er $p = \frac{S}{P} = \frac{14}{64}$, for n= 10 forsøk.

Sannsynlighet for Y=3 er

$$P(Y = 3) = \binom{n}{3} \cdot \left(\frac{S}{P}\right)^3 \cdot \left(1 - \frac{S}{P}\right)^{n-3} = \binom{10}{3} \cdot \left(\frac{14}{64}\right)^3 \cdot \left(1 - \frac{14}{64}\right)^{10-3}$$

$$= \frac{10!}{3!7!} \cdot \left(\frac{14}{64}\right)^3 \cdot \left(\frac{50}{64}\right)^7 \approx 0{,}223$$

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

$$P(Y = 2) = \binom{10}{2} \cdot \left(\frac{14}{64}\right)^2 \cdot \left(1 - \frac{14}{64}\right)^{10-2} = \frac{10!}{2!8!} \cdot \left(\frac{14}{64}\right)^2 \cdot \left(\frac{50}{64}\right)^8 \approx 0{,}299$$

$$P(Y = 1) = \binom{10}{1} \cdot \left(\frac{14}{64}\right)^1 \cdot \left(1 - \frac{14}{64}\right)^{10-1} = \frac{10!}{1!9!} \cdot \left(\frac{14}{64}\right)^1 \cdot \left(\frac{50}{64}\right)^9 \approx 0{,}237$$

$$P(Y = 0) = \binom{10}{0} \cdot \left(\frac{14}{64}\right)^0 \cdot \left(1 - \frac{14}{64}\right)^{10-0} = \frac{10!}{0!9!} \cdot \left(\frac{14}{64}\right)^0 \cdot \left(\frac{50}{64}\right)^{10} \approx 0{,}085$$

Dermed er $P(Y \leq 3) \approx 0{,}223 + 0{,}299 + 0{,}237 + 0{,}085 = 0{,}844$

I en annen, kjempestor skog er det 6400 elger, hvorav 1400 er merket. Alle jegerne i denne skogen feller til sammen 1000 elger. La X være antall merkede elger blant de felte.

b) Finn $P(X \leq 200)$ og $P(160 \leq X \leq 185)$.

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 (12%)

Gjennomsnittlig høyde for alle norske vernepliktige menn er 180 cm. Unge menn fra et fylke er tilsynelatende lavere. Et år var det 218 vernepliktige fra dette fylket, med gjennomsnittshøyde 177,5 cm. Anta at kroppshøyden til en tilfeldig ung mann fra dette fylket er normalfordelt med forventning μ og standardavvik $\sigma = 16$ cm. Anta også at det er tilfeldig hvem som utfører verneplikten. Er det grunnlag for å hevde at unge menn fra dette fylket er lavere enn landsgjennomsnittet? Utfør hypotesetest med signifikansnivå 0,01.

Gjennomfører en hypotesetest av gjennomsnitt

H_0: Menn fra dette fylket er normalt høye ($\mu \geq 180$), og

H_1: Menn fra dette fylket er lavere ($\mu \leq 180$),

Med $\mu_0 = 180,\ n = 218, \bar{X} = 177.5$, $\sigma = 16$ og $\alpha = 0{,}01 = 1\ \%$.

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

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{177{,}5 - 180}{16/\sqrt{218}} = -2{,}31$$

Da beholder vi H_0 siden $\alpha = 0{,}01$ er mindre enn $0{,}0107$ (dette finner en i vedlegg 3/4). Ifølge hypotesetesten er altså menn fra dette fylket normalt høye.