ECON2130: EKSAMEN 2017v

SENSORVEILEDNING.
Det anbefales at de 9 deloppgavene merket med A, B,… teller likt – uansett variasjon i vanskelighetsgrad. Svarene er gitt i << … >> Grensen til bestått bør ligge på ca 30-33%.

Oppgave 1

En vanlig kortstokk består av i alt 52 kort hvorav 12 er «billedkort». Billedkortene viser bilder av konge, dronning eller knekt, hver i fire «kortfarger», hjerter (♥) ruter (♦), kløver (♣) og spar (♠). De øvrige 40 kortene viser tallene 1 («Ess», som ikke betraktes som billedkort her), 2,…,10, som også er representert i de fire kortfargene.

Det trekkes 3 kort rent tilfeldig fra kortstokken. Kortene trekkes ett og ett uten tilbakelegging slik at i hver trekning har de gjenværende kortene samme sannsynlighet for å bli trukket. Det kan vises (som du ikke trenger å gjøre) at denne måten å trekke på leder til et rent tilfeldig utvalg slik at alle ikke-ordnete utvalg på tre kort har samme sjanse for å bli valgt.

La B_j betegne begivenheten at det trekkes et billedkort i trekning j (for j=1,2,3), og $\bar{B}_j$ komplementet til B_j - dvs. at et annet kort enn billedkort blir trukket.

A. **i)** Finn sannsynlighetene $P(B_1)$, $P(B_1 \cap B_2)$ og $P(B_1 \cap B_2 \cap \bar{B}_3)$.
[Hint: Bruk f.eks. multiplikasjonssetningen på de to siste.]

ii) Tegn et Venn-diagram som viser at B_2 kan skrives som en disjunkt union

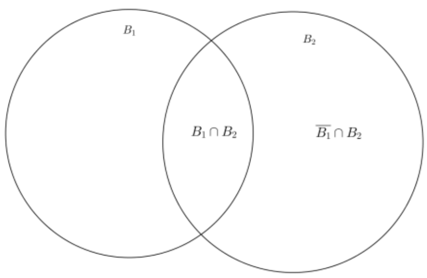
$$B_2 = (B_1 \cap B_2) \cup (\bar{B}_1 \cap B_2)$$

iii) Vis at $P(B_2) = \dfrac{3}{13}$. **[Hint:** Bruk **ii)**]

<< **Svar:** **i):** $P(B_1) = \dfrac{12}{52} = \dfrac{3}{13} = 0,231$, $P(B_1 \cap B_2) = P(B_1)P(B_2 \mid B_1) = \dfrac{12}{52} \cdot \dfrac{11}{51} = 0,050$,

$$P(B_1 \cap B_2 \cap \bar{B}_3) = P(B_1)P(B_2 \mid B_1)P(\bar{B}_3 \mid B_1 \cap B_2) = \frac{12}{52} \cdot \frac{11}{51} \cdot \frac{40}{50} = 0,040.$$

ii):



iii): $B_2 = (B_1 \cap B_2) \cup (\bar{B}_1 \cap B_2)$ $B_2 = (B_1 \cap B_2) \cup (\bar{B}_1 \cap B_2)$ disjunkt $\Rightarrow$

$$P(B_2) = P(B_1 \cap B_2) + P(\bar{B}_1 \cap B_2) = P(B_1)P(B_2 \mid B_1) + P(\bar{B}_1)P(B_2 \mid \bar{B}_1) =$$

$$= \frac{12}{52} \cdot \frac{11}{51} + \frac{40}{52} \cdot \frac{12}{51} = \frac{12}{52}\left(\frac{11}{51} + \frac{40}{51}\right) = \frac{12}{52} = \frac{3}{13}$$

>>

B. La X være antall billedkort som blir trukket ut i et rent tilfeldig utvalg på tre kort fra kortstokken.

i) Forklar kort hvorfor X er hypergeometrisk fordelt og sett opp en formel for $P(X = x), \quad x = 0,1,2,3$.

ii) Vis at $P(X = 0) = 0,447$ (med tre desimalers nøyaktighet). Finn også sannsynligheten for at det er minst ett billedkort i utvalget.

iii) Vis, ved bruk av relevante formler for den hypergeometriske fordelingen, at forventningen til X er 0,692, og at standardavviket ($SD(X) = \sqrt{\text{var}(X)}$) til X er 0,715 (begge med tre desimalers nøyaktighet).

<< **Svar:** **i):** Det trekkes et utvalg på $n = 3$ kort fra en endelig populasjon av størrelse, $N = 52$ kort, hvorav $M = 12$ har kjennetegnet «billedkort». Hvis X er antall billedkort i utvalget, er det nok å vise til at utvalget er rent tilfeldig for å garantere at X er hypergeometrisk fordelt.

Punktsannsynlighetene blir

$$P(X = x) = \frac{\binom{12}{x}\binom{40}{3-x}}{\binom{52}{3}}, \quad x = 0,1,2,3$$

ii)

$$P(X = 0) = \frac{\binom{12}{0}\binom{40}{3}}{\binom{52}{3}} = \frac{\binom{40}{3}}{\binom{52}{3}} = \frac{40 \cdot 39 \cdot 38 / 3!}{52 \cdot 51 \cdot 50 / 3!} = \frac{40 \cdot 39 \cdot 38}{52 \cdot 51 \cdot 50} = \frac{59280}{132600} = 0,44706...$$

Alternativ beregning er

$$P(X = 0) = P(\bar{B}_1 \cap \bar{B}_2 \cap \bar{B}_3) = P(\bar{B}_1)P(\bar{B}_2 \mid \bar{B}_1)P(\bar{B}_3 \mid \bar{B}_1 \cap \bar{B}_2) = \frac{40}{52} \cdot \frac{39}{51} \cdot \frac{38}{50} = 0,44706...$$

$P(\text{"minst ett billedkort"}) = P(X > 0) = 1 - P(X = 0) = 0,553$ med tre desimaler.

iii):

$$E(X) = n\frac{M}{N} = 3\cdot(0,2307...) = 0,692...$$

$$SD(X) = \sqrt{\text{var}(X)} = \sqrt{n\frac{M}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1}} = \sqrt{3\frac{12}{52}\left(1-\frac{12}{52}\right)\frac{49}{51}} = 0,7153..$$

>>

C. Følgende pengespill blir tilbudt: Spilleren betaler en spilleavgift på 75 kr. for hvert spill. Spillet består av at tre kort blir trukket rent tilfeldig fra en kortstokk som beskrevet i innledningen. For hvert billedkort som blir trukket ut får spilleren utbetalt 100 kr. Andre kort i utvalget som ikke er billedkort gir ingen utbetaling.

La U betegne spillerens gevinst etter et spill. For eksempel, hvis ingen billedkort blir trukket ut, taper spilleren spilleavgiften på 75 kr., som gir $U = -75$. Hvis alle tre kortene i utvalget blir billedkort, blir gevinsten 300 kr. – 75 kr. = 225 kr., som gir $U = 225$. Generelt kan vi uttrykke gevinsten som $U = 100X - 75$, der X er antall billedkort i utvalget som definert i punkt **B**.

i) Vis at sannsynligheten for at spilleren får positiv gevinst i et spill ($P(U > 0)$) er lik 0,553 (med tre desimalers nøyaktighet).

ii) Beregn forventningen og standardavviket for gevinsten (U) til spilleren etter et spill.

iii) Hvis spilleren taper 75 kr i et spill, tjener tilbyderen 75 kr. Hvis, på den annen side, spilleren får en gevinst på f.eks. 225 kr, taper tilbyderen samme beløp. La V betegne tilbyderens fortjeneste etter et spill. Vi kan dermed uttrykke tilbyderens fortjeneste etter et spill som $V = -U = 75 - 100X$.

Vis at tilbyderens fortjeneste i et spill (V) har forventning 5,8 og standardavvik 71,5 (begge med en desimals nøyaktighet).

<< **Svar: i):** Vi har $U > 0 \iff 100X - 75 > 0 \iff X > \frac{75}{100} = 0,75 \iff X \geq 1$ siden X bare kan ta hele verdier. Dermed

$$P(U > 0) = P(X \geq 1) = 1 - P(X = 0) \overset{\text{punkt B}}{=} 1 - 0,447 = 0,553$$

ii): $E(U) = E(100X - 75) \overset{\text{regel 4.12}}{=} 100E(X) - 75 \overset{\text{punkt A(iii)}}{=} 69,23... - 75 = -5,77... \approx -5,8$

$$SD(U) = \sqrt{\text{var}(100X - 75)} \overset{\text{regel 4.15}}{=} \sqrt{100^2\,\text{var}(X)} = 100\cdot SD(X) \overset{\text{punkt B(iii)}}{=} 71,53... \approx 71,5$$

iii): Siden $V = -U$, får vi $E(V) = -E(U) = 5{,}8$ og $SD(V) = \sqrt{\mathrm{var}(-U)} = \sqrt{(-1)^2 \mathrm{var}(U)} = SD(U) = 71{,}5$ (begge med en desimals nøyaktighet).

>>

D. La nå $V_{tot} = \sum_{i=1}^{k} V_i$ være total fortjeneste for tilbyderen etter *k* spill, der V_i er fortjenesten i spill nr. *i*. Vi antar $V_1, V_2, \ldots, V_k$ er uavhengige og identisk fordelte som *V* i punkt **C(iii).**

i) Hvis $k \geq 20$, kan vi regne at V_{tot} er tilnærmet normalfordelt. Hvilken normalfordeling er aktuell som tilnærmelse? [**Hint:** Bestem *a* og *b* (uttrykt ved *k*) i uttrykket $V_{tot} \overset{\text{tilnærmet}}{\sim} N(a,b)$].

ii) Tilbyderen er interessert i sannsynligheten for at den totale fortjenesten blir negativ ($P(V_{tot} < 0)$) for forskjellige verdier av *k*. Bruk tilnærmelsen til normalfordelingen til å vise at $P(V_{tot} < 0) \approx G\left(-\sqrt{k} \cdot \frac{5{,}8}{71{,}5}\right)$, der $G(z) = P(Z \leq z)$ er den kumulative fordelingsfunksjonen for *Z* som er standard normalfordelt ($Z \sim N(0,1)$). Forklar også hvorfor sannsynligheten $G\left(-\sqrt{k} \cdot \frac{5{,}8}{71{,}5}\right)$, som gjelder for en bestemt *k*, avtar når *k* øker.

iii) Beregn $P(V_{tot} < 0)$ tilnærmet når $k = 50$, $k = 500$ og $k = 1000$.

<< **Svar:** **i):** Siden $E(V_i) = E(V) = 5{,}8$ og $SD(V_i) = SD(V) = 71{,}5$, er det nok å vise til sentralgrenseteoremet formulert i regel 5.19 som impliserer at

$$V_{tot} \overset{\text{tilnærmet}}{\sim} N(k \cdot E(V), \sqrt{k} \cdot SD(V)) = N(k \cdot 5{,}8, \sqrt{k} \cdot 71{,}5).$$

ii): Normaltilnærmelsen impliserer at $\frac{V_{tot} - k \cdot 5{,}8}{\sqrt{k} \cdot 71{,}5} \overset{\text{tilnærmet}}{\sim} N(0,1)$. La $Z \sim N(0,1)$.

Dermed

$$P(V_{tot} < 0) = P\left(\frac{V_{tot} - k \cdot 5{,}8}{\sqrt{k} \cdot 71{,}5} < \frac{-k \cdot 5{,}8}{\sqrt{k} \cdot 71{,}5}\right) \approx P\left(Z < \frac{-k \cdot 5{,}8}{\sqrt{k} \cdot 71{,}5}\right) \overset{\text{kontinuerlig fordeling}}{=} P\left(Z \leq \frac{-k \cdot 5{,}8}{\sqrt{k} \cdot 71{,}5}\right) =$$

$$= G\left(\frac{-k \cdot 5{,}8}{\sqrt{k} \cdot 71{,}5}\right) = G\left(-\sqrt{k} \cdot \frac{5{,}8}{71{,}5}\right)$$

Hvis *k* øker, øker også $\sqrt{k}$ slik at $-\sqrt{k} \cdot \frac{5{,}8}{71{,}5}$ avtar. Siden $G(z)$ er strengt voksende i *z*, må dermed $G\left(-\sqrt{k} \cdot \frac{5{,}8}{71{,}5}\right)$ avta.

iii): Skriv kort $z_k = -\sqrt{k} \cdot \frac{5,8}{71,5}$. Siden $P(V_{tot} < 0) \approx G\left(-\sqrt{k} \cdot \frac{5,8}{71,5}\right) = G(z_k)$, får vi av tabell E3 i Løvås:

k	z_k	$G(z_k) \approx P(V_{tot} < 0)$
50	− 0,57	0,2843
500	− 1,81	0,0351
1000	− 2,57	0,0051

>>

E. Tilbyderen ønsker å bestemme en (tilnærmet) 95% sikker nedre grense for den totale fortjenesten, V_{tot}, når $k = 1000$ spill. Med andre ord, tilbyderen er ute etter en konstant, c, slik at $P(V_{tot} > c) \approx 0,95$ når $k = 1000$.

Bestem c ut fra normaltilnærmelsen for V_{tot}.

<< **Svar:**

Vi har $P(V_{tot} > c) = 1 - P(V_{tot} \le c) = 1 - P\left(\frac{V_{tot} - k \cdot 5,8}{\sqrt{k} \cdot 71,5} \le \frac{c - k \cdot 5,8}{\sqrt{k} \cdot 71,5}\right) \approx 1 - G\left(\frac{c - k \cdot 5,8}{\sqrt{k} \cdot 71,5}\right)$.

Bestemmer vi c slik at det siste uttrykket blir lik 0,95, blir $P(V_{tot} > c) \approx 0,95$. Altså

$$1 - G\left(\frac{c - k \cdot 5,8}{\sqrt{k} \cdot 71,5}\right) = 0,95 \quad \Leftrightarrow \quad G\left(\frac{c - k \cdot 5,8}{\sqrt{k} \cdot 71,5}\right) = 0,05 \quad \Leftrightarrow \quad \frac{c - k \cdot 5,8}{\sqrt{k} \cdot 71,5} = -1,645 \quad \Leftrightarrow$$

$$\Leftrightarrow \quad c = k \cdot 5,8 - (1,645)\sqrt{k} \cdot (71,5)$$

eller for $k = 1000$: $\quad c = 1000 \cdot 5,8 - (1,645)\sqrt{1000} \cdot (71,5) = 5800 - 3719,4 = 2080,6$.

Tilbyderen kan således være tilnærmet 95% sikker på å få en fortjeneste på over 2080 kr ved $k = 1000$ spill.
(I parentes bemerket er forventet fortjeneste 5800 kr.)
>>

Oppgave 2

To meningsmålingsinstitutter, her kalt MM1 og MM2, foretar på samme tidspunkt, og uavhengig av hverandre, en måling av støtten i befolkningen for et parti vi kaller B. Målingene er basert på to rent tilfeldige utvalg av stemmeberettigete. Den relative andelen av populasjonen av stemmeberettigete som ville stemme på B på det aktuelle tidspunktet kaller vi p, som er ukjent, men som estimeres både av MM1 og MM2 basert på sine utvalg.

Andelen målt i prosent kaller vi μ der $\mu = 100p$. Det er vanlig å publisere resultater fra meningsmålinger som prosentandeler (dvs. som estimater på μ) istedenfor relative andeler (estimater på *p*). I denne oppgaven vil vi derfor se på μ som parameteren av primær interesse.

Utvalgsstørrelsen til MM1 er $n_1 = 950$ og til MM2, $n_2 = 1055$. Begge utvalgene antas for enkelthets skyld i denne oppgaven som «rent tilfeldige», slik at alle mulige ikke-ordnete utvalg av stemmeberettigete har samme sannsynlighet for å bli valgt. Antallet som ville stemme på B i utvalg *j* (trukket av MM*j*) betegnes med X_j, for $j = 1,2$.

Som modellantakelse antar vi at X_1 og X_2 er uavhengige og binomisk fordelte stokastiske variable med parametere (n_1, p) og (n_2, p) henholdsvis (skrevet kort: $X_j \sim \text{bin}(n_j, p)$, $j = 1,2$), med samme *p* for $j = 1,2$. Prosentandelen i utvalg *j* som ville stemme på B betegnes med $Y_j = 100\frac{X_j}{n_j}$, $j = 1,2$. Y_1 og Y_2 vil da også være uavhengige stokastiske variable siden X_1, X_2 er det.

Resultatet av målingene er gitt i tabell 1:

Tabell 1

	MM1	MM2
Utvalgsstørrelse n_j	950	1055
Antall B-stemmer X_j	217	258
Relativ andel av B $\hat{p}_j$	0.228	0.245
Prosentandel Y_j %	22.8	24.5

A. **i)** Drøft kort om antakelsen at X_j er binomisk fordelt er rimelig hvis utvalget (*j*) er rent tilfeldig.

ii) Vis at $\hat{\mu} = \bar{Y} = \frac{Y_1 + Y_2}{2}$ er forventningsrett som estimator for μ, og beregn estimatet for μ basert på $\hat{\mu}$. [**Hint.** Gjør bruk av sammenhengen, $Y_j = 100\frac{X_j}{n_j}$, $j = 1,2$.]

iii) Vis at standardfeilen til $\hat{\mu}$ (dvs. standardavviket her) er

$$SE(\hat{\mu}) = 50 \cdot \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \cdot p(1-p)}$$

[**Hint:** Vis først at $\text{var}(Y_j) = 100^2 \frac{p(1-p)}{n_j}$ for $j = 1,2$]

<< **Svar:** **i):** Hvis utvalget er rent tilfeldig, er X_j hypergeometrisk fordelt som er tilnærmet binomisk (med $p = M/N =$ populasjonsandelen) hvis populasjonen er stor – som den åpenbart er i dette tilfellet.

ii): Siden $E(Y_j) = E\left(100\frac{X_j}{n_j}\right) = 100\frac{1}{n_j}E(X_j) = 100p = \mu$, blir

$E(\hat{\mu}) = \frac{E(Y_1)+E(Y_2)}{2} = \frac{2\mu}{2} = \mu$ **-** dvs. $\hat{\mu}$ forventningsrett.

Estimat: $\hat{\mu}_{obs} = \frac{22{,}8+24{,}5}{2} = 23{,}65$ %.

iii): Ved regneregler for varians får vi:

$$\text{var}(Y_j) = \text{var}\left(100\frac{X_j}{n_j}\right) = \frac{100^2}{n_j^2}\text{var}(X_j) = \frac{100^2}{n_j^2}n_j p(1-p) = \frac{100^2}{n_j}p(1-p)$$

hvorav, siden Y_1 og Y_2 er uavhengige,

$$\text{var}(\hat{\mu}) = \text{var}\left(\frac{Y_1+Y_2}{2}\right) = \frac{1}{4}\left(\text{var}(Y_1)+\text{var}(Y_2)\right) =$$

$$= \frac{1}{4}\left(\frac{100^2}{n_1}p(1-p) + \frac{100^2}{n_2}p(1-p)\right) = \frac{100^2}{4}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)p(1-p)$$

som gir

$$SE(\hat{\mu}) = \sqrt{\text{var}(\hat{\mu})} = \frac{100}{2}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)p(1-p)}$$

>>

B. **i)** Forklar, ved bruk av relevante regler, at $\hat{\mu}$ er tilnærmet normalfordelt, $\hat{\mu} \overset{\text{tilnærmet}}{\sim} N\left(\mu, SE(\hat{\mu})\right)$. [**Hint.** Det kan være lurt å skrive $\hat{\mu}$ uttrykt ved X_1 og X_2]

ii) **Innledning.** MM1 og MM2 oppgir begge $\pm 2{,}6$ prosentpoeng som (95%) feilgrenser for sine anslag på μ, der begrepet «feilgrense» refererer til et tilnærmet 95% konfidensintervall for μ. Når tallene fra MM1 (eller fra MM2) settes inn i konfidensintervallet, får intervallet formen « (anslag på μ) $\pm$ 2,6 » (som du ikke trenger å verifisere her), der $\pm 2{,}6$ ofte kalles feilgrenser. Det er grunn til å tro at feilgrensene blir mindre hvis vi kombinerer informasjonen fra de to målingene.

Spørsmål. Sett opp og beregn et tilnærmet 95% konfidensintervall for μ basert på punkt **B(i)**, og angi nye feilgrenser for estimatet $\hat{\mu}_{obs}$ basert på de kombinerte dataene fra MM1 og MM2.

[Hint: Du trenger et anslag på $SE(\hat{\mu})$ som avhenger av den ukjente p og er derfor ukjent. Som antydet under regel 6.8 og 6.12 l Løvås, kan den ukjente p erstattes med et estimat, $\hat{p}_{obs}$, (her f.eks. $\hat{\mu}_{obs}/100$, der indeksen *obs* indikerer at det er den observerte verdien som brukes) uten at konfidensgraden endres vesentlig (noe som kan vises i videregående teori)]

<< **Svar:**

i): Uttrykt ved X_1 og X_2, blir $\hat{\mu} = \dfrac{Y_1 + Y_2}{2} = \dfrac{100\dfrac{X_1}{n_1} + 100\dfrac{X_2}{n_2}}{2} = \dfrac{50}{n_1}X_1 + \dfrac{50}{n_2}X_2$.

X_1 og X_2 er uavhengige og tilnærmet normalfordelte ifølge regel 5.20. Da må en lineær-kombinasjon som $\hat{\mu}$ ifølge regel 5.17 også være tilnærmet normalfordelt. Løvås er litt vag i regel 5.17 siden han ikke sier hvilken normalfordeling som er aktuell, men det framgår av eksempelet etter regel 5.17 at det er forventning og standardavvik til $\hat{\mu}$ som skal brukes som parametere. På forelesningene har vi vært noe mer presis om dette, men slett ikke alle har fulgt forelesningene…
Alternativt vil noen bruke regel 5.17 direkte på Y_1 og Y_2 som de vil argumentere er uavhengige og tilnærmet normalfordelte siden Y_j er en konstant ganger X_j. En slik løsning er selvsagt akseptabel også.

ii): På grunn av punkt **i)** følger av regel 6.8 at et tilnærmet 95% konfidensintervall for μ er gitt ved $\hat{\mu} \pm (1{,}96)SE(\hat{\mu})$, der $1{,}96 = z_{0{,}025}$ er 0,025-kvantilen i N(0, 1). Siden $SE(\hat{\mu})$ avhenger av den ukjente p, foreslår Løvås å erstatte p med et estimat, $\hat{p}_{obs}$, og antyder således implisitt at dette ikke ødelegger konfidensgraden vesentlig. Dette poenget er presisert noe mer i et supplerende forelesningsnotat om konfidensintervall på kurssiden på nettet. Det er flere estimater for p man kan bruke (f.eks. et av de to estimatene produsert av MM1 og MM2 som bør aksepteres om kandidaten foreslår det), men det mest naturlige er vel å benytte det intuitivt beste anslaget,

$$\hat{p}_{obs} = \hat{\mu}_{obs}/100 = 23{,}65/100 = 0{,}2365$$

Med dette anslaget får vi estimert standardfeil:

$$SE(\hat{\mu}) \approx 50 \cdot \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \cdot \hat{p}_{obs}(1 - \hat{p}_{obs})} = 50 \cdot \sqrt{\left(\frac{1}{950} + \frac{1}{1055}\right) \cdot 0{,}2365 \cdot (1 - 0{,}2365)} = 0{,}9503$$

som gir tilnærmet 95% feilgrenser

$$\pm (1{,}96)SE(\hat{\mu}) \approx \pm 1.8626 \approx \pm 1{,}9$$

Det tilsvarende konfidensintervallet for μ blir $23,7 \pm 1,9 = [21,8,\ 25,6]$

>>

C. **Innledning.** Valgbarometre publiseres med jamne mellomrom der resultater fra flere meningsmålingsinstitutter samles. Tabell 2 viser[1] prosentandeler for partiet Høyre målt av 9 meningsmålingsinstitutter i januar 2017.

Tabell 2 Prosentandeler for Høyre målt i januar 2017

i	1	2	3	4	5	6	7	8	9
Institutt	Ipsos/MMI	Opinion	Gallup	Sentio	Norstat	Respons	Infact	Norstat	Norfakta
%-andel Høyre (Y_i)	22,9	23,3	24,5	22,8	24,5	23,3	20,9	22,6	25,3

Deskriptive størrelser for tallene i tabell 2 er gitt i tabell 3 produsert av Excel:

Tabell 3 Excel deskriptive størrelser for dataene i tabell 2

Y	
Mean	23.34
Standard Error	0.4334
Median	23.3
Mode	23.3
Standard Deviation	1.3001
Sample Variance	1.6903
Kurtosis	0.5
Skewness	-0.3
Range	4.4
Minimum	20.9
Maximum	25.3
Sum	210.1
Count	9

Tallene i tabell 2 representerer 9 anslag på prosentandelen, μ, for Høyre i populasjonen av stemmeberettigete i Norge i januar 2017. Tallene kan betraktes som observasjoner av uavhengige stokastiske variable, $Y_1, Y_2, \ldots, Y_n$, der $n = 9$. Utvalgsstørrelsene brukt av de forskjellige meningsmålingsinstituttene er ikke oppgitt i kilden, men ligger alle i

---

[1] **Kilde:** Bernt Aardals hjemmeside: http://www.aardal.info/

nærheten av 1000. Dette betyr at variansene til Y_i-ene, som egentlig ikke er helt like, på den annen side ikke varierer spesielt mye. For enkelthets skyld antar vi derfor i modellen at variansene er like uten vesentlig tap av realisme. Likeledes antar vi for enkelthets skyld i modellen at Y_i-ene er normalfordelte selv om de egentlig bare er tilnærmet normalfordelte.

Oppsummert: som **modell** antar vi at $Y_1, Y_2, \ldots, Y_9$ er uavhengige og identisk normalfordelte med ukjent forventning μ og (konstant og ukjent) varians, σ^2 (skrevet kort: $Y_i \sim N(\mu, \sigma),\ i = 1, 2, \ldots, 9$).

Spørsmål:

i) Estimer prosentandelen av stemmeberettigete i Norge som ville stemme på Høyre januar 2017 ved bruk av en forventningsrett estimator som utnytter informasjonen fra alle de 9 meningsmålingsinstituttene.

ii) I oktober 2016 lå Høyres prosentandel på ca. 24,5%. Sett opp en test med signifikansnivå 5% for nullhypotesen $H_0 : \mu \geq 24{,}5$ mot $H_1 : \mu < 24{,}5$ Bruk tallene i tabell 3 til å gjennomføre testen og formuler en konklusjon.

iii) Basert på tabellene i Løvås, er (a) p-verdien for testen i punkt **ii)** mindre enn 1%, (b) ligger den mellom 1% og 2,5%, (c) mellom 2,5% og 5% eller (d) er den større enn 5%? Begrunn svaret ditt.

<< **Svar:** **i):** Den foreslåtte modellen for $Y_1, Y_2, \ldots, Y_9$ er *uid* modellen som er relativt grundig behandlet i pensum, og der $\hat{\mu} = \bar{Y} = \frac{1}{9}\sum_{i=1}^{9} Y_i$ er den foreslåtte forventningsrette estimatoren. Tabell 3 gir estimatet $\hat{\mu}_{obs} = \text{"mean"} = 23{,}34$.

ii): Uid-modellen med 9 normalfordelte variable leder til en T-test for H_0:

Testobservator: $T = \frac{\bar{Y} - 24{,}5}{S}\sqrt{9}$, der $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$.

T er t(n-1) =t(8) fordelt hvis $\mu = 24{,}5$, og en 5% test for H_0 er gitt ved

Forkast H_0 hvis $T \leq -t_{8,\ 0{,}05} (= -1{,}86)$ (der $t_{8,\ 0{,}05}$ er 0,05-kvantilen i t(8)-fordelingen)

Observert gir tabell 3, $S_{obs} = \text{"standard deviation"} = 1{,}3001$ og

$$T_{obs} = \frac{23{,}34 - 24{,}5}{1{,}3001}\sqrt{9} = -2.68 \text{ mot kritisk verdi -1.86}$$

Konklusjon: Forkast H_0. Det er altså sterk evidens (nivå 5%) i data for at prosentandelen for Høyre i januar 2017 har sunket i forhold til nivået i oktober 2016.

iii): Siden p-verdien er det minste nivået som leder til forkastning, og nivå 5% leder til forkastning, følger at p-verdien er lavere enn 5%. De kritiske verdiene for nivå 2,5% og 1% er henholdsvis -2,306 og -2,896, slik at vi forkaster H_0 på 2,5% men ikke på 1%. Dette betyr at p-verdien ligger mellom 1% og 2,5%.

>>

D. Vi vender nå tilbake til situasjonen i punkt **2A**. En alternativ estimator for μ som kunne vært brukt, er

$$\mu^* = 100\frac{X_1 + X_2}{n_1 + n_2}$$

Det kan vises, som du ikke trenger å gjøre, at μ^* er en optimal estimator for μ i betydningen at den har minst varians blant alle forventningsrette estimatorer. Likevel viser det seg at når utvalgsstørrelsene er nær 1000 som her, så er den foreslåtte estimatoren, $\hat{\mu} = \overline{Y}$, nesten like god som μ^* i den forstand at $\hat{\mu}$ og μ^* har nesten samme standardfeil. Illustrer dette ved å beregne forholdet mellom standardfeilene

$$\frac{SE(\hat{\mu})}{SE(\mu^*)} = \sqrt{\frac{\text{var}(\hat{\mu})}{\text{var}(\mu^*)}}$$

<< **Svar:** Vi finner

$$\text{var}(\mu^*) = \frac{100^2}{(n_1+n_2)^2}\left(\text{var}(X_1) + \text{var}(X_2)\right) = \frac{100^2}{(n_1+n_2)^2}\left(n_1 p(1-p) + n_2 p(1-p)\right) = 100^2\,\frac{p(1-p)}{n_1+n_2}$$

hvorav

$$SE(\mu^*) = \sqrt{\text{var}(\mu^*)} = \frac{100}{\sqrt{n_1+n_2}}\sqrt{p(1-p)}$$

og

$$\frac{SE(\hat{\mu})}{SE(\mu^*)} = \frac{50\cdot\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\cdot\sqrt{p(1-p)}}{\frac{100}{\sqrt{n_1+n_2}}\sqrt{p(1-p)}} = \frac{1}{2}\sqrt{(n_1+n_2)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = 1{,}0014...$$

som betyr at $SE(\hat{\mu}) \approx SE(\mu^*)$

>>
