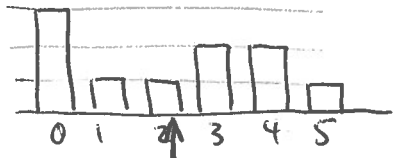
4.79

x	0	1	2	3	4	5	
$P(X=x)$	0.3	0.1	0.1	0.2	0.2	0.1	1

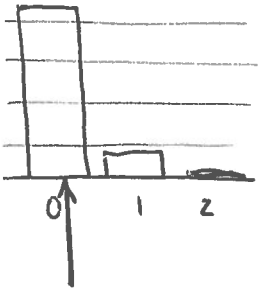


$$E(x) = \sum_{\text{alle } x} x \cdot P(X=x) = 0\cdot 0.3 + 1\cdot 0.1 + 2\cdot 0.1 + 3\cdot 0.2 + 4\cdot 0.2 + 5\cdot 0.1$$
$$= \underline{2.2}$$

tyngdepunktet

4.80 X = # ess når du bør gitt to kort

x	0	1	2	
$P(X=x)$	0.8507	0.1448	0.0045	1
	$\left(\frac{48}{52}\cdot\frac{47}{51}\right.$	$\frac{48}{52}\cdot\frac{4}{51}+\frac{4}{52}\cdot\frac{48}{51}$	$\left.\frac{4}{52}\cdot\frac{3}{51}\right)$	



$$E(x) = \sum_{\text{alle } x} x \cdot P(X=x) = 0\cdot 0.8507 + 1\cdot 0.1448 + 2\cdot 0.0045 = \underline{0.1538}$$

4.82

x	0	1	2	3	4	5	sum	
$P(X=x)$	0.3	0.1	0.1	0.2	0.2	0.1	1	
$x\cdot P(X=x)$	0	0.1	0.2	0.6	0.8	0.5	2.2	$=E(x)$
$x^2\cdot P(X=x)$	0	0.1	0.4	1.8	3.2	2.5	8.0	$=E(X^2)$

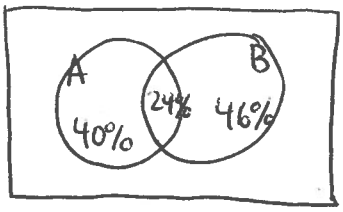
OBS: $E(X) = \sum_{\text{alle } x} x \cdot P(X=x) \longrightarrow \mu_x$

$$\begin{aligned}
Var(X) &= E\left((X-\mu)^2\right) \longrightarrow \sigma_x^2 \\
&= E(X^2 - 2\mu X + \mu^2) \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - (E(X))^2 \longrightarrow \sigma_x = \sqrt{E(X^2) - (E(X))^2} \\
&= \sqrt{8.0 - 2.2^2} = \underline{1.8}
\end{aligned}$$

4.81 $E(X^2) = \sum_{\text{alle } x} x^2 \cdot P(X=x) = 0.1628$

$Var(X) = 0.1628 - (0.1538)^2 = 0.139 \qquad \sigma_x = \sqrt{0.139} = \underline{0.37}$

4.111 + 112



A = nok søvn
B = nok trening

$P(A) = 0.4$
$P(B) = 0.46$
$P(AB) = 0.24$

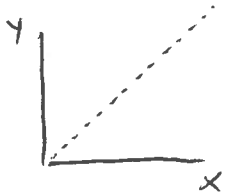
a) $P(AB^c) = P(A) - P(AB)$, fordi $P(A) = P(AB) + P(AB^c)$
b) $P(A^cB) = P(B) - P(AB)$, fordi $P(B) = P(BA) + P(AB^c)$
c) $P(A^cB^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1 - (P(A) + P(B) - P(AB))$

4.88 Hvis X & Y uavh ($\rho = 0$): $Var(X+Y) = Var(X) + Var(Y)$, ellers ikke.

Hvis $\rho = 1$: $Var(X+Y) = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X \cdot \sigma_Y$

$$= \sigma_X^2 + \sigma_Y^2 + 2\sigma_X\sigma_Y$$

$$= (\sigma_X + \sigma_Y)^2$$



$$\sigma_{X+Y} = \sqrt{Var(X+Y)} = \sqrt{(\sigma_X + \sigma_Y)^2} = \underline{\sigma_X + \sigma_Y}$$

4.90 X = høyde i cm $\quad E(X) = 176.8$ cm
$\sigma_X = 7.2$ cm

Y = høyde i tommer ; $Y = \frac{X}{2.54}$

$$E(Y) = E\left(\frac{X}{2.54}\right) = \frac{1}{2.54} \cdot E(X) = \frac{1}{2.54} \cdot 176.8 = \underline{69.6}$$

$$\sigma_Y = \sqrt{Var(Y)}$$

$$Var(Y) = Var\left(\frac{X}{2.54}\right) = \left(\frac{1}{2.54}\right)^2 \cdot Var(X) = \frac{7.2}{2.54^2} = 1.116$$

$$\sigma_Y = \sqrt{1.116} = \underline{1.06}$$

4.115 Bayes regel: Snu de betingede sannsynlighetene

$P(4\text{ årig}) = 0.61$
$P(2\text{ årig}) = 0.39$

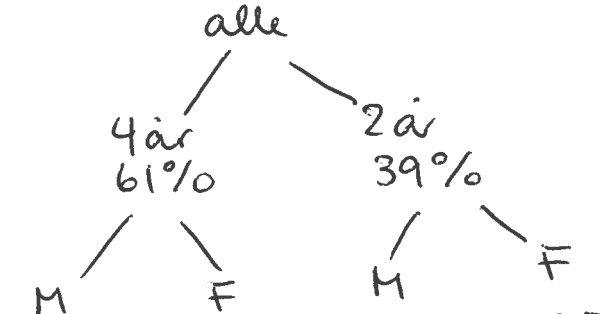
$P(M \mid 4\text{ år}) =$
$P(M \mid 2\text{ år})$

	M	F	
4år	0.44 →	0.56	0.61
2år	0.41 →	0.59	0.39
	0.4283	0.5717	1

$0.44 \cdot 0.61 + 0.41 \cdot 0.39$ $\qquad$ $0.56 \cdot 0.61 + 0.59 \cdot 0.39$

$$P(4\text{ år} \mid F) = \frac{P(F, 4\text{ år})}{P(F)} = \frac{P(F \mid 4\text{ år}) \cdot P(4\text{ år})}{P(F)} = \frac{0.56 \cdot 0.61}{0.5717} \approx \underline{0.60}$$

4.116



	M	F	M	F	Sum
	0.44	0.56	0.41	0.59	
sh :	0.2684	0.3416	0.1599	0.2301	1

4.131

Muskular dystropi : DMD

Recessiv kjønns-lenket mutasjon :

mor bærer, men frisk — $P(\text{sønn DMD} \mid \text{mor bærer}) = 0.5$

— $P(\text{datter bærer} \mid \text{mor bærer}) = 0.5$

$\frac{1}{3}$ av alle DMD-tilfeller : spontan mutasjon, altså

$P(\text{ikke bærer} \mid \text{DMD}) = \frac{1}{3}$

$P(\text{bærer} \mid \text{DMD}) = 1 - \frac{1}{3} = \frac{2}{3}$

Screeningtest : $P(\text{positiv test} \mid \text{bærer}) = 0.7$

$P(\text{positiv test} \mid \text{ikke bærer}) = 0.1$

$$P(\text{bærer} \mid \text{positiv test}) = \frac{P(\text{bærer \& positiv test})}{P(\text{positiv test})}$$

$$= \frac{P(+\text{test} \mid \text{bærer}) \cdot P(\text{bærer})}{P(+\text{test} \mid \text{bærer}) \cdot P(\text{bærer}) + P(+\text{test} \mid \text{ikke bærer}) \cdot P(\text{ikke bærer})}$$

$$= \frac{0.7 \cdot \frac{2}{3}}{0.7 \cdot \frac{2}{3} + 0.1 \cdot \frac{1}{3}} = \underline{0.9333}$$

5.14, 5.20, 21, 23

5.14 X = søvn per natt

$X \sim N(\mu_X, \sigma_X)$

$N(6.78, 1.24)$

SRS, size $n = 150 \Rightarrow X_1, \ldots, X_{150} \longrightarrow \bar{X} = \frac{1}{150}\sum_{i=1}^{150} X_i$

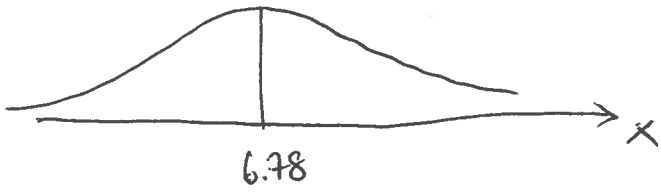
a) $\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{1.24}{\sqrt{150}} = 0.10$

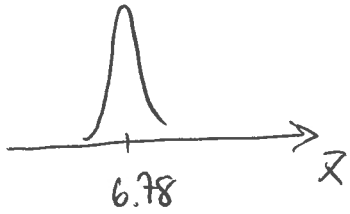
Bruk det du vet fra Ch 5.1 og s. 8 + 9 i de scannede forelesningsnotatene fra uke 39-40, som sier at når

$X_i \sim N$, så er også $\bar{X} \sim N$

$\downarrow$ $\mu_X = 6.78$, $\sigma_X = 1.24$

$\downarrow$ $\mu_{\bar{X}} = 6.78$, $\sigma_{\bar{X}} = 0.10$





$\downarrow$

$\bar{X} \pm 1 \cdot \sigma_{\bar{X}}$: $(6.68, 6.88)$

$\bar{X} \pm 2 \cdot \sigma_{\bar{X}}$: $(6.58, 6.98)$

$\bar{X} \pm 3 \cdot \sigma_{\bar{X}}$: $(6.48, 7.08)$

$$\underbrace{P(\bar{X} < 6.9)}_{N(6.78,\,0.10)} = P\left(\underbrace{Z}_{N(0,1)} < \frac{6.9 - 6.78}{0.1}\right) = 0.88$$

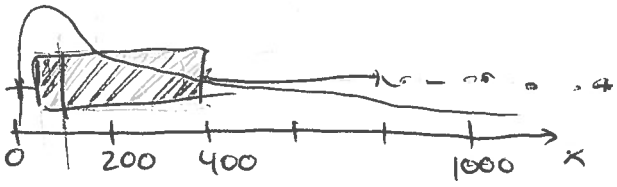
5.20 X = # venner på Facebook

$X \sim ?$

$\downarrow$

høyreskjev, median = 100
$\mu_X = 190$
$\sigma_X = 288$

jeg ser for meg noe sånt:



SRS, size $n = 70$: $X_1, \dots, X_{70}$

Bruk det du vet fra Ch 5.1 og s. 8 + 9 i de scannede forelesningsnotatene fra uke 39-40, som sier at hvis vi ikke kjenner fordelingen til X, <u>men n er stor</u>, så vil $\bar{X}$ allikevel være tilnærmet normalfordelt.

Dette kalles $\boxed{\text{sentralgrenseteoremet}}$ ❤❤❤

a) $\mu_{\bar{X}} = \mu_X = \underline{190}$

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{288}{\sqrt{70}} = \underline{34.4}$$

b) $\bar{X} \overset{\text{tilnærmet}}{\sim} N(190, 34.4)$. Da er

$$P(\bar{X} > 250) = 1 - P(\underbrace{\bar{X}}_{N(190,34.4)} < 250) = 1 - P\left(\underbrace{Z}_{N(0,1)} < \frac{250-190}{34.4}\right) \xrightarrow{\text{tabell A}} = 1 - 0.96 = \underline{0.04}$$

c) Y = totalt antall venner for de 70 facebook-brukerne

To fremgangsmåter; I og II

I $Y = \sum_{i=1}^{70} X_i = n \cdot \bar{X}$

$$\mu_{70\cdot\bar{X}} = E(70 \cdot \bar{X}) = 70 \cdot E(\bar{X}) = 70 \cdot 190 = \underline{13300}$$

$$\sigma_{70\bar{X}} = \sqrt{Var(70 \cdot \bar{X})} = \sqrt{70^2 \cdot Var(\bar{X})} = \sqrt{4900 \cdot 34.4^2} \approx \underline{2408}$$

II $Y = \sum_{i=1}^{70} X_i$

$$\mu_Y = E(Y) = E(\sum X_i) = \sum E(X_i) = 70 \cdot \mu_X = 70 \cdot 190 = \underline{13300}$$

$$\sigma_Y = \sqrt{Var(Y)}$$

$$Var(Y) = Var(\sum X_i) \overset{X_i \text{ uavh}}{=} \sum Var(X_i) = 70 \cdot \sigma_X^2 = 70 \cdot 288^2$$

$$\sigma_Y \approx \underline{2410}$$

d) Når $\bar{X} \overset{tilnn}{\sim} N$, så er også $Y = n \cdot \bar{X} \overset{tilnn}{\sim} N$

$\downarrow$ $\mu_{\bar{x}} = 190$, $\sigma_{\bar{x}} = 34.4$

$\downarrow$ $\mu_Y = 13300$, $\sigma_Y = 2410$

$$P(Y > 17.500) = 1 - P(\underbrace{Y}_{N(13300,\,2400)} < 17500) = 1 - P(\underbrace{Z}_{N(0,1)} < \frac{17500-13300}{2400}) \approx 1 - 0.96 = \underline{0.04}$$

Tabell A

5.21 SRS, måler kolesterolnivåer hos 13-16-åringer, og bruker $\bar{x}$ til å estimere μ, forventet kolesterolnivå i populasjonen.

a) "$\bar{X}$ is an unbiased estimator of μ"
betyr "$\bar{x}$ er en <u>forventningsrett</u> estimator for μ"
riktig forventning

Presist: Forventningsverdien til $\bar{X}$ er det samme som forventningen i populasjonen; $E(\bar{x}) = \mu$

Muntlig: Hvis n blir veldig stor (vi undersøker så mange 13-16-åringer at det nærmer seg uendelig), vil det observerte gjennomsnittet bli mer og mer likt gjennomsnittet i populasjonen, som vi kaller forventningsverdien i populasjonen.

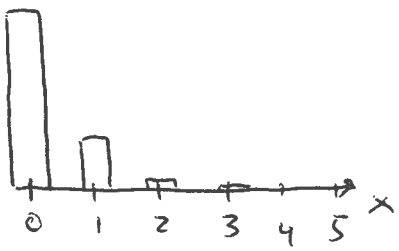
b) Hvorfor er $\bar{x}$ basert på et stort utvalg mer til å stole på enn et $\bar{x}$ basert på et lite utvalg?

Estimeringsusikkerheten, også kalt $\sigma_{\bar{x}}$, standardavviket til $\bar{x}$, eller estimeringsfeilen, minker når n øker; dermed blir estimatet (altså $\bar{x}$) sikrere jo større n vi har.

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

5.23 Setter opp feller for asiatisk askeprakttbille, og teller X = # biller i hver felle.

Diskret fordeling, svært skjev:
Majoriteten av fellene har 0 biller, få har flere enn 1 bille
Jeg ser for meg noe sånt:



Oppgis: $\mu_X = 0.3$, $\sigma_X = 0.8$

(jfr tidligere forelesninger om deskr. stat. Dette er ikke gode oppsummeringstall for slike data, men de kan allikevel regnes ut.)

a) Undersøker SRS, $n = 100$ feller.

$$\rightarrow X_1, \ldots, X_{100} \rightarrow \bar{X} = \frac{1}{100}\sum_{i=1}^{100} X_i$$

$$\mu_{\bar{X}} = \mu_X = \underline{0.3}$$

$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{0.8}{\sqrt{100}} = \underline{0.08}$$

b) Sentralgrenseteoremet sier at uansett hvilken fordeling X_i har, hvis $\bar{X}$ er stor nok, så er

$$\bar{X} \overset{\text{tilnærm}}{\sim} N$$

$$\downarrow$$

$$\mu_{\bar{X}} = 0.3$$
$$\sigma_{\bar{X}} = 0.08$$

Da er $P(\bar{X} > 0.5) = 1 - P(\underbrace{\bar{X}}_{N(0.3,\,0.08)} < 0.5) = 1 - P\left(\underbrace{Z}_{N(0,1)} < \frac{0.5-0.3}{0.08}\right) \approx 1 - 0.994$ (table A) $= \underline{0.006}$

c) Med en så skjev (og diskret) fordeling, er det usikkert/tvilsomt om $n = 100$ er stor nok n til at sentralgrenseteoremet egentlig har "begynt å virke". Antakelig vil ikke $\bar{X}$ være tilnærmet N-fordelt før n blir større. Det betyr i så fall at sannsynligheten vi regnet ut i b) er basert på feil premisser, og er feil.