CA 1 – STAT100

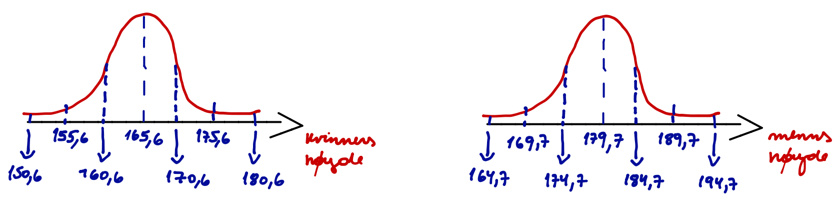
Fábio Rodrigues Pereira

fabio.rodrigues.pereira@nmbu.no

Oppgave 0

Ta utgangspunkt i at kvinners høyde er normalfordelt med forventningsverdi μ=165,6 cm og standardavvik σ=5 cm og at menns høyde er normalfordelt med forventningsverdi μ=179,7 cm og standardavvik σ=5 cm.

Tegn en skisse av de to fordelingene.



Regn ut z-scoren til deg selv og hvert av medlemmene i kollokviegruppa, og bruk normalfordelingstabellen (Tabell E3 i boka) til å beregne hvor stor andel av befolkningen som er lavere enn hver av dere, og hvor stor andel som er høyere enn hver av dere. Hvem har den laveste z-scoren på gruppa, og hvem har den høyeste?

Fábio = 186 = M;

$M \sim N(179{,}7\,;\,5) \longrightarrow Z = \dfrac{M-\mu}{\sigma} \longrightarrow Z = \dfrac{186 - 179{,}7}{5} = 1{,}26$ is the z-score of me

z	+0.00	+0.01	+0.02	+0.03	+0.04	+0.05	+0.06	+0.07	+0.08	+0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55966	0.56360	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147

'hvor stor andel av befolkningen som er lavere enn meg' ≅ 89,43 %

$\hookrightarrow P(Z < 1{,}26) = P(Z \leq 1{,}25) = 0{,}89435$ //

'hvor stor andel som er høyere enn meg' ≅ 10,38 %

$\hookrightarrow P(Z > 1{,}26) = 1 - P(Z \leq 1{,}26) = 1 - 0{,}89617 \cong 0{,}10383 \cong 10{,}38\,\%$

Oppgave 1

TV-programmet Folkeopplysningen på NRK tar i en episode for seg økologisk mat. Blant annet vil de teste om folk synes økologisk vin smaker bedre enn vin fra konvensjonell fremstilling. Ta en kikk på denne videosnutten, ca 20 minutter ut i programmet, innslaget varer i 2 min og 30 sek:

1. I innslaget ser vi 4 personer/grupper som smaker på vanlig hvitvin og på det de tror er økologisk hvitvin. La oss gå ut i fra at folk IKKE har preferanse hverken for økologisk eller ikke-økologisk vin. Hva er da sannsynligheten for at en tilfeldig person sier at den økologiske hvitvinen er best i denne testen?

$P(\text{en tilfeldig person sier at økologiske hvitvinen er best i testen}) = \frac{1}{2}$ $\frac{\longrightarrow 1 \text{ outcome possible: organic best}}{\longrightarrow \text{S with 2 options only: organic best or not best}}$

2. La den tilfeldige variabelen X være antall personer som synes økologisk vin er best blant de 4 som tar testen. Gi argumenter for at X er binomisk fordelt.

$X \sim bin(\#\text{forsøk} = 4,\ P(\text{suksess}) = 0,5)$ $*\ P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$

↳ n number of independent attempts where each attempt has the same probability of occurring;
↳ also each attempt has only 2 possible outcome.

3. I innslaget svarer alle fire at de synes den økologiske vinen er best. Finn sannsynligheten for at dette skal skje under antagelsene vi gjorde i oppgave a. På grunnlag av svaret ditt, tror du at folk er mer tilbøyelige til å mene at økologisk vin er best?

$P(\text{alle fire synes økologiske vinen er best}) = P(X=4;\ \overset{n}{4};\ \overset{p}{0,5}) = \left(\frac{n!}{x!(n-x)!}\right)\cdot p^x \cdot (1-p)^{n-x}$

$= \frac{\cancel{4!}^{1}}{\cancel{4!}\,0!}\cdot\left(\frac{1}{2}\right)^4\cdot\cancel{\left(\frac{1}{2}\right)^0}^{1} = \frac{1}{16}$

$\longrightarrow = 0,0625 = 6,25\%$

* Yes, it seems that the people under survey was influenced to say that the organic wine is the best, because the probability of all the 4 participants to pick the organic wine as the best is very low, about 6,25%.

4. Hva er sannsynligheten for at mer enn 2 personer svarer ja til at den økologiske vinen er best dersom antagelsen i 1. er oppfylt?

$P(X>2) = P(X\geq 3) = \sum_{x=3}^{4}\binom{4}{x}\left(\frac{1}{2}\right)^x\left(1-\frac{1}{2}\right)^{4-x} = \left(\frac{4\cdot\cancel{3!}^{1}}{\cancel{3!}\,1!}\cdot\left(\frac{1}{2}\right)^3\cdot\left(\frac{1}{2}\right)^1\right) + \left(\frac{\cancel{4!}^{1}}{\cancel{4!}\,0!}\cdot\left(\frac{1}{2}\right)^4\cdot\cancel{\left(\frac{1}{2}\right)^0}^{1}\right) = \frac{4}{16}+\frac{1}{16} = \frac{5}{16} = 0,3125 = 31,25\%$

or $P(X>2) = 1-P(X\leq 2) = 1-\left[\sum_{x=0}^{2}\binom{4}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{4-x}\right] = 31,25\%$

5. Anta at det i stedet for var 10 personer som ble spurt og at alle svarte at den økologiske var best. Kan dette med rimelig sannsynlighet skje dersom antagelsen i 1. holder?

$X \sim binom(10;\ 10;\ 0,5)$

$P(X=10) = \binom{10}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^0 = \frac{\cancel{10!}^{1}}{\cancel{10!}\,0!}\cdot\frac{1}{1024}\cdot\cancel{\left(\frac{1}{2}\right)^0}^{1} = \frac{1}{1024} \cong 0,00097 \cong 0,097\%$

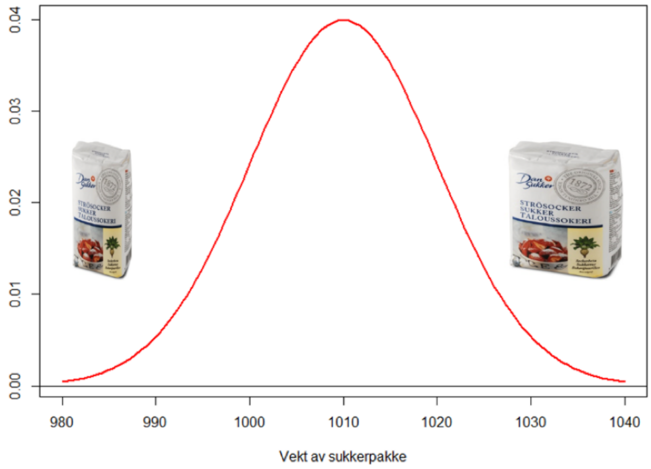
Yes, it is possible but with approx. 0,1% of chances.

Oppgave 2

Får vi det vi betaler for?

Tror du at du får akkurat 1 kg sukker når du kjøper en kilospakke i butikken? En sukkerprodusent produserer kilospakker som ideelt sett skal inneholde nøyaktig én kilo, men på grunn av unøyaktigheter i produksjonsprosessen vil det selvsagt variere litt. Utfra et stort antall målinger viser det seg at vekta på pakkene har forventning 1010 gram med et standardavvik på 10 gram. Vi antar at det er tilfeldige prosesser som gjør at vekten varierer.

$N(\mu = 1010 ; \sigma = 10)$



(Illustrasjon: Dansukker har ikke noe med tallene i denne oppgaven å gjøre)

1. Diskutér dette i gruppa:

a Hvorfor er det rimelig å anta at vekt av sukkerpakkene (X) er en *normalfordelt* tilfeldig variabel?

b Hva betyr det at X har forventning 1010 gram?

c Hvorfor tror dere at forventningen ligger på mer enn 1000 gram?

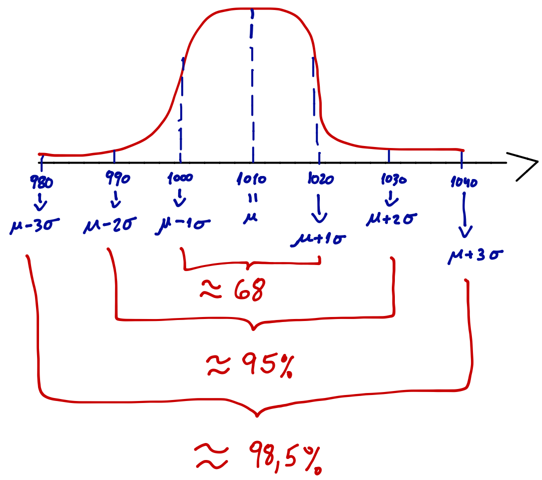
1a) The weight of a package of sugar (X) is a normal variable because its distribution's graphic is very symmetrical, with mean in the middle and standard deviation well defined.

1b) '$E[X]$=1010 gram' means the expected average over time.

1c) The expected value is above 1000 grams to ensure that most of the packages has the correct labeled weight (1000 gram).

2. Skissér følgende begivenheters utfallsrom og deres sannsynligheter i form av normalfordelingskurver med skraverte arealer under kurven og beregn sannsynlighetene for begivenhetene

$X > 1010$, $X < 1000$, $990 < X < 1020$



$*$ $P(X > 1010) = 1 - P(X \leq 1010) = 1 - 50\% = 50\%$

or $Z = \frac{1010 - 1010}{10} = 0$ is the z-score for $Z \sim N(0,1)$ $\therefore$ $P(X > 1010) = 1 - \overbrace{P(Z \leq 0)}^{\text{Table E.3}} = 1 - 0{,}5 = 50\%$

$*$ $P(X < 1000) = P(X \leq 999)$:

↳ $Z = \frac{999 - 1010}{10} = -1{,}1$ is the z-score for $Z \sim N(0,1)$ $\therefore$ $P(X < 1000) = \overbrace{P(Z \leq -1{,}1)}^{\text{Table E.3}} = 0{,}1357 = 13{,}57\%$

$*$ $P(990 < X < 1020) = P(990 < X \leq 1019) = P(X \leq 1019) - P(X \leq 990) = 0{,}8159 - 0{,}0228 \cong 0{,}7931 = 79{,}31\%$

↳ $Z_{1019} = \frac{1019 - 1010}{10} = 0{,}9 \longrightarrow P(Z \leq 0{,}9) = 0{,}8159$

$Z_{990} = \frac{990 - 1010}{10} = -2 \longrightarrow P(Z \leq -2) = 0{,}0228$

$\therefore$ $P(X \leq 1019) = P(Z \leq 0{,}9) = 0{,}8159$

$P(X \leq 990) = P(Z \leq -2) = 0{,}0228$

3. Hvor sannsynlig er det at du går hjem fra butikken med mindre sukker enn du betalte for hvis du kjøper en pakke?

$$P(X < 1000) = P(X \leq 999) \longrightarrow Z = \frac{999-1010}{10} = -1{,}1$$

From table E.3, for $z = -1{,}1$ the $P(X<1000) = P(Z \leq -1{,}1) \cong 13{,}57\%$

4. Dersom sukkerprodusenten bestemmer seg for at de 5 % letteste pakningene er for lette og skal trekkes tilbake, hva er grensa i gram for hvor lett sukkerposen kan være før den blir trukket tilbake?

$P(X \leq k) = 0{,}05 = P(Z \leq z) \longrightarrow$ from E.3, $z = -1{,}645$ for $P(Z \leq -1{,}645) = 5\% \longrightarrow -1{,}645 = \dfrac{k-1010}{10}$

$k = 993{,}55$ grams

5. Julebaksten står for døren og du kjøper like godt en 10-pakning med sukker, altså 10 kg sukker. Prøv, uten å regne, å gi en forklaring på hvorfor det er lite sannsynlig at du nå går hjem med mindre sukker enn det du betaler for i forhold det du fant i oppgave 3?

Kontrollér svaret ved regning.

$$\boxed{\text{normal distribution: } E[X] = \mu = 1010, \quad Var[X] = \sigma^2 = 100}$$

$W :=$ sum of the 10^{th} packing weights $= \sum_{i=1}^{10} X_i$

$$E[W] = E\left[\sum_{i=1}^{10} X_i\right] = E[X_1] + E[X_2] + \cdots + E[X_{10}] = 1010 \cdot 10 = 10100 \text{ grams}$$

$$Var[W] = Var\left[\sum_{i=1}^{10} X_i\right] = 10 \cdot 10^2 = 1000 \text{ grams}$$

$$SD[W] = \sqrt{1000} \cong 31{,}62$$

$$\therefore P(W < 10000) = P(W \leq 9999) = \frac{9999-10100}{31{,}62} \cong -3.19 \approx 0\%$$

6. Sukkerfirmaet kjøper en moderne fyllemaskin som gir vekt (V) som er fordelt slik: $V \sim N(1005, 5)$

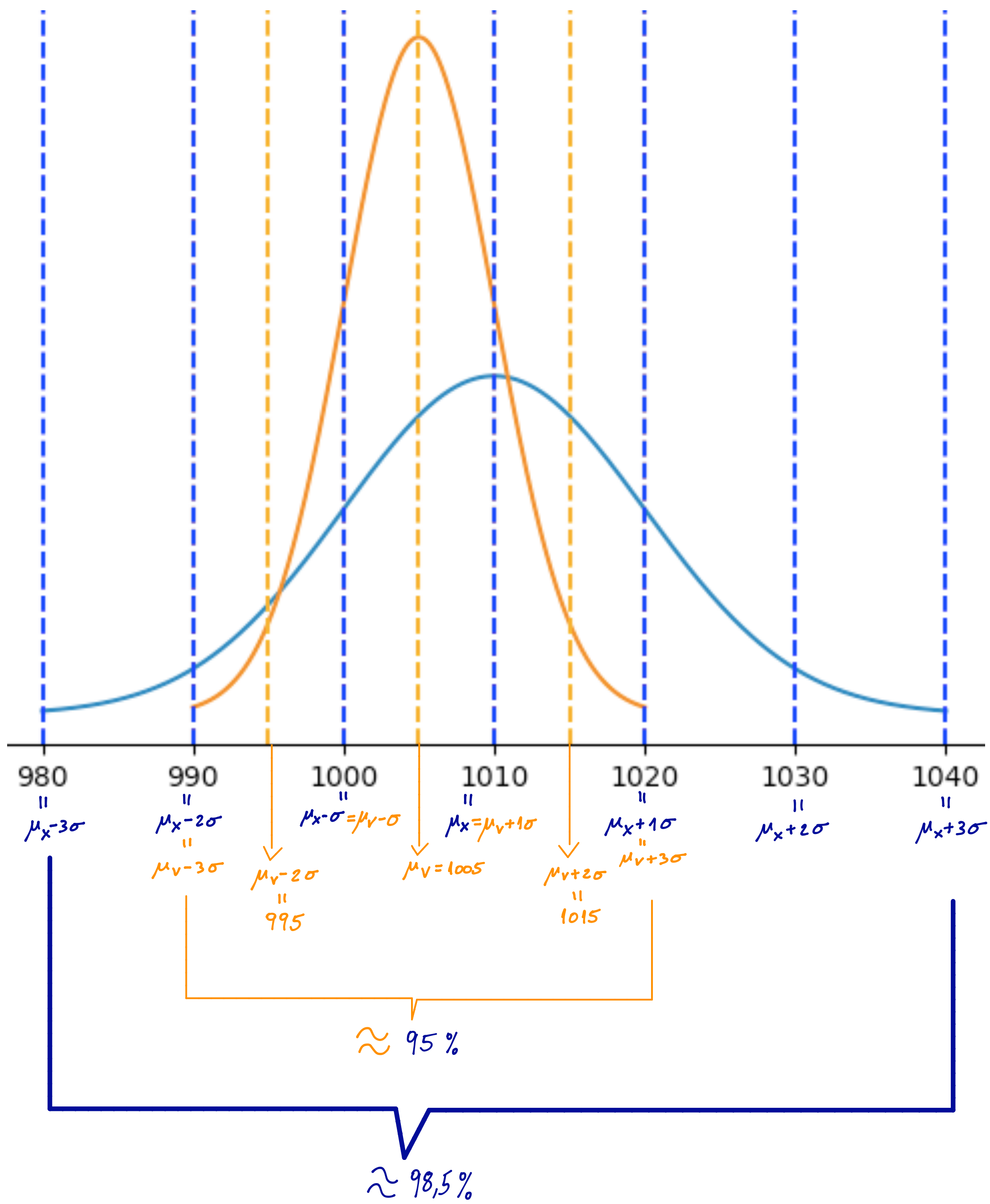
a) Hvorfor kan vi påstå at dette er en mer nøyaktig maskin enn den gamle?

b) Skissér fordelingene til X og V i samme figur. Hvor sannsynlig er det at du som kunde blir snytt på vekta ved kjøp av en pakke sukker etter at de byttet fyllemaskin?

cheated on

6a) We can suppose that the new machine is more precise because the mean is more near to the weight labeled and the standard deviation is less than the old machine.

6b)



$$P(V < 1000) = P(V \leq 999) \longrightarrow Z = \frac{999 - 1005}{5} = -1{,}2 \longrightarrow \therefore P(Z \leq -1{,}2) = 11{,}51\%$$