Kapittel 1

Beskrivende statistikk

Oppgave 1.1 Beregn empirisk middelverdi og standardavvik av følgende tall:

a) −1, 3, 8, 4, 13
b) −8, −9, −14, −11, 0

Løsningsforslag:

a) $\sum_{i=1}^{5} x_i = -1+3+8+4+13 = 27$, middelverdi: $\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = 27/5 = \underline{5.4}$

$\sum_{i=1}^{5} x_i^2 = (-1)^2 + 3^2 + 8^2 + 4^2 + 13^2 = 1+9+64+16+169 = 259$

standardavvik: $s = \sqrt{\frac{1}{5-1}\left(\Sigma x_i^2 - \frac{1}{5}(\Sigma x_i)^2\right)} = \sqrt{\frac{1}{4}(259 - \frac{1}{5}\cdot 27^2)} = \frac{1}{2}\sqrt{113.2} = \underline{5.32}$

b) $\sum_{i=1}^{5} x_i = (-8)+(-9)+(-14)+(-11)+0 = -42$,
middelverdi: $\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = -42/5 = \underline{-8.4}$

standardavvik: $s = \sqrt{\frac{1}{5-1}\left(\Sigma x_i^2 - \frac{1}{5}(\Sigma x_i)^2\right)} = \sqrt{\frac{1}{4}(462 - \frac{1}{5}\cdot (-42)^2)} = \frac{1}{2}\sqrt{109.2} = \underline{5.22}$

Oppgave 1.2 Gitt følgende tallmateriale:

1.43438 1.43443 1.43442 1.4344

a) La tallene ovenfor være x-verdier, og finn de tilsvarende z-verdier ved transformasjonen

$z = (x - 1.4344)/0.00001$

b) Beregn empirisk middelverdi, $\bar{z}$, og standardavvik, s_z, til z-verdiene.

c) Beregn middelverdi, $\bar{x}$, og standardavvik, s_x, til x-verdiene ved transformasjonen $s_x = 0.00001 s_z$ og

$\bar{x} = 1.4344 + 0.00001\cdot\bar{z}$

Løsningsforslag:

a) $z_1 = (x_1 - 1.4344)/0{,}00001 = (1.43438 - 1.4344)/0.00001 = \underline{-2}$

Tilsvarende finner vi $z_2 = \underline{3}$, $z_3 = \underline{2}$ og $z_4 = \underline{0}$

b) $\bar{z} = (-2+3+2+0)/4 = \underline{3/4}$, $\Sigma z^2 = 17$, $s_z = \sqrt{\frac{1}{3}(17-4\cdot(\frac{3}{4})^2)} = \sqrt{\frac{59}{12}} = \underline{2.22}$

c) $\bar{x} = 1{,}4344 + 0{,}00001\cdot 0{,}75 = \underline{1{,}4344075}$, $s_x = 0{,}00001\cdot 2{,}22 = \underline{0{,}0000222}$

Oppgave 1.3 Følg fremgangsmåten i oppgave 1.2 og beregn middelverdi og standardavvik til tallene

1221.3 1220.9 1221.7 1220.8 1221

Løsningsforslag: Vi søker først fornuftige verdier for *a* og *b* slik at transformasjonen $z = (x-a)/b$ fører til at $z_{max}-z_{min}$ verken blir for liten eller for stor (størrelsesorden lik 1). *a* er en såkalt lokaliseringsparameter som karakteriserer senter i tallmaterialet, og *b* er en såkalt skalaparameter som karakteriserer spredningen i tallmaterialet (rundt middelverdien). Vi velger skjønnsmessig a = 1221 og $b = 1$ og får:

$z_1 = (1221.3-1221)/1 = 0.3$, $z_2 = -0.1$, $z_3 = 0.7$, $z_4 = -0.2$ og $z_5 = 0$.

$$\bar{z} = (0.3-0.1+0.7-0.2+0)/5 = 0.7/5 = 0.14$$

$$\Sigma z^2 = 0.3^2 + (-0.1)^2 + 0.7^2 + (-0{,}2)^2 + 0^2 = 0{,}63,\ \ s_z = \sqrt{\tfrac{1}{4}(0.63 - \tfrac{1}{5}\cdot 0.7^2)} = 0.365$$

$$\bar{x} = a + b\cdot\bar{z} = 1221 + 0.14 = \underline{1221.14},\ \ s_x = b\cdot s_z = \underline{0.365}$$

Oppgave 1.4 I en bedrift med 14 ansatte er gjennomsnittsinntekten kr 158 000 pr. ansatt. Hvor store lønnsutgifter har bedriften?

Løsningsforslag: $\frac{1}{14}\Sigma x = \text{kr } 158000 \Rightarrow \Sigma x = 158000\cdot 14 = \underline{\text{kr } 2212000}$

Oppgave 1.5

a) Finn empirisk middelverdi og median for følgende tilfeldig målte utetemperaturer (^{o}C) som en person har notert i sin dagbok i løpet av ett år:

x: 14, –12, 0, 16, 7

b) En annen person har følgende måleresultater fra sin dagbok (samme sted og samme år):

y: 14, –12, 0, 16, 7, 112

Beregn empirisk middelverdi og median også i dette tilfellet. Ville du stole mest på middelverdien eller medianen?

Løsningsforslag:

a) $\bar{x} = (14 - 12 + 0 + 16 + 7)/5 = \underline{5\ ^{\text{o}}\text{C}}$. For å finne medianen ordner vi først x'ene i stigende rekkefølge: $x_{(1)} = -12, x_{(2)} = 0, x_{(3)} = 7, x_{(4)} = 14, x_{(5)} = 16$. Siden n =5 er et oddetall, blir medianen lik den midterste verdien:

$m_x = x_{(\frac{5+1}{2})} = x_{(3)} = \underline{7\ ^{\text{o}}\text{C}}$

b) $\bar{y} = (14 - 12 + 0 + 16 + 7 + 112)/6 = \underline{22.8\ ^{\text{o}}\text{C}}$.

For å finne medianen ordner vi første y'ene i stigende rekkefølge:

$y_{(1)} = -12, y_{(2)} = 0, y_{(3)} = 7, y_{(4)} = 14, y_{(5)} = 16, y_{(6)} = 112$

Siden n = 6 er et liketall, får vi to verdier på midten, og vi definerer medianen som middelverdien av disse: $m_y = \frac{1}{2}(y_{(6/2)} + y_{(6/2+1)}) = \frac{1}{2}(7 + 14) = \underline{10.5\ ^{\text{o}}\text{C}}$. Jeg ville stolt mer på medianen enn middelverdien, da målingen på 112 $^{\text{o}}$C er tvilsom.

Oppgave 1.6 Gitt følgende x-verdier:

1, –3, 7, 12, 6

Beregn 20- og 70-prosentilen.

Løsningsforslag: $n = 5,\ x_{(1)} = -3,\ x_{(2)} = 1,\ x_{(3)} = 6,\ x_{(4)} = 7,\ x_{(5)} = 12$

20-prosentil: $np = 5 \cdot 0.2 = 1$ (heltall) $\Rightarrow x_{0.2} = \frac{1}{2}(x_{(1)} + x_{(2)}) = \frac{1}{2}(-3 + 1) = \underline{-1}$

70-prosentil: $np = 5 \cdot 0.7 = 3.5$ (ikke heltall, forhøyes til 4) $\Rightarrow x_{0.7} = x_{(4)} = \underline{7}$

Oppgave 1.7 Gitt x-verdiene:

21, 17, 18, 17, 22, 21, 78, 22, 24

Beregn standardavvik og interkvartilbredde. Hvilket av de to målene på spredning ville du stole mest på, dersom du i ettertid fikk opplyst at en av verdiene var feil?

Løsningsforslag:

$\Sigma x = 21 + 17 + \cdots + 24 = 240, \quad \Sigma x^2 = 21^2 + 17^2 + \cdots + 24^2 = 9412$

standardavvik: $s = \sqrt{\frac{1}{9-1}(9412 - \frac{1}{9} \cdot 240^2)} = \underline{19.4}$

For å beregne interkvartilbredde ordner vi først dataene i stigende rekkefølge:

17, 17, 18, 21, 21, 22, 22, 24, 78

Nedre kvartil: $np = 9 \cdot 0.25 = 2.25 \Rightarrow Q_1 = x_{(3)} = 18$

Øvre kvartil: $np = 9 \cdot 0.75 = 6.75 \Rightarrow Q_{(3)} = x_{(7)} = 22$

Interkvartilbredden blir: $Q_3 - Q_1 = 22 - 18 = \underline{4}$.

Det er større grunn til å stole på <u>interkvartilbredden</u> enn på standardavviket, da «ekstrem»-verdien 78 i datamaterialet trolig er feil og bidrar sterkt til å øke standardavviket. Interkvartilbredden er imidlertid robust (lar seg lite påvirke) av tallet som er feil.

Oppgave 1.8 Tabellutsnittet til høyre viser 3 vektklasser. Angi nedre og øvre klassegrense, samt klassemidtpunkt og klassebredde.

Vekt [kg]
⋮
[10–20>
[20–30>
[30–40>
⋮

<u>Løsningsforslag</u>: Siden det er snakk om vekter er det fornuftig å tenke seg at disse er forhøyet. Nedre klassegrenser for de tre klassene som er angitt med tall blir da 9.5, 19.5, og 29.5. De tilsvarende øvre klassegrensene blir 19.499.., 29.499... og 39.499..., mens midtverdiene blir 14.5, 24.5 og 34.5 (midt mellom øvre og nedre klassegrense).

Oppgave 1.9 Neste tabell viser daglig inntekt av CD-salg i en musikkforretning, fordelt på 4 inntektgrupper. Anta at et dagssalg på kr. 9999 kommer i den første klassen.

a) Bestem klassemidtpunkt i hver klasse.
b) Bestem gruppert middelverdi.
c) Bestem gruppert standardavvik.
d) Bestem gruppert median.
e) Bestem gruppert interkvartilbredde.

Videosalg (kr 1000)	Frekvens (dager)
0–9	37
10–19	148
20–29	123
30–39	57

<u>Løsningsforslag</u>:

klasse:	frekvens f_i	relativ frekvens f_i/n	klasse-midtpkt m_i	$m_i \cdot f_i$	$m_i^2 \cdot f_i$	kumulativ frekvens F_i	relativ kum.frekv. F_i/n
0-9	37	0.1014	5	185	925	37	0.1014
10-19	148	0.4055	15	2220	33300	185	0.5068
20-29	123	0.3370	25	3075	76875	308	0.8438
30-39	57	0.1562	35	1995	69825	365	1.0000
sum:	365			7475	180925		

Kronebeløpene som inngår i tabellen (f.eks. m_i-verdiene) er angitt i enheter på kr 1000.

a) Siden det er pengebeløp forhøyer vi ikke, og klassemidtpunktene blir som angitt i tabellen.

b) Gruppert middelverdi: $\bar{x}_g = \frac{1}{n}\Sigma m_i f_i = \frac{7475}{365} = 20.480$, dvs. kr 20480

c) Gruppert standardavvik:

$$s_g = \sqrt{\tfrac{1}{n}\Sigma m_i^2 f_i - \bar{x}_g^{\,2}} = \sqrt{\tfrac{180925}{365} - 20.480^2} = 8.73, \text{ dvs. kr 8730}$$

d) Medianklassen er første klasse med relativ kumulativ frekvens større enn 0,5. Fra tabellen ser vi at dette er klassen 10-19 med nedre klassefrekvens x_1=10, klassefrekvens f_m=148, klassebredde $\Delta x_m = 10$ og med kumulativ frekvens F_1=37 i klassen forut for medianklassen. Vi får da:

$$m_g = x_1 + \frac{n/2 - F_1}{f_m} \cdot \Delta x_m = 10 + \frac{365/2 - 37}{148} \cdot 10 = 19.83, \text{ dvs. kr 19 830}$$

e) Nedre og øvre kvartilklasser er første klasse der relativ kumulativ frekvens overstiger henholdsvis 0.25 og 0.75. Klassen 10-19 er følgelig nedre kvartilklasse og klassen 20-29 er øvre kvartilklasse. Analogt med utregning av medianen får vi:

$$Q_{1g} = 10 + \frac{365 \cdot 0.25 - 37}{148} \cdot 10 = 13{,}666, \quad Q_{3g} = 20 + \frac{365 \cdot 0.75 - 185}{123} \cdot 10 = 27.215$$

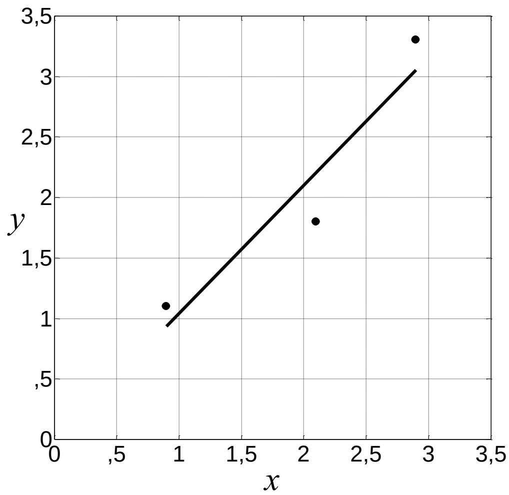
Gruppert interkvartilbredde: $Q_{3g} - Q_{1g} = 27.215 - 13.666 = 13.55$, dvs. kr 13 550

Oppgave 1.10 Gitt tallparene

(0,9 , 1,1), (2,1 , 1,8) og (2,9 , 3,3)

a) Velg 1 cm/enhet langs *x*- og *y*-aksen, og lag et spredningsdiagram.
b) Beregn korrelasjonskoeffisienten, *r*.
c) Beregn en regresjonslinje tilpasset dataene, og tegn den inn i spredningsdiagrammet.

Løsningsforslag: a) se figur til høyre



Hjelpestørrelser for b) og c):

$n = 3$

$\Sigma x = .9 + 2.1 + 2.9 = 5.9$

$\Sigma y = 1.1 + 1.8 + 3.3 = 6.2$

$\Sigma x^2 = .9^2 + 2.1^2 + 2.9^2 = 13{,}63$

$\Sigma y^2 = 1.1^2 + 1.8^2 + 3.3^2 = 15.34$

$\Sigma xy = .9 \cdot 1.1 + 2.1 \cdot 1.8 + 2.9 \cdot 3.3 = 14.34$

b) korrelasjonskoeffisient: $r = \dfrac{\Sigma\, xy - \frac{1}{n}\Sigma x \Sigma y\,/\,n}{\sqrt{(\Sigma x^2 - \frac{1}{n}(\Sigma x)^2)\cdot(\Sigma y^2 - \frac{1}{n}(\Sigma y)^2)}} = \underline{0.949}$

c) regresjonslinje

$$y_r^* = a^* + b^* x,\;\; b^* = \frac{\Sigma\, xy - \frac{1}{n}\Sigma x \Sigma y}{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2} = 1.06,\;\; a^* = \bar{y} - b^*\cdot \bar{x} = -0.016$$

Regresjonslinja $\underline{y^* = -.016 + 1.06 \cdot x}$ er tegnet inn i figuren mellom punktene

$(0.9,\, y_r^*(0.9))$ og $(2.9,\, y_r^*(2.9))$.

Oppgave 1.11 Gitt følgende samhørende *x*- og *y*-verdier:

x	.68	.52	.58	.72	.62
y	11	5	1	4	6

a) Beregn korrelasjonskoeffisienten, r.
b) Beregn en regresjonslinje tilpasset dataene, $y^* = a^* + b^* x$.
c) Tegn (x,y)-verdiene i et spredningsdiagram med 1 cm pr. enhet ***både*** langs *x*- og *y*-aksen, og tegn inn regresjonslinja. Gir figuren inntrykk av sterk korrelasjon?
d) Gjør c) på nytt, men velg selv en fornuftig skalering langs aksene.

Løsningsforslag: Hjelpestørrelser:

$$\Sigma\, x = .68 + .52 + \cdots + .62 = 3.12,\quad \Sigma\, y = 11 + 5 + \cdots + 6 = 27$$

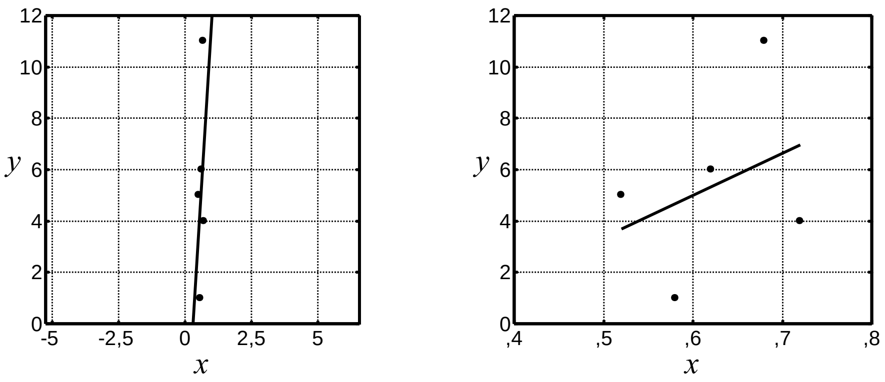
$$\Sigma\, x^2 = .68^2 + .52^2 + \cdots + .62^2 = 1.972,\quad \Sigma\, y^2 = 11^2 + 5^2 + \cdots + 6^2 = 199$$

$$\Sigma\, xy = .68\cdot 11 + .52\cdot 5 + \cdots + .62 \cdot 6 = 17.26$$

a) korrelasjonskoeffisient, $r = \dfrac{\Sigma xy - \frac{1}{5}\Sigma x \Sigma y}{\sqrt{(\Sigma x^2 - \frac{1}{5}(\Sigma x)^2)\cdot(\Sigma y^2 - \frac{1}{5}(\Sigma y)^2)}} = \underline{.356}$

b) $b^* = \dfrac{\Sigma xy - \frac{1}{5}\Sigma x \Sigma y}{\Sigma x^2 - \frac{1}{5}(\Sigma x)^2} = 16.40,\;\; a^* = \bar{y} - b^* \cdot \bar{x} = -4.834,\;\; \underline{y^* = -4.834 + 16.40 \cdot x}$

c), d) Spredningsplot og regresjonslinjer tegnet inn i figuren nedenfor. Skaleringen bør velges slik at utstrekningen av data i hver retning på figuren blir ca. like stor. I figuren til venstre har dataene for liten ustrekning i *x*-retning, og vi kan lett bli narret til å tro at vi har en sterk korrelasjon karakterisert ved en tilnærmet loddrett linje.



Oppgave 1.12 Størrelsen av en dyrebestand måles en gang i året. De 4 siste årene ble følgende observert:

år:	87	88	89	90
bestand:	123	237	471	982

La x = 1, 2, 3 og 4 betegne årstallene 87, 88, 89 og 90, og la y betegne de tilsvarende bestandene.

a) Beregn korrelasjonskoeffisienten.
b) Utfør transformasjonen $z = \ln y$, og beregn z-verdiene.
c) Bestem korrelasjonskoeffisienten for (x,z)-verdiene.
d) Tilpass en rett linje til (x,z)-dataene.
e) Bestem på basis av d) en regresjonsfunksjon, $y =$□ce^{dx}, tilpasset (x,y)-dataene.

<u>Løsningsforslag</u>: Hjelpestørrelser:

z-verdier: {ln(123), ln(237), ln(471), ln(982)} = {4.8122, 5.4681, 6.1549, 6.8896}

$$\Sigma x = 1 + \cdots + 4 = 10,\ \Sigma y = 123 + \cdots + 982 = 1813,\ \Sigma z = 4.8122 + \cdots + 6.8896 = 23.3247$$

$$\Sigma x^2 = 30,\ \ \Sigma y^2 = 1257463,\ \ \Sigma z^2 = 138.4055,\ \ \Sigma xy = 5938,\ \ \Sigma xz = 61.7712,\ n = 4.$$

a) Korrelasjonskoeffisient $r(x,y) = \dfrac{\Sigma xy - \frac{1}{n}\Sigma x \Sigma y}{\sqrt{(\Sigma x^2 - \frac{1}{n}(\Sigma x)^2)\cdot(\Sigma y^2 - \frac{1}{n}(\Sigma y)^2)}} = \underline{.9522}$

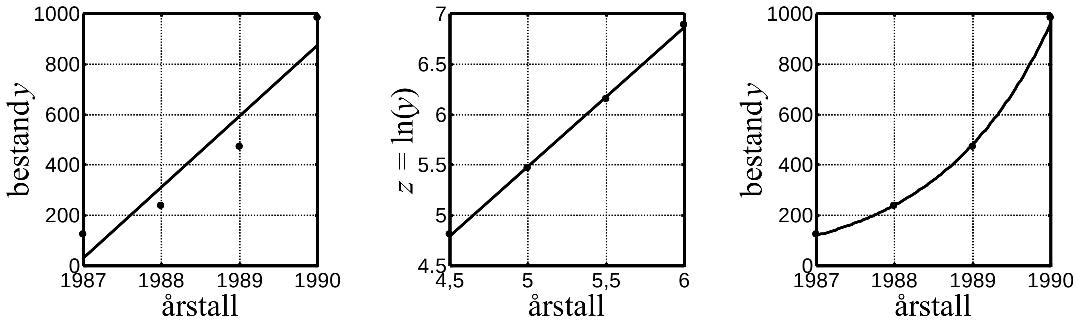
b) z-verdiene er beregnet først i løsningsforslaget.

c) Korrelasjonskoeffisient $r(x,z) = \dfrac{\Sigma xy - \frac{1}{n}\Sigma x \Sigma z}{\sqrt{(\Sigma x^2 - \frac{1}{n}(\Sigma x)^2)\cdot(\Sigma z^2 - \frac{1}{n}(\Sigma z)^2)}} = \underline{.9997}$

Sammenlignet med $r(x,y)$ ser vi at en rettlinjet tilpasning til (x,z)-dataene later til å være en bedre tilpasning enn en rettlinjet tilpasning til (x,y)-dataene.

d) Regresjonslinje,

$$z^* = a^* + b^* \cdot x,\ \ b^* = \frac{\Sigma xz - \frac{1}{4}\Sigma x \Sigma z}{\Sigma x^2 - \frac{1}{4}(\Sigma x)^2} = \underline{.6919},\ \ a^* = \bar{z} - b^* \cdot \bar{x} = \underline{4.1014}$$



e) $y = ce^{dx} \Rightarrow z = \ln(y) = \ln(c) + d \cdot x = a^* + b^* \cdot x$

$\Rightarrow a^* = \ln(c),\ c = e^{a^*} = 60.426,\ d = b^* = .6919,\ \underline{y = 60.426 \cdot e^{.6919 \cdot x}}$

Oppgave 1.13 (Vekt-data i tabell lenger ned).

a) Finn middelverdi og standardavvik til vekt-dataene på basis av rådataene.
b) Rangordne vekt-dataene og bestem median og interkvartilbredde.
c) Gruppér vekt-dataene og bestem gruppert middelverdi, gruppert standardavvik og gruppert median.
d) Fremstill de grupperte vekt-dataene i et relativ frekvens-histogram.

Løsningsforslag:

a) vekt: $\Sigma y = 55 + 59 + \cdots + 85 = 2386$, middelverdi: $\bar{y} = \frac{1}{35}\Sigma y = \underline{68.17 \text{ kg}}$

standardavvik: $s_y = \sqrt{\frac{1}{34}(\Sigma y^2 - \frac{1}{35}(\Sigma y)^2)} = \underline{10.92 \text{ kg}}$

b) Rangordna vektdata:

$x_{(1)} = 50,\ x_{(2)} = 52,\ x_{(9)} = 59,\ x_{(18)} = 67,\ x_{(27)} = 76,\ x_{(34)} = 85,\ x_{(35)} = 90$

Median: $n = 35 \Rightarrow n \cdot ,5 = 17{,}5$. Forhøyer til 18 og får: $m = x_{(18)} = \underline{67 \text{ kg}}$

Interkvartilbredde: $n \cdot ,25 = 8.75 \Rightarrow Q_1 = x_{(9)} = 59$, $n \cdot ,75 = 26.26 \Rightarrow Q_3 = x_{(27)} = 76$

$Q_3 - Q_1 = 76 - 59 = 17$, dvs. interkvartilbredden er $\underline{17 \text{ kg}}$

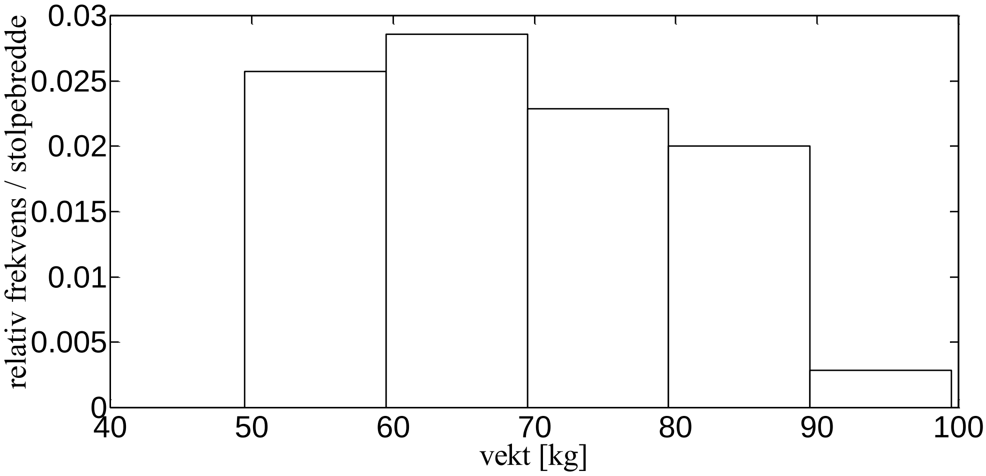
c) Minste vekt er på 50 kg og største på 90 kg. Vi velger klassebredde på 10 kg og får følgende tabell:

vekt-klasse: [kg]	frekvens f_i	relativ frekvens f_i/n	klasse-midtpkt m_i	$m_i \cdot f_i$	$m_i^2 \cdot f_i$	kumulativ frekvens F_i	relativ kum.frekv. F_i/n
[50-60>	9	0,257	54.5	490.5	26732	9	.257
[60-70>	10	,286	64.6	645	41602	19	.543
[70-80>	8	,229	74.5	596	44402	27	0.771
[80-90>	7	,200	84.5	591.5	49982	34	0.971
[90-100>	1	,029	94.5	94.5	8930	35	1.000
sum:	35	1.001		2417.5	171650		

Gruppert middelverdi: $\bar{y}_g = \frac{1}{35}\Sigma m_i \cdot f_i = \frac{2417{,}5}{35} = 69{,}07$, dvs. 69.07 kg

Gruppert standardavvik: $s_y = \sqrt{\frac{1}{35}\Sigma m_i^2 \cdot f_i - (\bar{x}_g)^2} = \sqrt{\frac{171650}{35} - 69.07^2} = \underline{11.55}$ kg

d) Relativ frekvenshistogram for de grupperte vektene:



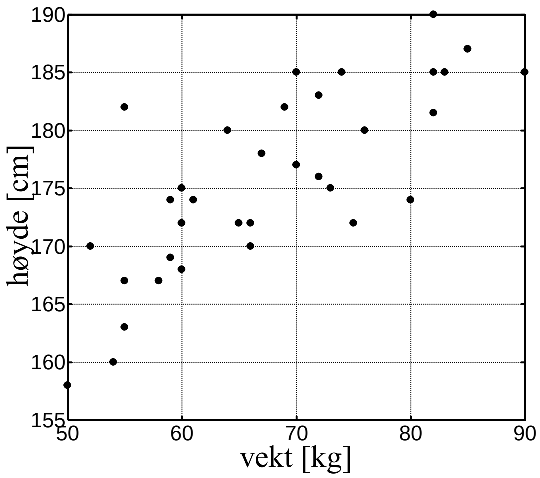
Oppgave 1.14 (data i tabell lenger ned).

a) Tegn et spredningsdiagram for samhørende vekt- og høyde-data.

b) Beregn korrelasjonskoeffisienten for vekt- og høyde-dataene, og kommenter spredningsdiagrammet i a).

c) Beregn korrelasjonskoeffisienten for høyde- og alders-dataene, og kommenter resultatet.

Løsningsforslag:



a) Se spredningsplot til venstre

b) Hjelpestørrelser:

$\Sigma x = 6160.5\,\text{cm},\ \Sigma y = 2386\ \text{kg}$

$\Sigma x^2 = 1.086500,\ \Sigma y^2 = 166700,$

$\Sigma xy = 422300,$

Korrelasjonskoeffisient: $r(x,y) =$

$$\frac{\Sigma xy - \Sigma x \Sigma y / 35}{\sqrt{(\Sigma x^2 - \frac{1}{35}(\Sigma x)^2)(\Sigma y^2 - \frac{1}{35}(\Sigma y)^2)}}$$

$= \underline{0.78}$

En korrelasjonskoeffisient på 0.78 indikerer en rimelig sterk positiv korrelasjon mellom vekt og høyde, hvilket stemmer overens med inntrykket fra figuren under pkt. a).

c) Ved tilsvarende regning som i b) finner vi korrelasjonskoeffisienten for (høyde,alders)- dataene: $r(x,z) = \underline{0.11}$. At denne indikerer en svak korrelasjon er intuitivt nokså logisk, da studenter stort sett er noskå fullt utvokste.

Tabell for oppgave 1.13 og 1.14

høyde [cm]	vekt [kg]	alder [år]	høyde [cm]	vekt [kg]	alder [år]	høyde [cm]	vekt [kg]	alder [år]
167	55	32	180	64	24	185	83	25
169	59	23	175	73	21	160	54	26
163	55	25	170	52	20	172	75	22
177	70	23	172	65	21	167	58	19
180	76	24	170	66	23	190	82	25
174	61	25	175	60	19	172	66	21
172	60	19	185	90	23	185	74	21
174	59	19	178	67	21	182	69	21
174	80	21	187	85	28	181.5	82	21
182	55	31	185	82	27	176	72	21
158	50	23	183	72	19	187	85	28
168	60	21	185	70	19			

Oppgave 1.15 (E) I et utvalg på 20 fire-barnsfamilier er antall gutter følgende:

3 2 3 2 3 1 3 3 2 3 1 3 0 3 2 1 4 2 3 2

a) Regn ut middelverdien for utvalget.
b) Regn ut utvalgsstandardavviket.
c) Sett opp en fordelingstabell for de absolutte og relative frekvensene.
d) Bestem median og interkvartilbredde.
e) Lag relativ frekvens stolpediagram.

<u>Løsningsforslag</u>:

a) middelverdi (antall gutter): $\bar{x} = \frac{1}{20}(3+2+\cdots+2) = \underline{2.30}$

b) $\Sigma x^2 = 3^2 + 2^2 + \cdots + 2^2 = 124,\ s_x = \sqrt{\frac{1}{19}(\Sigma x^2 - 20\bar{x}^2)} = \underline{0.979}$

c)

# gutter	0	1	2	3	4
frekvens	1	3	6	9	1
relativ frekvens	.05	.15	.30	.60	.05

d) Vi ordner rådataene i stigende rekkefølge:

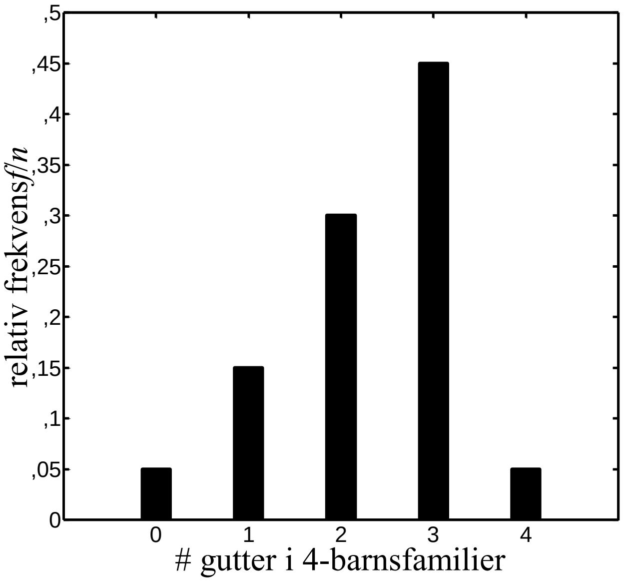
$$x_{(1)} = 0, x_{(2)} = 1, \ldots, x_{(5)} = 2, x_{(6)} = 2, \ldots, x_{(10)} = 2, x_{(11)} = 3, \ldots, x_{(15)} = 3, x_{(16)} = 3, \ldots$$

Median: $n \cdot 0.5 = 10 \Rightarrow m = \frac{1}{2}(x_{(10)} + x_{(11)}) = \underline{2.5}$

$n \cdot 0.25 = 5 \Rightarrow Q_1 = \frac{1}{2}(x_{(5)} + x_{(6)}) = 2,\ \ n \cdot 0.75 = 15 \Rightarrow Q_1 = \frac{1}{2}(x_{(15)} + x_{(16)}) = 3$

Interkvartilbredde: $Q_3 - Q_1 = 3 - 2 = \underline{1}$

e) Relativ frekvens stolpediagram:



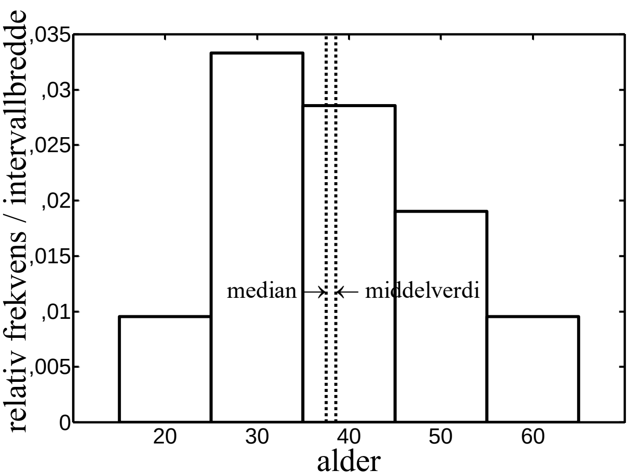
Oppgave 1.16 (E) Aldersfordelingen i en bedrift er:

Alder [år]	15-25	25-35	35-45	45-55	55-65
Frekvens	2	7	6	4	2

a) Tegn histogram over aldersfordelingen.
b) Beregn aritmetisk middelverdi, standardavvik og median.
c) Avmerk de beregnede størrelsene i histogrammet.

<u>Løsningsforslag</u>:

a)



b)

Klassemidtpunkt: 20, 30, 40, 50, 60

middelverdi: $\bar{x}_g = \frac{1}{21}\Sigma m_i f_i = \frac{1}{21}(20\cdot 2+\cdots+60\cdot 2) = 38.57$, dvs. <u>38.57 år</u>

standardavvik: $s = \sqrt{\frac{1}{21}\Sigma m_i^2 f_i - \bar{x}_g^2} = \sqrt{\frac{33900}{21} - 38.57^2} = 11{,}25$, dvs. <u>11.25 år</u>

Klassen 35-45 år blir medianklassen fordi dette er første klasse der relativ kumulativ frekvens er større enn 0,5. Nedre klassegrense i medianklassen er x_1=35, f_m = 6 er frekvensen til medianklassen, $\Delta x_m = 10$ er bredden til median-klassen og F_1 =9 er kumulativ frekvens til klassen forut for medianklassen.

Gruppert median: $m_g = x_1 + \dfrac{n/2 - F_1}{f_m}\cdot \Delta x_m = 35 + \dfrac{10.5-9}{6}\cdot 10$ år $=$ <u>37.5 år</u>

c) Median og middelverdi er inntegnet i figuren under pkt. a).

Oppgave 1.17 (E) En måleserie resulterte i følgende klassedelte observasjons-materiale:

intervall	frekvens
[195, 200>	2
[200, 205>	4
[205, 210>	10
[210, 215>	11
[215, 220>	9
[220, 225>	4

a) Utvid tabellen med relative frekvenser, kumulative frekvenser og relative kumulative frekvenser.

Lag et histogram for observasjonsmaterialet.

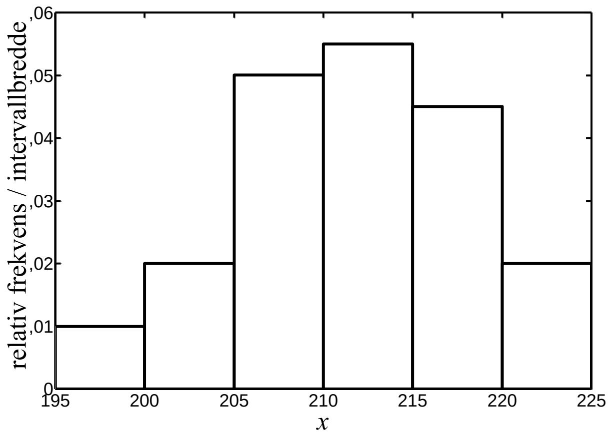
b) Beregn tilnærmet aritmetisk middelverdi og standardavvik for observasjonsmaterialet.

c) Beregn tilnærmet median og interkvartilbredde for observasjonsmaterialet.

<u>Løsningsforslag</u>:

a)

intervall	frekvens f_i	relativ frekvens	kumulativ frekvens	relativ kum. frekvens
[195, 200>	2	.050	2	.050
[200, 205>	4	.100	6	.150
[205, 210>	10	.250	16	.400
[210, 215>	11	.275	27	.675
[215, 220>	9	.225	36	.900
[220, 225>	4	.100	40	1.000
sum:	$n = 40$	1		



b) Klassemidtpunkter: m_i: 197.5, 202.5, 207.5, 212.5, 217.5, 222.5

middelverdi: $\bar{x}_g = \frac{1}{40}\Sigma m_i f_i = \frac{1}{40}(197.5 \cdot 2 + \cdots + 222.5 \cdot 2) = \underline{211.6}$

standardavvik: $s = \sqrt{\frac{1}{40}\Sigma m_i^2 f_i - \bar{x}_g^2} = \sqrt{\frac{1793100}{40} - 211.6^2} = \underline{6.5}$

c) Klassen 210-215 blir medianklassen fordi dette er første klasse der relativ kumulativ frekvens er større enn 0.5. Nedre klassegrense i medianklassen er x_1=210, $f_m = 11$ er frekvensen til medianklassen, $\Delta x_m = 5$ er bredden til medianklassen og $F_1 = 16$ er kumulativ frekvens til klassen forut for medianklassen.

$$\text{Gruppert median: } m_g = x_1 + \frac{n/2 - F_1}{f_m} \cdot \Delta x_m = 210 + \frac{20-16}{11} \cdot 5 = \underline{211.8}$$

Analogt til median er nedre og øvre interkvartilklasser de første klassene der relativ kumulativ frekvens overstiger henholdsvis 0,25 og 0,75, dvs. klassene 205-210 og 215-220. Vi får:

$$Q_1 = 205 + \frac{40 \cdot 0.25 - 6}{10} \cdot 5 = 207, \quad Q_3 = 215 + \frac{40 \cdot 0.75 - 27}{9} \cdot 5 = 216.7$$

Interkvartilbredde: $Q_3 - Q_1 = 216.7 - 207 = \underline{9.7}$