ØVINGER 2017

Løsninger til oppgaver

Øving 1

2.1.

Frekvenstabell

For å lage en frekvenstabell må vi telle antall observasjoner av hvert antall henvendelser. Siden antall henvendelser på en gitt dag alltid ligger mellom 0 og 7 trenger vi en tabell med åtte rader. Videre lar vi tabellen ha tre kolonner. I den venstre kolonnen lister vi variabelen vi ønsker å tabulere frekvensen av, altså antall henvendelser. I den midterste kolonnen skriver vi frekvensen vi har observert hvert antall henvendelser med, altså hvor mange ganger hvert av tallene opptrer i tabellen i oppgaveteksten. I den høyre kolonnen skriver vi relativ frekvens, altså hvor stor andel av observasjonene hver observerte verdi utgjør. Vi finner de relative frekvensene ved å normalisere frekvensene i den midterste kolonnen. Siden det totalt er 100 observasjoner, normaliserer vi frekvensene ved å dele på 100.

Antall henvendelser	Frekvens	Relativ frekvens
0	9	0.09
1	20	0.20
2	26	0.26
3	18	0.18
4	11	0.11
5	8	0.08
6	5	0.05
7	3	0.03

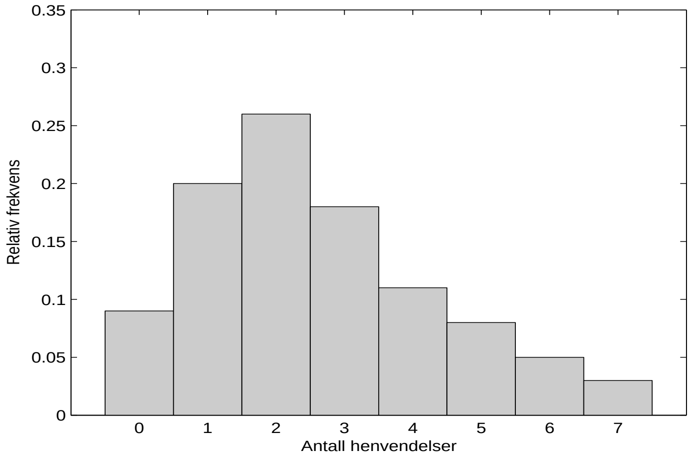
For å kontrollere utregningene kan vi sjekke at summen av frekvensene blir lik antall observasjoner, alstå 100, og at summen av de relative frekvensene blir 1.

Merk at en frekvenstabell ikke nødvendigvis trenger å ha med både frekvenser og relative frekvenser.

Histogram

Histogrammet skal være et søylediagram som visualiserer innholdet i frekvenstabellen. Søylene bør være brede nok til at de er helt inntil hverandre. I tillegg ønsker vi at det totale arealet av søylene skal være lik 1. Hvis hver rad i frekvenstabellen representeres av en rektangulær søyle med bredde 1 og høyde lik den relative frekvensen, da vil arealet av søylen være lik relativ frekvens, og sammenlagt vil arealet av alle søylene være lik summen av de relative frekvensene, som er 1.

Det kan være praktisk å sentrere hver søyle horisontalt på verdien den representerer, slik at søylen for null henvendelser strekker seg fra -0.5 til 0.5 på x-aksen, søylen for en henvendelse strekker seg fra 0.5 til 1.5, og så videre.



Histogrammet har en skjev form. De høyeste søylene finner vi rundt to henvendelser. Høyden avtar gradvis når man beveger seg til høyre, dvs. mot flere henvendelser. Histogrammet har ikke noen tilsvarende hale på venstre side, siden antall henvendelser alltid er større enn eller lik null. En slik skjev form er typisk for histogrammer av ikke-negative variable.

2.5. Vi bør unngå å bruke gjennomsnitt som sentralmål når vi studerer inntektsforhold. Dette fordi fordelingen av inntekt ofte er svært skjev. Gjennomsnittsinntekten er gjerne mange ganger større enn inntekten til en typisk arbeidstaker.

Medianinntekten er derimot et informativt sentralmål for inntekt, og er bedre skikket til å fortelle oss hvor inntektsnivået faktisk ligger. Sammenlignet med gjennomsnittet påvirkes medianen lite av den store skjevheten i utvalget, altså det at noen få arbeidstakere har svært høy inntekt.

Modus kan også brukes som sentralmål for inntekt, men fungerer best dersom observasjonene er gruppert i intervaller, slik at det blir mange nok observasjoner av hvert intervall. Modus forteller da hvilket intervall en typisk inntekt ligger innenfor.

2.7.
Median
For å finne medianen begynner vi med å sortere tallene i stigende rekkefølge:

$$2 \quad 3 \quad 4 \quad 7 \quad 8 \quad 12.$$

De to midterste tallene er 4 og 7, så medianen er $(4+7)/2 = 5.5$.
Gjennomsnitt

Gjennomsnittet av tallene er

$$\frac{2+3+4+7+8+12}{6} = \frac{36}{6} = 6.$$

Standardavvik
For å finne standardavviket regner vi først ut variansen,

$$s^2 = \frac{1}{6-1}\left[(2-6)^2 + (3-6)^2 + (4-6)^2 + (7-6)^2 + (8-6)^2 + (12-6)^2\right] =$$
$$= \frac{1}{5}\left[4^2 + 3^2 + 2^2 + 1^2 + 2^2 + 6^2\right] =$$
$$= \frac{1}{5} \cdot 70 = 14.$$

Standardavviket er da $s = \sqrt{14} = 3.74$.

2.8.
Median
Medianen er gjennomsnittet av de to midterste tallene,

$$\frac{1\,000\,004 + 1\,000\,007}{2} = 1\,000\,005.5.$$

Gjennomsnitt
Summen av tallene er 6 000 036, og gjennomsnittet er summen delt på 6,

$$\frac{6\,000\,036}{6} = 1\,000\,006.$$

Standardavvik
Vi skal regne ut standardavviket, og ser først på de individuelle kvadratavvikene

$$(1\,000\,002 - 1\,000\,006)^2 = (2-6)^2$$
$$(1\,000\,003 - 1\,000\,006)^2 = (3-6)^2$$
$$(1\,000\,004 - 1\,000\,006)^2 = (4-6)^2$$
$$(1\,000\,007 - 1\,000\,006)^2 = (7-6)^2$$
$$(1\,000\,008 - 1\,000\,006)^2 = (8-6)^2$$
$$(1\,000\,012 - 1\,000\,006)^2 = (12-6)^2.$$

Dette er de samme kvadratavvikene som i oppgave 2.7, slik at variansen også i dette tilfellet blir

$$\begin{aligned} s^2 &= \frac{1}{6-1}\left[4^2+3^2+2^2+1^2+2^2+6^2\right] \\ &= \frac{1}{5}\cdot 70 = 14, \end{aligned}$$

og standardavviket blir $\sqrt{14} = 3.74$.

Tallene i denne oppgaven er forskjøvet oppover med 1 000 000 i forhold til tallene i oppgave 2.7. Dermed er medianen og gjennomsnittet forskjøvet tilsvarende i forhold til sentralmålene i oppgave 2.7. Spredningen til tallene er imidlertid uendret, slik at standardavviket er det samme som før.

2.9

Median

Her er de to midterste tallene 400 og 700, så medianen er

$$\frac{400+700}{2} = 550.$$

Gjennomsnitt

Gjennomsnittet blir

$$\frac{200+300+400+700+800+1200}{6} = \frac{3600}{6} = 600.$$

Standardavvik

Variansen er

$$\begin{aligned} s^2 = \frac{1}{6-1}\big[&(200-600)^2+(300-600)^2+(400-600)^2+ \\ &+(700-600)^2+(800-600)^2+(1200-600)^2\big] = \\ &= \frac{1}{5}\left[400^2+300^2+200^2+100^2+200^2+400^2\right] = \\ &= \frac{1}{5}\cdot 700\,000 = 140\,000, \end{aligned}$$

slik at standardavviket blir $\sqrt{140\,000} = 374.17$.

Tallene som er gitt i denne oppgaven er de samme som i oppgave 2.7, men multiplisert med 100. Medianen og gjennomsnittet blir her 100 ganger tilsvarende sentralmål i oppgave 2.7. Til forskjell fra translasjon, som i oppgave 2.8, vil multiplikasjon endre spredningen til utvalget, så her finner vi også at standardavviket er 100 ganger større enn standardavviket i oppgave 2.7.

2.22 (19). Definisjon 2.3 gir oss følgende formel for varians:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$

og vi skal vise at dette er det samme som

$$\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right).$$

Vi følger anbefalingen i oppgaveteksten, og bruker nederste linje i regel 2.2 til å utvide summen,

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2\right).$$

Videre merker vi oss at siden $\bar{x} = 1/n\sum_{i=1}^{n} x_i$, så kan vi bytte ut $\sum_{i=1}^{n} x_i$ med $n\bar{x}$. Dermed er

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - 2\bar{x}(n\bar{x}) + n\bar{x}^2\right) = \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2\right) = \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right). \end{aligned}$$