Sentralmål

La $x_1, x_2, \ldots, x_n$ vere ei mengde med n observasjonar, eit utval av storleik n.

Utvalsgjennomsnitt $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

Legg merke til $\sum_{i=1}^{n} x_i = n\bar{x}$. Kva blir $\sum_{i=1}^{n} (x_i - \bar{x})$

Gjennomsnittet kan og oppfattast som eit balansepunkt for dataane.

Gå ut i frå at kvar observasjon veg 1 kg.

Moment er $\overset{1\,kg}{\text{kraft}} \cdot \overset{x_i - \bar{x}}{\text{arm}}$. $\quad \sum_{i=1}^{n} (x_i - \bar{x}) = 0$

eller $\sum_i (x_i - \bar{x})_+ = -\sum_i (x_i - \bar{x})_-$

Utvalsmedian

La oss tenkje oss at vi ordnar dataane etter algebraisk storleik

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$$

Observasjonen som ligg i midten blir kalla utvalsmedianen.

Eks. Treningstider for eit fotball-lag (1 mnd).

7 9 12 13 14.15 16 18 19 20 21 21

$\tilde{x} = 15$

Men laget har og ein reserve som trenar 20 timar i månaden.

Då blir $\tilde{X} = \frac{15+16}{2} = 15.5$

Difor er utvalsmedianen definert som

$$\tilde{X} = \begin{cases} X_{(\frac{n+1}{2})}, & \text{dersom } n \text{ er odde} \\ \frac{1}{2}\left[X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}\right], & \text{dersom } n \text{ er jamn.} \end{cases}$$

Medianen er meir robust mot avvikande observasjonar enn gjennomsnittet. Den representerer den verdien som har 50% av observasjonane mindre enn seg. Og blir og kalla Q_2 [andre utvals kvartil].

Første kvartil Q_1 representerer verdien som har 25% av observasjonane mindre enn seg.

Tilsvarande Q_3 (75% mindre).

Ein vanleg variant er å la

Q_1 ~ vere verdien i posisjon $\frac{n+1}{4} = 3.25$

Q_2 ~ vere verdien i — " — $\frac{n+1}{2} = 6.5$

Q_3 ~ vere verdien i — " — $3 \cdot \left(\frac{n+1}{4}\right) = 9.75$.

$Q_1 = \frac{3}{4} \cdot 12 + \frac{1}{4} \cdot 13 = 12.25$, $Q_2 = 15 \cdot \frac{1}{2} + 16 \cdot \frac{1}{2} = 15.5$

$Q_3 = 19 \cdot \frac{1}{4} + 20 \cdot \frac{3}{4} = \underline{19.75}$

Boksplott

Rekn ut Q_2 = medianen

Rekn ut $Q_3 - Q_1 = h$ = høgda på boksen

Observasjonar lenger borte enn $\frac{3}{2}h$ frå nedre og øvre side på boksen blir kalla uteliggarar. Det går ein strek frå nedre og øvre side på boksen til den ~~største~~ minste og største observasjonen som ikkje er uteliggar.

Spreiingsmål

Utvala: 17, 18, 19 og 10, 18, 26

er begge svært forskjellige, men har same utvalsgjennomsnitt.

Som mål på spreiing brukas ein ofte utvals-variansen definert som Def (1.3)

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

eller utvals standardavviket

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

som har same eining som observasjonane.

Variasjonsbredda = $x_{max} - x_{min}$.

Typer av data

1. Data frå planlagte eksperiment
2. Observerte data.

<u>TMA 4240</u>

Statistikk er vitenskapen om korleis vi skal

1. Samle inn
2. Presentere
3. Trekkje informasjon ut frå data.

I kurset skal vi lære om:

Sannsynsrekning
Statistiske modellar
Statistisk Inferens (Trekkje informasjon ut frå data).

<u>Kap 2. Sannsyn</u>

Eks. Chevalier de Méré's Problemstilling (Fermat og Pascal)

To spelarar A og B spelte eit rettferdig spel

$P(A \text{ vinn}) = \frac{1}{2}$ $\quad P(B \text{ vinn}) = \frac{1}{2}$

Begge spelararane sette inn ein like stor innsats og den 1. til å vinne 6 spel fekk potten. Spelet blei avbrote då A hadde vunne 5 gonger og B hadde vunne 3. Korleis skal dei dele potten?