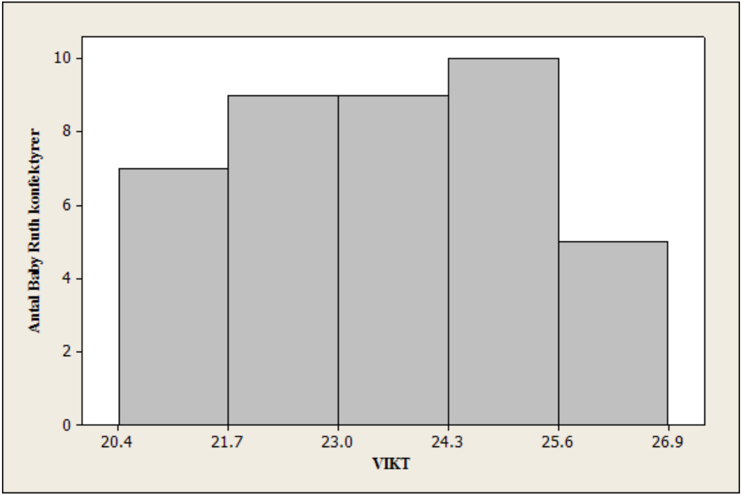
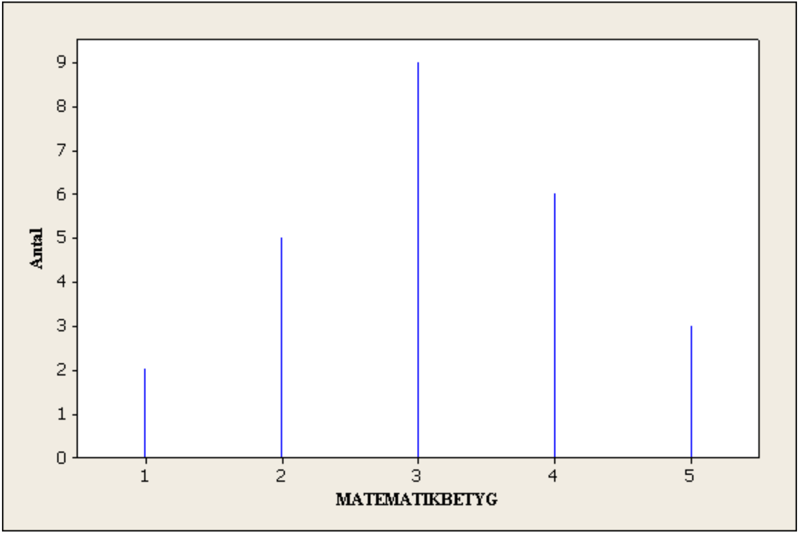
F6
Att dela upp ett datamaterial i delar.
Median, kvartiler och percentiler

Intuitiv idé: dela fördelningen för ett datamaterial i två lika delar!




Antal
9
8
7
6
5
4
3
2
1
0
1
2
3
4
5
MATEMATIKBETYG

Olika indelningar

Percentilerna $P_{01}, P_{02}, \ldots, P_{99}$ delar ett datamaterial i hundra lika stora delar.

Percentilerna P_{25}, P_{50}, P_{75} delar fördelningen i fyra lika stora delar och brukar betecknas q_1, q_2, q_3.

Ordnade stickprovet

Antag att vi har ett stickprov med $n = 4$ observationer, t ex

$$x_1 = 9.34 \quad x_2 = 4.22 \quad x_3 = 0.56 \quad x_4 = 22.33.$$

Det *ordnade stickprovet* åstadkommer vi genom att ordna observationerna i storleksordning från den minsta till den största. Då har vi

$$x_3 = 0.56 \quad x_2 = 4.22 \quad x_1 = 9.34 \quad x_4 = 22.33.$$

Vi döper men minsta till y_1, den näst minsta till y_2, o s v ända till den största y_4.
I det här datamaterialet har vi alltså att

$$y_1 = x_3 \quad y_2 = x_2 \quad y_3 = x_1 \quad y_4 = x_4.$$

Approximativ lösning för att hitta percentiler

Om vi har ett datamaterial $x_1, \ldots, x_n$ med n observationer. Låt p vara ett tal mellan 0 och 1. Vi önskar ett tal sådant att antalet np observerade värden är mindre och antalet $n(1-p)$ är större. (Andelen p mindre och andelen $1-p$ större).

Beräkna talet $(n+1)p$. Om det är ett heltal t ex r, tag då det observerade värde som har detta ordningsnummer y_r som $(100p)$ percentilen.

Om $(n+1)p$ inte är ett heltal, utan kan skrivas

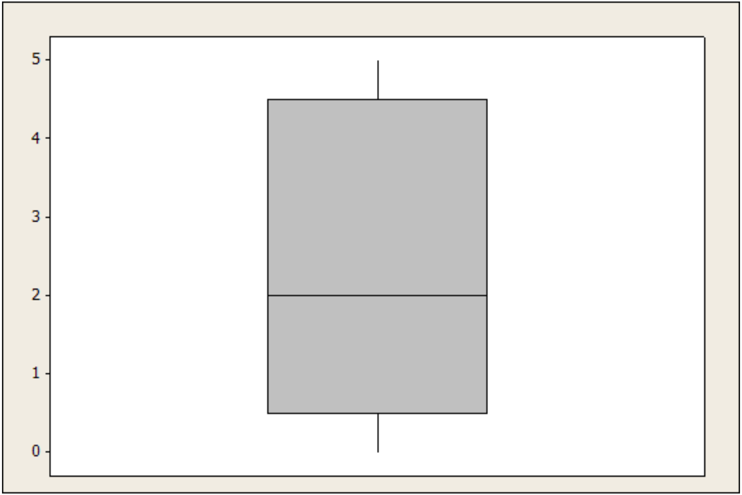
$$(n+1)p = r + \frac{a}{b},$$

så hamnar $(100p)$ percentilen mellan det r:te och det $(r+1)$:ta ordnade värdena, d v s y_r och y_{r+1}.
Exakt var bestäms av en formel som använder sig av resultatet $r + a/b$. $(100p)$ percentilen ges som

$$\left(1 - \frac{a}{b}\right) y_r + \left(\frac{a}{b}\right) y_{r+1}.$$

Lådagram (Boxplot)

$$\text{Kvartilavståndet} = q_3 - q_1.$$



Vägt medelvärde

Antag att vi har ett datamaterial $x_1, \ldots, x_n$ med n observationer. Låt $w_1, \ldots, w_n$ vara tal som anger den relativa vikt vi tillmäter värdena $x_1, \ldots, x_n$ när vi bedömer helhetsresultatet. Då anger formeln

$$\overline{x}_w = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

det *vägda medelvärdet* av observationerna $x_1, \ldots, x_n$.