Beskrivande statistik

Grafiska metoder

Innehåll

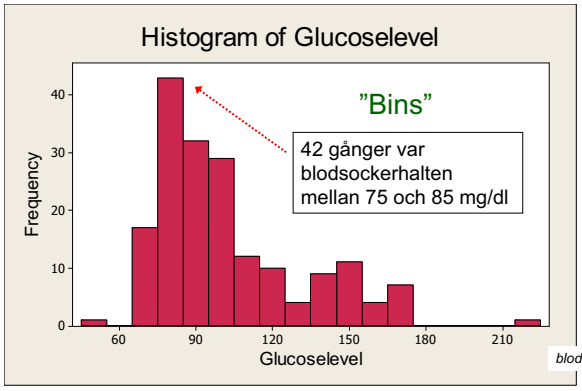
- Histogram
- Spridningsdiagram (Scatterplot)
- Lådagram (Boxplot)
- Missbruk av diagram

Histogram - Data

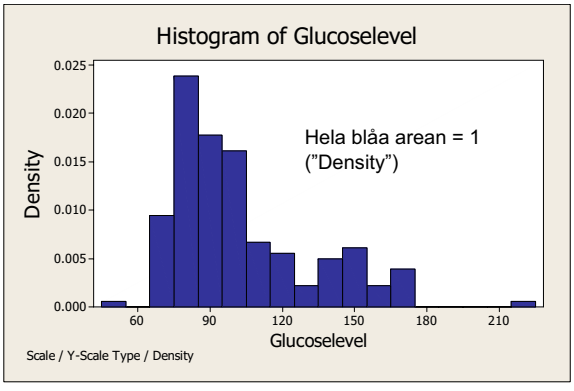
Glukosehalt, 180 prov:

85, 87, 150, 100, 100, 90, 70, 72, 75, 70, 85, 143, 100, 121, 92, 66, 70, 69, 75, 80, 140, 92, 130, 83, 70, 68, 67, 75, 83, 149, 95, 130, 80, 68, 85, 75, 73, 78, 140, 90, 124, 86, 69, 70, 75, 77, 110, 165, 110, 150, 110, 115, 80, 75, 75, 98, 172, 110, 145, 110, 95, 52, 80, 96, 110, 168, 110, 145, 110, 80, 80, 75, 89, 95, 170, 110, 145, 120, 89, 72, 79, 75, 95, 220, 100, 149, 100, 110, 80, 85, 80, 90, 165, 103, 135, 95, 77, 76, 85, 80, 88, 155, 103, 120, 85, 79, 78, 82, 75, 85, 150, 103, 135, 90, 75, 85, 78, 75, 88, 150, 95, 130, 90, 70, 76, 89, 82, 95, 145, 100, 133, 90, 77, 89, 79, 80, 90, 165, 103, 135, 95, 77, 86, 80, 85, 100, 160, 120, 140, 100, 90, 79, 92, 70, 100, 165, 120, 140, 100, 120, 86, 71, 95, 100, 155, 120, 139, 100, 89, 86, 78, 78, 110, 158, 122, 145, 108, 95, 95, 78

Histogram



blodsocker.MPJ



File: Blodsugar.txt
Graph / Histogram / Simple
Graph variable = *Glucoselevel*
Scale / Y-Scale Type: Density

Normalisering

$$FR_i = \frac{FA_i}{180}$$

$$H_i = \frac{FR_i}{B}$$

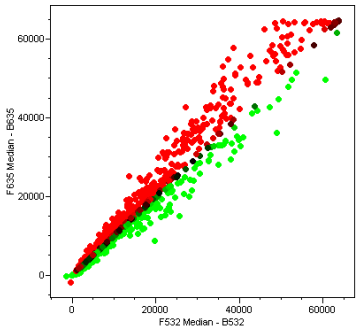
$$A_i = H_i \cdot B$$

Arean under kurvan = 1

$$\sum A_i = \sum H_i \cdot B = \sum \frac{FR_i}{B} \cdot B = \sum FR_i = 1$$

Bin	Bredd	abs. F	rel. F	Höjd	Area
50	10	0	0.000	0.000	0.000
60	10	1	0.006	0.001	0.006
70	10	13	0.072	0.007	0.072
80	10	43	0.239	0.024	0.239
90	10	33	0.183	0.018	0.183
100	10	28	0.156	0.016	0.156
110	10	16	0.089	0.009	0.089
120	10	7	0.039	0.004	0.039
130	10	6	0.033	0.003	0.033
140	10	9	0.050	0.005	0.050
150	10	12	0.067	0.007	0.067
160	10	4	0.022	0.002	0.022
170	10	6	0.033	0.003	0.033
180	10	1	0.006	0.001	0.006
190	10	0	0.000	0.000	0.000
200	10	0	0.000	0.000	0.000
210	10	0	0.000	0.000	0.000
220	10	1	0.006	0.001	0.006
		180	1	0.1	1

Scatterplot



Scatterplot Data

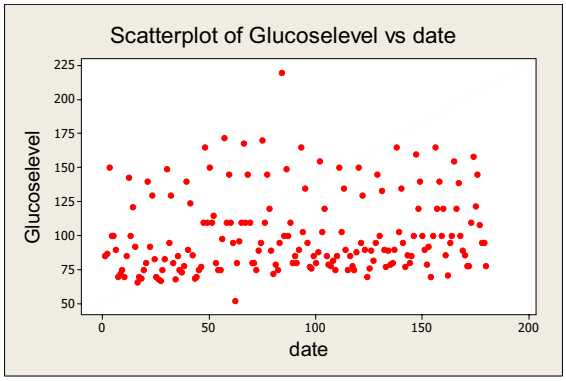
(minst) två kolumner behövs (x och y)

Worksheet 3 ***

	C1	C2	C3
	date	Glucoselevel	
1	1	85	
2	2	87	
3	3	150	
4	4	100	
5	5	100	
6	6	90	
7	7	70	
8	8	72	
9	9	75	
10	10	70	
11	11	85	
12	12	143	
13	13	100	
14	14	121	
15	15	92	

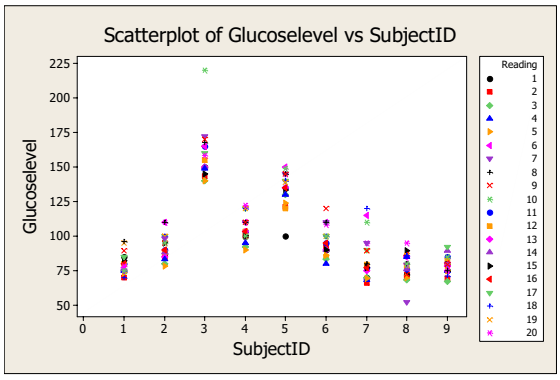
Glucose_vs_date.MTW

Scatterplot



Scatterplot – grupperade data

Reading	SubjectID	Glucoselevel
1	1	85
2	1	70
3	1	75
4	1	75
5	1	73
6	1	77
7	1	75
1	2	87
2	2	85
3	2	80
4	2	83
5	2	78



blodsocker.MPJ

Graph / Scatterplot – With groups
Y-variables = *Glucoselevel*
X-variables = *SubjectID*
Categorical variables for grouping: *Reading*

Lådagram (Boxplot)

Lådagram Data

Glukose.txt

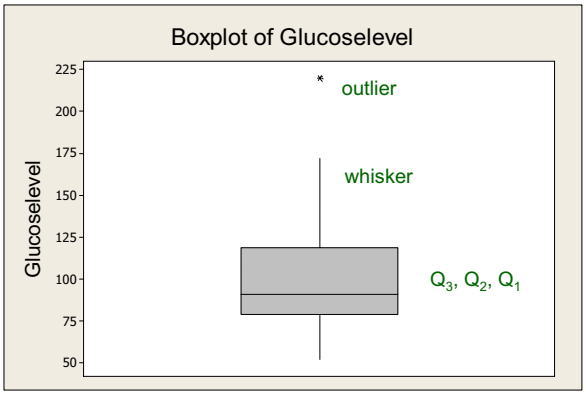
Worksheet 2 ***

	C1	C2	C3
	Glucoselevel		
1	85		
2	87		
3	150		
4	100		
5	100		
6	90		
7	70		
8	72		
9	75		
10	70		
11	85		
12	143		
13	100		
14	121		
15	92		

Lådagram

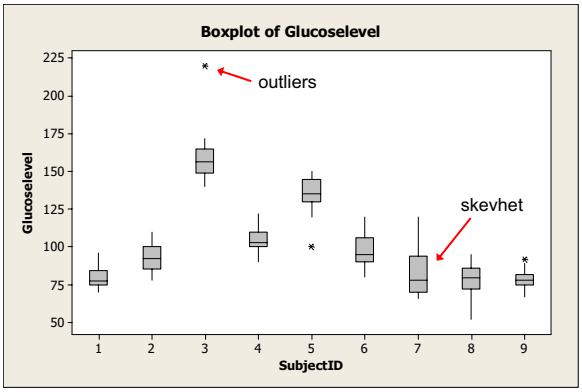
- Datamaterialet beskrivs med några få tal
  - Median
  - Undre och övre kvartil (Q_1, Q_3)
  - Min och max (whisker-längd) (R_l, R_u)
  - Outliers

Lådagram



Lådagram – grupperade data

Reading	SubjectID	Glucoselevel
1	1	85
2	1	70
3	1	75
4	1	75
5	1	73
6	1	77
7	1	75
1	2	87
2	2	85
3	2	80
4	2	83
5	2	78



blodsocker.MPJ

Graph / Boxplot – With groups
Graph variables = Glucoselevel
Categorical variables for grouping: *SubjectID*

Flera definitioner av kvartilerna

The following table summarizes a number of common methods for computing the position of the first and third quartiles from a sample size n (P. Stikker, pers. comm., Jan. 24, 2005). In the table, $[x]$ denotes the nearest integer function.

method	1st quartile	1st quartile	3rd quartile	3rd quartile
	n odd	n even	n odd	n even
Minitab	$\frac{n+1}{4}$	$\frac{n+1}{4}$	$\frac{3n+3}{4}$	$\frac{3n+3}{4}$
Tukey (Hoaglin et al. 1983)	$\frac{n+3}{4}$	$\frac{n+2}{4}$	$\frac{3n+1}{4}$	$\frac{3n+2}{4}$
Moore and McCabe (2002)	$\frac{n+1}{4}$	$\frac{n+2}{4}$	$\frac{3n+3}{4}$	$\frac{3n+2}{4}$
Mendenhall and Sincich (1995)	$\left\lfloor\frac{n+1}{4}\right\rfloor$	$\left\lfloor\frac{n+1}{4}\right\rfloor$	$\left\lfloor\frac{3n+3}{4}\right\rfloor$	$\left\lfloor\frac{3n+3}{4}\right\rfloor$
Freund and Perles (1987)	$\frac{n+3}{4}$	$\frac{n+3}{4}$	$\frac{3n+1}{4}$	$\frac{3n+1}{4}$

From: *Wolfram Research*

Q1, Q3 - Minitab

$n = 20 \quad (20\ värden)$

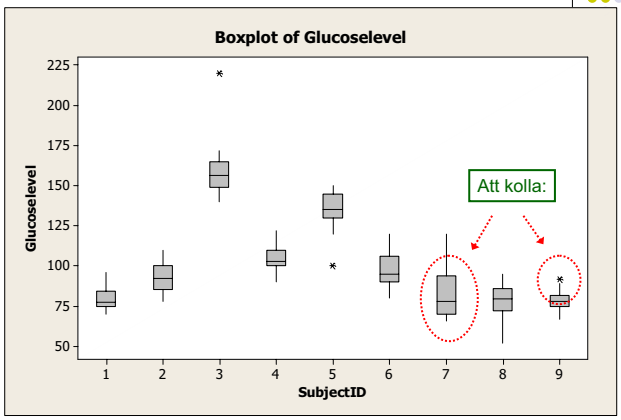
$n_1 = \frac{1}{4}(n+1) = \frac{1}{4}\cdot 21 = 5.25$ ← närmare 5

$Q_1 = 0.25\cdot v_6 + 0.75\cdot v_5$

$n_3 = \frac{3}{4}(n+1) = \frac{3}{4}\cdot 21 = 15.75$ ← närmare 16

$Q_3 = 0.75\cdot v_{16} + 0.25\cdot v_{15}$

Lådagram



Outlier (person nio)?

67 69 71 75 75 75 75 75 78 78 | 78 79 79 80 80 82 85 85 89 92

$n = 20 \quad (20\ värden)$

$n_1 = \frac{1}{4}(n+1) = \frac{1}{4}\cdot 21 = 5.25$

$Q_1 = 0.25\cdot v_6 + 0.75\cdot v_5 = 0.25\cdot 75 + 0.75\cdot 75 = 75$

$n_3 = \frac{3}{4}(n+1) = \frac{3}{4}\cdot 21 = 15.75$

$Q_3 = 0.75\cdot v_{16} + 0.25\cdot v_{15} = 0.75\cdot 82 + 0.25\cdot 80 = 81.5$

$\Delta Q_{1,3} = Q_3 - Q_1 = 81.5 - 75 = 6.5$ *interkvartilavståndet*

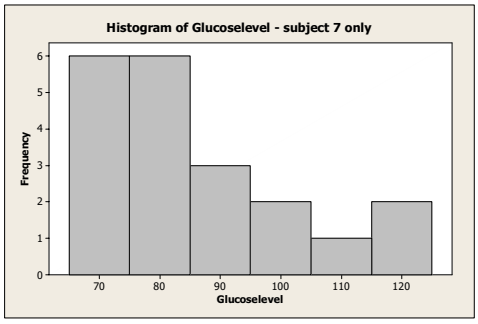
$R_l = Q_1 - 1.5\cdot\Delta Q_{1,3} = 75 - 1.5\cdot 6.5 = 65.25$ *inga outliers*

$R_u = Q_3 + 1.5\cdot\Delta Q_{1,3} = 81.5 + 1.5\cdot 6.5 = 91.25$ *en outlier*

Whiskers tecknas till högsta värdet som inte är någon outlier → 67; 89
Jämför med den information man får när man sätter musen på en box!!

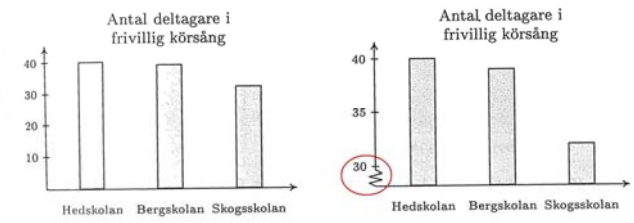
Skevhet (person sju) ?

66 68 69 70 70 70 75 77 77 77 | 79 80 89 89 90 95 95 110 115 120

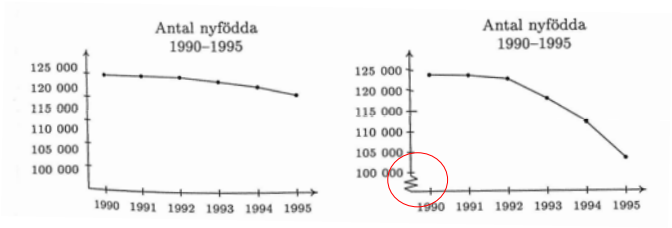


Missbruk av diagram

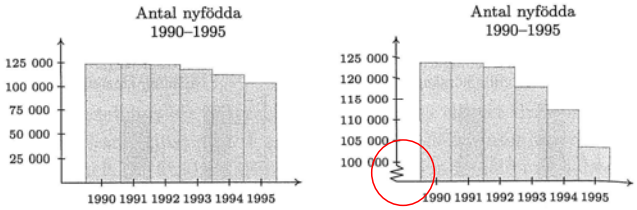
Missbruk av diagram



Missbruk av diagram



Missbruk av diagram



Missbruk av diagram

Missbruk av diagram

Under 1999 spelade det svenska folket uppskattningsvis för drygt 31 miljarder kronor