

Stolpdiagram

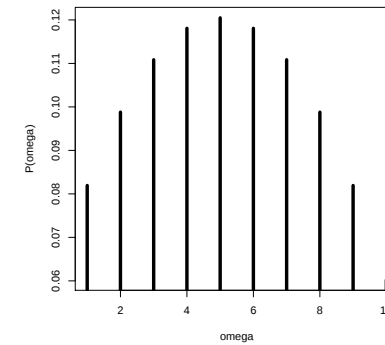
Sannolikhet och statistik

Beskrivande statistik

HT 2008

Uwe.Menzel@math.uu.se

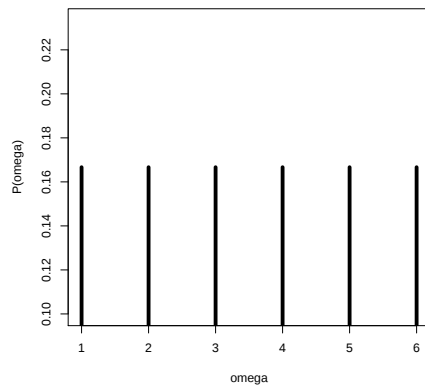
<http://www.math.uu.se/~uwe/>



Figur: Diskret fördelning

Stolpdiagram

x	1	2	3	4	5	6
$P(x)$	1/6	1/6	1/6	1/6	1/6	1/6



Kontinuerligt utfallsrum

```

1.79504290  0.30389744 -0.27868565  1.53528391  4.27525921
1.99335216 -0.11332719 -2.76141900  4.79268510  3.77099826
4.46985339 -0.31218207 -0.80597373  4.19811668 -2.37991388
4.52382691 -2.80171944 -1.11885446  1.80338641  3.40319091
0.82440971  2.10905489  2.12794223  2.64901903 -1.00863476
2.28696205  1.14821861 -0.03020109  2.54841542  0.10218402
2.31106440 -2.47308445  1.90527120 -0.62079369 -1.49561715
-0.05906434  2.86264583 -1.02218255 -1.12347129  3.85684637
2.90077282 -0.71179195

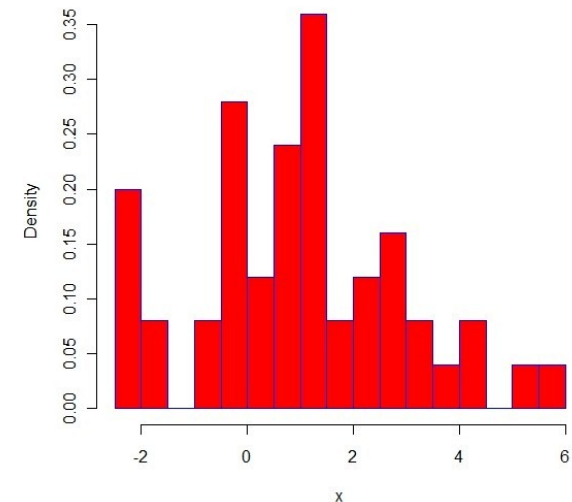
```

Kontinuerligt utfallsrum

Klassgränser	Klassmitt	Absolut frekvens	Relativ frekvens
g_i g_{i+1}	y_i	f_i	p_i
2.30-2.32	2.31	2	0.0032
2.32-2.34	2.33	7	0.0111
2.34-2.36	2.35	9	0.0143
2.36-2.38	2.37	22	0.0349
2.38-2.40	2.39	14	0.0222
2.40-2.42	2.41	30	0.0476
2.42-2.44	2.43	28	0.0444
2.44-2.46	2.45	42	0.0667
2.46-2.48	2.47	45	0.0714
2.48-2.50	2.49	81	0.1286
2.50-2.52	2.51	90	0.1429
2.52-2.54	2.53	59	0.0937
2.54-2.56	2.55	63	0.1000
2.56-2.58	2.57	50	0.0794
2.58-2.60	2.59	31	0.0492
2.60-2.62	2.61	25	0.0397
2.62-2.64	2.63	10	0.0159
2.64-2.66	2.65	10	0.0159
2.66-2.68	2.67	7	0.0111
2.68-2.70	2.69	5	0.0080

Figur: Frekvenstabell för data från $\mathbb{R}$

Histogram



```
hist(x, breaks = 20,freq=FALSE,col=2,border=4)
```

Medelvärde, Median, Kvartiler

Medelvärde

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Median

1.192 2.125 2.781 3.561 **3.910** 3.967 4.781 5.125 6.410

Undre kvartil

1.192 2.125 **2.781** 3.561 **3.910**

Övre kvartil

3.910 3.967 **4.781** 5.125 6.410

Sammanfattning

1.192 2.125 **2.781** 3.561 **3.910** 3.967 **4.781** 5.125 6.410

Stickprovsvarians, -standardavvikelse

Stickprovsvariansen

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

Stickprovets standardavvikelse

$$s = \sqrt{\frac{Q}{n-1}}$$

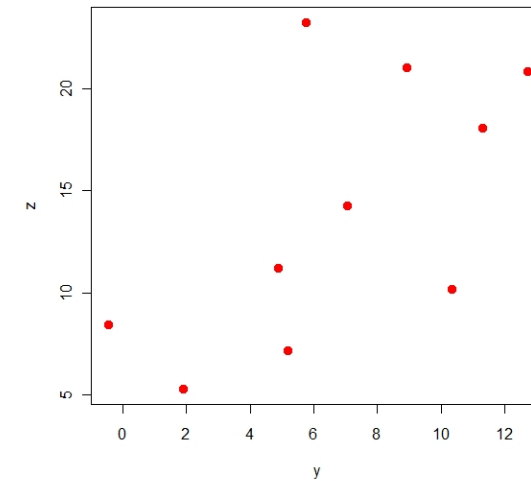
Kvartilavståndet

$$\Delta = \tilde{x}_{\text{ovre}} - \tilde{x}_{\text{undre}}$$

Beroendemått

- ▶ parade observationer (längd och vikt för varje insekt)
- ▶ samband ?

-0.456	1.910	4.875	10.326	11.304	7.060	8.920	5.750
8.42	5.29	11.24	10.17	18.08	14.28	21.03	23.26



Kovarians och Korrelation

$$c_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

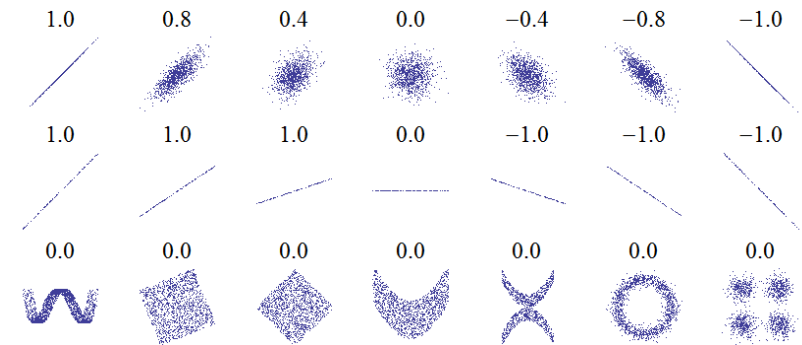
$$r = \frac{c_{xy}}{s_x s_y}$$

$$-1 \leq r \leq 1$$

- ▶ $r = +1$ ideal rät linje med positiv lutning
- ▶ $r = -1$ ideal rät linje med negativ lutning

Beroendemått

Korrelationskoefficienten mäter linjär sammanhang !



Källa: Wikipedia

Räkneregler

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = S_{xy} = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}$$

Rangkorrelation

Elev	A	B	C	D	E	F	G	H	I	J
Matte	7	4	3	10	6	2	9	8	1	5
Musik	5	7	3	10	1	9	6	2	8	4

Kan användas för data som inte kan kvantifieras men kan “sorteras” (rankas) på något sätt

Annat exempel: smak och lukt av vinsorter