

## Lektion 2

2006-11-01  
Kapitel 2 - 3

---

---

---

---

---

---

---

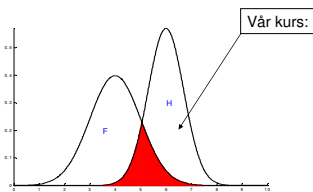
---

## Variation

- Deterministisk
- Slumpmässig
- För en produkt:

$$\left( \text{Kvalitet} \sim \frac{1}{\text{variationen}} \right)$$

Villkor:  
 $F < H$



---

---

---

---

---

---

---

---

## Modeller av data

- "Alla modeller är felaktiga men vissa är användbara." (George Box 1977)
- Använd rätt modell:
  - Svara på intressanta frågor.
  - Gäller under de förutsättningar den var utvecklad!
  - T.ex. Newton – Einstein.

---

---

---

---

---

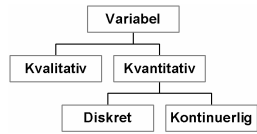
---

---

---

## Variabler och data

- Variabel beskriver en intressant egenskap.
  - Antal (st)
  - Vikt (kg)
  - Funktion (Ja/Nej)



- Olika metoder för olika typer av data.

---

---

---

---

---

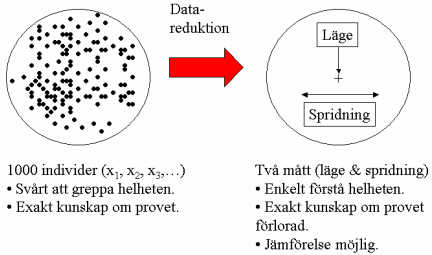
---

---

---

## Datareduktion – en modell

### Varför använda mått?



---

---

---

---

---

---

---

---

## Numerisk summering av data

- Summa
- Medelvärde
- Median och percentiler
- Variationsvidd (Range): Max – Min
- Varians och standardavvikelse

---

---

---

---

---

---

---

---

## Grafiska metoder

Table 2-1 Days to Pay Employee Health Insurance Claims

Claim	Days	Claim	Days	Claim	Days	Claim	Days
1	48	11	35	21	37	31	16
2	41	12	34	22	43	32	22
3	35	13	36	23	17	33	33
4	36	14	42	24	26	34	30
5	37	15	43	25	28	35	24
6	26	16	36	26	27	36	23
7	36	17	56	27	45	37	22
8	46	18	32	28	33	38	30
9	35	19	46	29	22	39	31
10	47	20	30	30	27	40	17

Stem-and-Leaf Display: Days  
 Stem-and-leaf of Days  
 N = 40  
 Leaf Unit = 1.0

```

3 | 1 6 7 7
8 | 2 2 2 2 3 4
13 | 2 6 6 7 7 8
(8) | 3 0 0 0 1 2 3 3 4
19 | 3 5 5 5 6 6 6 6 7 7
10 | 4 0 2 2 2 4
6 | 4 5 6 6 7 8
1 | 5
1 | 5 6
    
```

Figure 2-1 Stem-and-leaf plot for the health insurance claim data.

Table 2-4 Hole Diameters (in mm) in Wing Leading Edge Ribs

120.5	120.4	120.7
120.9	120.2	121.1
120.3	120.1	120.9
121.3	120.5	120.8

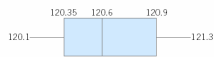
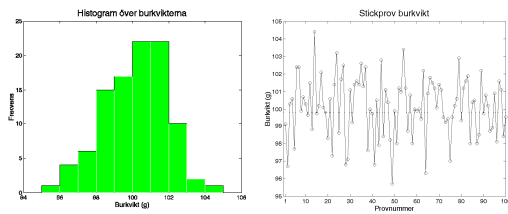


Figure 2-7 Box plot for the aircraft wing leading edge hole diameter data in Table 2-4.

## Histogram och tidsplot



Rita ALLTID dessa två diagram!

## Stokastiska variabler

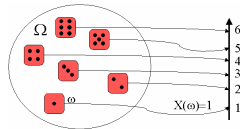
- Kontinuerliga eller diskreta

Väntevärde:

$$\mu = \begin{cases} \int_{-\infty}^{\infty} xf(x)dx \\ \sum_{i=1}^{\infty} xp(x_i) \end{cases}$$

Varians:

$$\sigma^2 = \begin{cases} \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \\ \sum_{i=1}^{\infty} (x_i-\mu)^2 p(x_i) \end{cases}$$



## Hypergeometrisk

### Definition

The hypergeometric probability distribution is

$$p(x) = \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}} \quad x = 0, 1, 2, \dots, \min(n, D) \quad (2-8)$$

The mean and variance of the distribution are

$$\mu = \frac{nD}{N} \quad (2-9)$$

and

$$\sigma^2 = \frac{nD}{N} \left(1 - \frac{D}{N}\right) \frac{N-n}{N-1} \quad (2-10)$$

---

---

---

---

---

---

---

---

## Binomialfördelningen

### Definition

The binomial distribution with parameters  $n \geq 0$  and  $0 < p < 1$  is

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \dots, n \quad (2-11)$$

The mean and variance of the binomial distribution are

$$\mu = np \quad (2-12)$$

and

$$\sigma^2 = np(1-p) \quad (2-13)$$

---

---

---

---

---

---

---

---

## Poissonfördelningen

### Definition

The Poisson distribution is

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!} \quad x = 0, 1, \dots \quad (2-15)$$

where the parameter  $\lambda > 0$ . The mean and variance of the Poisson distribution are

$$\mu = \lambda \quad (2-16)$$

and

$$\sigma^2 = \lambda \quad (2-17)$$

---

---

---

---

---

---

---

---

**Normal-fördelningen**

**Definition**

The normal distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty \quad (2-21)$$

The mean of the normal distribution is  $\mu$  ( $-\infty < \mu < \infty$ ) and the variance is  $\sigma^2 > 0$ .

Symmetrisk klockformad

medelvärde  
median  
typvärde

---

---

---

---

---

---

---

---

---

---

**Centrala gränsvärdesatsen**

---

---

---

---

---

---

---

---

---

---

**Övriga fördelningar**

- Exponentialfördelningen
- Gamma
- Weibull
- Lognormal

Hitta rätt (användbar) modell!

*(Det finns oändligt många fördelningar och ganska många är namngivna.)*

---

---

---

---

---

---

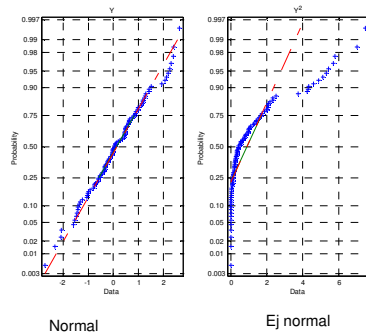
---

---

---

---

## Probability plots



Visuellt test

Matlab:  
normplot(x)  
Probplot  
wblplot

---

---

---

---

---

---

---

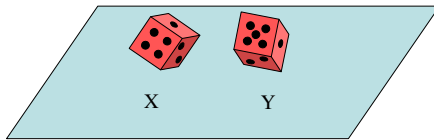
---

---

---

## Oberoende

- Om X ej ger information om Y och om Y ej ger information om X, så är X och Y *oberoende* av varandra.



- Processindustri, batcher, verktygsslitage,...

---

---

---

---

---

---

---

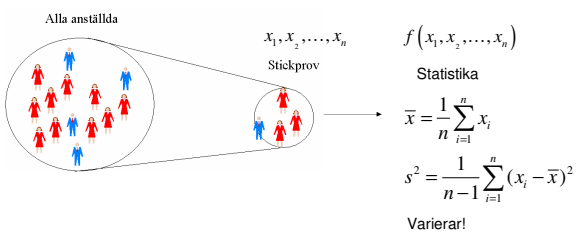
---

---

---

## Inferens

- Modellerar helheten m.h.a ett stickprov!




---

---

---

---

---

---

---

---

---

---

## Stickprov från normalfördelningen

- Antag oberoende och normalfördelade s.v.

$$X_1, \dots, X_n \sim N(\mu, \sigma^2)$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad (\text{Medelvärdet varierar mindre än de enskilda värdena})$$

$$\frac{\bar{X} - \mu}{\sigma} \sim N\left(\mu - \mu, \frac{\sigma^2}{\sigma^2}\right) = N(0,1)$$

---

---

---

---

---

---

---

---

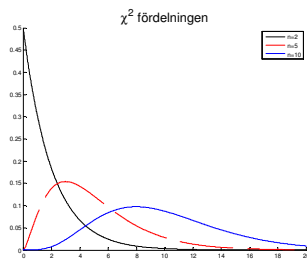
## Chi2-fördelningen

$$X_1, \dots, X_n \sim N(0,1)$$

$$Y = X_1^2 + \dots + X_n^2$$

$$Y \sim \chi_n^2, \text{ n frihetsgrader}$$

$$Y = \frac{(n-1)S^2}{\sigma^2} \sim N(0,1)$$



---

---

---

---

---

---

---

---

## t- och F-fördelningen

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

$$\frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}} \sim F_{n_1-1, n_2-1}$$

---

---

---

---

---

---

---

---

## Binomial och Poisson

- Binomial:  $X \sim Bin(n, p)$

$$\begin{cases} \mu_{\bar{x}} = p \\ \sigma_{\bar{x}}^2 = \frac{p(1-p)}{n} \end{cases}$$

- Poisson:  $X \sim Poi(\lambda)$

$$\begin{cases} \mu_{\bar{x}} = \lambda \\ \sigma_{\bar{x}}^2 = \frac{\lambda}{n} \end{cases}$$

---

---

---

---

---

---

---

---