Lektion 1
part 2

2007-10-31

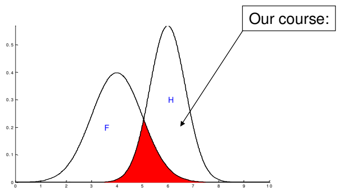
Chapters 2 - 3

Variation

- Deterministic
- Random
- For a product

$$\left(Quality \sim \frac{1}{\text{variation}} \right)$$

$$Villkor: \quad F < H$$

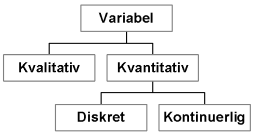


Models of data

- ”All models are wrong but some are useful” (George Box 1977)
- Use the right model:
  - Answer interesting questions
  - Valid under the conditions it was developed!
  - E.g. Newton – Einstein.

Variables and data

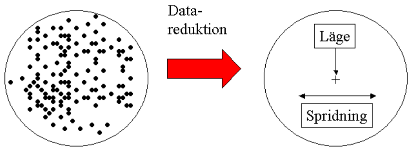
- Variabel describes an interesting characteristics
  - Number of..
  - Weight (kg)
  - Function (yes/No)



- Different methods for different types of data.

Data reduction – a model

Varför använda mått?



1000 individer (x_1, x_2, x_3,…)
• Svårt att greppa helheten.
• Exakt kunskap om provet.

Två mått (läge & spridning)
• Enkelt förstå helheten.
• Exakt kunskap om provet förlorad.
• Jämförelse möjlig.

Numerical data reduction

- Sum
- Average
- Median and percentiles
- Range (variationsvidd): Max – Min
- Variance and standard deviation

Grafical methods

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  1 677
   8  2 22234
  13  2 66778
 (8)  3 00012334
  19  3 555666677
  10  4 02224
   6  4 56678
   1  5
   1  5 6
```

Figure 2-1 Stem-and-left 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

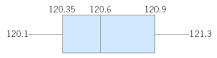
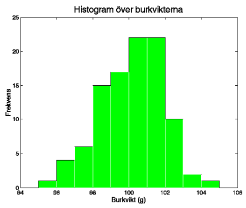
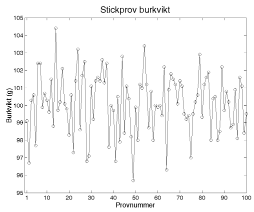


Figure 2-7 Box plot for the aircraft wing leading edge hole diameter data in Table 2-4.

Histogram and time plots

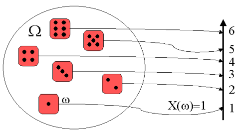




Always draw these diagrams!

Stokastic variables

- Continous or discrete



Expected value:

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

Variance:

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

Hypergeometric

Definition

The **hypergeometric probability distribution** is

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

The mean and variance of the distribution are

$$\mu = \frac{nD}{N} \tag{2-9}$$

and

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

Binomial

Definition

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

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \qquad x = 0,1,\ldots,n \tag{2-11}$$

The mean and variance of the binomial distribution are

$$\mu = np \tag{2-12}$$

and

$$\sigma^2 = np(1-p) \tag{2-13}$$

Poisson

Definition

The **Poisson distribution** is

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

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

$$\mu = \lambda \tag{2-16}$$

and

$$\sigma^2 = \lambda \tag{2-17}$$

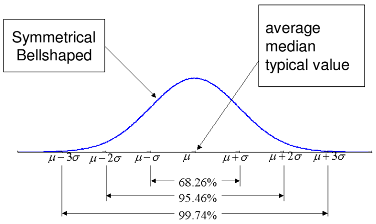
Normal distribution

Definition

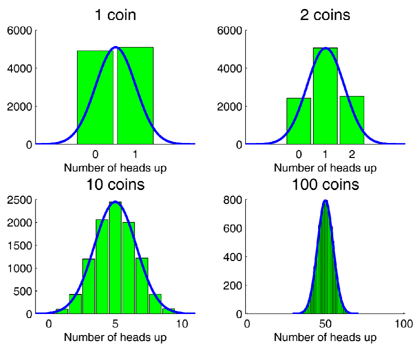
The **normal distribution** is

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

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



Central limit theorem



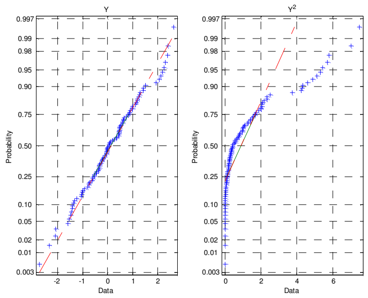
Other distributions

- Exponential
- Gamma
- Weibull
- Lognormal
- Empirical

Find the right (useful) model!

(There are infinite many distributions and many of them are baptized and examined.)

Probability plots



Normal

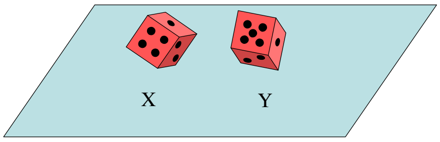
Ej normal

Visual test

Matlab:
normplot(x)
Probplot
wblplot

Independence

- Independence important!

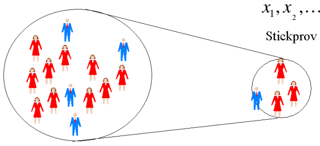


- Process industry, batcher, worn tools,…

Inferens

- Model the population with aid of a sample!

Alla anställda



$x_1, x_2, \ldots, x_n$

Stickprov

$f(x_1, x_2, \ldots, x_n)$

Statistic

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

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

Population

Sample

Varies!

Sample from normal distribution

- Assume independent and normally distributed s.v.

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

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$ (The average varies less than the individuals)

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

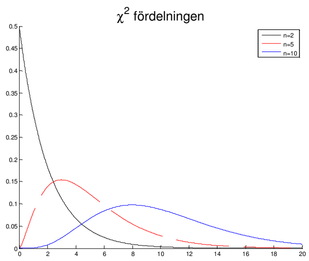
Chi2-fördelningen

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

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

$$Y \sim \chi_n^2, \text{ n degrees of freedom}$$

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



t- and F-distribution

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

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

Binomial and Poisson

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

$$\begin{cases} \mu_{\bar{X}} = p \\ \sigma^2_{\bar{X}} = \dfrac{p(1-p)}{n} \end{cases}$$

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

$$\begin{cases} \mu_{\bar{X}} = \lambda \\ \sigma^2_{\bar{X}} = \dfrac{\lambda}{n} \end{cases}$$