The normal distribution

Ib Skovgaard & Claus Ekstrøm
E-mail: ims@life.ku.dk

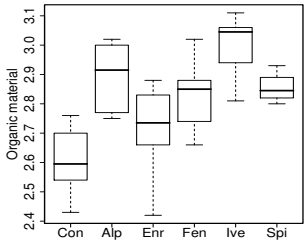
Program

- The normal distribution
  - Histogram and density
  - Probabilities
  - Symmetry, center and variation

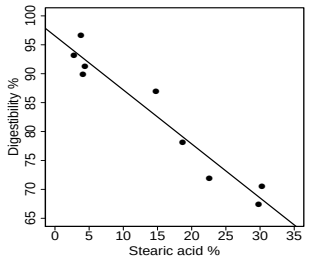
The normal (or Gaussian) distribution

The normal distribution is used to model residual variation

One-way ANOVA



Linear regression

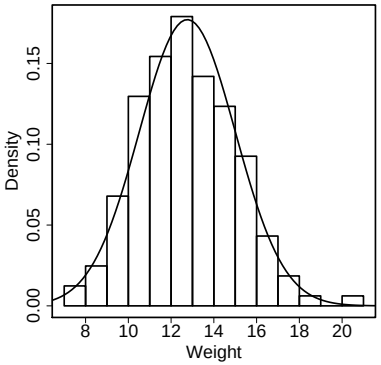


One sample:

Blood pressure								
96	119	119	108	126	128	110	105	94

Weights of crabs

- Weights of 162 crabs of a certain age: $y_1, \ldots, y_{162}$.
- R: $\bar{y} = 12.76$, $s = 2.25$
- Histogram normalized to have total area 1
- Curve for f, where f is the density of the normal distribution



$$f(y) = \frac{1}{\sqrt{2\pi \cdot 2.25^2}} \exp\left(-\frac{1}{2 \cdot 2.25^2}(y - 12.76)^2\right)$$

The curve fits nicely to the histogram.

Probabilities

For a standardized histogram, "relative frequency = area of rectangle", for example

$$\frac{\text{number of crabs between 14 g og 15 g}}{162} = 0.12$$

Similarly for the density: probability that an observation is between a and b equals the area under the curve, for example,

$$P(14 < Y < 15) = \int_a^b f(y)\,dy = 0.13$$

The two probabilities are not the same: the population is not the sample.

- If the density describes the population the histogram of the sample should look like the density
- Normal distribution density is a model of the histogram.

Normal distribution with mean μ and standard deviation σ

Replace the numbers 12.76 and 2.25 by μ and $\sigma > 0$:

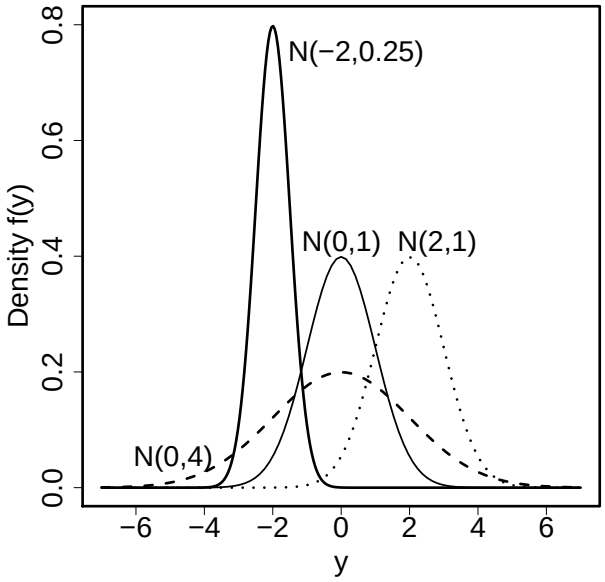
$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right)$$

A variable, Y, is normally distributed with mean μ and standard deviation σ if

$$P(a < Y < b) = \int_a^b f(y)\,dy.$$

for any a and b, that is, for all intervals. We write $Y \sim N(\mu, \sigma^2)$.

Symmetry — center — variation



Summary

- Population and density vs. sample and histogram
- Probability equals area under density curve
- The density for the normal distribution: symmetry, center and variation

Next week:

- Properties of the normal distribution
- How do we check that data follow a normal distribution?
- Why is the normal distribution so special?