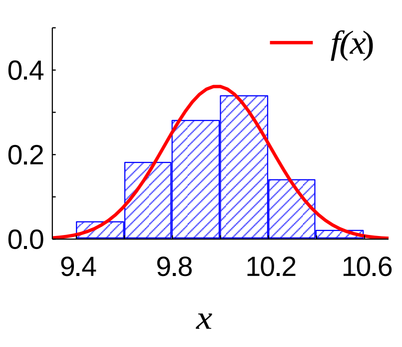
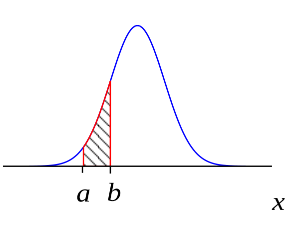
STATISTICAL BACKGROUND

CONTINUOUS RANDOM VARIABLE

The continuous random variable can take any numerical value in a given interval with some probability.

Probability density function

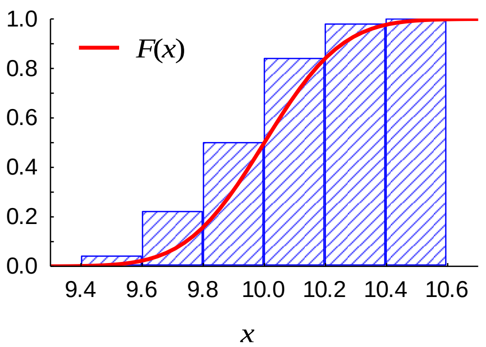


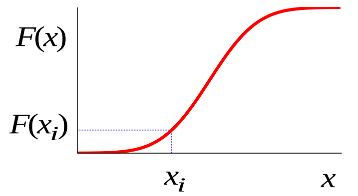


$$P(a < x \leq b) = \int_a^b f(x)dx$$

Continuous Random Variable

Cumulative distribution function





$$F(x_i) = P(x \le x_i) = \int_{-\infty}^{x_i} f(x)dx$$

Parameter and Statistic

<u>Parameter</u>

Characteristic of the **population**

Typically constant

expected value:

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx$$

variance

$$Var(x) = \int_{-\infty}^{\infty} [x - E(x)]^2 f(x)dx$$

<u>Statistic</u>

Characteristic of the **sample**

random variable

sample mean:

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

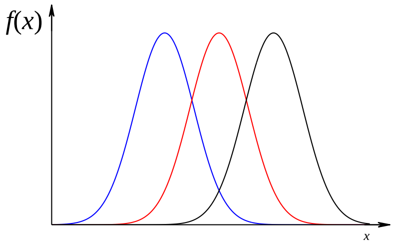
sample variance

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

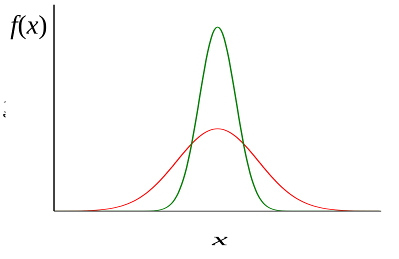
Gauss (Normal) Distribution

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

Two parameters: μ and σ^2



μ is different



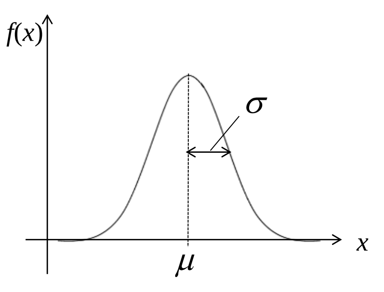
σ^2 is different

Gauss (Normal) Distribution

Short notation $N(\mu,\sigma^2)$ e.g. $N(0,1)$

Expected value $E(x) = \mu$

Variance $Var(x) = \sigma^2$



Standardisation

$$z = \frac{x-\mu}{\sigma}$$

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

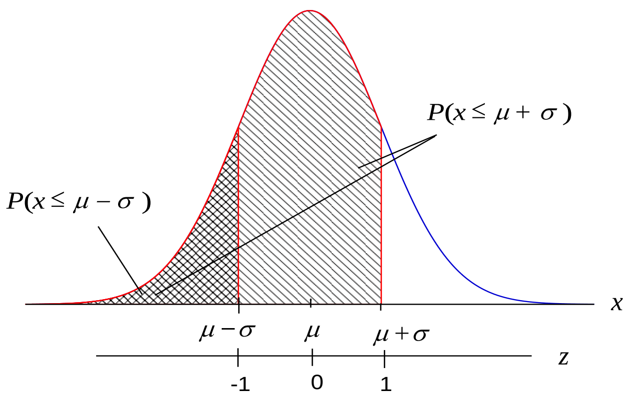
$$\mu = E(z) = 0$$

$$\sigma^2 = Var(z) = 1$$

GAUSS (NORMAL) DISTRIBUTION

What is the probability of finding the *x* Gauss d. random variable in the $(\mu-\sigma,\mu+\sigma)$ range?

$$P(\mu-\sigma < x \leq \mu+\sigma) = F(\mu+\sigma) - F(\mu-\sigma)$$



$$z = \frac{x-\mu}{\sigma}$$

GAUSS (NORMAL) DISTRIBUTION

$$z_{lower} = \frac{\mu-\sigma-\mu}{\sigma} = -1 \qquad z_{upper} = \frac{\mu+\sigma-\mu}{\sigma} = 1$$

Width of the interval	$\pm\sigma$	$\pm 2\sigma$	$\pm 3\sigma$
P	0.68268	0.9545	0.9973

Interpretation of the results

The variance of a measurement system is 0.09(mg/l)^2.

The measured concentration of the same sample vary in a ±2*0.3mg/l= ±0.6mg/l range with ~96% probability.

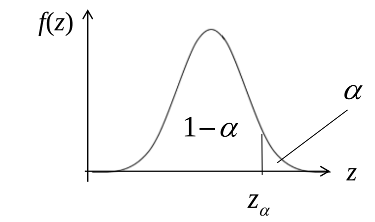
The maximum bite force of a spotted hyena is 4500N, the sigma is 45N.

The maximum bite force of a spotted hyena is between 4365-4635N with ~99.7% probability.

NOTATION

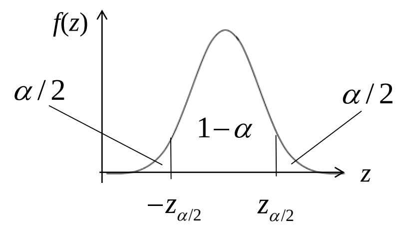
One sided

$$P\left(z \le z_\alpha\right) = 1-\alpha$$



Two sided

$$P\left(-z_{\alpha/2} < z \le z_{\alpha/2}\right) = 1-\alpha$$



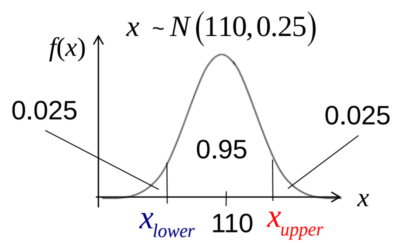
<u>Example 1</u>

The expected weight of a chocolate bar is 110 g, the variance of the weight is σ^2 =0.25g ^{2}. In which range will the weight of a chocolate with 95% probability?

The question „translated to maths”:

$$P\left(x_{lower} < x \le x_{upper}\right) = 0.95$$

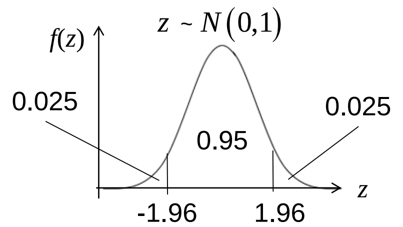
x_{lower} = **?** x_{upper} = **?**



This is known from the distribution function (table):

$$P\left(-z_{\alpha/2} < z \le z_{\alpha/2}\right) = 1-\alpha$$

$$P\left(-1.96 < z \le 1.96\right) = 0.95$$



Question:

$$P\left(x_{lower} < x \le x_{upper}\right) = 0.95$$

Known:

$$P(-1.96 < z \le 1.96) = 0.95$$

Connection:

$$z = \frac{x-\mu}{\sigma}$$

$$\mu = 10$$
$$\sigma = \sqrt{0.25}$$

$$P\left(-1.96 < \frac{x-110}{\sqrt{0.25}} \le 1.96\right) = 0.95$$

$$P(\overbrace{110-1.96\cdot 0.5}^{x_{lower}} < x \le \overbrace{110+1.96\cdot 0.5}^{x_{upper}}) = 0.95$$

$$P(109.02 < x \le 110.98) = 0.95$$

INTERPRETING THE RESULTS

Example 1
The expected weight of a chocolate bar is 110 g, the variance of the weight is $\sigma^2 = 0.25 g^2$. In which range will be the weight of a chocolate with 95% probability?

$$P(109.02 < x \le 110.98) = 0.95$$

The measured weight will be between 109.02 and 110.98 with 95% probability.

How would you explain this result?

Generalization of Example 1

Question: $P\left(x_{lower} < x \le x_{upper}\right) = 1-\alpha$ $\qquad$ $x_{lower} = ?\quad x_{upper} = ?$

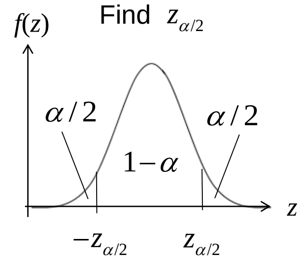
Known: $P\left(-z_{\alpha/2} < z \le z_{\alpha/2}\right) = 1-\alpha$ $\qquad$ $1-\alpha$ is given

$$z = \frac{x-\mu}{\sigma}$$

$$P\left(-z_{\alpha/2} < \frac{x-\mu}{\sigma} \le z_{\alpha/2}\right) = 1-\alpha$$

$$P\left(\mu - z_{\alpha/2}\sigma < x \le \mu + z_{\alpha/2}\sigma\right) = 1-\alpha$$

Find $z_{\alpha/2}$



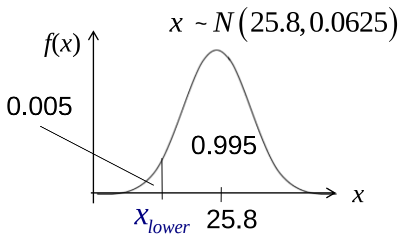
In which range will be the weight of the chocolate in Example 1 with 99% probability?

Example 2
The volume of gas bottles are normally distributed with 25.8 dm^3 expected value and 0.0625 (dm^3)^2 variance. What is that minimum volume, that 99.5% of the bottles exceed?

The question „translated to maths”:

$$P\left(x_{lower} < x\right) = 0.995$$

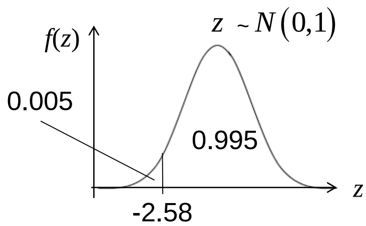
$$x_{lower} = ?$$



This is known from the distribution function (table):

$$P\left(-z_{\alpha} < z\right) = 1-\alpha$$

$$P\left(-2.58 < z\right) = 0.995$$



Question:

$$P\left(x_{lower} < x\right) = 0.95$$

Known:

$$P\left(-2.58 < z\right) = 0.995$$

Connection:

$$z = \frac{x-\mu}{\sigma}$$

$$\mu = 25.8$$

$$\sigma = \sqrt{0.0625}$$

$$P\left(-2.58 < \frac{x-25.8}{\sqrt{0.0625}}\right) = 0.995$$

$$P\big(\overbrace{25.8-2.58\cdot 0.25}^{x_{lower}} < x\big) = 0.995$$

$$\boxed{P\left(25.16 < x\right) = 0.995}$$

GENERALIZATION OF EXAMPLE 2

Question: $\boxed{P\left(x_{lower} < x\right) = 1-\alpha}$ $\boxed{x_{lower} = \textbf{?}}$ $1-\alpha$ is given

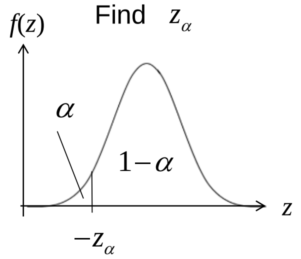
Known: $P\left(-z_\alpha < z\right) = 1-\alpha$

$$z = \frac{x-\mu}{\sigma}$$

$$P\left(-z_\alpha < \frac{x-\mu}{\sigma}\right) = 1-\alpha$$

$$\boxed{P\left(\mu - z_\alpha \sigma < x\right) = 1-\alpha}$$

Find z_α



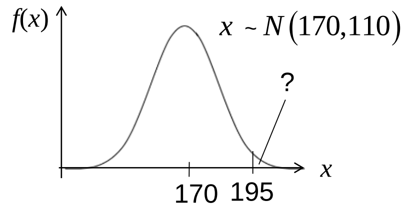
Give a 99% upper limit for the volume of the bottles in Example 2!

Example 3

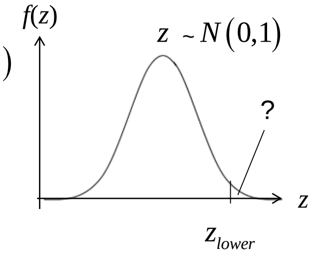
Harry is 204 cm tall and he is fond of tall girls (preferably above 195 cm). The average height of girls at his age is 170 cm with a variance of 110 cm^2. A random girl sits next to him in the library at the university. What is the probability that the girl meets his standard i.e. she is above 195 cm?

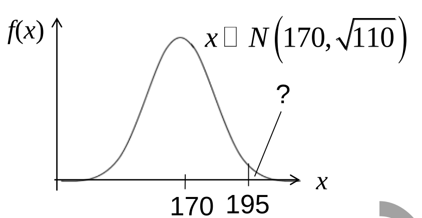
The question „translated to maths”:

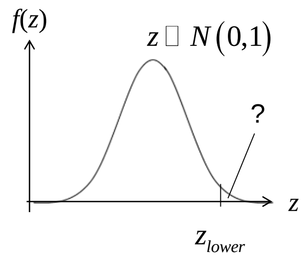
$$P\left(195 < x\right) = ?$$



If we knew z_{lower} we could read $P\left(z_{lower} < z\right)$ from the distribution function (table)







$$z = \frac{x - \mu}{\sigma}$$

$$z_{lower} = \frac{x_{lower} - \mu}{\sigma} = \frac{195 - 170}{\sqrt{110}} = 2.34$$

$$P\left(2.34 < z\right) = 0.00964 = P\left(195 < x\right)$$

Interpreting the Results

Example 3

Harry is 204 cm tall and he is fond of tall girls (preferably above 195 cm). The average height of girls at his age is 170 cm with a variance of 110 cm^2. A random girl sits next to him in the library at the university. What is the probability that the girl meets his standard i.e. she is above 195 cm?

$$P(195 < x) = 0.00964$$

Harry doesn't know a thing about statistics and probability. How would you explain this result to him?

What assumptions were made during this calculation? What if they are not true?

Generalization of Example 3

Question: $P(x_{lower} < x) = ?$

x_{lower} is given

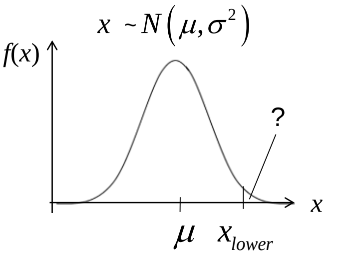
The question is transformed to z-scale:

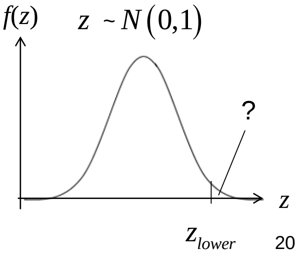
$$P\left(\frac{x_{lower} - \mu}{\sigma} < \frac{x - \mu}{\sigma}\right) = ?$$

$$P(z_{lower} < z) = ?$$

$$z = \frac{x - \mu}{\sigma}$$

$$P(x_{lower} < x) = P\left(\frac{x_{lower} - \mu}{\sigma} < z\right)$$





How many percent of the girls are between 170 and 185 cm?

EXAMPLES

1. The variance of a measurement is $\sigma^2 = 4g^2$. The bias of the measurement is 2 g. We measure an object, its weight is 200 g.
a) In which range will be the outcome of the measurement with 99% probability?
b) With what probability will be the measured weight above 205 g?
c) With what probability will be the measured weight below 200 g?
d) Give a 90% upper limit for the measured weight! (The measured weight will be below this value with 90% probability.)

2. The volume of gas bottles are normally distributed with 25.8 dm^3 expected value and 0.0625 $(dm^3)^2$ variance.
a) How many percent of the bottles will be in the 25.8±0.3 dm^3 interval?
b) In what volume-interval will be 99% of the bottles?

THE SAMPLE MEAN

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + ... + x_n) = \frac{1}{n}\sum x_i$$

$$E(\bar{x}) = \frac{1}{n}[nE(x)] = E(x) = \mu \qquad \sigma_{\bar{x}}^2 = Var(\bar{x}) = \frac{Var(x)}{n} = \frac{\sigma_x^2}{n}$$

$$x \sim N(\mu, \sigma^2) \qquad z = \frac{x-\mu}{\sigma}$$

$$\bar{x} \sim N(\mu, \sigma^2/n) \qquad z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

Sampling distribution: probability distribution of a statistic (e.g. sample mean or sample variance)

$N(\mu, \sigma^2/n)$ is the sampling distribution of the sample mean

CENTRAL LIMIT THEOREM

The mean of sample elements taken from any distribution approximately follows Gauss distribution around the expected value of the original distribution with variance σ^2/n. Where n is the sample size.

Sum as well $\sum_{i=1}^{N} x_i \sim N\left(n\mu, n\sigma^2\right)$

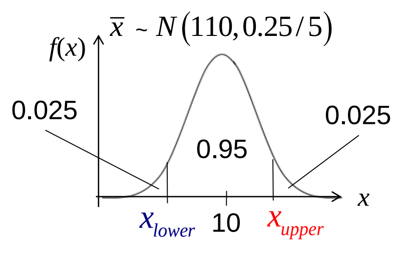
Based on the Central Limit Theorem: $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

<u>Example 4</u>
The expected weight of a chocolate bar is 110 g, the variance of the weight is σ^2 =0.25g ^{2}. In which range will be the average weight of 5 chocolate bar with 95% probability?

The question „translated to maths":

$$P\left(\bar{x}_{lower} < \bar{x} \le \bar{x}_{upper}\right) = 0.95$$

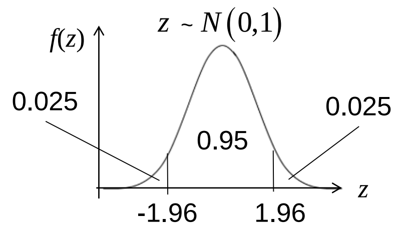
x_{lower} = **?** x_{upper} = **?**



This known from the distribution function (table):

$$P\left(-z_{\alpha/2} < z \le z_{\alpha/2}\right) = 1 - \alpha$$

$$P\left(-1.96 < z \le 1.96\right) = 0.95$$



Question:

$$P\left(\bar{x}_{lower} < \bar{x} \le \bar{x}_{upper}\right) = 0.95$$

Known:

$$P(-1.96 < z \le 1.96) = 0.95$$

Connection:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$\mu = 10$$
$$\sigma = \sqrt{0.25}$$
$$n = 5$$

$$P\left(-1.96 < \frac{\bar{x} - 110}{\sqrt{0.25} / \sqrt{5}} \le 1.96\right) = 0.95$$

$$P\Big(\overbrace{110 - 1.96 \cdot \sqrt{0.5}}^{\bar{x}_{lower}} < \bar{x} \le \overbrace{110 + 1.96 \cdot \sqrt{0.5}}^{\bar{x}_{upper}}\Big) = 0.95$$

$$P(109.56 < \bar{x} \le 110.44) = 0.95$$

Interpreting the Results

Example 1
The expected weight of a chocolate bar is 110 g, the variance of the weight is σ^2 =0.25g ^{2}. In which range will be the weight of a chocolate with 95% probability?

$$P(109.02 < x \le 110.98) = 0.95$$

Example 4
The expected weight of a chocolate bar is 110 g, the variance of the weight is σ^2 =0.25g ^{2}. In which range will be the average weight of 5 chocolate bar with 95% probability?

$$P(109.56 < \bar{x} \le 110.44) = 0.95$$

In general: $P\left(\mu - z_{\alpha/2}\sigma/\sqrt{n} < \bar{x} \le \mu + z_{\alpha/2}\sigma/\sqrt{n}\right) = 1 - \alpha$