Statistics / Lecture1: Probability distributions, expected value, decision trees

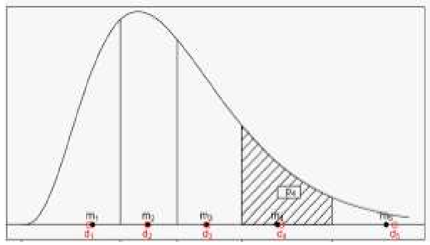
1. Discrete probability distributions
2. Parameters: Expected Value, standard deviation
3. Continuous probability distributions

   - example: Normal distribution

4. Decision trees

   - example: Newsboy Problem

discrete distribution



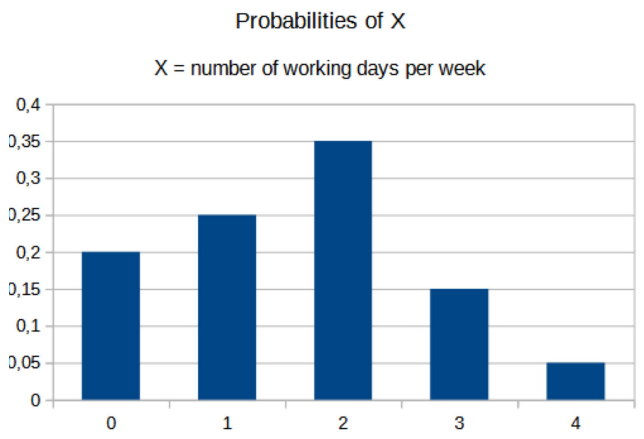
continuous distribution

1. Discrete probability distributions

- A discrete probability distribution shows all possible values of a discrete random variable x and their associated probabilities

Example1: Anne is a nursing student. She works occasionally as a sick leave substitute in a health centre. The number of weekly working days per week vary between 0 to 4. Following table shows the possible values of x and their probabilities. Visualization is bar chart.

X	P
0	0.2
1	0.25
2	0.35
3	0.15
4	0.05



Expected value E(x)

- Expected value describes the long-term average level of a random variable x based on its probability distribution.

Expected value **$E(x) = \Sigma\, p_i\, x_i$**

(that is: $E(x) = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n$)

X	P
0	0.2
1	0.25
2	0.35
3	0.15
4	0.05

What is the expected value of the example?

The expected value of Anne's working days per week
E(x) = 0.2*0 +0.25*1+0.35*2+ 0.15*3+0.05*4 = 1.6

In other words the long term average is 1.6 working days per week.

Variance Var(x), Standard deviation

Variance and its square root are parameters, which describe how widely values of x are distributed around the average (expexted return). The formulas for variance and standard deviation in case of discrete probability distribution are following.

Variance $\quad Var(\mathbf{x}) = \boldsymbol{\Sigma}\, \mathbf{p_i}\, (\mathbf{x_i} - \boldsymbol{\mu})^2$, where $\boldsymbol{\mu} = \mathbf{E(x)}$ (expected value)

Standard deviation Std(x) = √(Var(x)) *)

X	x-µ =x-1.6	P
0	-1.6	0.2
1	-0.6	0.25
2	0.4	0.35
3	1.4	0.15
4	2.4	0.05

Calculate the variance and standard deviation of the example

$Var(x) = 0.2*1.6^2+0.25*0.6^2+0.35*0.4^2+0.15*1.4^2+0.05*2.4^2= 1.24$

$Std(x) = \sqrt{1.24} = 1.11$

*) $(-1.6)^2 = 1.6^2$ => minus signs are not needed

Expected Return (= expexted value of profit)

Ex. Midsummer music festival profitability analysis
The profit or loss of organizing the festival depends on weather conditions in the following way. Calculate the expected return. What aree your conclusions.

weather	probability	Return (profit or loss)
Sunny weather	0.40	+12000
Cloudy, chilly, but no rain	0.35	+ 2000
rain	0.25	- 50000

Expexted return (long term average result) is following

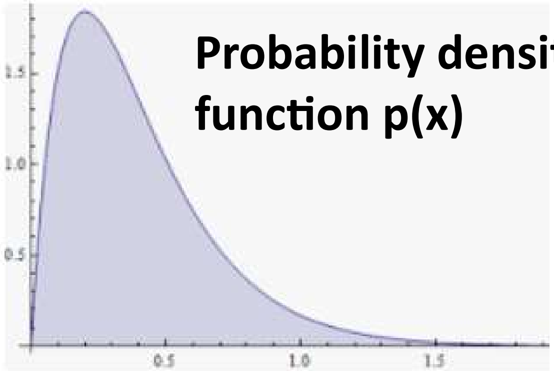
ER = 0.4*12000+0.35*2000+ 0.25*(-50000) = -7000

Conclusion: Organizing the festival is not profitable.

2. Continuous probability distributions

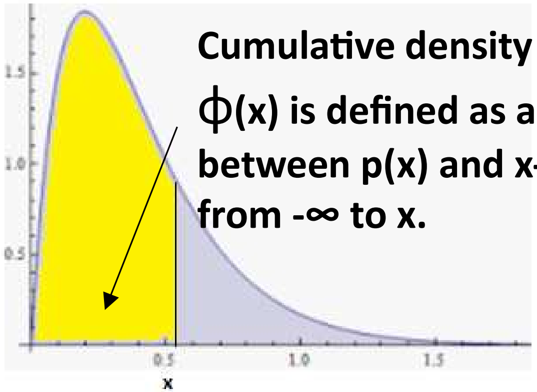
A continuous probability distribution is one in which a continuous random variable X can take on any value within a given range of values (which can be infinite). Probabilities of values of x are given by a function p(x), called **probability density function** or **probability mass function.**

p(x)



Probability density function p(x)

x



Cumulative density function $\phi(x)$ is defined as area between p(x) and x- axis from -∞ to x.

Value p(x) describes the probability of value of x

$p(x) \geq 0$ (function is always non-negative)

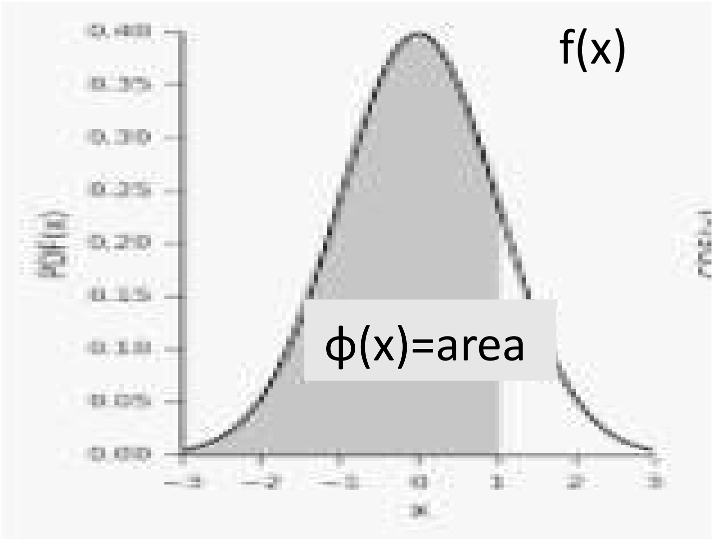
Area between f(x) and x–axis equals **1**

Interpreation: $\phi(x_0) = P(x \leq x_0)$
Value of cumulative density function at x_0 gives the probability that the variable value $x \leq x_0$

Normal distribution

<u>The most common continuous distribution is ”Gaussian” normal distribution.</u>

It has two parameters: mean μ and standard deviation σ



$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

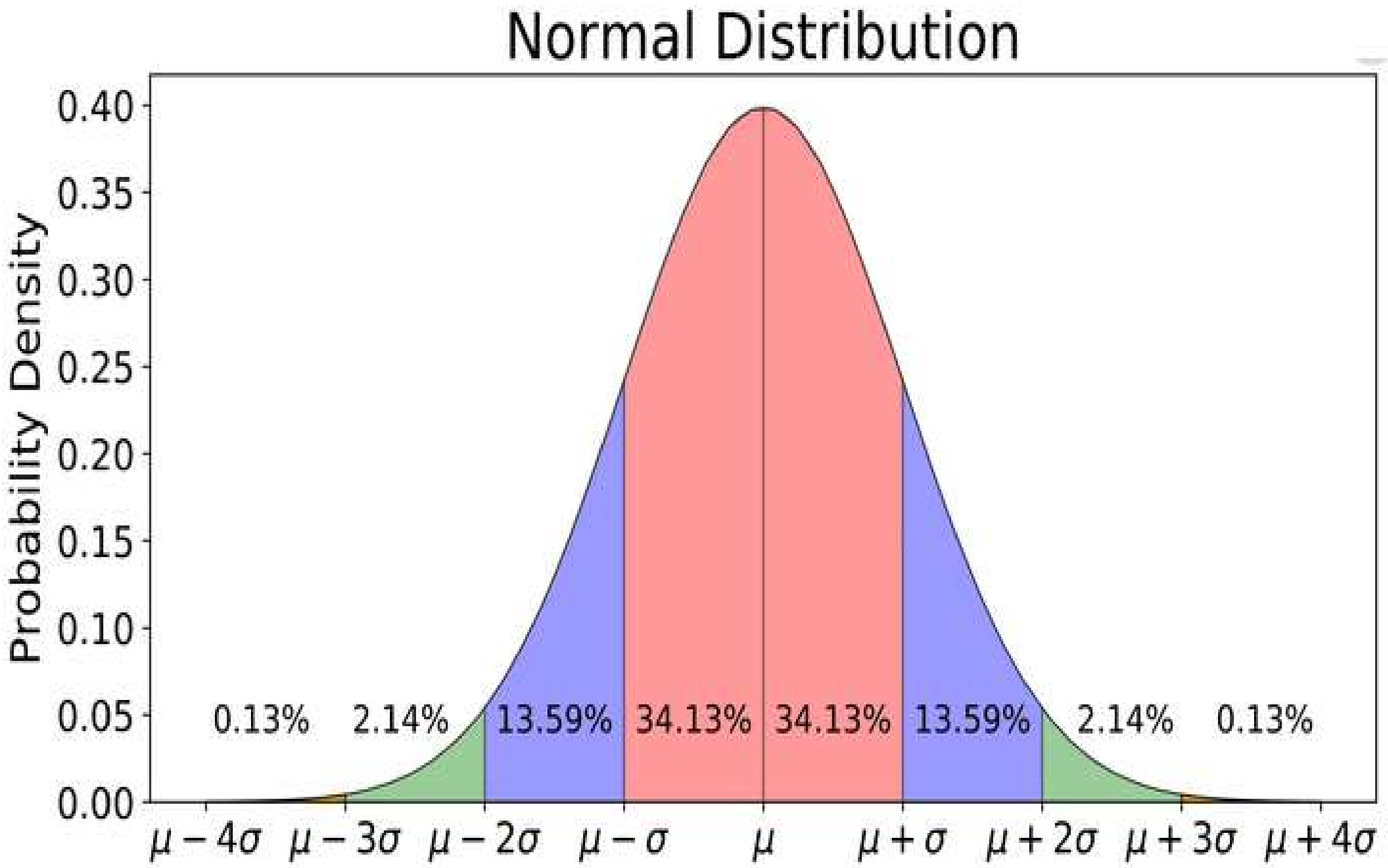
$$\phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

Both functions look complicated, but they are luckily found in calculators. In EXCEL function NORM.DIST can be used to calculate both p(x) and ϕ(x)

Notation. **x ~ N(μ,σ) means that** variable x follows normal distribution with mean = **μ and standard deviation σ.**

For example if variable x is hemoglobin of male person, notation x ~ N(153,9) would mean that x is normally distributed with mean of 154 and standard deviation of 9.

Picture of the probabilities of variable X in intevals limited by μ + k*σ , where k=0,±1,….

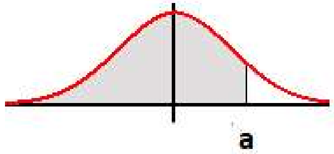


Around 68 % of values lie in range [μ-σ,μ+σ]

Around 95% of values lie in range [μ-2σ,μ+2σ]

Usage of cumulative density function **$\phi(x)$ and its inverse function $\phi^{-1}(P)$**

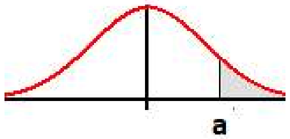
1. Calculation of probability that value of $x \leq a$



$P(x \leq a) = \phi(a)$

Excel: =NORM.DIST (x;μ;σ;1)

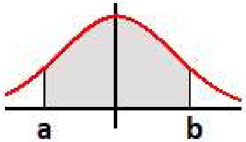
2. Calculation of probability that value of $x \geq a$



$P(x \geq a) = 1 - \phi(a)$

Excel: = 1- NORM.DIST (x;μ;σ;1)

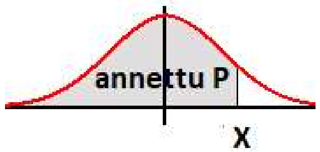
3. Calculation of probability that $a \leq \mathbf{x} \leq b$



$P(a \leq x \leq b) = \phi(b) - \phi(a)$

Excel: = NORM.DIST (b,.) - NORM.DIST (a,.)

4. Calculation of value of **x** corresponding probability P



$\phi(x) = P \Rightarrow x = \phi^{-1}(P)$

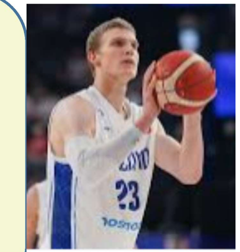
Excel:
NORM.INV(P; μ;σ)

Ex3. Basket ball player. The mean height of Finnish adult men is 180.7 cm ja standard deviation is 7.4 cm. Assume that height is normally distributed.

a) Calculate the probability that height ≥ 213 cm (Lauri Markkanen's height is 213)

b) Calculate the "P90 height". (a height that only 10% exceeds)

c) Calculate the probability that height is between 190 cm – 200 cm?



a) **P(x≥ 213)** = 1 - ϕ(213) = 1-0,99999364= **0,00000636**

6.4 miljoonasta

Excel ϕ(213):
=**NORM.DIST**(213;180,7;7,4;1)
which gives 0,99999517 =>
1- ϕ(213) = 0.00000636

b) P90 height requires solving x from ϕ(x) = 0.90
Using inverse function we get x = $\phi^{-1}(0.9)$ = **190.2 cm**
=> Conclusion there are 90% with h <= 190.2 cm and 10% with h>190.2.

Excel's ϕ^{-1} is NORM.INV:
=NORM.INV(90%;180,7;7,4)
(arguments: percentage, mean, st.dev). It gives 190,2 cm

c) P(190≤x≤200) = ϕ(200)-ϕ(190) =0.996 – 0.899 = 0.097 = **9.7%**

=**NORM.DIST**(200;180,7;7,4;1) -
NORM.DIST(190;180,7;7,4;1)

Ex4. *A gravel is suitable for its purpose, if only 2.5% of its stones have diameter greater than 30 mm. In the table there are diameters of stones of a random sample of 80 stones. Assume that diameters are normally distributed. Does the gravel meet the requirements?*

	A									
1	17,0	19,3	26,7	17,0	13,3	11,8	9,4	6,9	30,9	12,2
	12,0	16,7	17,9	5,8	23,4	25,4	10,5	14,2	15,5	25,2
	13,0	11,0	19,9	21,1	25,6	25,5	25,4	31,1	19,2	21,8
	18,5	24,5	17,9	19,2	24,8	23,0	27,3	29,6	27,3	16,1
	29,3	24,2	16,2	19,9	10,4	16,8	17,9	19,8	22,1	20,8
	30,8	24,8	33,7	29,0	12,8	20,0	15,6	24,7	11,7	16,0
	15,6	16,7	20,0	28,0	20,4	11,8	19,1	13,7	17,4	25,6
	24,4	22,8	12,7	45,0	24,9	27,8	21,7	18,7	17,7	18,0

1. Calculate with Excel μ and σ

=average(A1:H10) gives μ = 20.1

=stdev.s(A1:H10) gives σ = 6.8

2. Calculate using cumulative density function

P(x<30) (=probability that diameter < 30)

= NORM.DIST(30;20,1;6,8;1)

gives 0,927 = **92,7%**

Conclusion: From gravel 92,7% has diameter < 30 mm

=> 7.3% has diameter > 30 mm, which is too much (maximum accepted share is 2.5%)

Another method would be to calculate the value of x corresponding probability P = 97.5%

=NORM.INV(97,5%; 20,1 ; 6,8) gives 33,4 mm , which exceeds 30 mm limit.

3. Decision tree

- A decision tree is a decision support tree-like model of decisions and their possible consequences
- A decision tree consists of three types of nodes:
  - Decision nodes – typically represented by squares
  - Chance nodes – typically represented by circles
  - End nodes – typically represented by triangles

Decision trees are used in statistics, data mining and machine learning

”Newsboy problem” (classical decision tree example)

Ex5. How many papers the newsboy should buy from the publisher to maximize his profit? Probability distribution of number of sold papers are given.

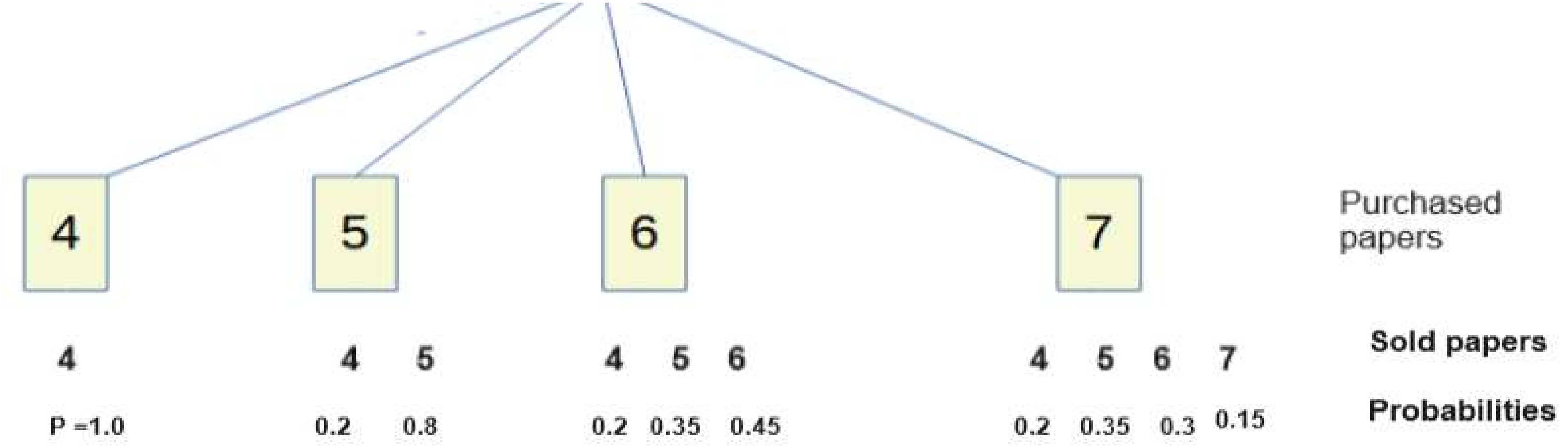
Newsboys were the main distributors of newspapers to the general public in USA 100-150 years ago. Youngest were 5 – 6 years old. They bought papers from the publisher at their own risk and sold them on the streets with a small profit. From unsold papers they got no compensation.

Demand of papers	Proba-bility
4	0.2
5	0.35
6	0.3
7	0.15

Purchase price: 1 $ per paper
Selling price: 2 $ per paper

How many papers should the newsboy purchase daily to maximize his profit?

(Almost all shopkeepers have to solve problems of type: How many items he/she should order? Especially question is critical when we talk about products, which cannot be sold later (milk, fish, etc.)



Demand of papers	Proba-bility
4	0.2
5	0.35
6	0.3
7	0.15

Next we calculate expected returns (profit) in all four cases: **Profit for sold paper = 1$, loss for unsold is -1$**

Case1: Newsboy N buys 4 papers Everything is sold. **Return (profit)** = 4*1=**4$**

Case2: N buys 5 papers. With P = 0.2 he sells 4 and with P =0.8 he sells all 5:
ER = 0.2*(4-1) + 0.8*5 = **4.6$**

Case3: N buys 6 papers: With P = 0.2 he sells 4. 2 remains unsold. Profit = 4-2 = 2$. With P = 0.35 he sells 5. 1 remains unsold. Profit = 5 – 1 = 4$. With P = 0.45 he sells all 6: Profit = 6$
ER = 0.2*2 + 0.35*4+0.45*6 = **4.5$**

Case4: N buys 7 papers: With P = 0.2 sells 4, unsold 3, profit=1. With P=0.35 sells 5 unsold 2, profit = 3. With P=0.3 sells 6 unsold 1, profit 5 and with P = 0.15 sells 7. profit = 7$. ER = 0.2*1+0.35*3+0.3*5+0.15*7 = **3.8$**

Best long term strategy is to buy 5 papers every day which gives 4.6$ average income.

2. Lecture 20.3.24 topics

- Concepts and terminology

- Descriptive statistics
  - - Parameters
  - - Grouped data
  - - Graphical presentation

A. Basic concepts

1. Population
=the set of objects of interest, which we will to research
Example: Students of Lapland UAS

2. Sample
=proportion of population, which is researched

- usually randommly chosen subset of population
Example: Student, who were asked to fill a questionary

3. Variables
The properties of objects of the population, which we collect information about.
Example: Degree programme, age, gender,...

Example:

Customer feedback questionary

Population: All customers of the super market

Sample: Customers, who participated in the questionary

Variables:
- age
- gender
- household size
- total of purchases

4. Discrete and continuous variables

Variable is **discrete**, if the set of possible values is finite or enumerable
(example: number of children of a family)

Variable is **continuous**, if the set of possible values is continuous, infinite.
(Example: monthly salary)

5. Statistical parameters

= numerical measures which charaterize the distribution of variable values
(Example: mean, variance)

Notation convention:
Latin letters are used for parameters of a sample, greek letters for parameters of the population.

(Example: population mean = μ
Sample mean = x)

Discrete variables:
age in years
Gender
size of the houshold

satisfaction to service (scale 0 - 5)

Continuous variables:

total value of purchases

6. Levels of measurement

Statistical variables can be divided into four categories, which are called the levels of measurements

level of measurement	description	examples
nominal level variable or "class variable"	variable values are just names or labels, which define classes	-gender (male, female) -marital status (married, single, divorced, widow) -color (red,blue,...)
ordinal level variable	values are labels, but an order can be added to the values	-quality class of potatoes -a players location on the ranking list
interval level variable	-a numerical variable, where it is meaningful to compare differences, but not ratios - absolute zero does not exist	- Celsius temperature
ratio level variable	a numerical variable where also ratios are comparable	-Kelvin temperature -monthly salary - age in years

Nominal level variables:
- gender

Ordinal level
- satisfaction to service (scale 0 - 5)

Ratio level:
- age
- Household size
- total of purchases

B. Descriptive statistics

- Statistical parameters:
  - Measures of central tendency
  - Mesures of variation
- Grouping data
  - Calculation of parameters from grouped data
- Graphical presentations

1. Measures of central tendency

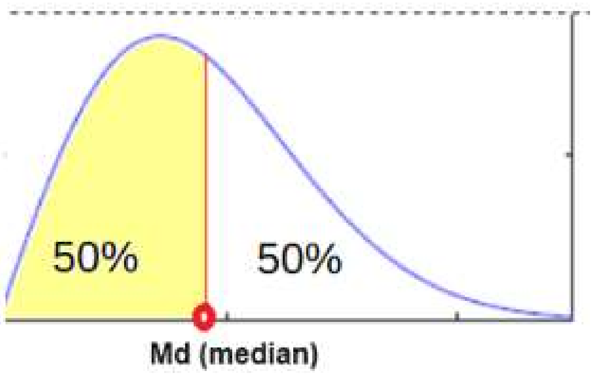
	A
1	19
2	11
.	16
.	12
.	7
	11
	10
	15
	7
	11
	17
	9
	7
	14
	14
	13
	8
	16
	12
	19
	8
	12
	14
	5
	16
	18
	13
28	11

Mean value (or average)

Population Mean	Sample Mean
$\mu = \dfrac{\sum x}{N}$	$\bar{X} = \dfrac{\sum x}{n}$

(Notice the convention of using greek letters for parameters of population. Formulas are the same: mean = sum of values / count of values)

Median



Excel- formulas

=average(A1:A28)

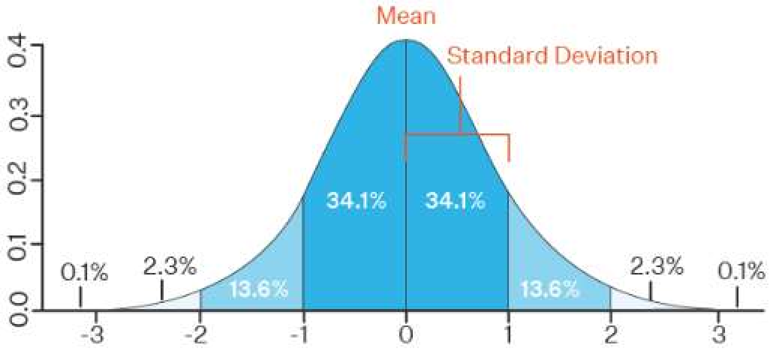
=median(A1:A28)

Rule for manual calculation of median:

1. Sort data by magnitude
2. Median is the value in the middle of sorted data. Half of the values are less than median, the other half is greater than median.

(If number of values is even, median is the average of the two values in the middle)

2. Measures of variation



Standard deviation

= the square root of the mean of squared distances from average: $(x_i - \mu)^2$.

Formulas are slightly different for whole population and sample.

Population	Sample
$\sigma = \sqrt{\dfrac{\sum(X-\mu)^2}{N}}$	$s = \sqrt{\dfrac{\sum(X-\bar{x})^2}{n-1}}$
X - The Value in the data distribution μ - The population Mean N - Total Number of Observations	X - The Value in the data distribution $\bar{x}$ - The Sample Mean n - Total Number of Observations

Example: Calculate the standard deviation of data in cells A1:A28 of the previous slide:

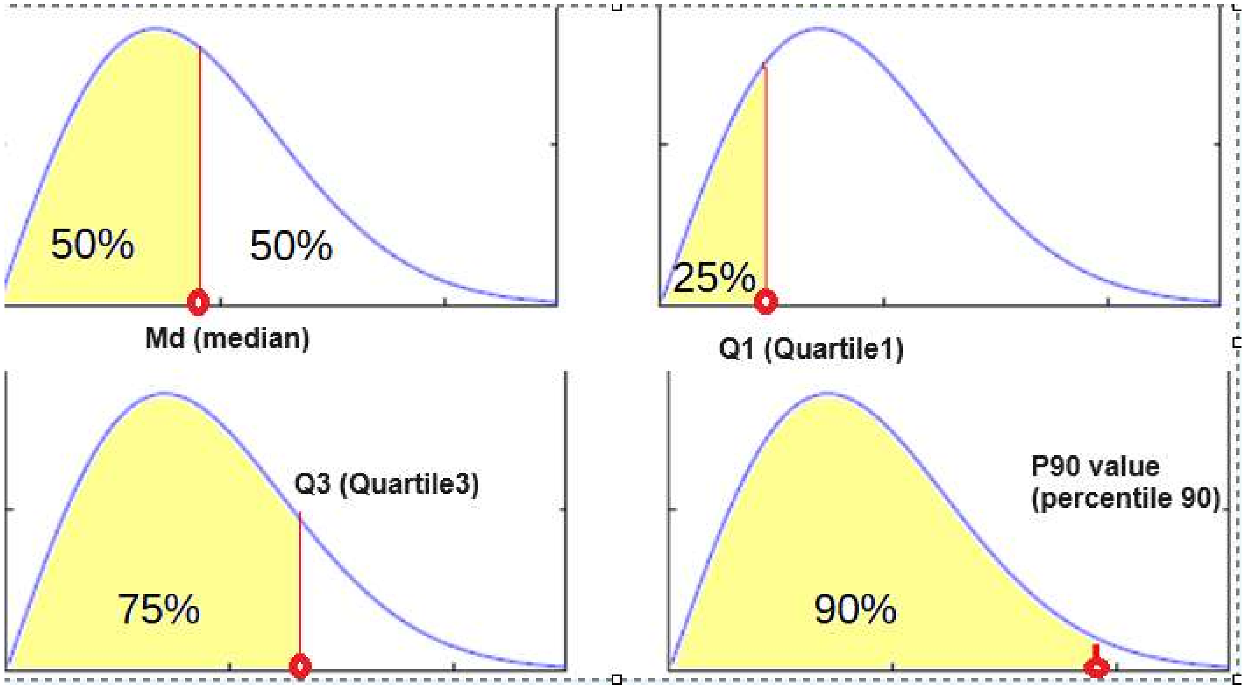
= STDEV.P(A1:A28)

Percentiles

= values of x which divide the variable data into two fractions according to the given percentages.

Most common percentiles are Quartiles 1 and 3 and P90 – value.

(P90- values are commonly used as reference values in medical measurements like blood pressure, cholesterol, etc.)



Usage of Excel functions

Percentile(A1:A60; 90%)
(or Percentile(A1:A60; 0,9)

Median(A1:A60)
(or Percentile(A1:A60; 50%)

Quartile(A1:A60, 1)
(or Percentile(A1:A60; 25%)

Quartile(A1:A60, 3)
(or Percentile(A1:A60; 75%)

Mean and standard deviation can be calculated also from a frequency table,

in which individual variable values are replaced by intervals.

Calculation of mean and standard deviation can be estimated using class centers as variable values.

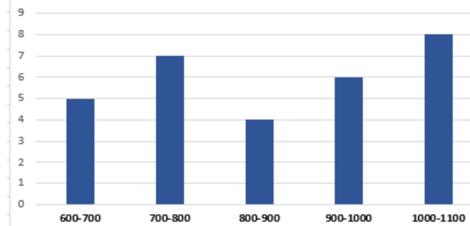
Formulas: **MEAN** $$\mu = \frac{\sum f_i x_i}{N}$$

STDEV.P $$\sigma = \sqrt{\frac{\sum f_i (x_i - \mu)^2}{N}}$$

STDEV:S $$s = \sqrt{\frac{\sum f_i (x_i - \overline{x})^2}{N-1}}$$

reaction time	frequency
600 – 699	5
700 – 799	7
800 – 899	4
900 – 999	6
1000 – 1099	8

Frequency table



Bar char presentation

Example. In a psychological test the reaction times of 30 participants were measured. Results in form of a frequency table are on the right. (unit is millisecond).

In our example the formulas give following results:

$$\mu = \frac{5 \cdot 650 + 4 \cdot 750 + 7 \cdot 850 + 6 \cdot 950 + 9 \cdot 1050}{30} = 867$$

$$\sigma = \sqrt{\frac{5 \cdot (650 - 867)^2 + 4 \cdot (750 - 867)^2 + 7 \cdot (850 - 867)^2 + 6 \cdot (950 - 867)^2 + 9 \cdot (1050 - 867)^2}{30}} = 146$$

Comment: The parameter values from grouped data are approximations. The exact mean and stdev calculated from original data were in this case 861 and 153.

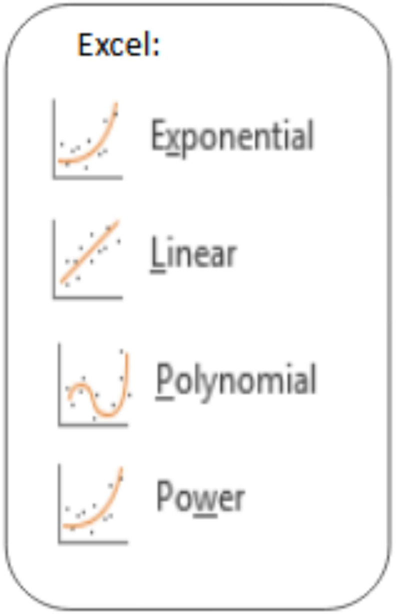
Common mathematical models:

Linear model $y = ax + b$

Exponential model $y = ae^{bx}$

Power model $y = ax^{b}$

Polynomial model $y = ax^{2} + bx + c$



Fitting a mathematical model to observed data

Linear model $y = ax + b$

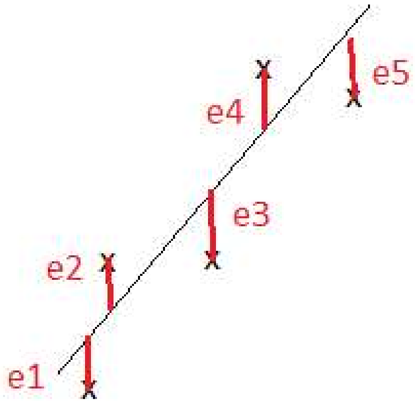
Principle: Find parameters slope a and constant b, that the sum of squares of differences between observed y – values and those calculated from expression a x + b is at minimum

Example of observed data Consists of x, y pairs

x	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**
y	**42.5**	**42.8**	**43.2**	**43.5**	**43.9**	**44.2**

WolframAlpha function finds the minimum of the sum of squares $\Sigma (a x_i + b - y_i)^2$ with following command line.

minimize (a*1.0+ b-42.5)² + (a*1.5+ b-42.8)²+ (a*2+ b-43.2)²+ (a*2.5+ b-43.8)²+ (a*3+ b-43.9)²+ (a*3.5+ b-44.3)²



(* "differences" are the vertical lines in the picture

result: $(a, b) \approx (0.691429, 41.7943)$

Method is known as **linear regression analysis** or the **method of least square sum**

Calculators have functions for regression analysis.

Excel graphics trendline

Easiest way to find parameters of

a mathematical model is TRENDLINE

Function of Excel Graphics

Linear model using trendline property

x	y
1	42,5
1,5	42,8
2	43,2
2,5	43,5
3	43,9
3,5	44,2

Data

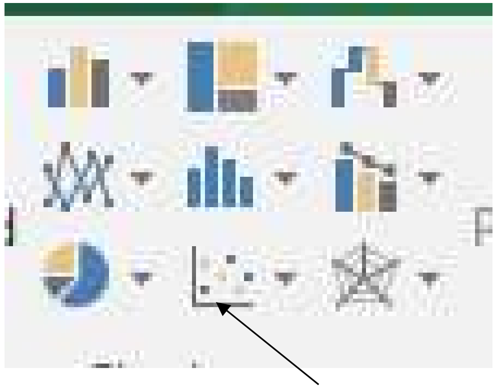
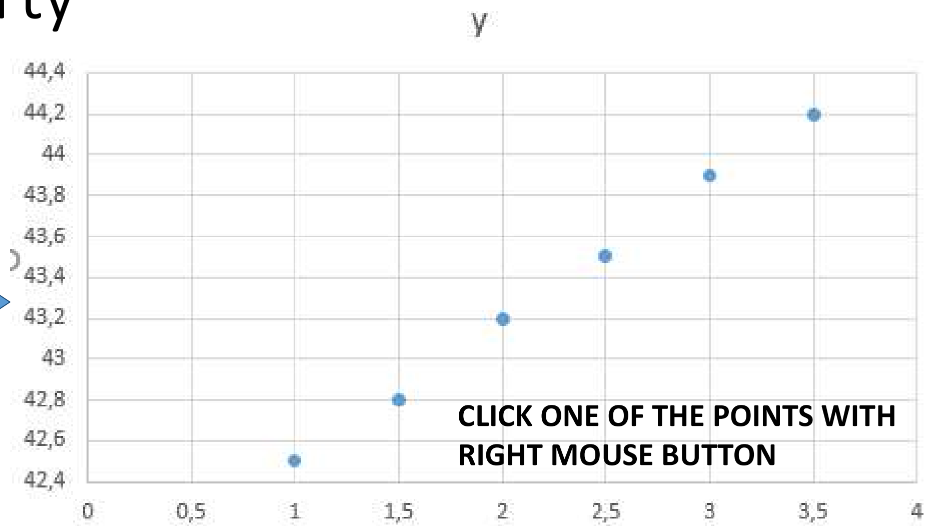
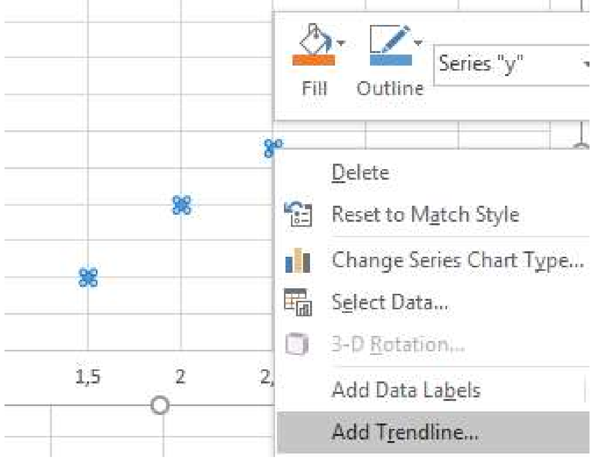


Chart TYPE

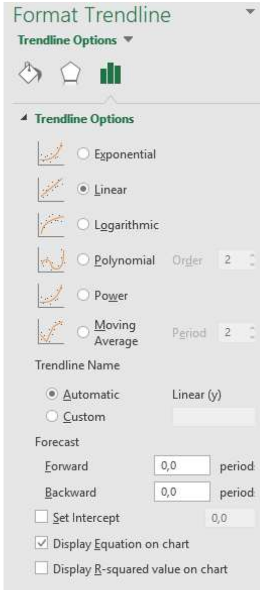


CLICK ONE OF THE POINTS WITH RIGHT MOUSE BUTTON

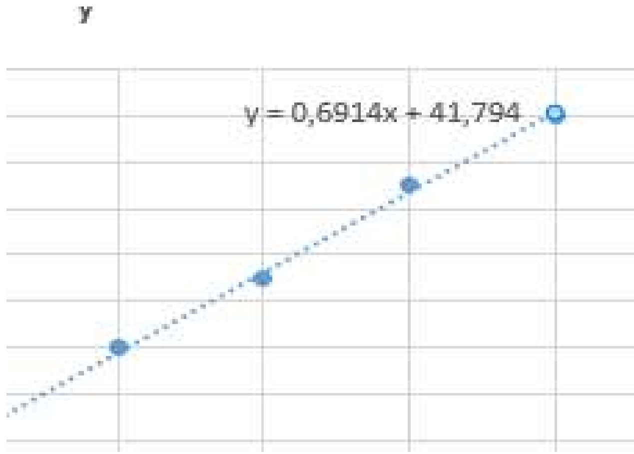
ADD TRENDLINE



TYPE = LINEAR

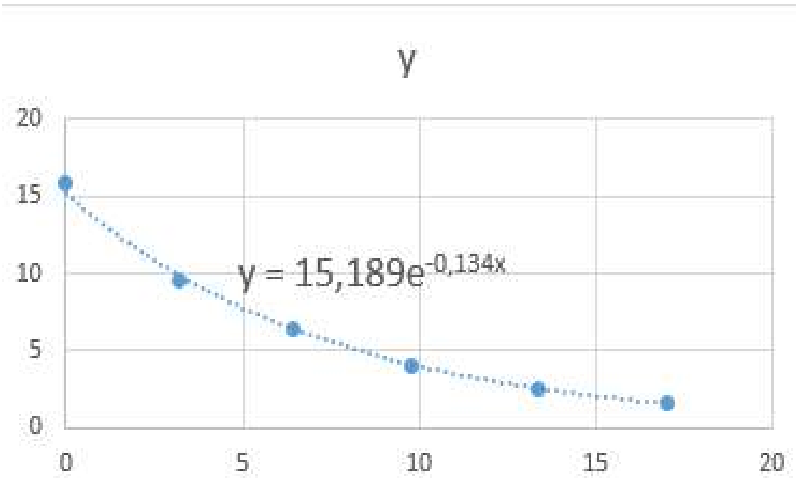


RESULT: EQUATION OF THE MODEL



Other mathematical models (in Excel graphics trendline)

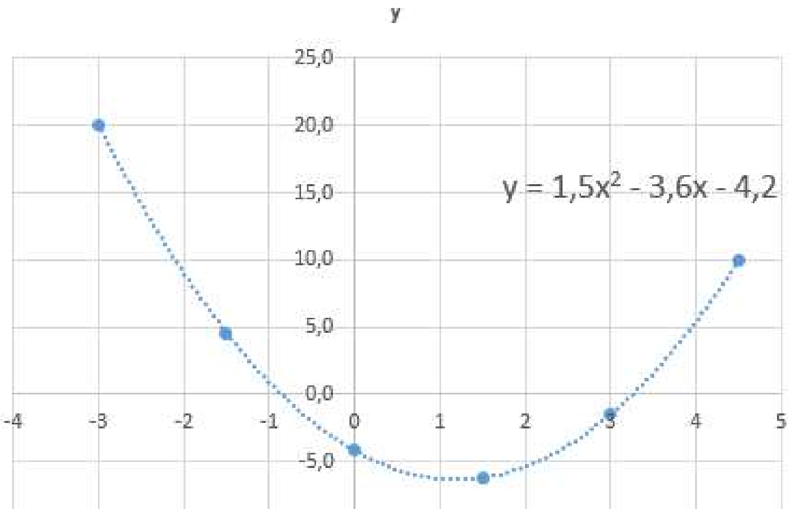
x	y
0	15,8
3,2	9,6
6,4	6,4
9,8	4
13,4	2,5
17	1,6



Exponent model

$$\mathbf{y = a\, e^{-b\,x}}$$

x	y
-3	20,1
-1,5	4,6
0	-4,2
1,5	-6,2
3	-1,5
4,5	10,0



Polynomial model

$$\mathbf{y = a\, x^2 + b\, x + c}$$

Topics / 8.3.24 lesson

1. Estimation of parameters

-error margins of parameters

-confidence margins and confidence intervals

2. Confidence intervals of regression analysis

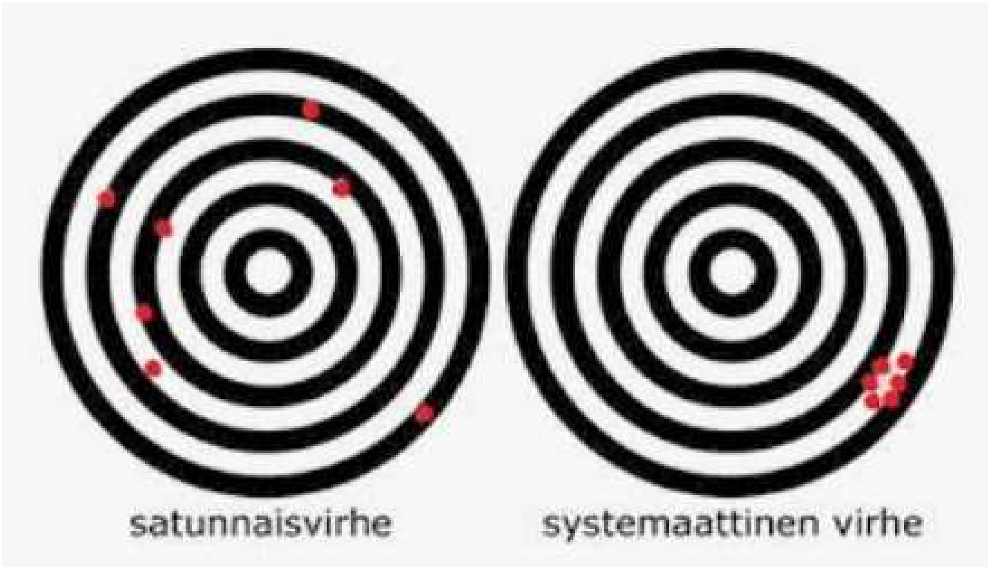
parameters, R^2 value = measure of goodness of the model

Error types: Random error and systematic error

Type A: Random error

is an error caused by random factors in measurement. It appears in the way that measured values are randomly distributed around the mean.

By increasing the number of measurements, error of type A can be eliminated.



Type B: Systematic error

Errors appears always in the same direction.

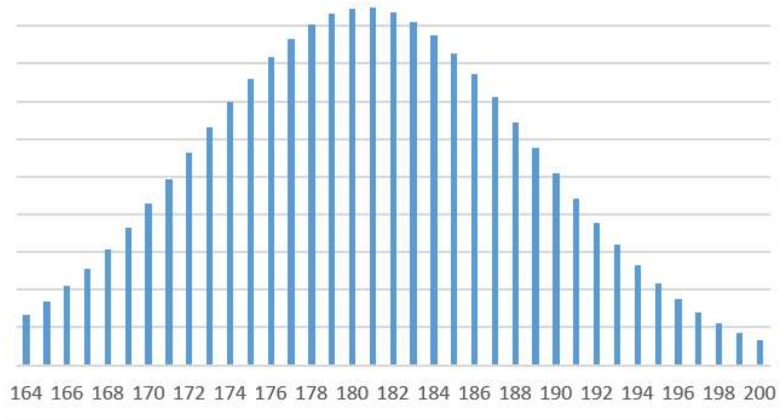
Statistical methods cannot be used to eliminate systematic error.

Error margin of a sample research

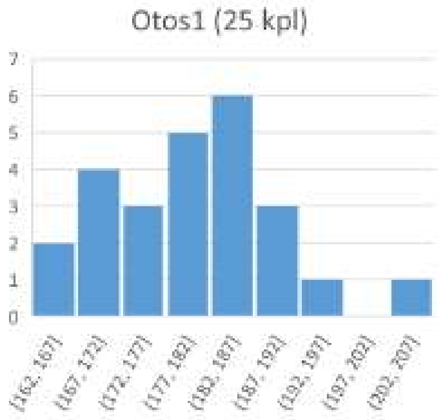
Confidence intervals

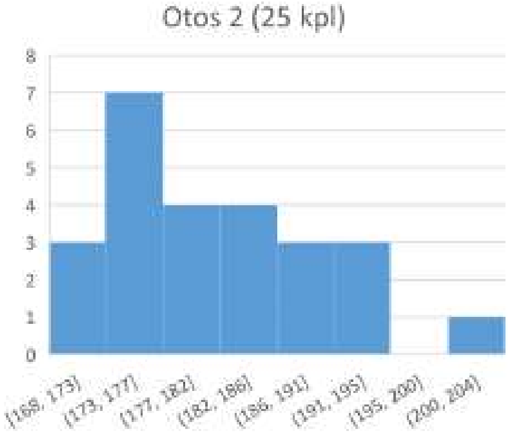
SAMPLING ERRORS

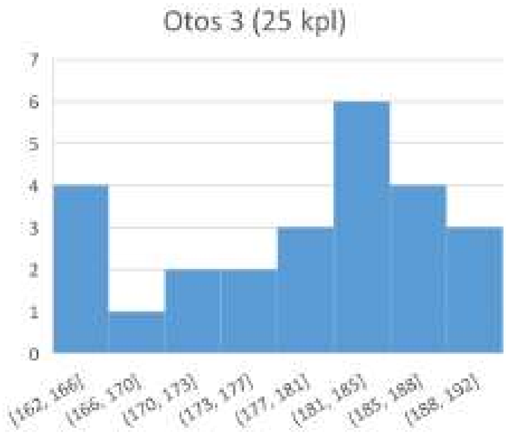
Height of 18 year old men follows normal distribution with mean = 180 cm and standar deviation = 8 cm.



When you measure heights of men of samples of size 25, all samples look different. By increasing sample size sample distribution would get closer to the normal distribution.

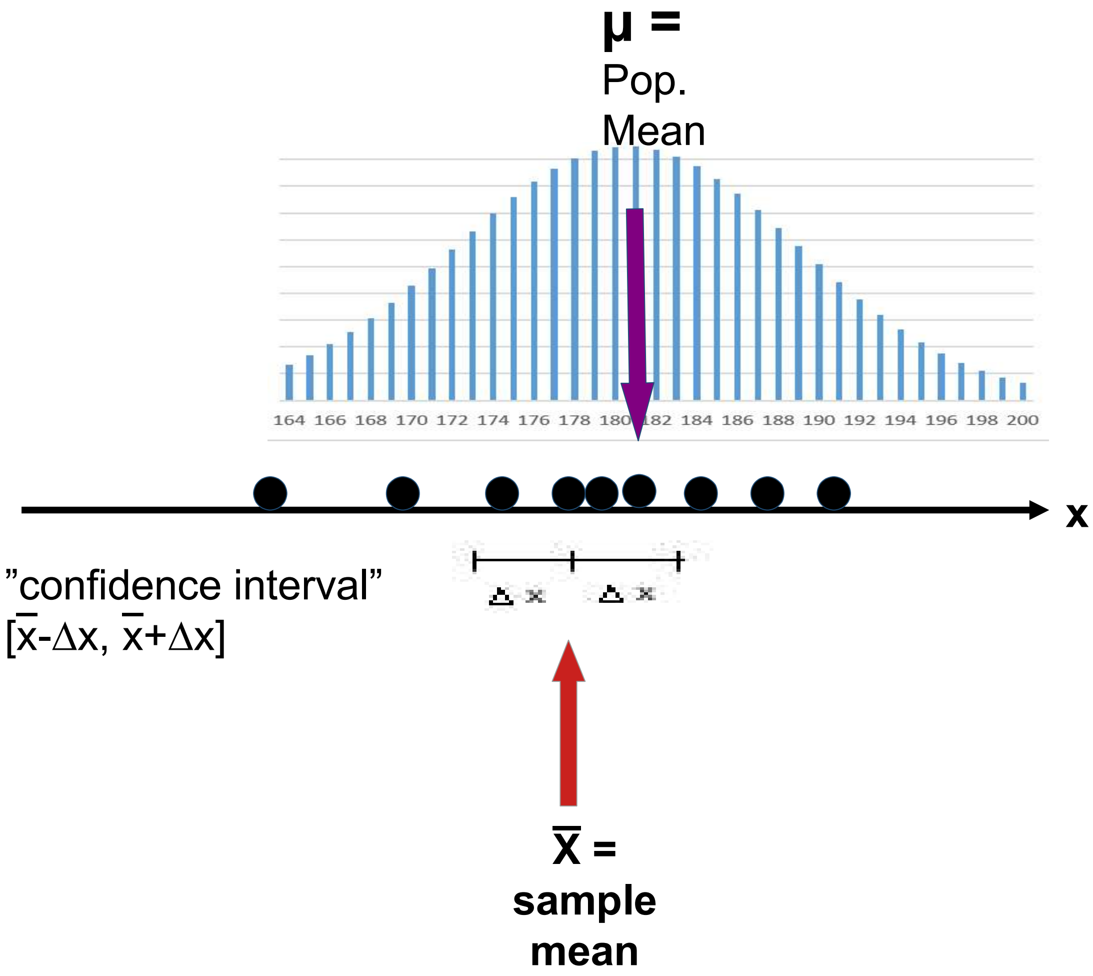






However sample means 180.3 cm, 180.7 cm and 181.3 cm are not very far from the population mean.

Goal is to determine the population mean μ of variable x using sample.



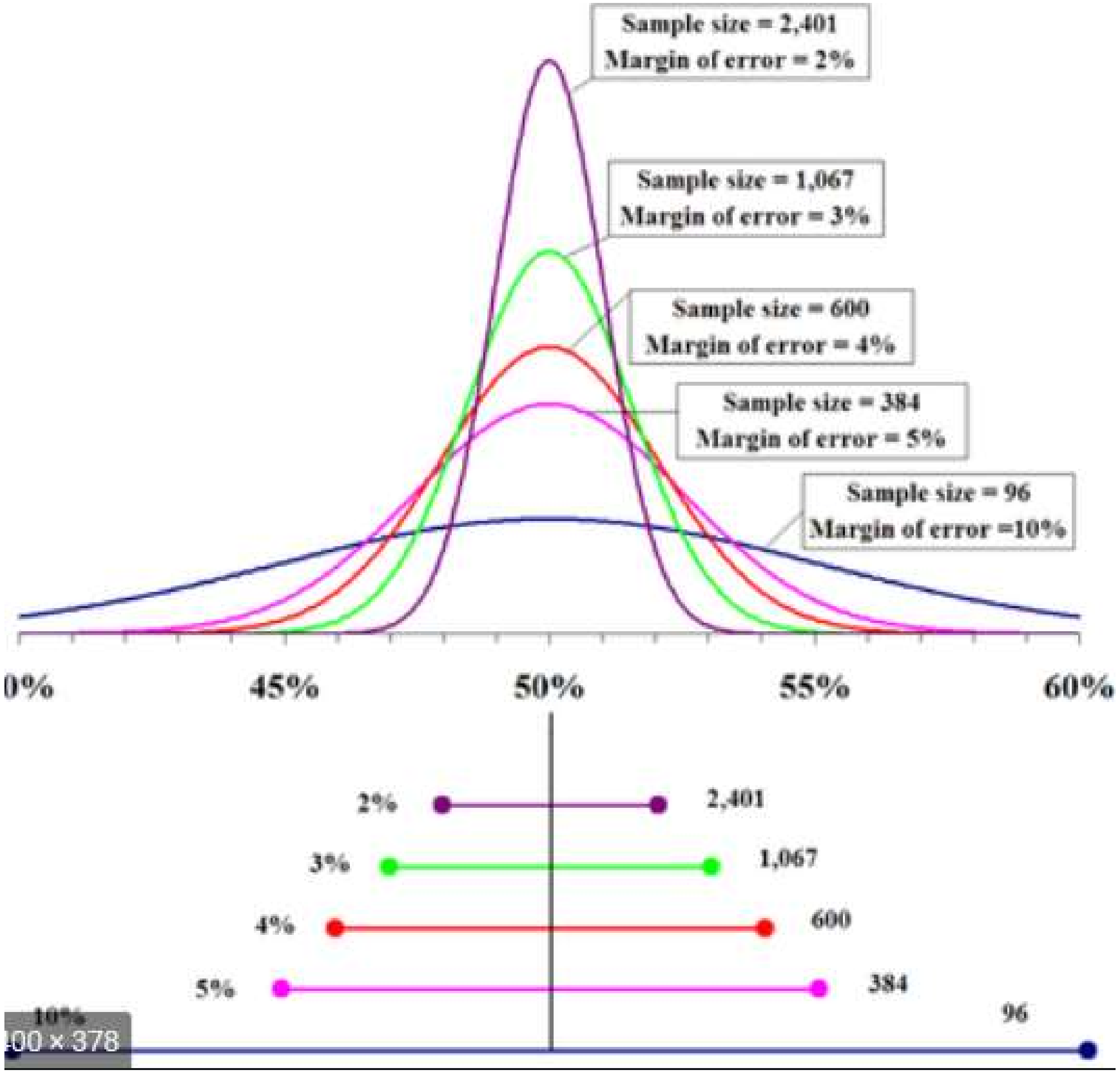
How far from the sample mean x can be from the population mean μ?

If x is normally distributed, sampling error could be in theory very large, but probablility for large error is close to zero.

Next slides show formulas of estimation of sampling error. Formulas gives error margin Δx based on desired certainty levels, which we call significance levels.

Standard significance levels are 95% (or sometimes 99%)

Significance level means, that in 95% of random samples the population mean μ lies in interval $[\bar{x}-\Delta x, \bar{x}+\Delta x]$, which is called confidence interval.



The best estimate for population mean μ is sample mean $\bar{x}$.

Picture shows how variation of sample means decreases, when sample size n gets higher

=> The bigger is the number of measurement, the more accurate is the result.

Picture shows the error margins with different sample sizes in candidate support polls of presidential election.
To reach 2% error margin, sample size should be about 2500.

How far can the population mean be from the sample mean?

1) Best estimate for the population mean μ is the sample mean $\bar{x}$.

2) Theoretical result: If the standard deviation of variable X of population **= σ ,**
then **the standard deviation of sample means = σ/√n**. If σ is not known, we use **s**.

3) To calculate the error margin we need to multiply s/√n with a coverage factor t,
4) which depends on the confidence level we aim to reach.

$$\Delta x = t\frac{s}{\sqrt{n}}$$

s = sample standard deviation
n = sample size
t = coverage coefficient called critical value

4) If sample size ≥ 500, we can use coverage coefficients t based on normal distribution:

- **at confidence level of 95%, coefficient t = 1.96,**
- **at confidence level of 99%, coefficient t = 2.57**

5) If sample size is smaller than 500, values of t are based on Student's distribution, in which t values are larger. (table on the next slide)

6) Interpretation: population mean μ is at 95% probability in **interval**

$$\overline{x} - t\frac{s}{\sqrt{n}} < \mu < \overline{x} + t\frac{s}{\sqrt{n}}$$

Student's distribution: Table of t – values

called ”critical values” or ”coverage factors”

”Degrees of freedom” f = n-1	95% significance level	99% significance level
6	2.447	3.707
7	2.365	3.499
8	2.306	3.355
9	2.262	3.250
10	2.228	3.169
11	2.201	3.106
12	2.179	3.055
13	2.160	3.012
14	2.145	2.977
15	2.131	2.947
16	2.120	2.921
17	2.110	2.898
18	2.101	2.878
19	2.093	2.861
20	2.086	2.845
21	2.080	2.831
30	2.042	2.750
40	2.021	2.704
50	2.010	2.679
60	2.000	2.660
∞	1.960	2.576

n = sample size

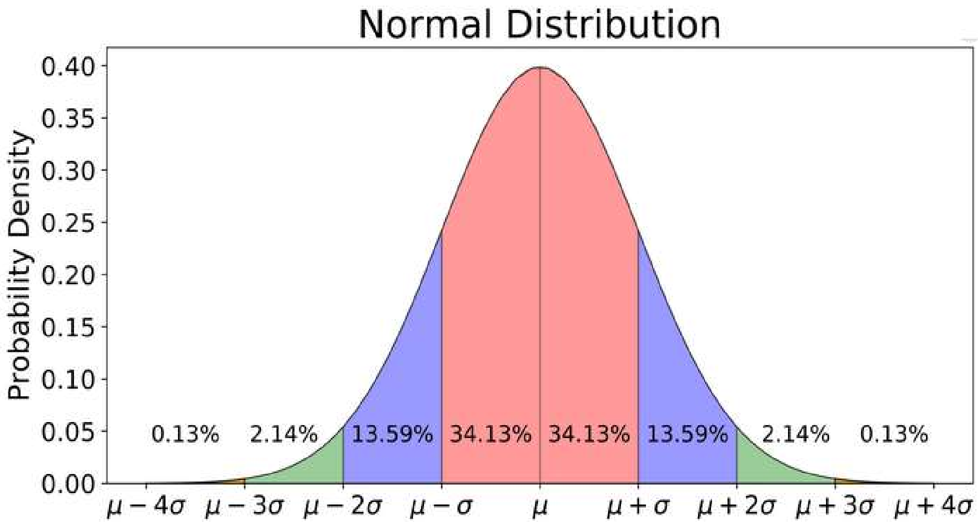
Use of the table:

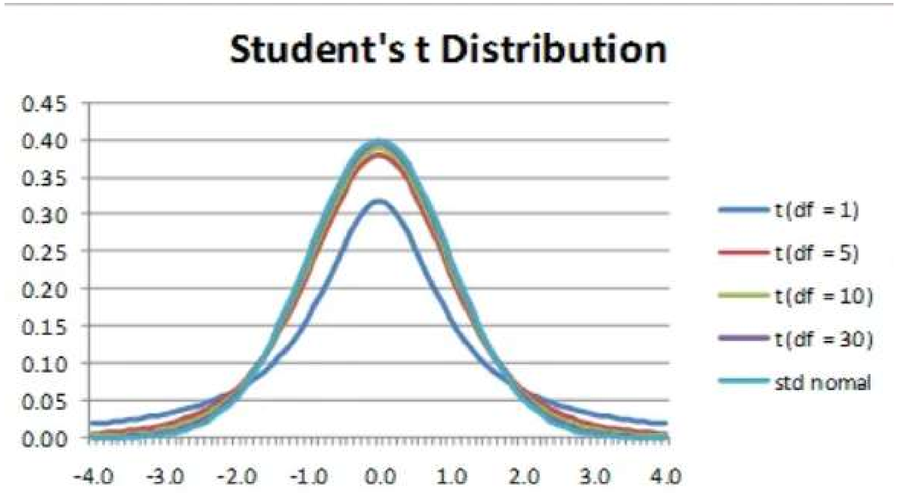
If sample size is large (> 500), use t = 1.96 at 95% siginificance level

If sample size is smaller, look the t- value from row, which corresponds the sample size.

For example: If sample size is about 30, use t = 2.042

Excel functions confidence.norm and confidence.t





For large sample size (n > 500), sample means follow normal distribution

If n is large, we could use **Excel function confidence.norm to evaluate Δx using fixed t = 1.96 in formula**

$$\Delta x = t\frac{s}{\sqrt{n}}$$

For (n <500), sample means follow t – distribution, which depends on n (see picture)
For n > 500, Student's distribution and normal distributions are nearly equal.

If n < 500 , use Excel function confidence.t
Notice: confidence.t can be used in all cases. Its result are more accurate and better.

Example. Health institute wanted to determine the average blood sugar level of 70 year old citizens. In their research sample size was 200.

6,61	6,26	7,13	3,31	5,91	4,86	5,15	6,37	6,45	6,14	4,95	7,30	5,92	5,64	4,55	6,30	6,12	7,75	5,91	5,69
5,64	7,04	6,35	7,06	3,31	6,26	7,13	6,10	7,97	5,33	6,03	4,95	6,88	7,67	6,49	6,82	6,55	5,09	4,56	6,01
5,58	8,36	7,83	6,41	7,59	4,41	7,23	5,92	6,98	4,73	6,07	5,47	7,02	6,27	6,04	6,95	6,25	4,38	5,61	6,36
4,61	5,45	6,49	5,82	5,02	7,58	6,45	7,42	5,37	6,51	6,92	6,35	5,65	5,78	5,17	5,11	6,10	5,66	6,52	5,34
5,33	5,85	6,15	5,28	5,75	5,68	5,90	5,77	7,44	6,36	6,05	4,88	7,46	7,40	5,64	5,36	7,03	5,80	6,79	4,55
4,98	6,69	6,14	5,32	7,21	5,16	7,26	5,94	4,83	6,73	5,68	5,79	6,46	5,92	6,41	4,50	5,98	7,63	8,09	6,10
4,64	5,91	5,99	5,32	5,18	6,40	5,92	5,35	5,93	5,90	5,69	5,12	6,35	6,59	6,00	7,31	6,37	4,95	5,74	6,82
5,68	6,93	6,54	6,50	6,10	6,51	4,76	5,66	7,33	6,04	4,67	6,70	4,69	5,73	4,56	7,37	5,47	6,64	6,10	4,84
6,06	5,68	3,76	7,42	6,13	4,98	7,00	5,59	5,47	6,51	5,01	5,88	5,80	5,29	7,11	6,50	6,57	5,02	7,44	5,20
5,04	6,82	6,60	6,89	6,42	5,16	8,09	5,50	3,82	4,58	3,91	6,13	6,65	5,32	5,49	6,34	7,27	6,39	4,11	6,18

average	6,01
standard deviation	0,94

What is a) the average blood sugar level of population and b) its error margin at 95% confidence level?

a) Average blood sugar = 6.01
b) error margin = $t*s/\sqrt{n}$ = $1.96*0.94/\sqrt{200}$ = 0.13

=> **average blood sugar level = 6.01 ± 0.13**

Using confidence inteval, this can be written

$$5.88 \le \mu \le 6.14$$

Error margin Excel function.

=confidence.t(5%;0,94;200)

Arguments:
Significance level = 100%- confidence level = 100%-95% = 5%
Sample mean = 0.94
Sample size = 200

Confidence margin of proportions

In opinion surveys the error margin depends on the sample size according to the following formula.

$$\Delta p = t\sqrt{\frac{p(1-p)}{n}}$$

p = measured proportion in decimal form
n = sample size
t = coverage factor = 1.96 at confidence level of 95% (most usual)

A newspaper The Times made a research of support of Brexit in 2016. Brexit got 58.0% support among 1000 participants of the survey. Calculate the error margin of the measured value.
(Brexit = leaving EU)

$$\Delta p = t\sqrt{\frac{p(1-p)}{n}} = 1.96\sqrt{\frac{0.58(1-0.58)}{1000}} = 0.031 = 3.1\%$$

Answer: Error margin was 3.1%

Interpretation: "Support percent of Brexit lies in interval [58.0 – 3.1, 58.0 + 3.1] at 95% probabillity"

Formula for calculation of the sample size for a given error margin

Solving n from $\Delta p = t\sqrt{\frac{p(1-p)}{n}}$ gives

Sample size formula: $$\boxed{n = \frac{t^2 p(1-p)}{\Delta p^2}}$$

How big sample size would give 2.0 % error margin in the previous example, where p = 58% ?

$$n = \frac{t^2 p(1-p)}{\Delta p^2} = \frac{1.96^2 \cdot 0.58(1-0.58)}{0.02^2} = 2340$$

Sample size should be close to 2500.

Solved examples

200 randomly chosen customers of CityMarket were asked about their age.
Sample mean was 41.3 years and sample standard deviation was 23.5 years.
Estimate the mean age of all customers. Give also the error margin

Excel- function = CONFIDENCE.T(5% ; 23.5 ; 200) gives 3.3

Or alternatively = CONFIDENCE.NORM(5% ; 23.5;200)

Result: Mean age = 41.3 ± 3.3 years

500 randomly chosen individuals were asked, whether they use the bank card's contactless payment feature at the store's checkout. 47.5 % answered YES. Estimate how many percents of all customers use that feature. Give also the error margin

$$\Delta p = t\sqrt{\frac{p(1-p)}{n}} = 1.96\sqrt{\frac{0.475(1-0.475)}{500}} = 0.044 = 4.4\%$$

Answer: 47.5% ± 4.4 %

Error margins of linear regression model parameters

Find a linear model y = a x + b based on the table below. Variables: x = temperature, y = heating costs. Determine the confidence margins of parameters a and b. Calculate the coefficient of determination R^2, which is a statistical measure of how well regression line approximates the observed data.

Mean temperature in Rovaniemi	Households heating costs (€)
-17,3	338
-12,2	294
-5,2	260
-1,2	203
9,7	128
13,4	99
17,4	58
18,2	65
9,3	134
1,7	180
-3,5	234
-9,2	256

1. Paint 3 x 3 cell area for results

-7,73	201,00
0,24	2,77
99,0 %	9,49

2. Insert function LINEST

3. Fill in the argument list

Known y's	D$4:D$15
Known x's	C$4:C$15
Type	1
Stats	1

paint y values

paint x values

4. Finally PRESS SHIFT-CTRL-ENTER combination to get all the answers:

1st row: parameters a and b
2nd row: standard errors of a and b
3rd row: R -squared value and standard error of predicted y

Summary of formulas of estimation the error of mean

$$\Delta x = t\frac{s}{\sqrt{n}}$$

Excel: confidence.t

Error margin of mean of numeric variable
n = sample size, s = sample standard deviation
t depends on sample size and significance level

$$\Delta p = t\sqrt{\frac{p(1-p)}{n}}$$

Error margin of proportion
n = sample size, **p** = proportion (measured of sample)
t depends on significance level, for 95% level t = 1.96

$$n = \frac{t^2 p(1-p)}{\Delta p^2}$$

Formula for calculation of adequate sample size
p = proportion (measured), Δp = error of proportion
t depends significance level, for 95% level t = 1.96

Estimation of error margins of parameters **a** and **b** of linear model **y = a x + b** is performed using Excel function **LINEST**

Explanation of the results given by LINEST function

-7,73	201,00
0,24	2,77
99,0 %	9,49

Linear model: $y = -7.73*x + 201.0$

(If mean temperature of some month = 10°C, then heating costs are -7.73*10+201 = 193.27)

Standard errors of a and b are: s_a= 0.24 and s_b=2.77.

The sample size = 12 => we need to use coverage factor value t = 2.2 to get the confidence margins for a and b:

See slide 7 table of t – values.

N – 1	t
10	2.228
11	2.201
12	2.179

Confidence margins of a and b :

$\Delta a = t*s_a = 2.2*0.24 = 0.52$ and $\Delta b = t*s_b = 2.2*2.77 = 6.1$

Coefficient R^2 = 99% => Model explains 99% of the variation of y values. (The observed points lie almost perfectly on the regression line. If all points would lie exactly on the line, R^2 would be 100%). **=> Model is excellent.**

Standard error for the y -estimate = 9.5

The standard devitation of differences between observed y -values and y – values calculated from the model y = a x + b

Optimization / lec1

1. Limit values of a function f(x)
2. Derivative at a given point $f'(x_0)$
3. Derivative function $f'(x)$
4. Derivation rules of simple functions
5. Derivative in calculators

1. Limit values of a function

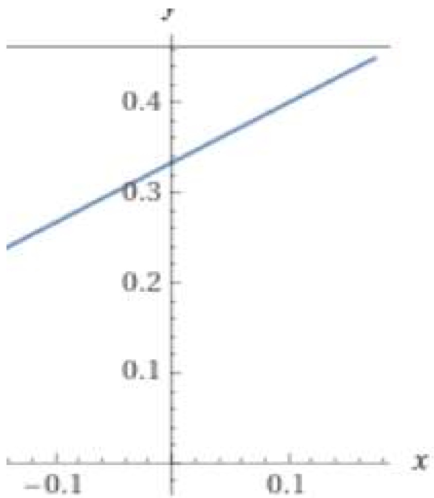
Consider a rational function $f(x) = \dfrac{2x^2 + x}{3x}$

1. Function f(x) is not defined at point x = 0

2. If $x \neq 0$, f(x) can be simplified as $\dfrac{2x^2+x}{3x} = \dfrac{x(2x+1)}{3x} = \dfrac{(2x+1)}{3}$

3. Graph of f(x) is the same as graph of (2x+1)/3 ,
with the exception that at one point at x = 0 is missing

plot	$\dfrac{2x^2+x}{3x}$



4. The graph shows, that the value of f(x) approaches
a specific value, as x approaches 0.

We call this value the limit value of f(x) as x tends to 0.

Notation for limit value is $\lim\limits_{x \to 0} \dfrac{2x^2+x}{3x}$

Calculation of the **limit value of a rational function at a zero of its denominator.**

"Rational function" means a function $f(x)=\dfrac{P(x)}{Q(x)}$, where P(x) and Q(x) are polynomials

At a point $x = x_0$, where both $P(x_0)$ and $Q(x_0)$ are zero, f(x) can be simplified by reducing by common factor $x - x_0$. After reduction limit value at $x = x_0$ can be calculated by substitution.

Ex. Calculate: $$\lim_{x \to 0} \frac{2x^2 + x}{3x}$$

Simplification: $$\frac{2x^2 + x}{3x} = \frac{x(2x+1)}{3x} = \frac{(2x+1)}{3}$$

Substitution x = 0 gives: $$\frac{2*0+1}{3} = \frac{1}{3}$$

Ex2. Calculate limit value of $\dfrac{x^2-4}{x-2}$ **, as x -> 2**

Direct substitution of x = 2 gives 0/0 .

Formula
$(a^2-b^2) = (a-b)(a+b)$

Simplification : $$\frac{x^2-4}{x-2} = \frac{(x-2)(x+2)}{x-2} = x+2$$

Substitution x = 2 gives 2 + 2 = 4

Limits of other functions can be calculated with calculator

Ex3. Calculate limit value of $\dfrac{e^x - 1}{x}$ **as** $x \to 0$

Direct substitution of x = gives $(e^0 - 1)/0 = (1-1)/0 = 0/0$

Simplification by simple methods is not possible =>

The limit value is calculated using WolframAlpha calculator

$$\boxed{\text{limit } \frac{e^x - 1}{x} \text{ as } x \to 0}$$

returns $\displaystyle \lim_{x \to 0} \frac{e^x - 1}{x} = 1$

Notation convention

$$\lim_{x->0}\frac{3x^2+2x}{2x}=\lim_{x->0}\frac{x(3x+2)}{2x}=\lim_{x->0}\frac{(3x+2)}{2}=\frac{0+2}{2}=1$$

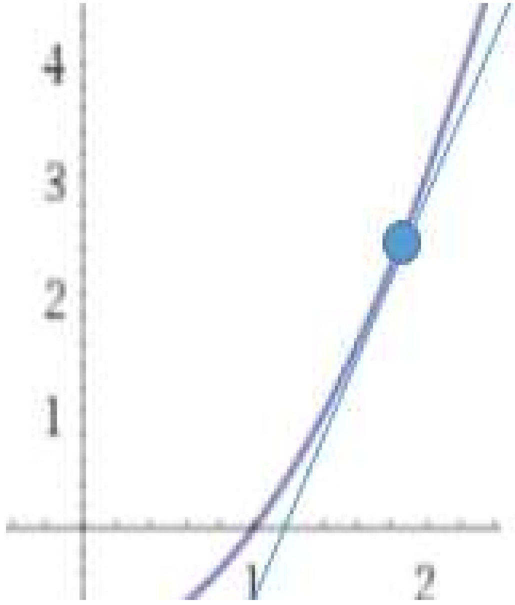
Read: ” (3x2+2x)/(2x) tends to 1 as x tends to 0 ”

$$\lim_{x->2}\frac{x^2-4}{x-2}=\lim_{x->2}\frac{(x-2)(x+2)}{x-2}=\lim_{x->2}(x+2)=2+2=4$$

Read: ” (x^2-4)/(x-2) tends to 4 as x tends to 2 ”

Definition of the derivative

The value of the derivative of f(x) at $x = x_0$ is a measure of rate of change of f(x) at that point



<u>Graphical definition of derivative</u>

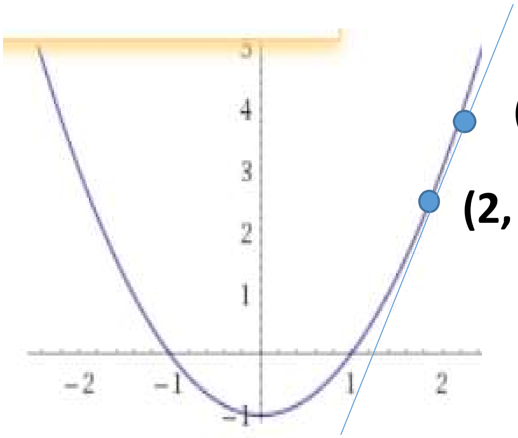
The derivative of function f(x) at $x = x_0$
is the slope of the tangent line at $x = x_0$.

The notation of the derivative is $f'(x_0)$

Calculation of derivative at a given point

Example. Determine the value of derivative of function $y = x^2 - 1$ at $x = 2$

Consider the "secant line" between points of the curve



x	$y = x^2 - 1$
2	3
2 + h	$(2+h)^2 - 1$

Slope $$k = \frac{y_2 - y_1}{x_2 - x_1} = \frac{(2+h)^2 - 1 - 3}{h} = \frac{(2+h)^2 - 4}{h}$$

The limit value of k as h → 0 is the derivative (=the slope of the tangent)

$$\lim_{h \to 0} \frac{(2+h)^2 - 4}{h} = 4$$

Answer: f' (2) = 4

Using W.A

limit ((2+h)^2-1-3)/h as h->0

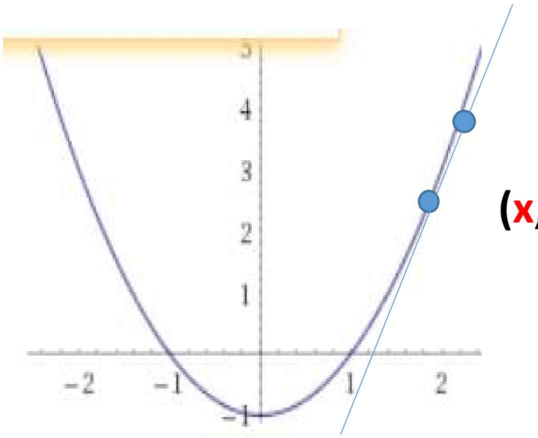
$$\lim_{h \to 0} \frac{(2+h)^2 - 1 - 3}{h} = 4$$

Derivative function

Example: Determine the derivative function of $y = x^2 - 1$ = general expression for calculation of the derivative at any point x.

Consider points at x and x + h

x	$y = x^2 - 1$
x	$x^2 - 1$
x + h	$(x+h)^2 - 1$



$(x+h, (x+h)^2 - 1)$

$(x, x^2 - 1)$

slope $$k = \frac{y_2 - y_1}{x_2 - x_1} = \frac{(x+h)^2 - 1 - (x^2 - 1)}{h} = \frac{(x+h)^2 - x^2}{h}$$

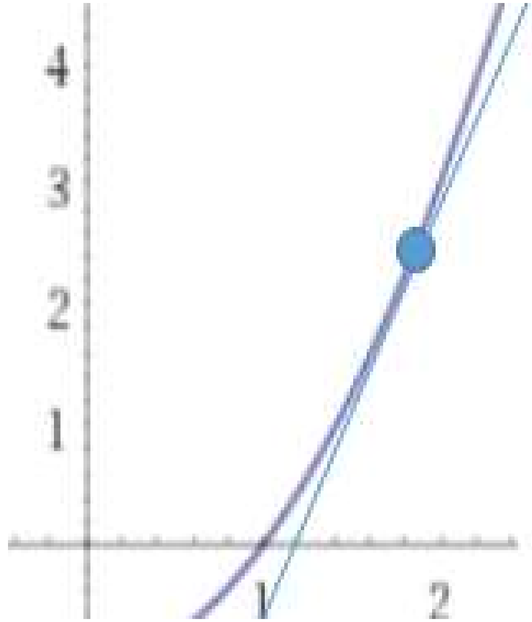
Derivative is the limit value of k as h → 0

WolfamAlpha $$\lim_{h \to 0} \frac{((x+h)^2 - 1) - (x^2 - 1)}{h} = 2\,x$$

Answer: f ’ (x) = 2 x

Algebraic definition of derivative function

$$f'(x) = \lim_{h->0} \frac{f(x+h) - f(x)}{h}$$



All following notations mean the derivative of function y = f(x)

$$f'(x) \qquad y' \qquad \frac{df(x)}{dx} \qquad \frac{dy}{dx} \qquad Df(x)$$

From now on we use mostly notation Df(x)

D(x^2 – 1) = 2 x

Derivatives are mostly calculated either using formulas of derivation or calculators

Derivation formulas for some basic functions:

1. Derivative of a constant function y = c:

$$Dc = 0$$

2. Derivative of power function

$$Dx = 1$$

$$Dx^n = n \cdot x^{n-1}$$

3. Power function with negative exponent.

$$D\frac{1}{x^n} = -\frac{n}{x^{n+1}}$$

Solved examples

1 a) Determine the derivate function of $f(x) = x^5$

a) b) Calculate the slope of the tangent of $f(x)$ at $x = 2$

b) c) Calculate the slope of the tangent of $f(x)$ at $x = -1$

a) $Dx^5 = 5x^4$

b) $f'(2) = 5*2^4 = 80$

c) $f'(-1) = 5*(-1)^4 = 5$

2. Calculate following derivates:

a) $D\, x^{2015}$

b) $D\, \frac{1}{x}$

c) $D\frac{1}{x^5}$

a) $2015\, x^{2014}$

b) $-\frac{1}{x^2}$

c) $-\frac{5}{x^6}$

Derivative of a polynomial function

Laws for derivation of linear combinations of basic functions:

1) D a f(x) = a Df(x)
2) D (f(x) + g(x)) = D f(x) + D g(x)

Example calculate the derivative

$D(-3x^3 + 5x^2 - 4x + 7)$

Applying above rules we get

$= -3*Dx^3 + 5*Dx^2 - 4*Dx$

$= -3*3x^2 + 5*2x - 4*1 + 0$

$= -9x^2 + 10x - 4$

Check result with W.A

D (- 3 x^3+ 5 x^2- 4 x + 7)

$= -9x^2 + 10x - 4$

Derivation rules of special functions

$D\ e^{x} = e^{x}$

$D \ln(x) = 1/x$

$D \sin(x) = \cos(x)$

$D \cos(x) = -\sin(x)$

We do not need or use these rules in this course

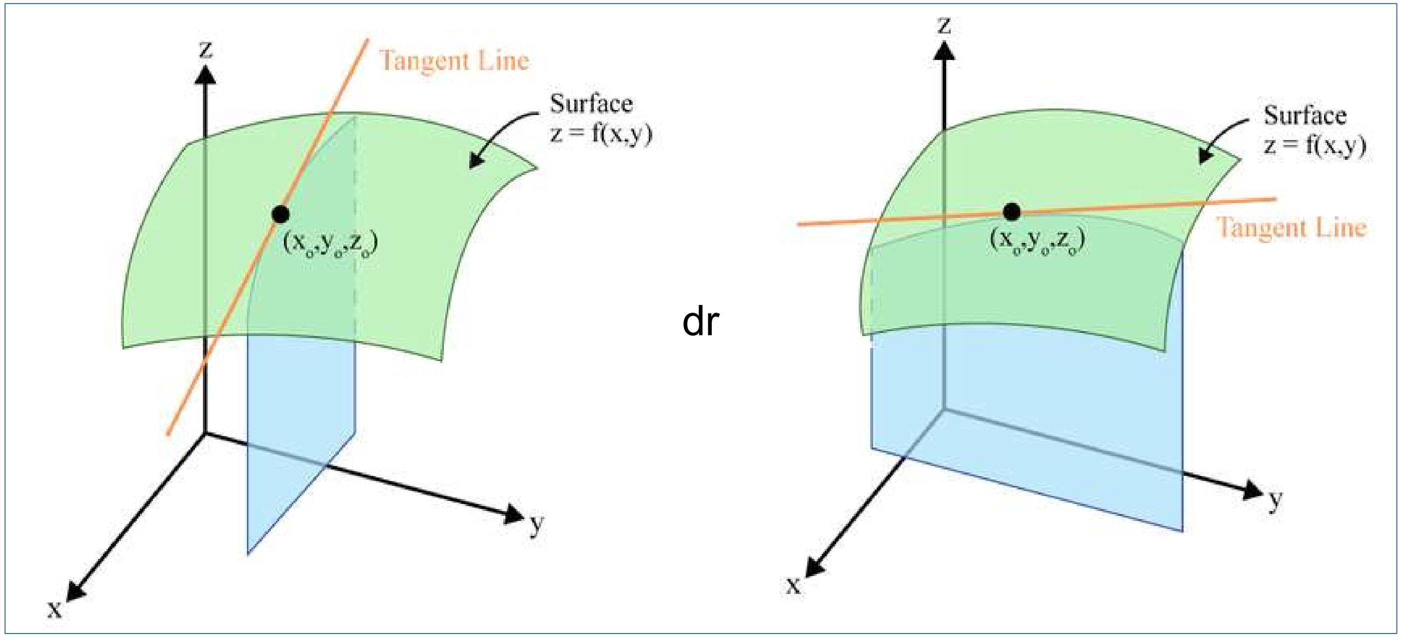
Functions of several variables

Partial derivatives

Graphical interpretation

Gradient vector

A function of two variables z = f(x,y) presents a surface in 3D space.



At any point of the surface we can draw tangent lines to directions of both coordinate axes.

The slopes of these tangent lines are called partial derivatives of f(x,y).

Calculation of partial derivatives

Derivation is applied to one variable at the time, while the other variables are considered as constants

A function of many variables has derivatives with respect to all variables

D (y*x^2 + 4 x + 5 y , x) = y*2x + 4*1 + 0 = 2 x y + 4

D (y*x^2 + 4 x + 5 y , y) = 1*x^2 + 0 +5*1 = x^2 +5

Another notation of partial derivatives: $\dfrac{\partial f}{\partial x}$ $\dfrac{\partial f}{\partial y}$ or simply f_x, f_y

In these slides we use mostly notation D(f,x), D(f,y)

Also on WolframAlpha you can write D (y*x^2 + 4 x + 5 y , x) , which gives $2 x y + 4$

Example. Linear function f(x,y) = 2 x + 3 y – 1 represents a plane

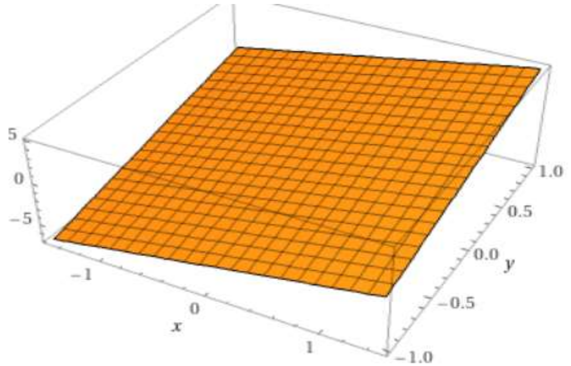
Calculate the partial derivatives of f(x,y)

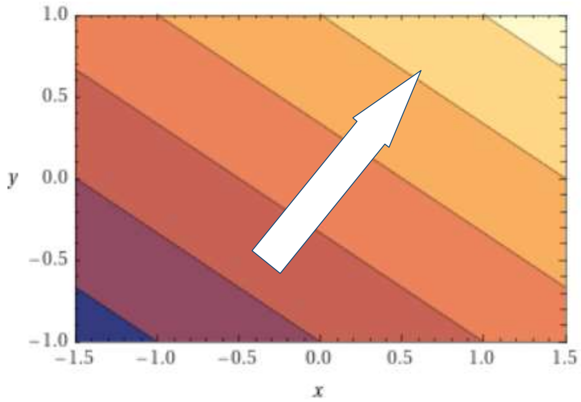
D(2 x + 3 y – 1,x) = 2

D(2 x + 3 y – 1,y) = 3

The vector of partial derivatives
(f_x, f_y) = (2, 3) gives the direction in which the function grows fastest

plot 2 x + 3 y − 1



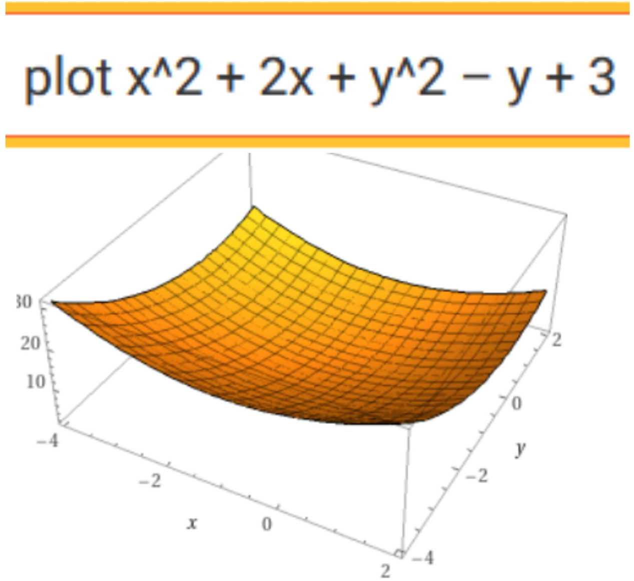


Example. Function $f(x,y) = x^2 + 2x + y^2 – y + 3$ represents a paraboloid surface

Calculate the partial derivatives of f(x,y)

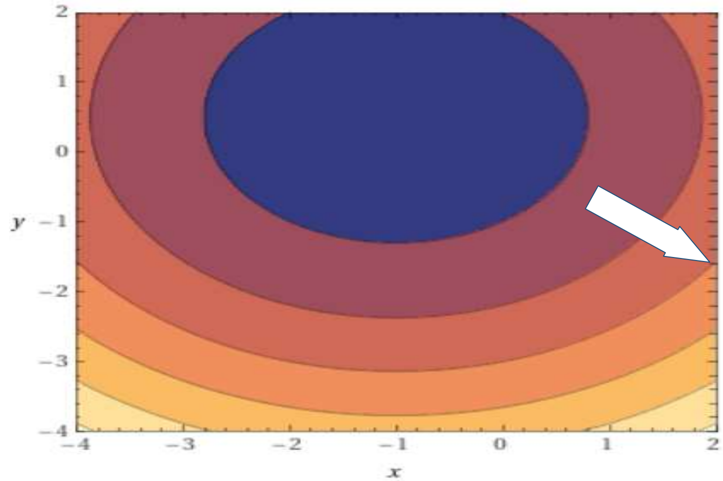
$$D(x^2 + 2x + y^2 – y + 3,x) = 2x + 2$$

$$D(x^2 + 2x + y^2 – y + 3,y) = 2y -1$$



Calculate the values of partial derivatives at point x = 1 , y = 0

f_X (1, 0) = 2*1 + 2 = 4
f_Y (1, 0) = 2*0 - 1 = -1



At point (1,0) gradient vector (f_x,f_y) = (4,-1)
It shows the direction in which function grows fastest

Gradient vector of a multivariate function

Grad f(x,y) = (f_x , f_y) is the vector of partial derivatives

It is denoted often in mathematics as $\nabla f(x, y)$

Operator ∇

is called "nabla"

WolframAlpha command

gradient of x^2 +2x +y^2 - y + 3

gives (2 (1 + x), -1 + 2 y)

It can be simplified to normal form (2x + 2, 2 y - 1)

gradient of x^2 +2x +y^2 - y + 3 where x = 1, y = 0

calculates the gradient vector at point (1,0). Answer is (4,-1)

Functions of more than 2 variables cannot be visualized.

Example: $f(x,y,z) = x^2 + y^2 + z^2$ is a function of 3 variables.

The partial derivatives $f_x=2x$, $f_y = 2y$ and $f_z = 2z$ measure the rate of change of function f in the directions of coordinate axes.

The gradient vector is (2x,2y, 2z)

Gradient vector at point (3,2,-1) = (6, 4, -2)

=> At point x=3, y=2 , z=-1 vector (6,4,-2) gives the direction in which function value increases most rapidly.

Extreme value problems

The goal is to find on the miminum or maximum value of a function.

Applications:

A) optimization of parameters of a model to find the best fit

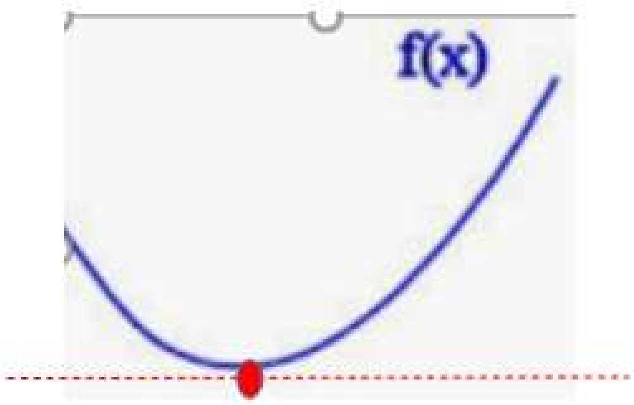
(examples: regression analysis, machine learning)

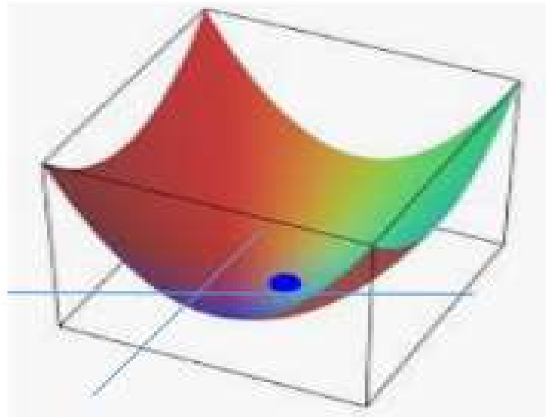
B) maximize profit

C) minimize costs, time or use of materials

Concept: The objective function

(= the function of which we wish to find minimum or maximum)





Extreme values of one variable function are often found at points, where the derivative is zero.

(called critical points)

Extreme values of a many variable function are mostly found at points, where both partial derivatives vanish. (derivatives to directions of both axes) are zero.

Methods of optimization

1. ANALYTICAL METHOD

Extreme values of a single variable function f(x) are often found at points where derivative f'(x) = **0**

Extreme value of a multivariable function f(x,y,...) are often found at points, where all **partial derivatives D(f(x),x), D(f(x),y),...** are zero.

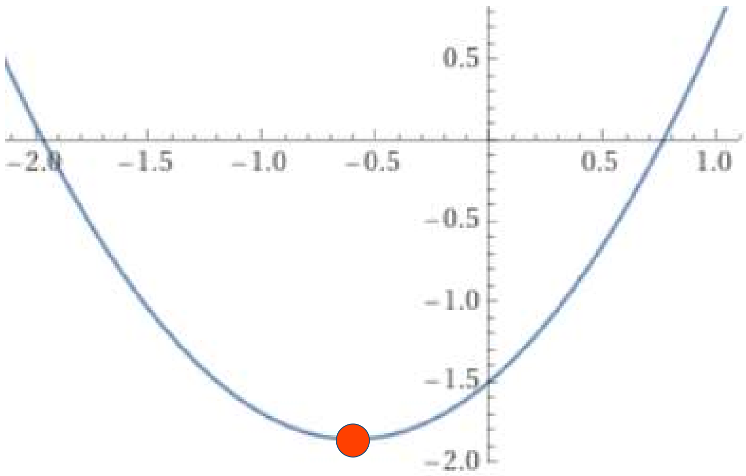
2. ITERATIONAL METHODS

Extreme values are found by moving in steps in direction of gradient vector until maximum is found (or in direction of -gradient vector until minimum is found).

Algorithm "Gradient Descend Method" for finding minimum is used in Machine Learning

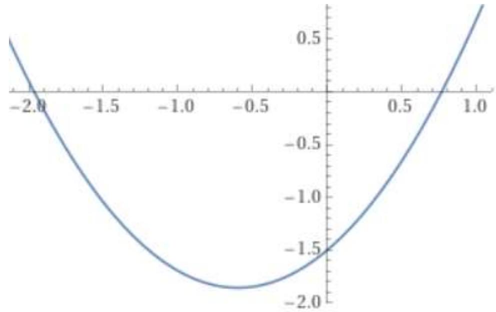
Example1. Find the minimum value of $f(x) = x^2 + 1.2x - 1.5$

Next slides show different methods to solve the problem



Find the minimum value of $f(x) = x^2 + 1.2x - 1.5$

plot x^2+1.2x-1.5



Method A: Analytical method (manual version)

1. Calculate the derivative function

$$f'(x) = 2x + 1.2$$

2. Solve the zero of derivative

$$2x + 1.2 = 0 \Rightarrow 2x = -1.2 \Rightarrow x = -0.6$$

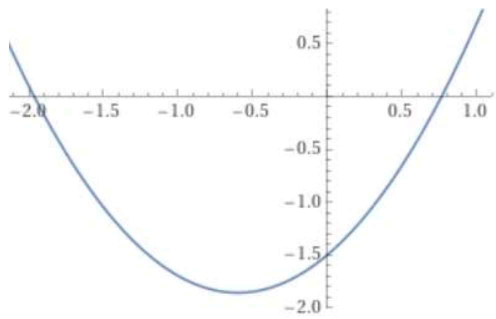
3. Calculate f(-0.6)

$$f(-0.6) = (-0.6)^2 + 1.2*(-0.6) - 1.5 = -1.86$$

Answer: f(x) has a minimum value -1.86 at x = -0.6

Find the minimum value of $f(x) = x^2 + 1.2x - 1.5$

plot x^2+1.2x-1.5



Analytical method using WolframAlpha calculator

Calculate zero of derivative (solve f'(x) = 0)

solve D(x^2 +1.2*x -1.5)=0

Result: x = -0.6

Calculate f(x) at point x = -0.6

x^2 +1.2*x -1.5 where x=-0.6

Result: x =-1.86

Direct solution using WolframAlpha's minimize command

minimize x^2 +1.2*x -1.5

Global minimum

$\min\{x^2 + 1.2\,x - 1.5\} = -1.86$ at $x = -0.6$

Find the minimum value of $f(x) = x^2 + 1.2x - 1.5$

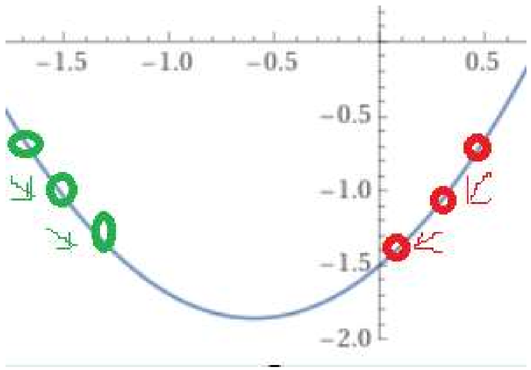
Method B. Iteration
Gradient descent method

1.Input start value for x

2. If f'(x) > 0, step dx to the left
If f'(x) < 0 , step dx to the right

step size = dx = - f '(x)*t
where t is a coefficent (for ex. 0.2)

Near minimum |dx| decreases and tends to zero.




```
# Python code
# 1. Define function f(x)
def f(x):
   return x**2 + 1.2*x-1.5
# Define derivative function fx(x)
def fx(x):
   return 2*x + 1.2

# 2. Set an arbitrary initial value for variable x
x = 2.0
# 3. Define step coefficient t and initialize step dx=-fx(x)*t
t = 0.2
dx=-fx(x)*t
#4. Create a loop to update x until |dx| <0.01
while abs(dx)>0.01:
    dx=-fx(x)*t
    x=x+dx
    print(f"x= {x:5.2f}, y = {f(x):7.4f}")
#5. Print coordinates of minimum
print(f"\nMinimum : x= {x:5.2f}, f(x) = {f(x):5.2f}")
```

Find the minimum value of $f(x) = e^x – 2.5 x + 1$

Method A: Analytical method (manual version)

1. Calculate the derivative function

$$f'(x) = e^x – 2.5$$

2. Solve the zero of derivative

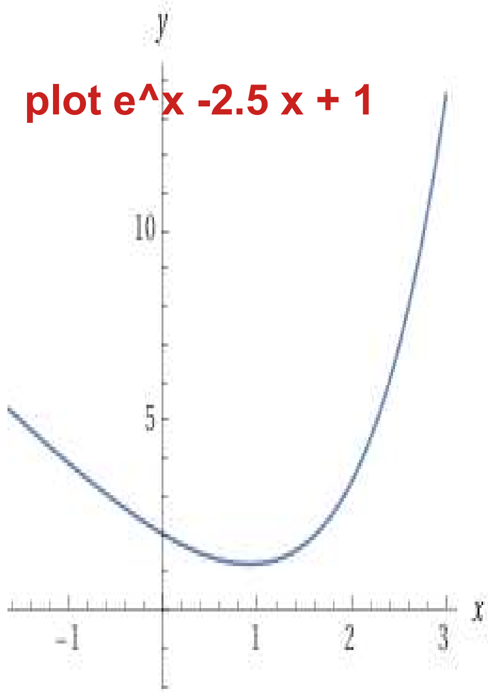
$$e^x – 2.5 = 0 \Rightarrow e^x = 2.5 \Rightarrow x = \ln(2.5) = 0.916$$

3. Calculate f(0.916)

$$f(0.916) = e^{0.916} – 2.5*0.916 + 1 = 1.21$$

Answer: Minimum 1.21 at x = 0.916

plot e^x -2.5 x + 1



EXCEL – version of Gradient Descend Method

	A	B	C		
1	*step coefficient t*	**0,2**			
2					
3		**X**			
4	*initial value*	**0**			
5		0,300	iteration formula	= B4 – (exp(B4)-2,5)*B1	
6		0,530			
7		0,690			
8		0,791			
9		0,850			
10		0,882			
11		0,899			
12		0,908			
13		0,912			
14		0,914			
15		0,915			
16		0,916			
17		0,916	minimum value at 0,916		
18		0,916			

1. Write step coefficient t and initial value for x in cells marked yellow

2. Iteration formula

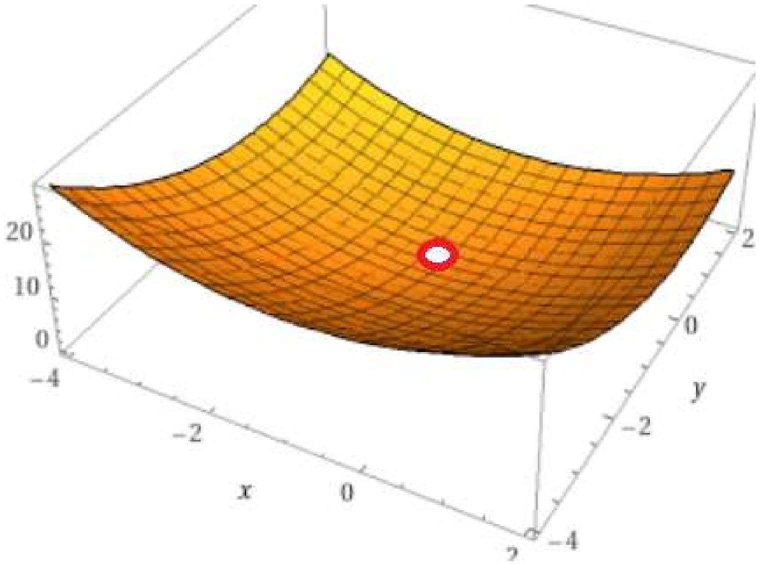
New x value = x – f'(x)*t
In this example is:
New x value = x – (e^x-2,5)*t

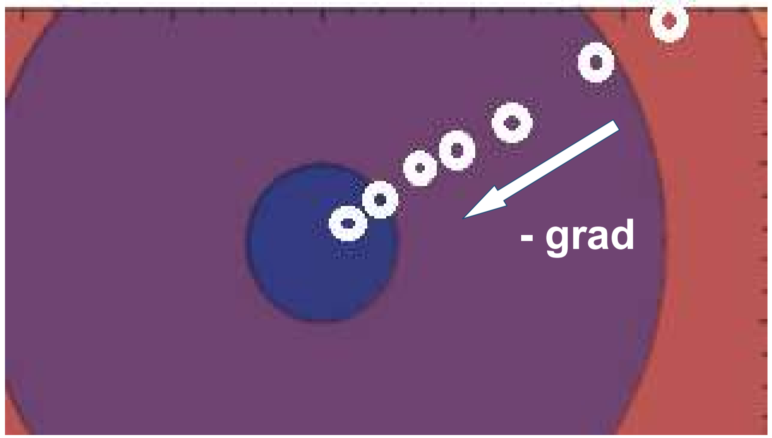
is written using cell addressess

The iteration formula is copied down.

The values stabilize to the x value, which corresponds the minimum

Example2. Find the minimum of $f(x,y) = x^2 + y^2 + 2x - y + 1$



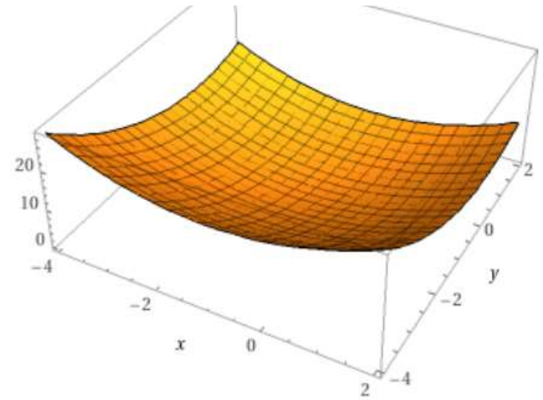


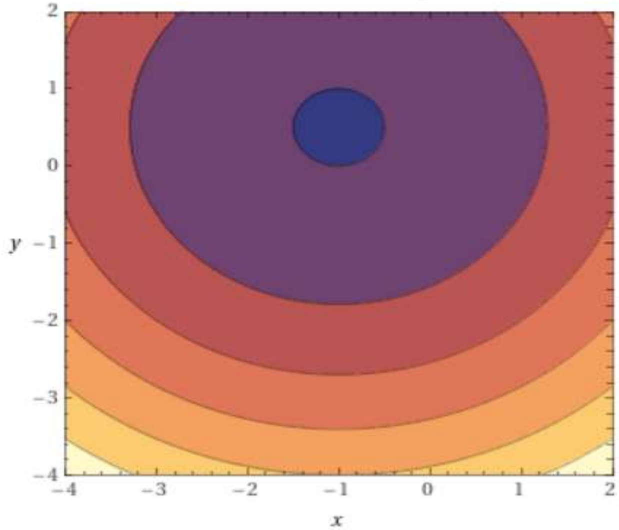
Principle: Calculate the gradient (=vector of partial derivatives)

Step in the opposite direction of gradient until minimum is reached

Find the minimum of $f(x,y) = x^2 + y^2 + 2x - y + 1$

plot x^2 + y^2 + 2x - y +1 from -4 to 2





Analytical method (manual version)

1. Calculate both partial derivatives

$$f_x(x,y) = 2x + 2$$
$$f_y(x,y) = 2y - 1$$

2. Solve point where both partial derivatives are zero.

$$2x + 2 = 0 \Rightarrow x = -1$$
$$2y - 1 = 0 \Rightarrow y = \tfrac{1}{2}$$

3. Calculate f(-1, 1/2)

$$f(-1,1/2) = (-1)^2 + (1/2)^2 + 2*(-1) - 1/2 + 1 = -1/4$$

Answer: Function has a minimum value -1/4 at point (-1, 1/2)

Find the minimum of $f(x,y) = x^2 + y^2 + 2x - y + 1$

Analytical method using WolframAlpha calculator

```
solve D(x^2 + y^2 + 2x - y +1, x)=0 , D(x^2 + y^2 + 2x - y +1, y)=0
```

Result: x = -1, y = 1/2

```
x^2 + y^2 + 2x - y +1 where x=-1, y=1/2
```

Result: -1/4

Direct solution using WolframAlpha's minimize command

```
minimize x^2 + y^2 + 2x - y +1
```

$$\min\{x^2 + y^2 + 2x - y + 1\} = -\frac{1}{4} \text{ at } (x, y) = \left(-1, \frac{1}{2}\right)$$

Find the minimum of $f(x,y) = x^2 + y^2 + 2x - y + 1$

Gradient descent method

Iteration formula:

(x,y) = (x,y) – grad(f)*t

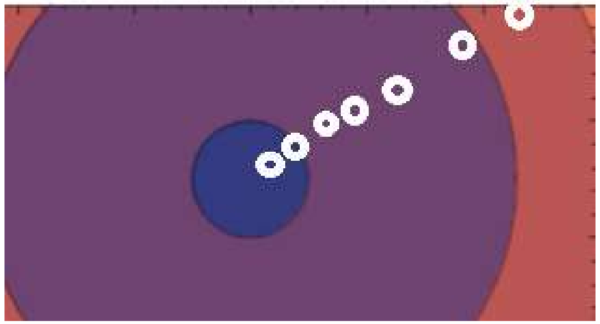
Iteration can be written also:

x = x - fx*t

y = x – fx*t

(fx and fy are partial derivatives,

t = coefficient defining the step)



```
# Python code
# 1. Define function f(x) and partial derivatives fx(x,y) and fy(x,y)
def f(x,y):
  return x**2 + y**2 + 2*x – y + 1
def fx(x,y):
  return 2*x + 2
def fy(x,y):
  return 2*y -1
# 2.Give an arbitrary initial value for variable x
x = 1.5
y = 2.0
# 3. Define step coefficient t and initialize step dx=-Df(x)*t
t = 0.2
dx=-fx(x,y)*t
dy=-fy(x,y)*t
#4. Create a loop to update x as long as |dx| >0.005
while abs(dx)>0.005:
   dx=-fx(x,y)*t
   dy=-fy(x,y)*t
   x = x+dx
   y = y+dy
   print(f"x={x:5.2f}, y ={y:5.2f}, f(x,y) = {f(x,y):7.4f} ")
#5. Print minimum value and coordinates of minimum point
print(f”Minimum value {f(x,y):5.2f} at x= {x:5.2f}, y = {y:5.2f}”)
```

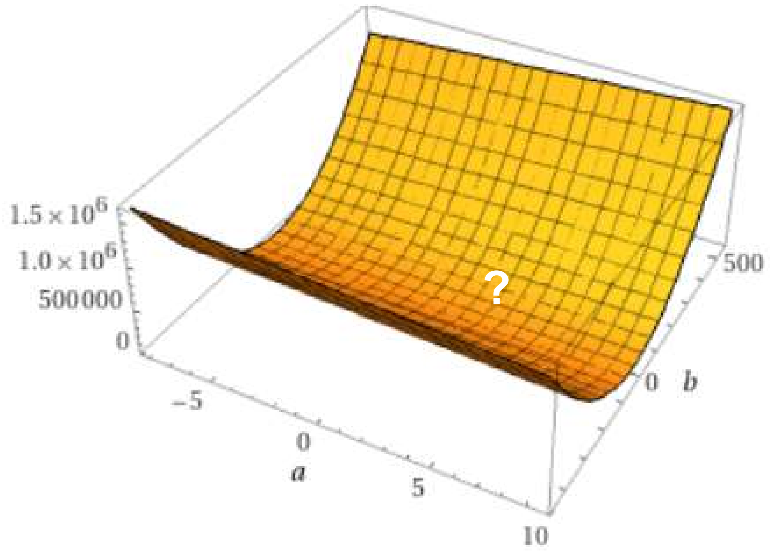
Regression example

Fit a linear model y = a x + b to following observed data

x	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**
y	**42.5**	**42.8**	**43.2**	**43.5**	**43.9**	**44.2**

Method is to find a minimum of the square sum

$(a*1.0+ b -42.5)^2 + (a*1.5+ b -42.8)^2 + (a*2+ b -43.2)^2 + (a*2.5+ b -43.8)^2 + (a*3+ b -43.9)^2 + (a*3.5+ b -44.3)^2$



From plot it is difficult to locate, where In the "walley" is the minimum

Analytical method finds a point (a,b) where both partial derivatives of the square sum are zero.

WolframAlpha command minimize automatizes this procedure.

minimize $(a*1.0+ b -42.5)^2 + (a*1.5+ b -42.8)^2+ (a*2+ b -43.2)^2+ (a*2.5+ b -43.8)^2+ (a*3+ b -43.9)^2+ (a*3.5+ b -44.3)^2$

Result: 0.0510476 at $(a, b) \approx (0.737143, 41.7581)$

On the next slide we try to get the same result by writing a Python code, which uses iteration

Code for descend method for finding linear model y = a x + b

```
def f(a,b):
    return (a*1.0+b-42.5)**2 +(a*1.5+b-42.8)**2+(a*2+b-43.2)**2+(a*2.5+b-43.8)**2+(a*3+b-43.9)**2+(a*3.5+b-44.3)**2

def fa(a,b):
    return 2*(a*1.0+b-42.5)+3*(a*1.5+b-42.8)+4*(a*2+b-43.2)+5*(a*2.5+b-43.8)+6*(a*3+b-43.9)+7*(a*3.5+b-44.3)

def fb(a,b):
    return 2*(a*1.0+b-42.5)+2*(a*1.5+b-42.8)+2*(a*2+b-43.2)+2*(a*2.5+b-43.8)+2*(a*3+b-43.9)+2*(a*3.5+b-44.3)

# initial values
a=1.0
b=40.0
z=f(a,b)
nr=1

# iteration coefficient and iteration step
t = 0.02;
da=-fa(a,b)*t
db=-fb(a,b)*t


# loop
while abs(da)>0.0003:
    da= -fa(a,b)*t  # iteration step in x-direction
    db= -fb(a,b)*t  # iteration step in y-direction
    #print(f"nr {nr:2d}    a = {a:5.2f}   b ={b:5.2f}   f(a,b)={z:9.5f} ")
    a=a+da        # new x
    b=b+db        # new y
    z=f(a,b)      # function value
    nr+=1         # increase round nr


print(f"\nMinimum found by iteration at a ={a:6.3f}   b ={b:6.3f}")
```

```
Minimum found by iteration at a = 0.748   b =41.730

Compare: True minimum found with W.A minimize at  [0.74, 41.8]
```

Comments:

Coding this has some challenges

1. Long function and long partial derivatives

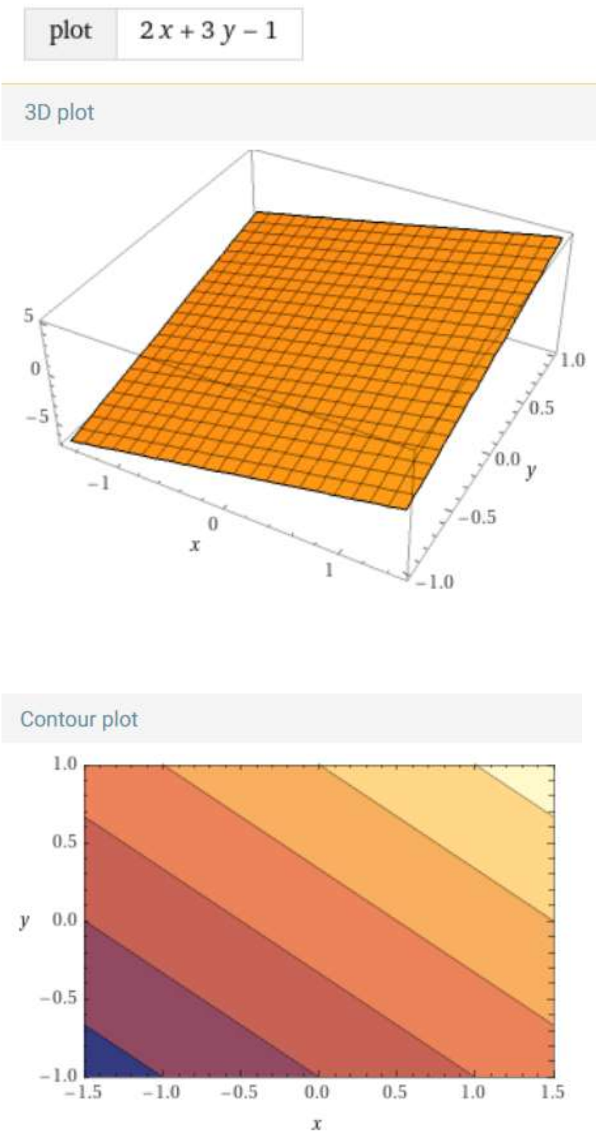
2. Finding suitable value for coefficient t needs several trials

3. Finding suitable value for loop conditions
abs(da)>0.0003
needs several trials

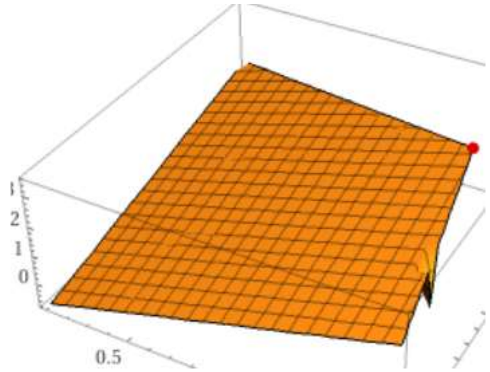
(Wrong values lead to deadlock)

Linear multivariable functions, for example f(x,y) = 2 x + 3y – 1 have constant gradients. In our example df/dx = 2 and df/dy = 3 => Gradient vector = (2,3).

Equation f(x,y) = 2 x + 3y – 1 represents a plane.
Function has neither minima nor maxima.

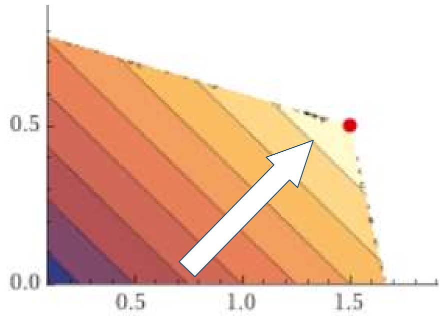


"Linear optimization" = finding a minimum (or maximum) of a linear function within a closed domain



If we restrict the domain with inequalities to a closed area, maximum is found at the corner of the domain that is farthest in the direction of the gradient

Example: Find the maximum of function f(x,y) = 2x + 3y – 1 inside the polygon, which is defined with conditions
x>=0
y>=0
3x+y<=5
x+5y<=4



Gradient vector (2,3)

maximize 2x+3y-1 where x>=0,y>=0, 3x+y<=5, x + 5y<=4

$$\max\{2x+3y-1 \mid x \geq 0 \land y \geq 0 \land 3x+y \leq 5 \land x+5y \leq 4\} = \frac{7}{2} \text{ at } (x, y) = \left(\frac{3}{2}, \frac{1}{2}\right)$$

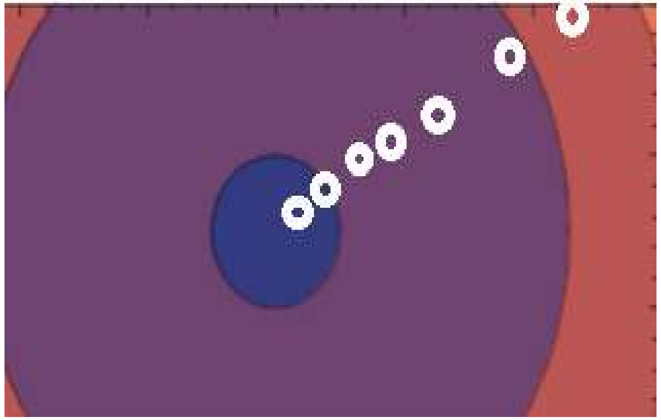
Appendix: Excel version of minimizing $f(x,y) = x^2 + y^2 + 2x - y + 1$

Gradient descent method

Iteration for x and y
x = x – fx*t = x-(2x+2)*t
y = y – fy*t = y-(2y-1)*t

(fx and fy are partial derivatives,
t = coefficient defining the step)

	A	B	C		
1	*step coefficient t*	**0,25**			
2					
3		***x***	***y***		
4	*initial value*	**2**	**1**		
5	*iteration formulas*	0,500	0,750	=B4-(2*B4+2)*B1	=C4-(2*C4-1)*B1
6		-0,250	0,625		
7		-0,625	0,563		
8		-0,813	0,531		
9		-0,906	0,516		
10		-0,953	0,508		
11		-0,977	0,504		
12		-0,988	0,502		
13		-0,994	0,501		
14		-0,997	0,500		
15		-0,999	0,500		
16		-0,999	0,500		
17		-1,000	0,500		
18		-1,000	0,500		



Found minimum at (-1, 0.5)

Previous example Python code

```python
def f(x,y):     # define function f
   return x**2 + y**2 + 2*x - y + 1

def f_x(x,y):    # define partial derivative fx
   return 2*x+2

def f_y(x,y):    # define partial derivative fy
   return 2*y-1

# give initial values for iteration
x=1.5
y=2.0
z=f(x,y)
nr=1  # counter for number of iterations

# iteration coefficient and iteration step
t = 0.2;
dx=-f_x(x,y)*t
dy=-f_y(x,y)*t

print("Iteration steps: \n")

# loop which prints all the steps of iteration
while abs(dx)>0.0005:
   dx=-f_x(x,y)*t  # iteration step in x-direction
   dy=-f_y(x,y)*t  # iteration step in y-direction
   print(f"nr {nr:2d}    x = {x:5.2f}   y ={y:5.2f}   f(x,y)={z:9.5f} ")
   x=x+dx        # new x
   y=y+dy        # new y
   z=f(x,y)        # function value
   nr+=1          # increase round nr

# print results
print(f"\nMinimum found by iteration at (x,y) =[{x:7.4f},{y:7.4f}] ")
print(f"\nMinimum value by iteration = {z:7.4f}")
print("\nCompare: True minimum at zero of gradient ",[-1,0.5,f(-1,0.5)])
```

Output of the Python program

```
Iteration steps:

nr  1    x =  1.50   y = 2.00   f(x,y)=  8.25000
nr  2    x =  0.50   y = 1.40   f(x,y)=  2.81000
nr  3    x = -0.10   y = 1.04   f(x,y)=  0.85160
nr  4    x = -0.46   y = 0.82   f(x,y)=  0.14658
nr  5    x = -0.68   y = 0.69   f(x,y)= -0.10723
nr  6    x = -0.81   y = 0.62   f(x,y)= -0.19860
nr  7    x = -0.88   y = 0.57   f(x,y)= -0.23150
nr  8    x = -0.93   y = 0.54   f(x,y)= -0.24334
nr  9    x = -0.96   y = 0.53   f(x,y)= -0.24760
nr 10    x = -0.97   y = 0.52   f(x,y)= -0.24914
nr 11    x = -0.98   y = 0.51   f(x,y)= -0.24969
nr 12    x = -0.99   y = 0.51   f(x,y)= -0.24989
nr 13    x = -0.99   y = 0.50   f(x,y)= -0.24996
nr 14    x = -1.00   y = 0.50   f(x,y)= -0.24999
nr 15    x = -1.00   y = 0.50   f(x,y)= -0.24999
nr 16    x = -1.00   y = 0.50   f(x,y)= -0.25000

Minimum found by iteration at (x,y) =[-0.9993, 0.5004]

Minimum value by iteration = -0.2500

Compare: True minimum at zero of gradient  [-1, 0.5, -0.25]
```