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

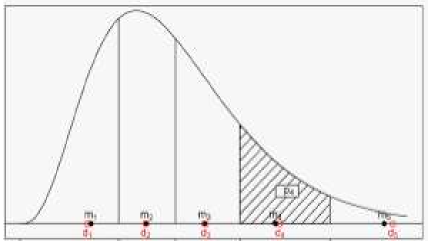
1. Discrete probability distributions
2. Parameters: Expected Value, standard deviation
3. Continuous probability distributions

   - example: Normal distribution

4. Decision trees

   - example: Newsboy Problem

discrete distribution



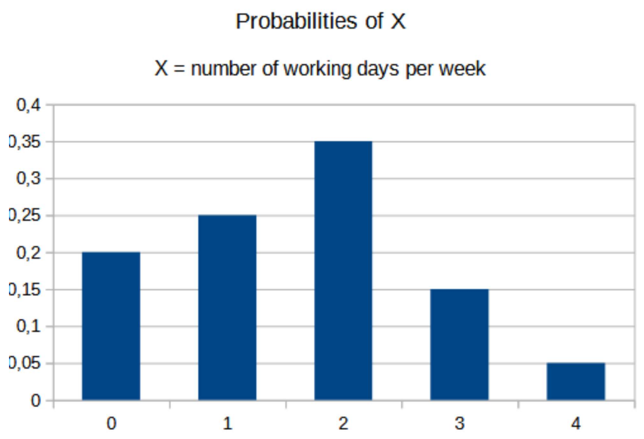
continuous distribution

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

Standard deviation $Std(x) = \sqrt{(Var(x))}$ *)

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

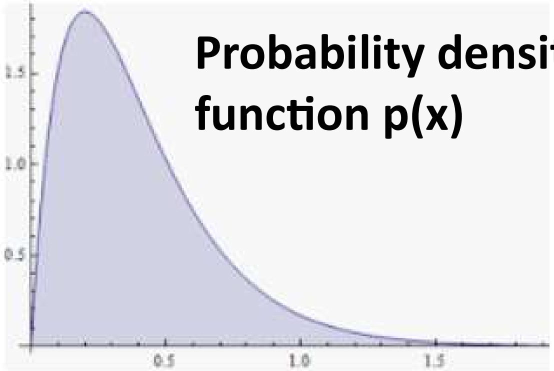
Expexted return (long term average result) is following

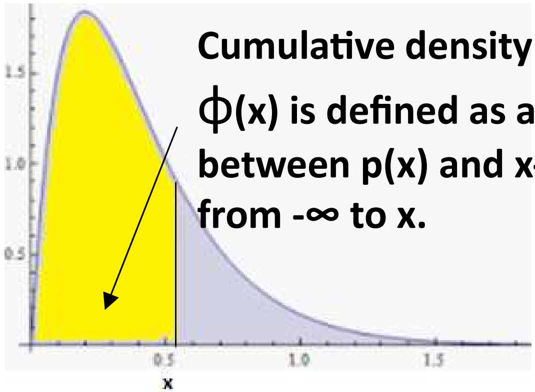
ER = 0.4*12000+0.35*2000+ 0.25*(-50000) = -7000

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.**





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**

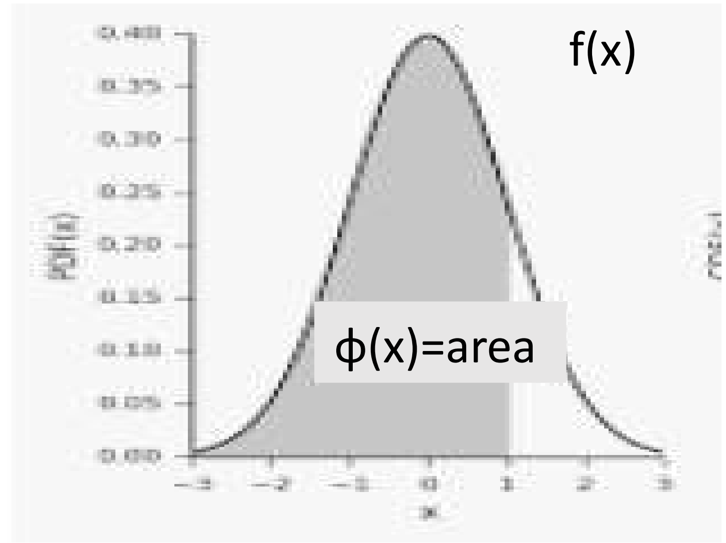
Interpreation: $\phi(x_0) = P(x \leq x_0)$

Value of cumulative density function at x_0 gives the probability that the variable value $x \leq x_0$

Normal distribution

The most common continuous distribution is ”Gaussian” normal distribution.

It has two parameters: mean μ and standard deviation σ



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

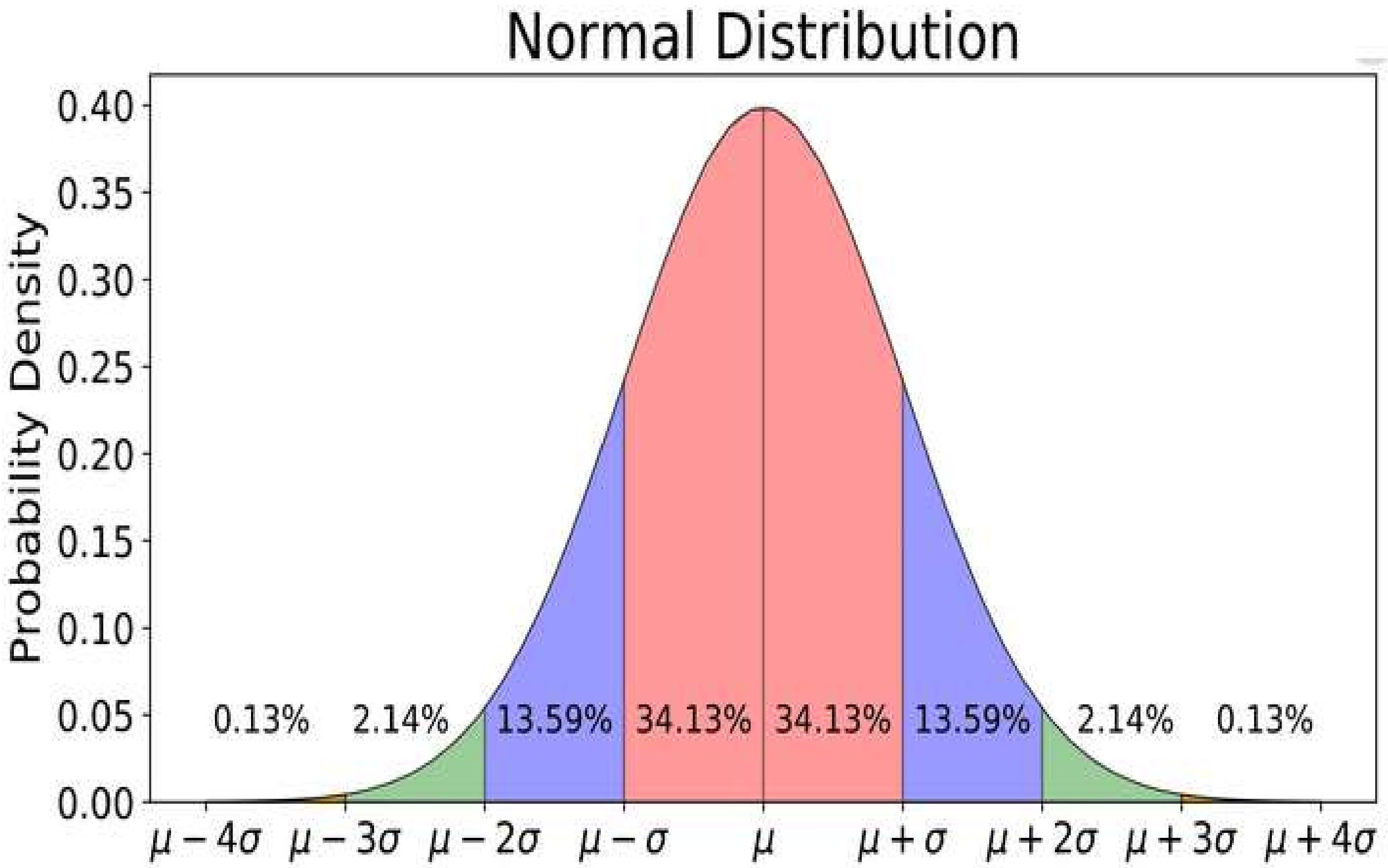
$$\phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

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,….

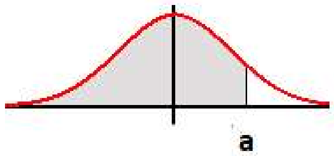


Around 68 % of values lie in range [μ-σ,μ+σ]

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

Usage of cumulative density function **ϕ(x) and its inverse function ϕ⁻¹(P)**

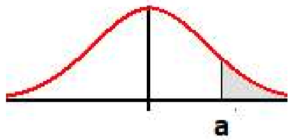
1. Calculation of probability that value of $x \le a$



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

Excel: =NORM.DIST (x;μ;σ;1)

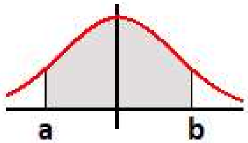
2. Calculation of probability that value of $x \ge a$



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

Excel: = 1- NORM.DIST (x;μ;σ;1)

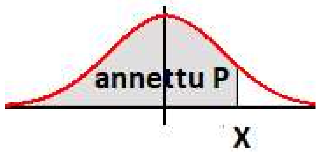
3. Calculation of probability that $a \le \mathbf{x} \le b$



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

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

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

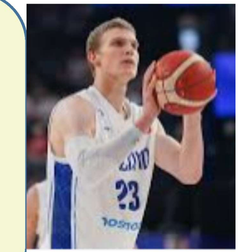


$\phi(x) = P \quad \Rightarrow \quad x = \phi^{-1}(P)$

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 = ϕ^{-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.

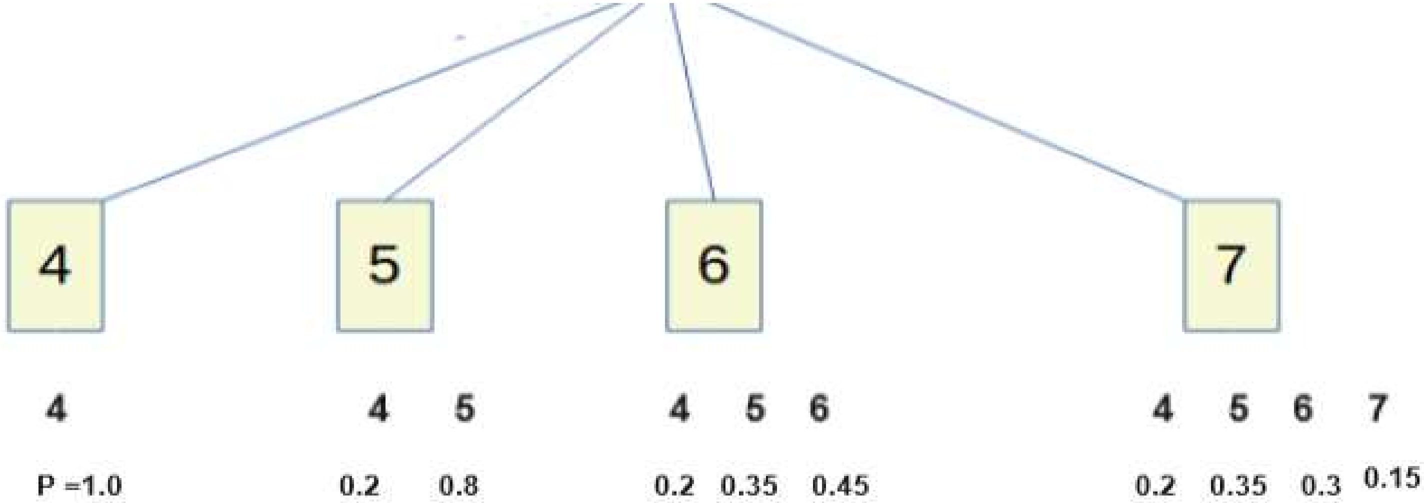
Newsboys were the main distributors of newspapers to the general public in USA 100-150 years ago. Youngest were 5 – 6 years old. They bought papers from the publisher at their own risk and sold them on the streets with a small profit. From unsold papers they got no compensation.

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

$$\mu = \frac{\sum x}{N}$$

Sample Mean

$$\bar{X} = \frac{\sum x}{n}$$

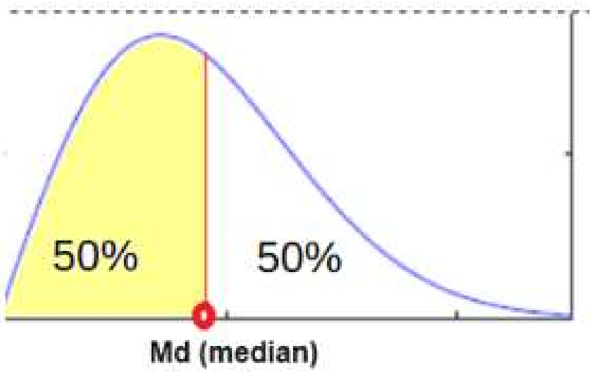
(Notice the convention of using greek letters for parameters of population. Formulas are the same: mean = sum of values / count of values)

Excel- formulas

=average(A1:A28)

=median(A1:A28)

Median

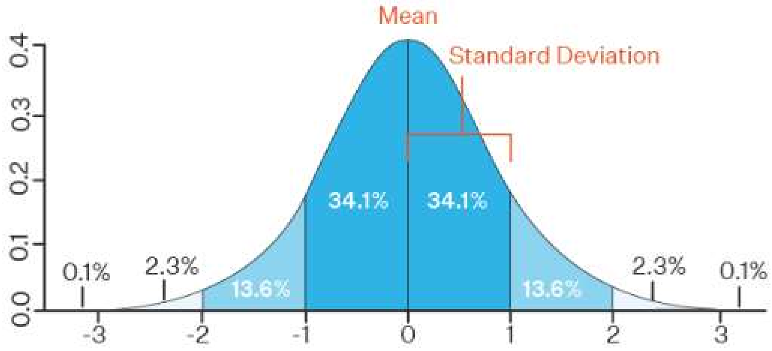


Rule for manual calculation of median:

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
$\sigma = \sqrt{\frac{\sum(X - \mu)^2}{N}}$	$s = \sqrt{\frac{\sum(X - \bar{x})^2}{n - 1}}$
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:

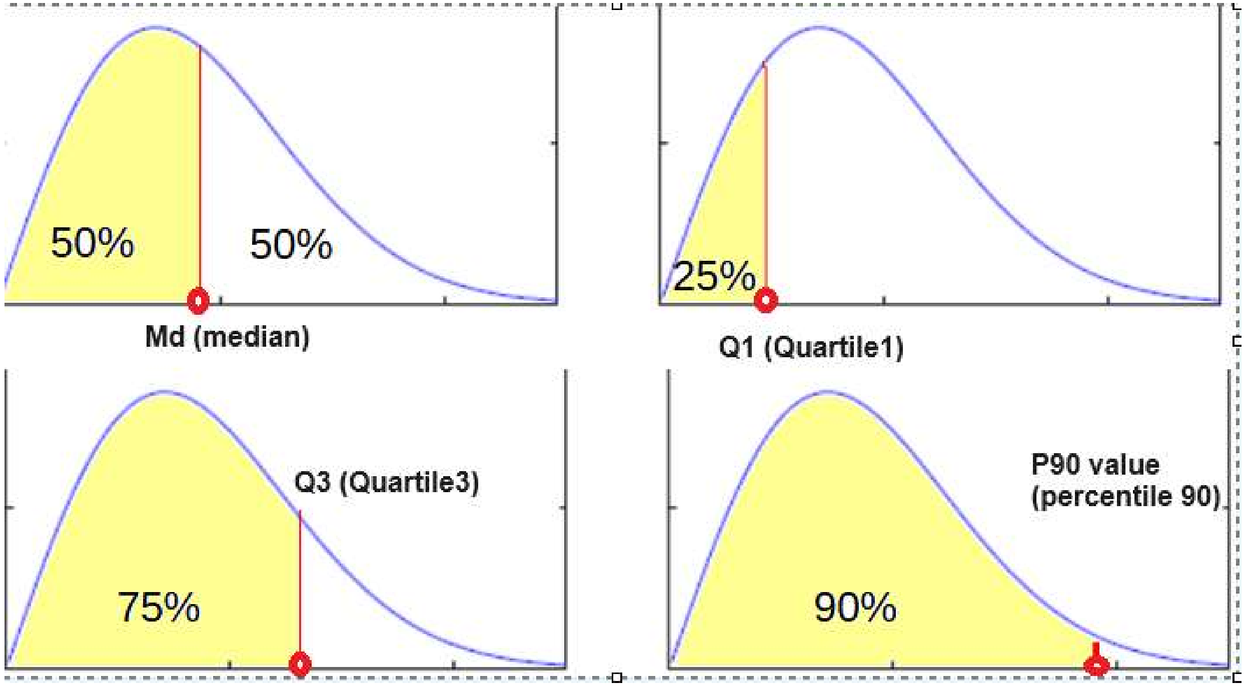
= STDEV.P(A1:A28)

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.

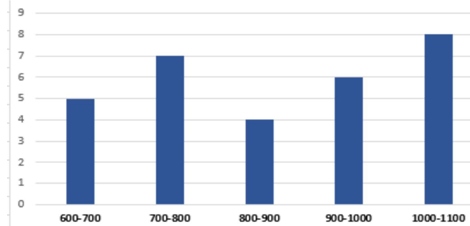
Formulas: **MEAN** $$\mu = \frac{\sum f_i x_i}{N}$$

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.

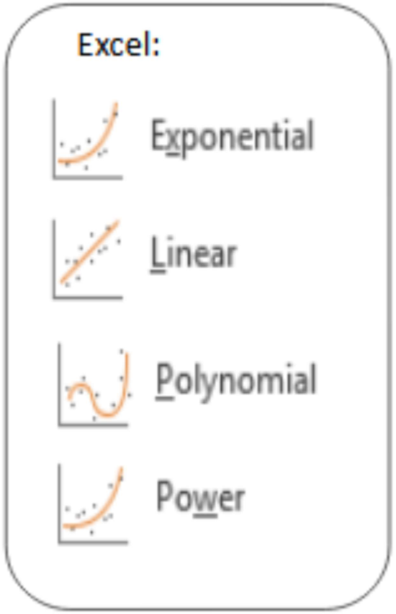
Common mathematical models:

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

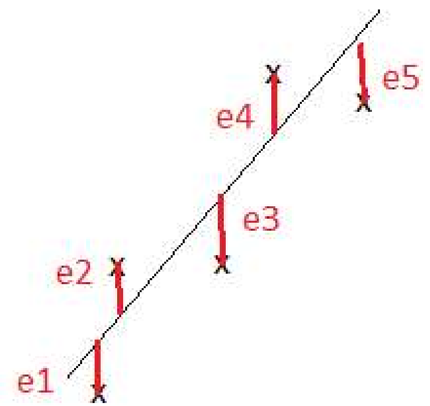
Example of observed data
Consists of x, y pairs

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

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



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

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

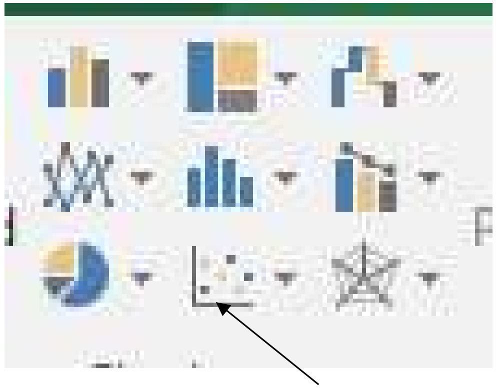
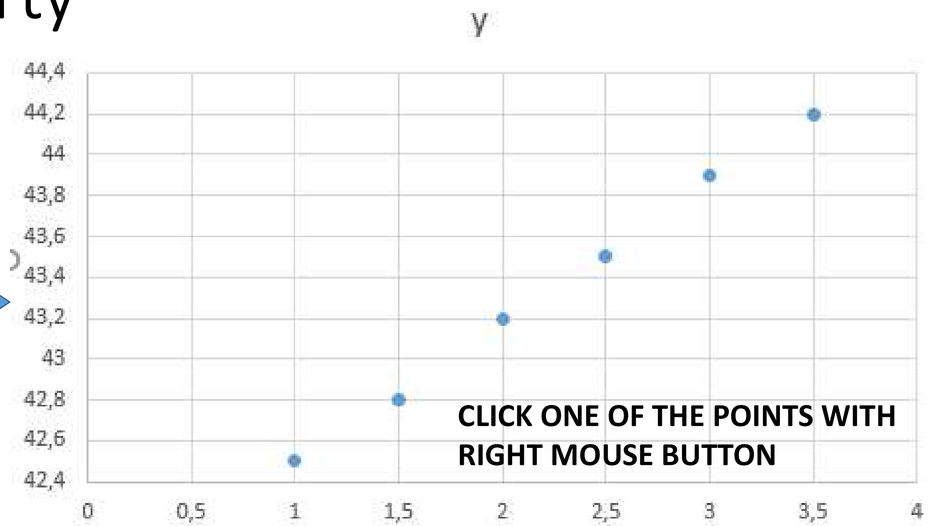
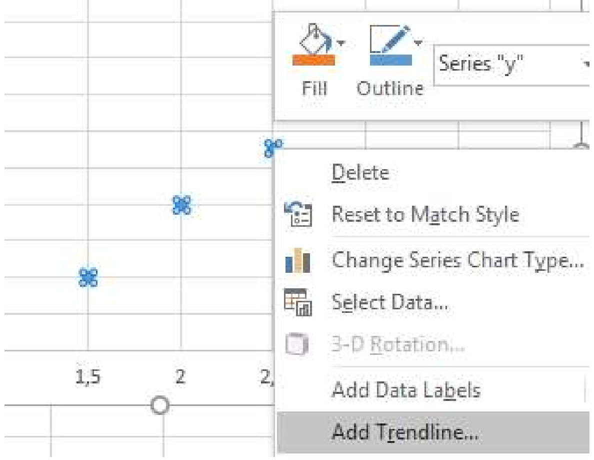


Chart TYPE

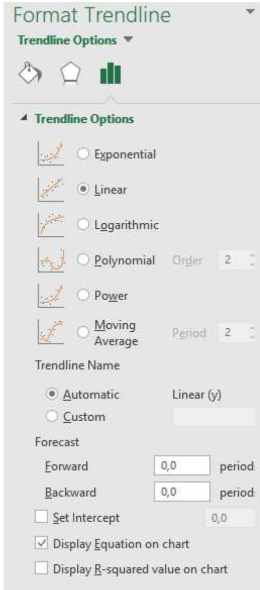


CLICK ONE OF THE POINTS WITH RIGHT MOUSE BUTTON

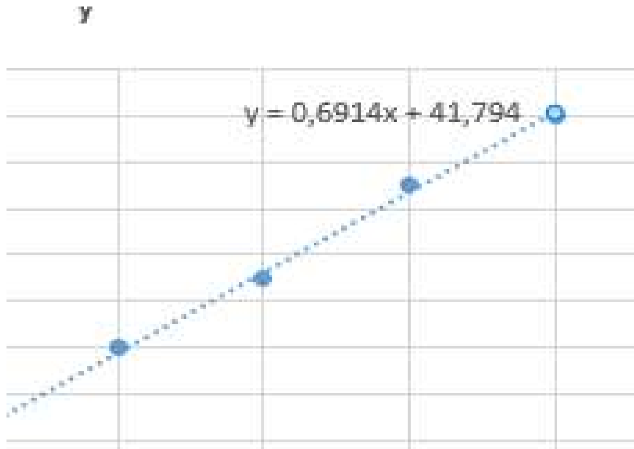
ADD TRENDLINE



TYPE = LINEAR

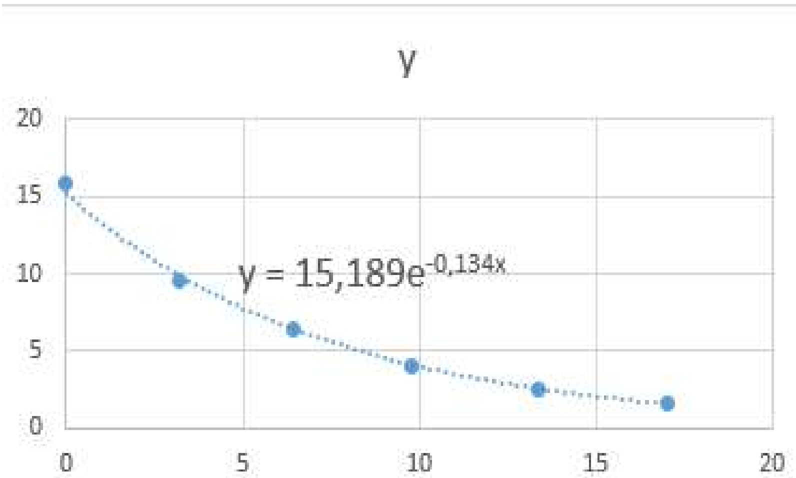


RESULT: EQUATION OF THE MODEL



Other mathematical models (in Excel graphics trendline)

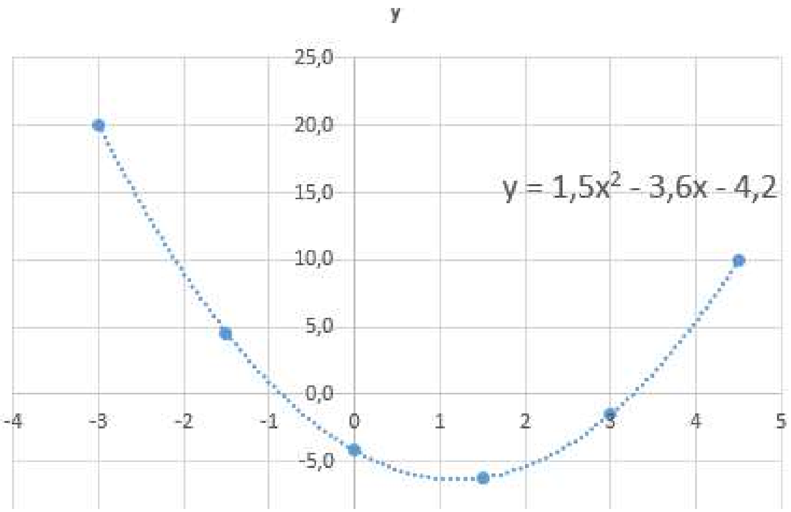
x	y
0	15,8
3,2	9,6
6,4	6,4
9,8	4
13,4	2,5
17	1,6



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

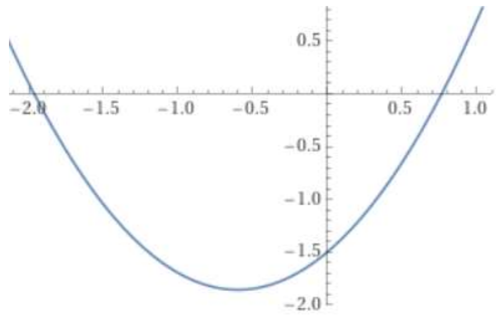
-confidence margins and confidence intervals

2. Confidence intervals of regression analysis

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

CĂyŘ űKŜ YĂyĂYdzY ǾŀfdzŜ 2Ŧ ŦóÈ̀ó Ґ È̀ Ƅмϕнȅ πмϕp

plot x^2+1.2x-1.5



Method A: Analytical method (manual version)

1. Calculate the derivative function

fô(x) = 2x + 1.2

2. Solve the zero of derivative

2x + 1.2 = 0 => 2x = - 1.2 => x = -0.6

3. Calculate f(-0.6)

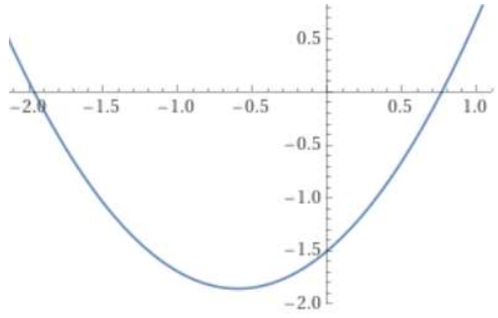
f(-0.6) = (-0.6)ᴴ ƄмϕнϜóπлϕcóπмϕp = -1.86

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

CǐyŘ űKŜ YǐyǐYdzY ǾŀfdzŜ 2Ŧ ŦóÈó Ґ Ȅª ЬмΦнȄ πмΦp

Analytical method using WolframAlpha calculator

plot x^2+1.2x-1.5



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 WolframAlphaǿs minimize command

minimize x^2 +1.2*x -1.5

Global minimum

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

CǐyŘ ǘKŜ YǐyǐYdzY ǿŀfdzŜ 2Ŧ ŦóÈó Ґ Èᴴ ѢмǿнÈ πмǿp

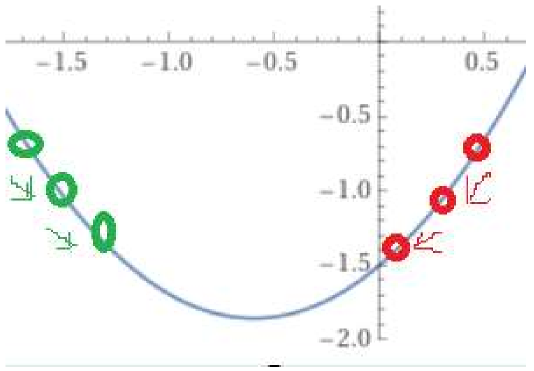
Method B. Iteration
Gradient descent method

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(fò\nMinimum : x= {x:5.2f}, f(x) = {f(x):5.2f}ò)
```

CĂyŘ ǘKŜ YĂyĂYdzY ǾŀfdzŜ ƻŦ ŦóȄó Ґ Ŝȅ ς нǂp Ȅ Ђ м

Method A: Analytical method (manual version)

1. Calculate the derivative function

fô(x) = e ï 2.5

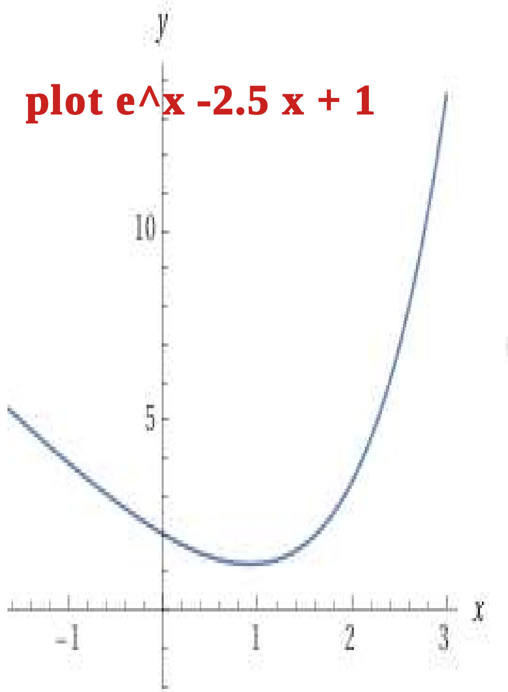
2. Solve the zero of derivative

e ï 2.5 = 0 => e = 2.5 => x = ln(2.5) = 0.916

3. Calculate f(0.916)

f(0.916) = e^0.916 ï 2.5*0.916 + 1 = 1.21

Answer: Minimum 1.21 at x = 0.916



EXCEL ï 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

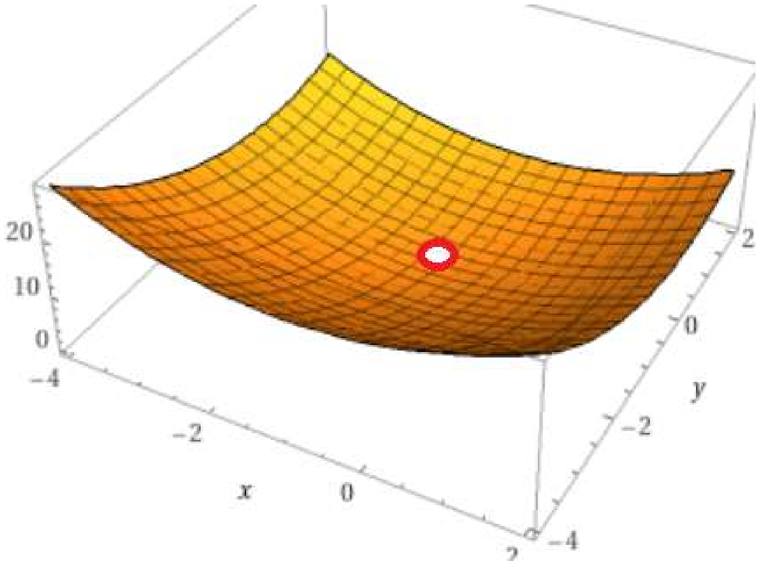
New x value = x ï fô(x)*t
In this example is:
New x value = x ï (e^x-2,5)*t

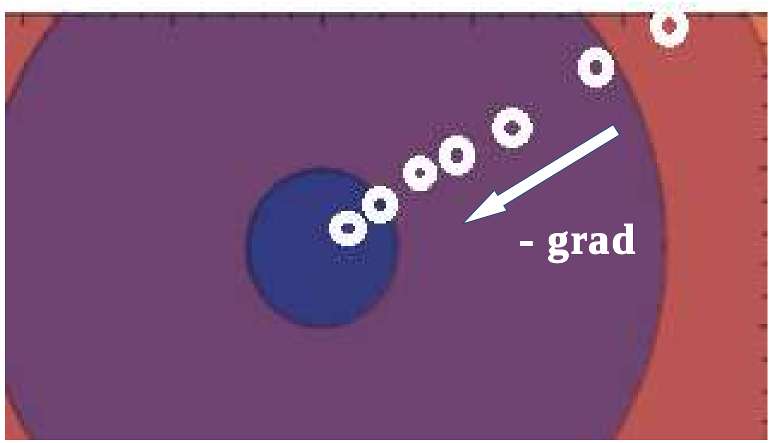
is written using cell addressess

The iteration formula is copied down.

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

[illegible]





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

Step in the opposite direction of gradient until minimum is reached

CĺyŘ ǿKŜ YĺyĺYǳY 2Ŧ ŦóÈŁèó Ґ Èª Ѣ èª Ѣ нÈ π è Ѣм

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

Analytical method (manual version)

1. Calculate both partial derivatives

f (x,y) = 2x + 2

f (x,y) = 2y - 1

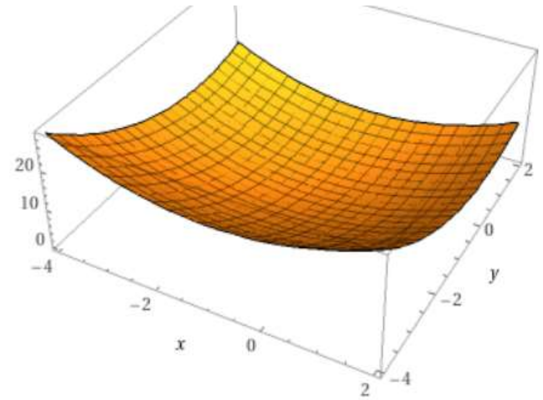
2. Solve point where both partial derivatives are zero.

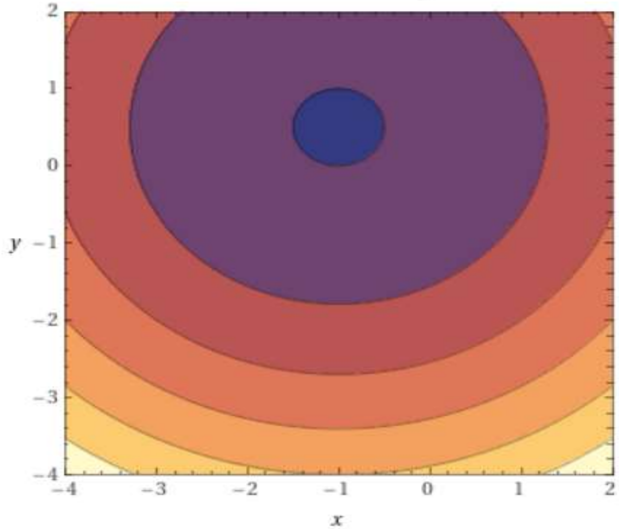
2x + 2 = 0 => x = -1

2y ï 1 = 0 => y = İ

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

f(-1,1/2) = (-1)ª Ѣ óмкнóª Ѣ нҒóπмó π мкн Ѣ м Ґ πмкп





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

CÏyŘ ǘKŜ YÏyÏYǆY 2Ŧ ŦóÈŁèó Ґ Èª b èª b нÈ π è bм

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 WolframAlphaôs minimize command

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

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

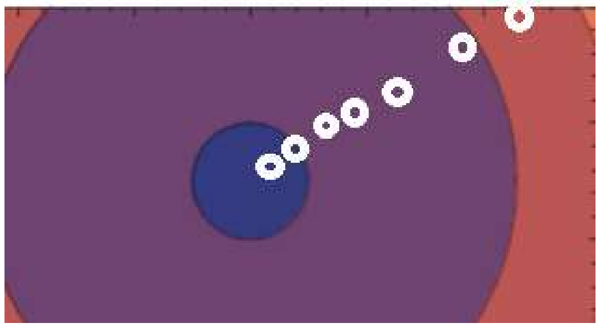
CĭyŘ ŕKŜ YĭyĭYďzY 2Ŧ ŦóÈĽèó Ґ Èª Ƅ èª Ƅ ʜÈ π è Ƅм

Gradient descent method

Iteration formula:
(x,y) = (x,y) ï grad(f)*t

Iteration can be written also:
x = x - fx*t
y = x ï 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 ï 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(fòMinimum value {f(x,y):5.2f} at x= {x:5.2f}, y = {y:5.2f}ò)
```

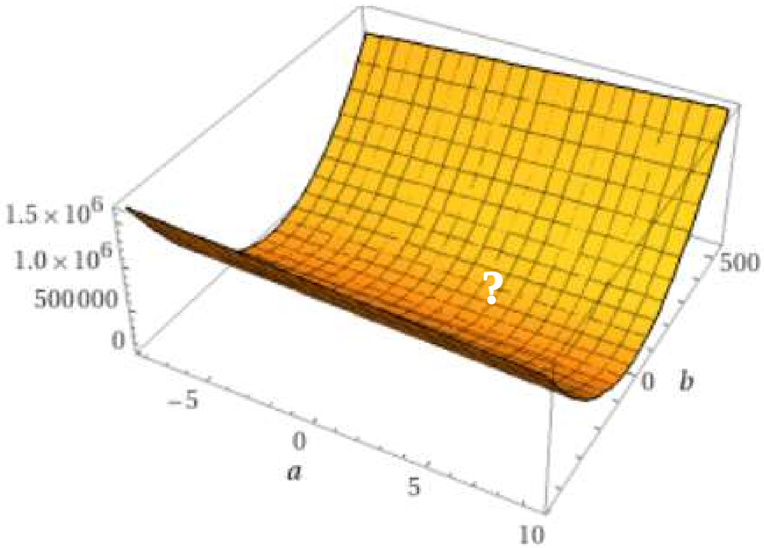
Regression example

Cǐú Ɩ· ƒǐyŜƖ·Ŋ ȲƧŘŜſ è Ґ Ɩ· È ƀ ō úƧ ŦƧffƧ ǿǐyƎ ƧōǎŜŊǾŜŘ ŘƖ·úƖ·

È	мφл	мφр	нφл	нφр	оφл	оφр
è	пнφр	пнφу	поφн	поφр	поφф	ппφн

a ŜúҞƧŘ ǐǎ úƧ ŬyŘ Ɩ· ȲǐyǐȲǆȲ Ƨŧ úҞŜ ǎǉǆƖ·ŊŜ ǎǆȲ

όƖ·ғмφлƀ ō ппнφрό ƀ όƖ·ғмφрƀ ō ппнφуό ƀ όƖ·ғнƀ ō ппоφнό ƀ όƖ·ғнφрƀ ō ппоφуό ƀ όƖ·ғоƀ ō ппоφфό ƀ όƖ·ғоφрƀ ō пппφоό



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

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

WolframAlpha command <u>minimize</u> automatizes this procedure.

YλyλYλÎŜ όl·ҒмΦлѢ ō πпнΦpύ Ѣ όl·ҒмΦpѢ ō πпнΦyύ Ѣ όl·ҒнѢ ō πпоΦнύ Ѣ όl·ҒнΦpѢ ō πпоΦyύ Ѣ όl·ҒоѢ ō πпоΦфύ Ѣ όl·ҒоΦpѢ ō πппΦоύ

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}")
```

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

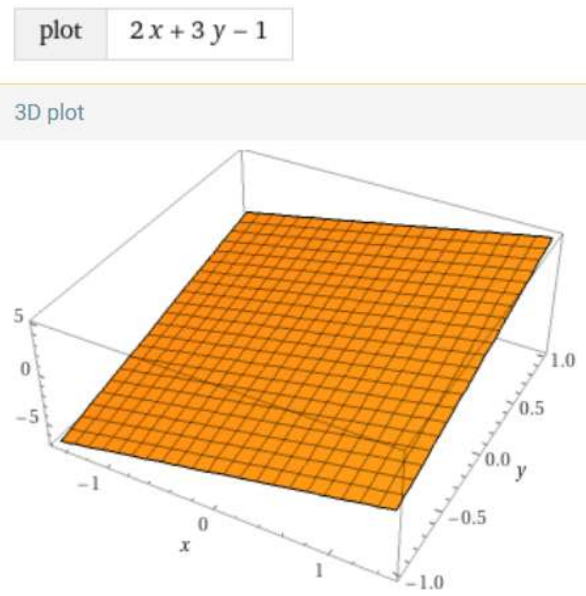
(Wrong values lead to deadlock)

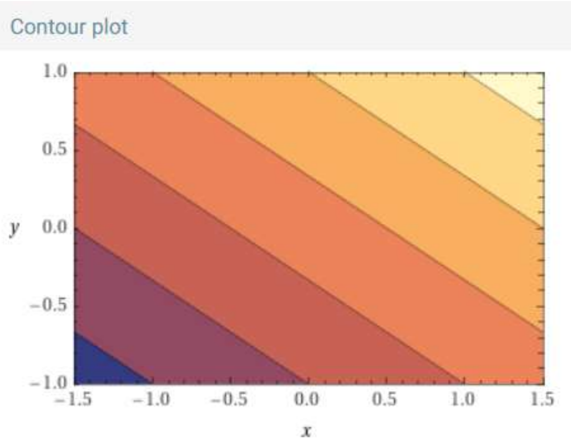
```
Minimum found by iteration at a = 0.748   b =41.730

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

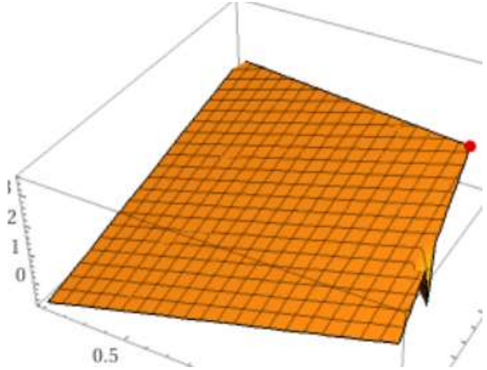
Linear multivariable functions, for example f(x,y) = 2 x + 3y ï 1 have constant gradients. **In our example df/dx = 2 and df/dy = 3 => Gradient vector = (2,3).**

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



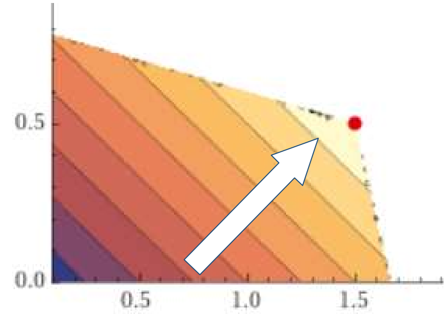


òLinear optimizationò = finding a minimum (or maximum) of a linear function within a closed domain



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 ï 1 inside the polygon, which is defined with conditions
x>=0
y>=0
3x+y<=5
x+5y<=4



Gradient vector (2,3)

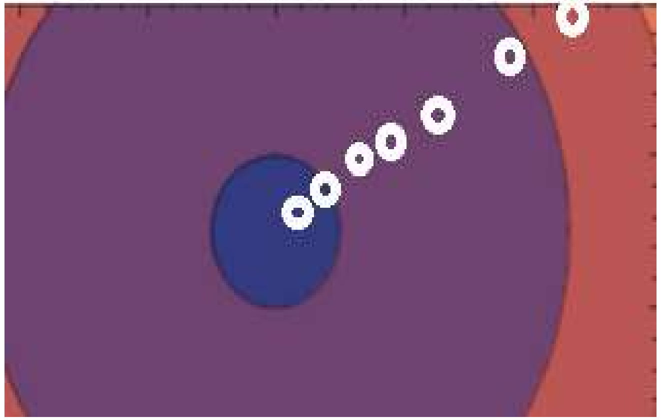
maximize 2x+3y-1 where x>=0,y>=0, 3x+y<=5, x + 5y<=4

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

Gradient descent method

Iteration for x and y
x = x ï fx*t = x-(2x+2)*t
y = y ï 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]
```