Vesalius College
Brussels

STA-101P: Quantitative Methods

Fall 2013

Lecturer: Koen Lefever

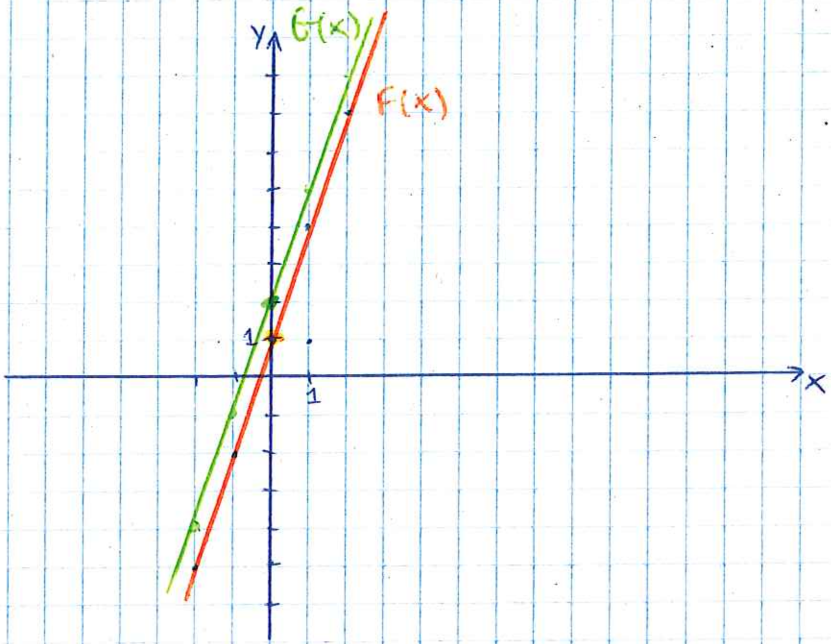
Class notes by Sophie Vériter

28.02.13

I. LINEAR EQUATIONS.

$F(x) = y = 3x + 1$

x	y = 3x + 1
-2	-5
-1	-2
-0	1
1	4
2	7
3	10
4	13
5	16
6	19



$G(x) = 3x + 2$
$F(x) = 3x + 1$
} same slope → lines are parallels.

$$y = \underbrace{a}_{\text{parameter}}\ \underbrace{x}_{\text{variable}} + \underbrace{b}_{\text{parameter}}$$

y: variable; a: SLOPE (parameter); x: variable; b: INTERCEPT with the y axis (parameter)

$a > 0$: upward line.
$a = 0$: parallel (⋮ or ---).
$a < 0$: downward line.

(1) The equation of the y axis is $x = 0$
↳ if $x = 0$: $y = ax + b$
$y = b$
↳ <u>intercept with the y axis</u> $= b$.

(2) The equation of the x axis is $y = 0$
↳ if $y = 0$: $y = ax + b = 0$
$x = -b/a$
↳ <u>intercept with the x axis</u> "the root" $= -b/a$.
⚠ some lines have no root (⋮ ---→).

II. QUADRATIC EQUATIONS parabols.

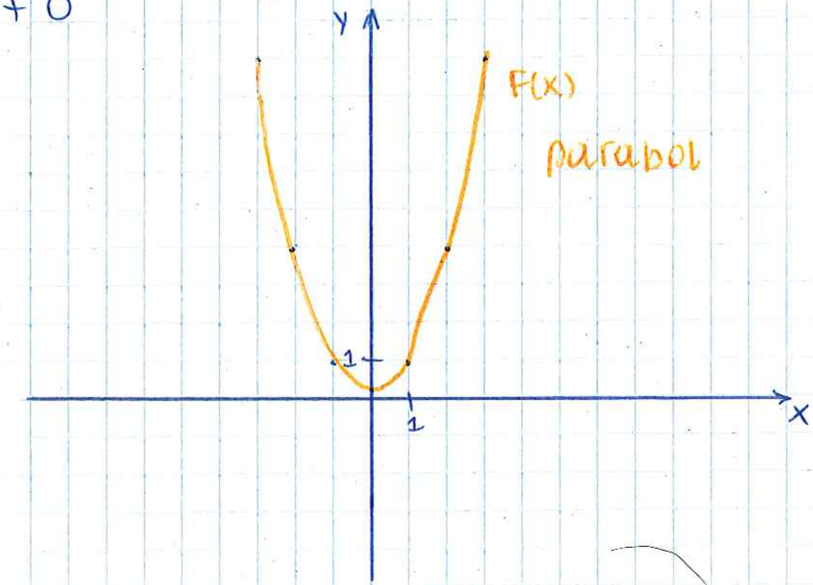
$F(x) = y = ax^2 + bx + c$

SLOPE — INTERCEPT

⚠ if the exponent is a variable, it becomes a linear equation.

$f(x) = y = 1x^2 + 0x + 0$
$= x^2$

x	$y = x^2$
-3	9
-2	4
-1	1
0	0
1	1
2	4
3	
4	16
5	25



$F(x) = y = ax^2 + bx + c$

∩ or ∪ ? — how high is the parabol ?

→ if the slope is positive : ∪
→ if the slope is negative : ∩
→ if the slope is 0, it's a linear equation.

$a^n \times a^m = a^{n+m}$

$a^{1/n} = \sqrt[n]{a}$

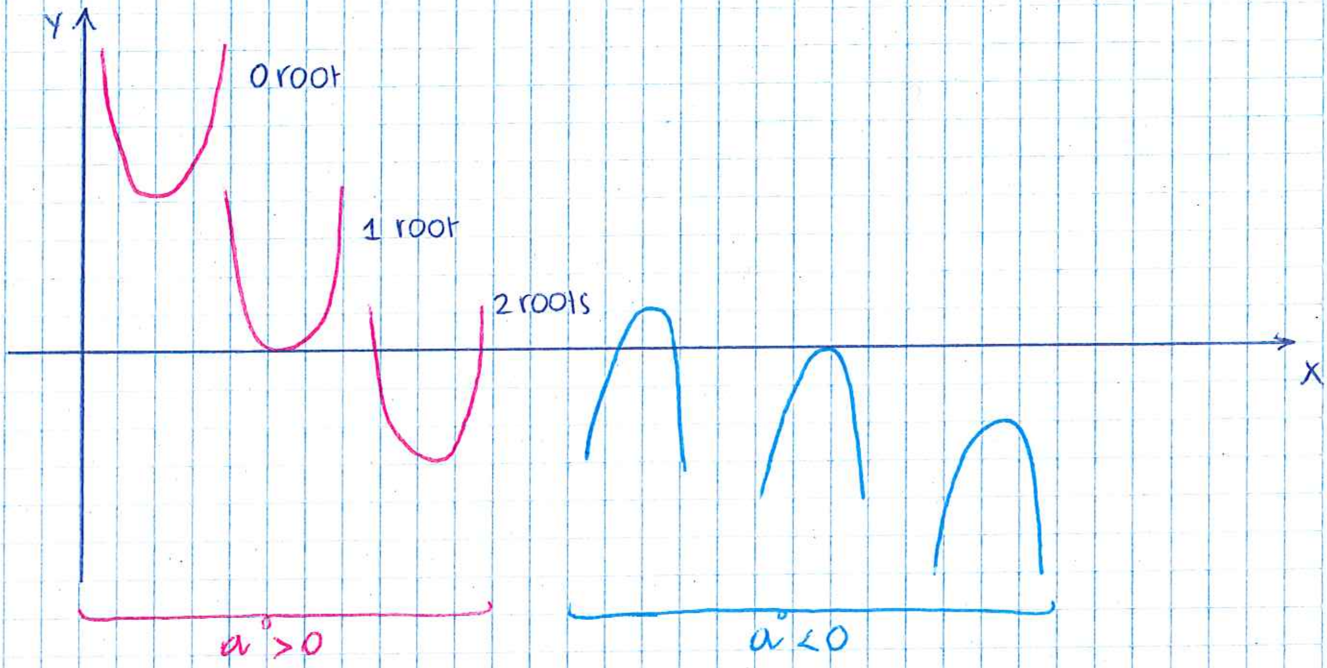
$a^{-n} = 1 / a^n$

$(a^n)^m = a^{n \cdot m}$

$\frac{a^n}{a^m} = a^{n-m}$

Quantitative Methods 02.09.13

1. Find the Discriminant, the root(s) & the min/max.



The Discriminant:

the Discriminant indicates how many roots have the parabol.

$$\boxed{D = b^2 - 4ac}$$

D > 0 : 2 roots.
D = 0 : 1 root.
D < 0 : no root.

The root(s) - intercept(s) with the x axis:

the root(s) are the intercept(s) of the parabol with the x axis.

x_1, x_2

$$\boxed{\text{root(s)} = \frac{-b \pm \sqrt{D}}{2a}}$$

if there are 2 roots, Do the calculations with + and then with −
if there is 1 root, only do the +
↳ if there is only 1 root, we Do the operation with + or − ?

the minimum or maximum:

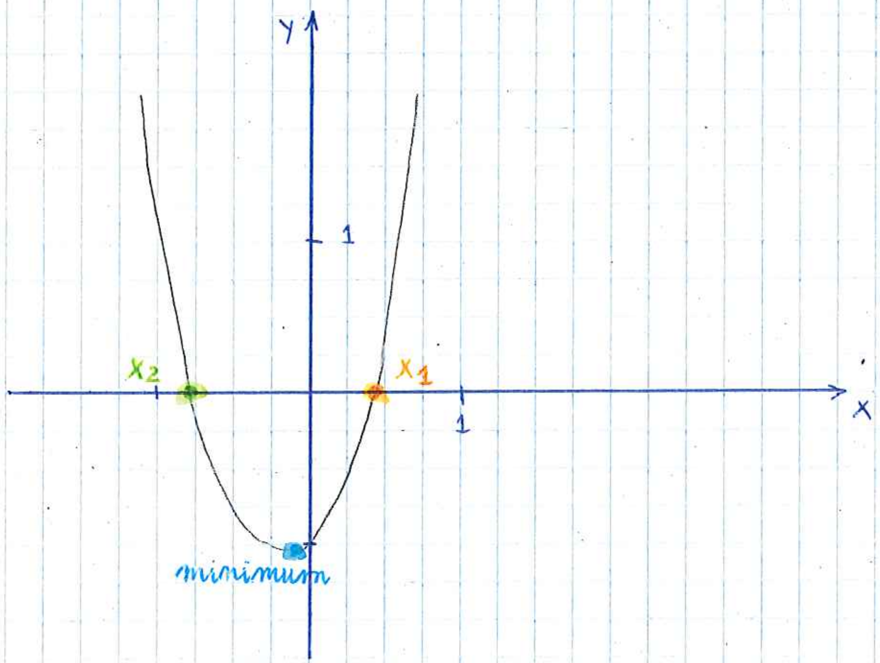
the minimum/maximum is the lowest/highest point of the parabol. It is in the middle Between the 2 roots (the 2 results of the operation are necessary).
↳ what if there is just 1 root? no root? we still Do the same operation?

$$\boxed{m = \frac{x_1 + x_2}{2}}$$

→ x value of the min/max.
To find the y value of the min/max, solve the equation with the x value.

EXAMPLE:

$$\boxed{y = 3x^2 + x - 1}$$



- the slope (a) > 0 : U parabol.
- the intercept with the y axis is -1.
- the Discriminant ($D = b^2 - 4ac$):

$$D = 1^2 - 4(3)(-1) = 1 - (-12) = 13$$

13 > 0
D > 0
→ 2 roots.

- the roots ($\frac{-b \pm \sqrt{D}}{2a}$):

$$x_1 = \frac{-1 + \sqrt{13}}{2.3} = 0,43$$

$$x_2 = \frac{-1 - \sqrt{13}}{2.3} = -0,77$$

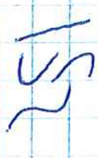
- the minimum ($m = \frac{x_1 + x_2}{2}$):

$$m = \frac{0,43 + 0,77}{2} = 0,17$$ → x value of the min.

$$y_m = 3(-0,17)^2 + (-0,17) - 1 = -1,08$$ → y value of the min.

minimum (0,17 ; -1,08)

III. CUBIC EQUATIONS.



linear	$y = ax + b$	graph
quadratic	$y = ax^2 + bx + c$	Graph
cubic	$y = ax^3 + bx^2 + cx + d$	Graph

$$\boxed{y = ax^3 + bx^2 + cx + d}$$

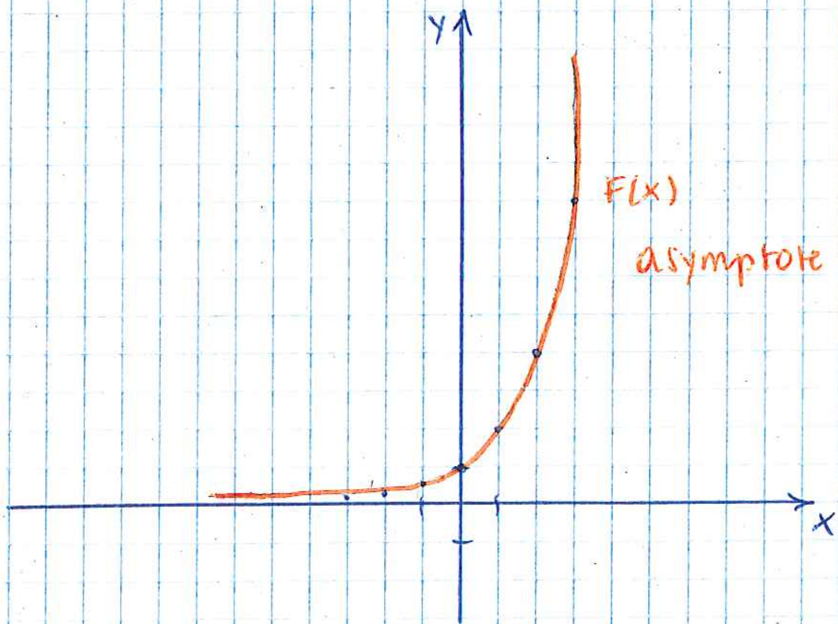
IV. EXPONENTIAL EQUATIONS.

$$\boxed{y = a^x + b}$$

⚠ the intercept with the y axis is $y = a^0 + b \rightarrow 1 + b$

$F(x) = 2^x$

x	$y = 2^x$
−3	$2^{-3} = 1/2^3 = 1/8$
−2	$2^{-2} = 1/2^2 = 1/4$
−1	$2^{-1} = 1/2$
0	$2^0 = 1$
1	$2^1 = 2$
2	$2^2 = 4$
3	$2^3 = 8$



if the parameter b increases, the asymptote wil goes up on the graph ( ). ex: $y = x^2 + 5$
conversely when the parameter b decrease.

1. Logarithmic operations.

$\boxed{M = b^a}$ → $16 = 4^2$

$\boxed{b = \sqrt[a]{M}}$ → $4 = \sqrt[2]{16}$

$\boxed{a = \log_b M}$ → $2 = \log_4 16$

a is the logarithm in b base of M.

- EXAMPLES:

(1) $\log_{10} 1000 = 3$
$10^3 = 1000$
$\sqrt[3]{1000} = 10$

(2) $\log_2 512 = 9$
$2^9 = 512$
$\sqrt[9]{512} = 2$

⚠ on the calculator $\boxed{\log} = \log_{10}$

→ $\boxed{\log}_{10}\ 1010 = 3{,}004.$ → $10 = \sqrt[3{,}004]{1010}$ → $1010 = 10^{3{,}004}$

→ $\boxed{\log}_{10}\ 1000 = 3.$ → $10 = \sqrt[3]{1000}$ → $1000 = 10^3$

⚠ on the calculator $\boxed{\ln} = \log_e$ ($e = 2{,}718$).

→ $\boxed{\ln}\ 1 = \log_e 1 = 0$ → $e = \sqrt[0]{1}$ → $1 = e^0$

- FORMULAS:

$\boxed{b^m \cdot b^n = b^{n+m} \;=\; \log_b(n \cdot m) = \log_b n + \log_b m}$

$\boxed{\frac{b^m}{b^n} = b^{m-n} \;=\; \log_b\left(\frac{m}{n}\right) = \log_b m - \log_b n}$

⚠ $\boxed{(b^n)^m = b^{n \cdot m} \;=\; \log_b n^m = m \log_b n}$

HOW TO change the base OF a logarithm?

$$\boxed{\log_b K = \frac{\log_m K}{\log_m b}}$$

HOW TO change the Graph result with logarithm?

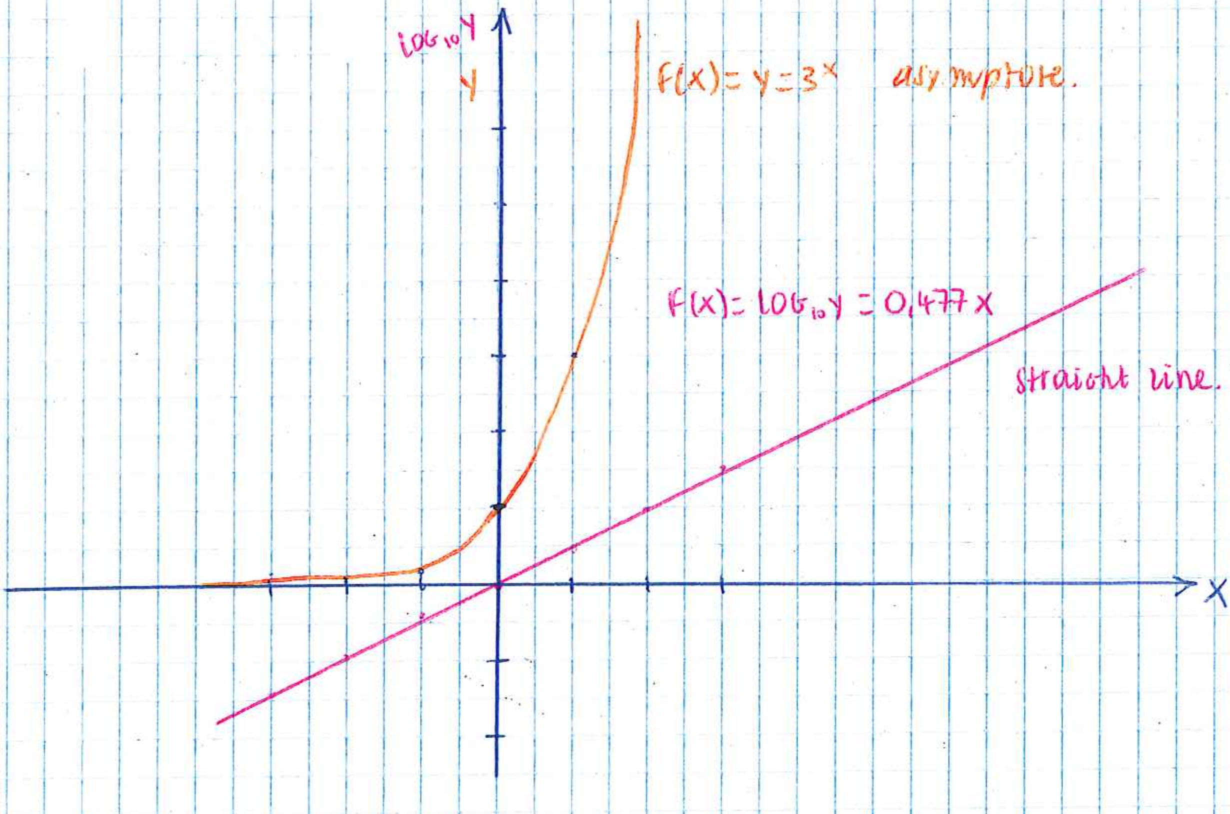
if you change the scale of the y axis in a logarithmic scale, the asymptote will become a straight line.

$\boxed{F(x) = y = 3^x}$ $\longrightarrow$ $\log_{10} y = \log_{10} 3^x$

$= x \underbrace{\log_{10} 3}_{\text{number: } 0{,}477}$

$= 0{,}477\,x$

X	$y = 3^x$	$\log_{10} y = 0{,}477\,x$
−3	$3^{-3} = 1/27$	$0{,}477 \cdot (-3) = -1{,}431$
−2	$3^{-2} = 1/9$	$0{,}477 \cdot (-2) = -0{,}954$
−1	$3^{-1} = 1/3$	$0{,}477 \cdot (-1) = -0{,}477$
0	$3^0 = 1$	$0{,}477 \cdot 0 = 0$
1	$3^1 = 3$	$0{,}477 \cdot 1 = 0{,}477$
2	$3^2 = 9$	$0{,}477 \cdot 2 = 0{,}954$
3	$3^3 = 27$	$0{,}477 \cdot 3 = 1{,}431$



V. PERCENTAGES.

$$\boxed{\% = \frac{1}{100}}$$

$$\boxed{ab + ac = a(b+c)}$$

↳ distribution.

Calculate a margin:

1. 500% of $2 = 2 \times 500\% = 2 \times \frac{500}{100} = 10$

2. ADD 10 % to 20 € :

10% of $20 = 20 \times \frac{10}{100} = 2$

$\rightarrow 20 + 2 = 22$ €

3. ADD 50% to 300 € :

50% of $300 = 300 \times \frac{50}{100}$

$\rightarrow 300 + 300 \times \frac{50}{100} = 300\left(1 + \frac{50}{100}\right) = 300\,(1 + 0{,}5)$

$= 300 \cdot 1{,}5$

ADD a % to b :

$a\%$ of $b = b \times \frac{a}{100}$

$\rightarrow \left(b + b \cdot \frac{a}{100}\right) = b\left(1 + \frac{a}{100}\right).$

ADD a% to b:

$$b\left(1 + \frac{a}{100}\right)$$

Find the original price & the margin:

selling price : \$ 280 margin : 40 %

$280 / 1{,}4 = 200.$

Calculate a discount:

Remove 20 % to \$ 280 :

$\$280 \cdot \left(1 - \frac{20}{100}\right) = \$280 \cdot (1 - 0{,}2)$

$= \$280 \cdot 0{,}8 = 224\$.$

Remove a% to b.

$$b\left(1 - \frac{a}{100}\right)$$

Calculate your interests:

I invest \$ 100 for 1 year at 10 %

$\$100 + 10\%$ of $\$100$
$\$100 \times 1{,}10 = \$100 \times (1{,}10)^1 = \$110$.

I invest \$ 100 for 2 years at 10 %

$(\$100 + 10\%) + 10\%$
$\$110 + 10\%$ of $\$110$
$\$110 \times 1{,}10 = \$121$
$\$100 \cdot 1{,}10 \cdot 1{,}10 = \$121$
$\$100 \cdot (1{,}10)^2 = \$121$

I invest \$ 100 for 10 years at 10 %.

$\$100 \cdot (1{,}10)^{10} = \$259$

to find the total sum with interests.

$$\boxed{\underbrace{S}_{\text{total sum.}} = \underbrace{P}_{\text{principal.}} \Big(1 + \underbrace{\frac{r}{100}}_{\text{interest.}} \Big)^{\overbrace{n}^{\text{number of years.}}}}$$

to find the original principal.

$$\boxed{P = \frac{S}{\left(1 + \frac{r}{100}\right)^n}}$$

HOW LONG TO INVEST \$ 100 AT 10% TO GET \$ 1000 ?

$$S = P\left(1 + \frac{r}{100}\right)^n$$

$$\underbrace{\log S}_{\text{number}} = \log\left(P\left(1 + \frac{r}{100}\right)^n\right)$$

$$\log S = \underbrace{\log P}_{\text{number}} + \log\left(1 + \frac{r}{100}\right)^n \quad \text{see formulas}$$

$$\log S = \log P + n \log\left(1 + \frac{r}{100}\right) \quad \text{see formulas}$$

$$\log S - \log P = n \log\left(1 + \frac{r}{100}\right)$$

to find the number of years.

$$\boxed{n = \frac{\log S - \log P}{\log\left(1 + \frac{r}{100}\right)}}$$

$$\hookrightarrow \; n = \frac{\log 1000 - \log 100}{\log(1+0{,}1)}$$

$$= \frac{\log(1000/100)}{\log 1{,}1}$$

see formulas

$$= \frac{\log 10}{\log 1{,}1} = \frac{1}{\log 1{,}1} = 24{,}15$$

↳ After 25 years, I will have more than \$ 1000.

<u>CHECK:</u> $100 \cdot (1+0{,}1)^{25} = 100 \cdot (1{,}1)^{25}$

$= 1083{,}47$ \$

- <u>Calculate Depreciation</u>:

<u>A company buy a car for 50 000 €. The depreciation is of 10% per year. What is its value after 5 years?</u>

↳ $50\,000\,(1-0{,}1)^5 = 50\,000 \cdot (0{,}9)^5$

$= 50\,000 \cdot 0{,}59 = 29\,500$ €.

to calculate depreciation.

$$\boxed{S = P\left(1 - \frac{r}{100}\right)^n}$$

- <u>Calculate net present value (NPV)</u>:

<u>I want 200 € in 2 years. What should I invest, at 8%?</u>

↳ $S = P\left(1 + \frac{r}{100}\right)^n$

$200 = P\left(1 + \frac{8}{100}\right)^2$

$P = \frac{200}{(1+0{,}08)^2} = \frac{200}{(1{,}08)^2} = \frac{200}{1{,}1164} = 171{,}5$

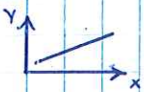
↳ If I invest now 171,5 €, at 8%, I will get 200 € in 2 years.

Find the net present value.

$$\boxed{NPV = \frac{S}{\left(1 + \frac{r}{100}\right)^n}}$$

VI. SERIES & PROGRESSIONS.

• arithmetic serie:



ADDITION

	1	2	3	4	5
values.	1	3	5	7	9

→ index.
→ values.
→ rule

+2 +2

$$S \ ; \ S+1d \ ; \ S+2d \ ; \ S+3d \ ; \ S+4d$$

S = starting value.
d = difference (rule).

→ in the example, $d = 2$.

→ the 100^{th} value would be $(S + 99d)$, that is to say, $(1 + 99.2)$.

to find the n^{th} value.

$$\boxed{n^{th} \text{ value} = S + (n-1)d}$$

to find the sum of n values.

$$\boxed{SUM_n = \frac{n}{2}\left(2S + (n-1)d\right)}$$

→ the sum of the 100 first values would be:

$$SUM_{100} = \frac{100}{2}\left(2.1 + (100-1)2\right)$$
$$= 50 \ (2 + 99.2) = 50 . 200 = 10\,000.$$

1	+	199	=	200
3	+	197	=	200
5	+	195	=	200
7	+	193	=	200
9	+	191	=	200
...		...	=	200
99	+	101	=	200

+2 +2

SUBSTRACTION

d is negative!

$$n^{th} \text{ value} = S + (n-1)(-d)$$

$$SUM_n = \frac{n}{2}\left(2S + (n-1)(-d)\right)$$

- ## Geometric serie

MULTIPLICATION

1	2	3	4	5
1	2	4	8	16

index. values. rule.

$1 \xrightarrow{\times 2} 2 \xrightarrow{\times 2} 4$

$S \; ; \; S.r \; ; \; S.r.r \; ; \; S.r.r.r \; ; \; S.r.r.r.r$

$S \; ; \; S.r^1 \; ; \; S.r^2 \; ; \; S.r^3 \; ; \; S.r^4$

S = starting value.
r = rule.

→ in the example, $r = 2$.

→ the 100^{th} value would be $(S.r^{99})$, that is to say, $(1 \cdot 2^{99})$.

to find the n^{th} value.

$$\boxed{n^{th} \text{ value} = S.r^{n-1}}$$

⚠ on the calculator:

6,34 ... E +29
= $6{,}34 \cdot 10^{29}$

to find the sum of n values.

$$\boxed{SUM_n = \frac{S\,(r^n - 1)}{r - 1}}$$

→ <u>the sum of the 100 first values would be:</u>

$$SUM_{100} = \frac{1\,(2^{100} - 1)}{(2 - 1)} = 1{,}27 \cdot 10^{30}$$

DIVISION

$9 \; ; \; 3 \; ; \; 1 \; ; \; \frac{1}{3} \; ; \; \frac{1}{9} \; ; \; \frac{1}{27}$

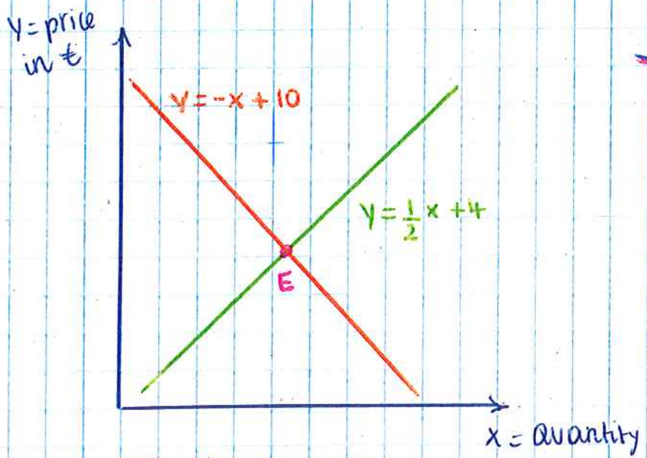
$9 \xrightarrow{:3 \,=\, \times \frac{1}{3}} 3$

→ 50^{th} value $= 9\left(\frac{1}{3}\right)^{50-1} = 9\left(\frac{1}{3}\right)^{49} = 3{,}76 \cdot 10^{-23}$

→ $SUM_{50} = 9\,\dfrac{((1/3)^{50} - 1)}{(1/3 - 1)} = 13{,}5$

VII. SYSTEMS OF EQUATIONS.

• Straight line + straight line:



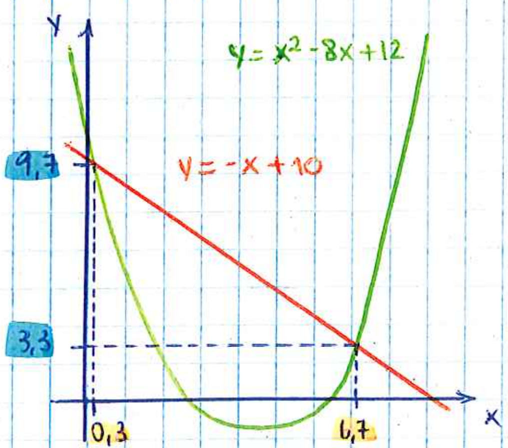
— demand curve.
— supply curve.
E market equilibrium.

→ How to find E ?

$$-x + 10 = \tfrac{1}{2}x + 4$$
$$-x - \tfrac{1}{2}x = -10 + 4$$
$$-3/2\,x = -6$$
$$3x = 12$$
$$x = 4$$

market equilibrium price.

• Straight line + Parabol:



→ Find the intercept(s) of both.

$$\begin{cases} y = -x + 10 \\ y = x^2 - 8x + 12 \end{cases}$$

$$-x + 10 = x^2 - 8x + 12$$
$$-x^2 - x + 8x - 12 + 10 = 0$$
$$x^2 - 7x + 2 = 0$$

$$D = b^2 - 4ac$$
$$= (-7)^2 - 4.1.2$$
$$= 49 - 8 = 41$$

> 0 : 2 intercepts.

$$x_1 = \frac{7 + \sqrt{41}}{2} = 6{,}7$$
$$x_2 = \frac{7 - \sqrt{41}}{2} = 0{,}3$$

} x values.

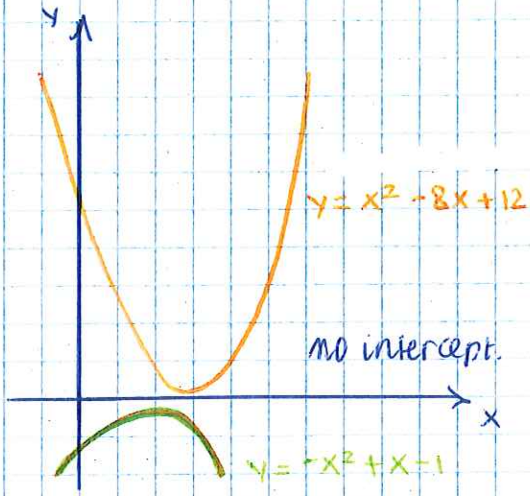
To find the y values, fill the equations:

↳ $x_1 = 6{,}7$: $y_1 = -(6{,}7) + 10 = 3{,}3$
$y_1 = (6{,}7)^2 - 8.(6{,}7) + 12 = 3{,}3$

↳ $x_2 = 0{,}3$: $y_2 = -(0{,}3) + 10 = 9{,}7$
$y_2 = (0{,}3)^2 - 8.(0{,}3) + 12 = 9{,}7$

} y values

Parabol + Parabol:

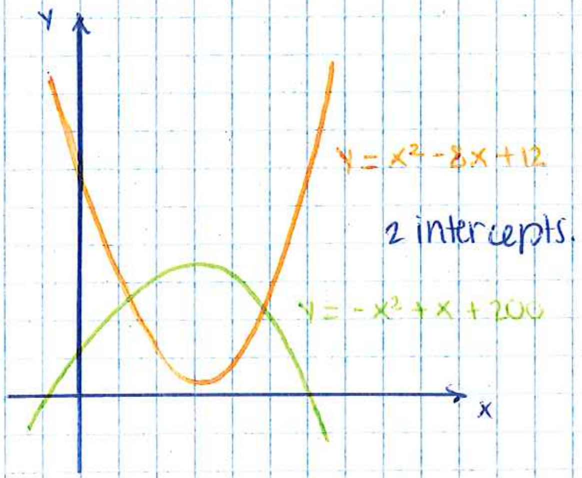


→ Find the intercepts lel both:

$$\begin{cases} y = x^2 - 8x + 12 \\ y = -x^2 + x - 1 \end{cases}$$

$$-x^2 + x - 1 = x^2 - 8x + 12$$
$$-2x^2 + 9x - 13 = 0$$

$$D = b^2 - 4ac$$
$$= 9^2 - 4.(-2).(-13)$$
$$= -23$$ < 0: no intercept.



→ Find the intercepts lel both:

$$\begin{cases} y = x^2 - 8x + 12 \\ y = -x^2 + x + 200 \end{cases}$$

$$-x^2 + x + 200 = x^2 - 8x + 12$$
$$-2x^2 + 9x + 188 = 0$$

$$D = b^2 - 4ac$$
$$= 9^2 - 4.(-2).188$$
$$= 1585$$ > 0. 2 intercepts.

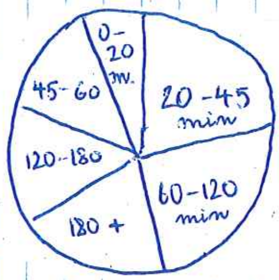
$$\left.\begin{aligned} x_1 &= \frac{-9 + \sqrt{1585}}{-4} = -7,7 \\ x_2 &= \frac{-9 - \sqrt{1585}}{-4} = 12,2 \end{aligned}\right\}$$ x values.

STAT 101 - Quantitative Methods 11.09.13

How many minutes do you call per month?

120 minutes.

minutes	people	%
0-20	1	1/23 x 100% = 4,35
20-45	5	5/23 x 100% = 21,74
45-60	4	17,4
60-120	5	21,74
120-180	4	17,4
180 +	4	17,4
	23	100 %



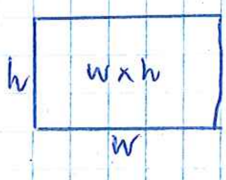
Make a sample: Take random people in the concerned area.

Histogram:

(min) → on the graph

minutes	people	%	width	height (% per min).
0-20	1	4,35	20-0=20	4,35/20 = 0,22
20-45	5	21,74	45-20=25	21,74/25 = 0,87
45-60	4	17,4	15	17,4/15 = 1,16
60-120	5	21,74	60	21,74/60 = 0,36
120-180	4	17,4	60	17,4/60 = 0,29
180-600	4	17,4	420	17,4/420 = 0,04

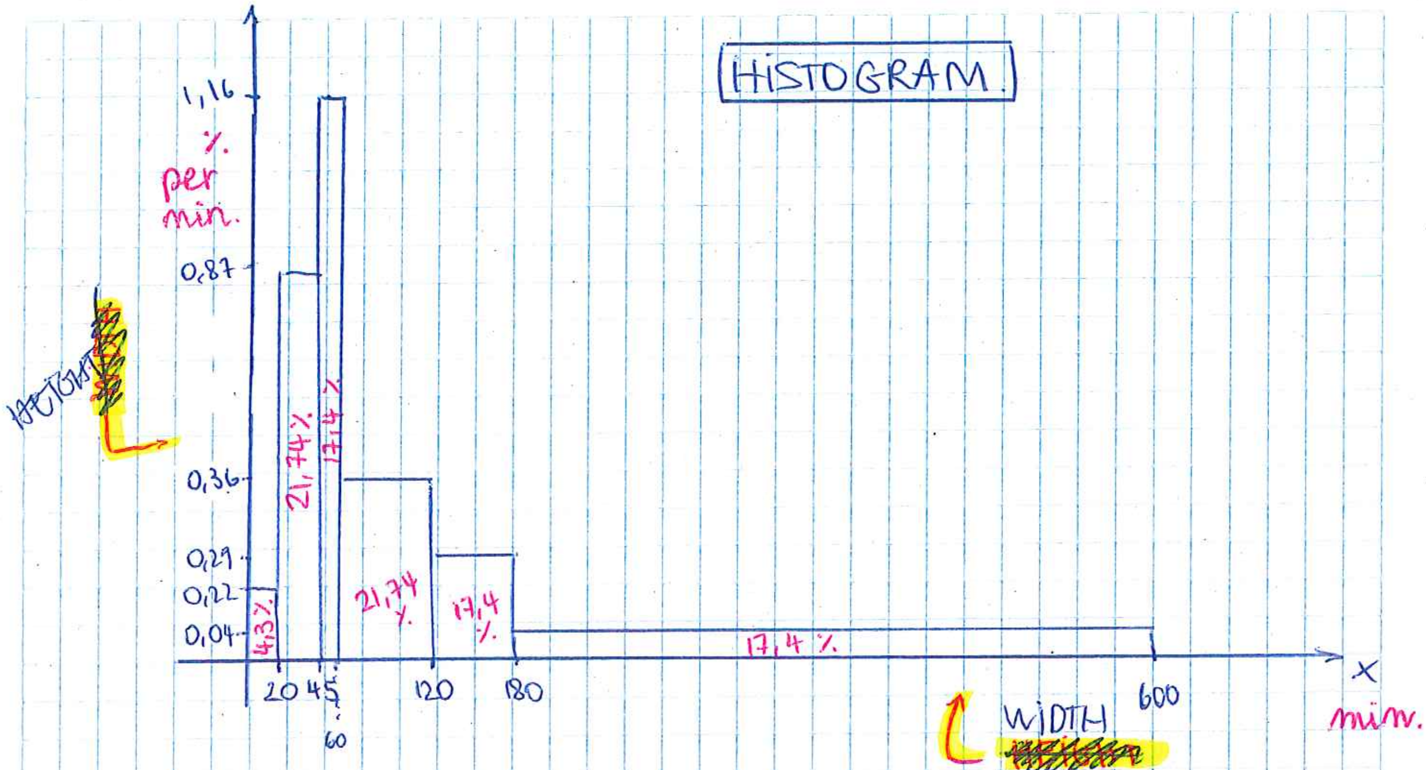
1. Convert all {counts / frequences} into %
2. Calculate the width of class intervals
   max - min = interval



3. Calculate height = $\frac{\text{area (\%)}}{\text{width}}$
4. Draw axis & label them.
5. Draw blocks of histogram.

STAT 101 — 11.09.13

HISTOGRAM



Bar chart ≠ histogram:

Bar chart	histogram
Total area ≠ 100%	Total area = 100%

its better to use intervals & no fix number, because it shows better what the result is.

AVERAGE

		middle
0–20	1	$\frac{0+20}{2}=10$
20–45	5	65/2 = 32,5
45–60	4	105/2 = 52,5
60–120	5	180/2 = 90
120–180	4	300/2 = 150
180–600	4	780/2 = 390
	23	

MODE — 20-45 min / 60-120 min

Average: add all values and divide by the number of values.

ex.: $\frac{1+2+5+7+10}{5} = \frac{25}{5} = 5$

$\frac{1}{5}\sum_{i=1}^{5} x_i$

WHEN YOU HAVE FIX NUMBERS

middle x frequency

10 x 1 = 10
32,5 x 5 = 210
52,5 x 4 = 162,5
90 x 5 = 450
150 x 4 = 600
390 x 4 = 1560
= 2995,5

AVERAGE = 2995,5 / 23 = 30 mins

WHEN YOU DON'T HAVE FIX NUMBERS BUT INTERVALS.

STAT 101 – Quantitative Methods. 16.09.13

$\boxed{h \times w = a}$

Price in €	count or frequency	area %	(€) width of each interval	height = area/width
0 - 50	10	11,1 %	50	0,222
50 - 100	20	22,2 %	50	0,444
100 - 200	30	33,3 %	100	0,333
200 - 300	20	22,2 %	100	0,222
300 - 500	10	11,1 %	200	0,055
intervals	90	100 %	x axis	y axis

SUMMARY STATISTICS.

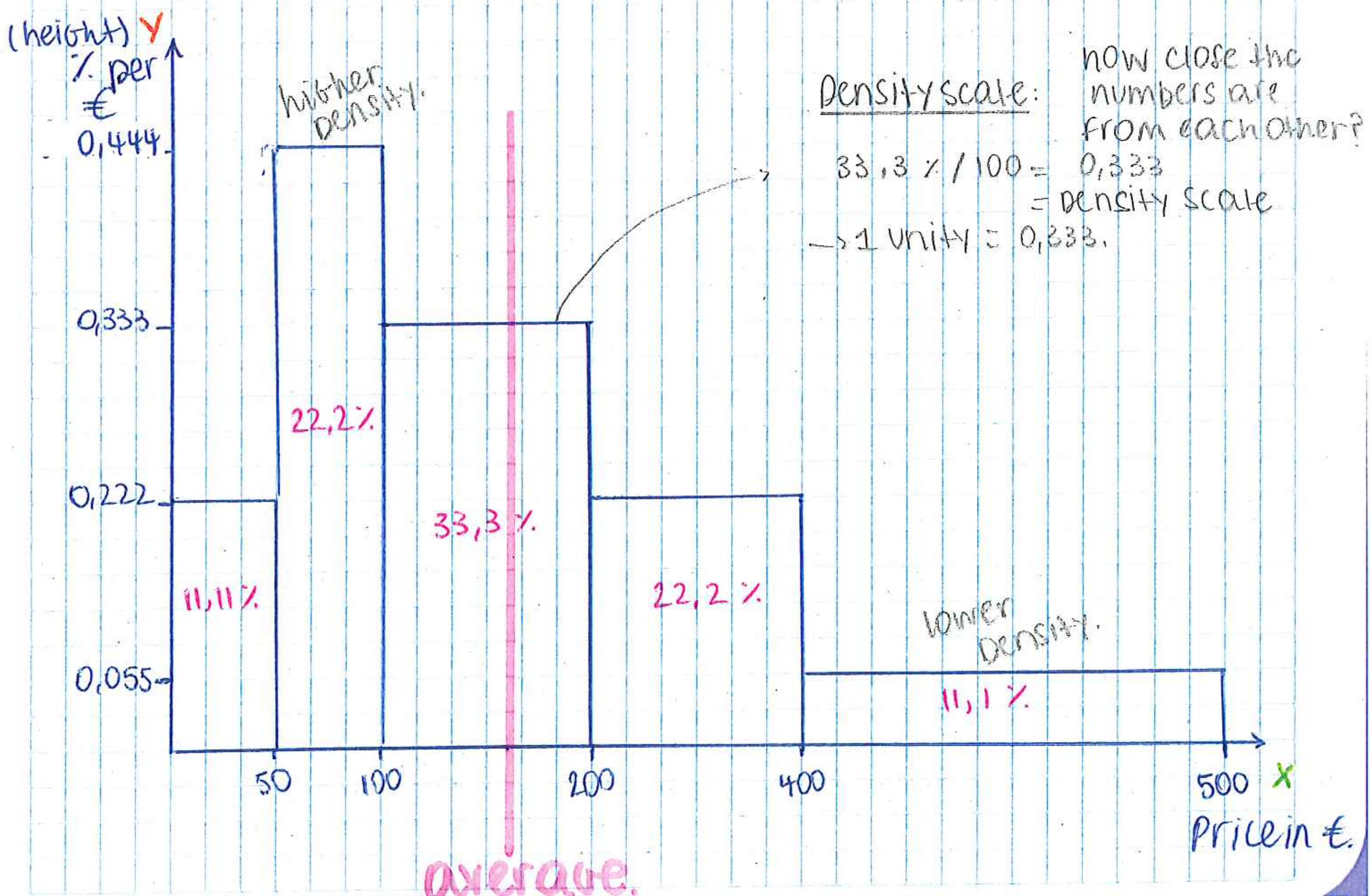
- LEFT ENDPOINT CONVENTION.

inthe interval 0 - 50 , 0 is included but 50 is excluded.

↳ [0 ; 50[

- HISTOGRAM :

100 % Below the curve.



- MODE (highest peak): the mode is 50 - 100 interval.

it is not always the middle.

- the average : $\frac{\text{sum of all values}}{\text{number of values.}}$

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

Example: 1 ; 3 ; 5 ; 7 ; 10 — median (5); 5,2 : average.

x_1 x_2 x_3 x_4 x_5

average of $x_i = \overline{x}$

$= \mu$ (greek M "mean")

$$\overline{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{n}$$

$$= \frac{1 + 3 + 5 + 7 + 10}{5} = \frac{26}{5} = 5{,}2$$

$$= \frac{1}{5} \sum_{i=1}^{5} x_i$$

- the median :

50% is smaller ← median → 50 % is higher

is the example serie, the median is 5.

it is a number of the list.

if there is un ~~ ~~ pair number list:

1 ; 3 ; 5 ; 7 ; 10 ; 11

↓ median

$\frac{5+7}{2} = 12/2 = 6$

⚠ if I change an extreme number, the average will change, BUT the median is still the same.

→ 1 ; 3 ; 5 ; 7 ; 100 — average: 23,2. median: 5.

⚠ in an asymetric histogram, look at the median more than the average, cause the extreme value influences the average.

in a complete symetric histogram, the average & the median are the same.

16.09.13

→ when you have intervals instead of numbers:

how to calculate the average?
→ Find the middle of the intervals.

intervals:	count:	middle of intervals:	middle × count
0 - 50	10	25	250
50 - 100	20	75	1500
100 - 200	30	150	4500
200 - 300	20	250	5000
300 - 500	10	400	4000
	90		15250

$\underbrace{10+20}_{30}$

average:

$$\frac{15250}{90} = 169{,}4\ €$$

balanceing point of the histogram.

median: 50th percentile.

median interval: 100-200

median exact number: formula p. 175-187.

$$\boxed{L + i\left(\frac{\frac{n}{2} - F}{f}\right)}$$

L = left boundary of the median interval. 100
i = width of median interval. 100
n = total count 90
F = cumulative frequency of all intervals below the median interval. 10+20 = 30
f = frequency inside the median interval. 30

$$100 + 100\left(\frac{\frac{90}{2} - 30}{30}\right)$$
$$= 100 + 100 \times 0{,}5$$
$$= 100 + 50 = 150$$

$$L + i\left(\frac{\frac{n}{4} - F}{f}\right)$$

→ we will find the 25th percentile.

$$\frac{90}{4} = 22{,}5$$

25th percentile:

$$50 + 50\left(\frac{\frac{90}{4} - 10}{20}\right) = 81{,}25$$

L = 50-100 | 20 $\frac{90}{4} = 22{,}5$
= 50

in this interval

i = 50
n = 90
F = 10
f = 20

because:
10 → 10
20 → 30
30 → 60
20 → 80
10 → 90

22,5 is between 10 and 30.

75th percentile: $\frac{3.90}{4} = 67,5$ in this interval

0 - 50	10
50 - 100	20 → 30
100 - 200	30 → 60
200 - 300	20 → 80
300 - 500	10 → 90

67,5 is between 60 & 80

median interval = 200-300 | 20

L = 200 F = 10+20+30
i = 100 = 60
n = 90 f = 20

$$75^{th} = L + i\left(\frac{\frac{3.n}{4} - F}{f}\right) = 200 + 100\left(\frac{\frac{3.90}{4} - 60}{20}\right)$$
$$= 200 + (100 . 0,375)$$
$$= 200 + 37,5 = 237,5$$

- INTERQUARTILE RANGE (IQR) = middle 50%

IQR = 237,5 - 67,5 = 170 €

IQR = the ≠ between the 25th percentile & the 75th percentile is 170 €.

AVERAGE = 169,4 €
MEDIAN = 150 €
IQR = 170 €
TOTAL RANGE = 500 € (maximum).

if average = median => symetrical histogram.

with the MEDIAN
IQR
TOTAL RANGE
→ general idea of the histogram.

50% => median.
25% & 75% => IQR.
0% & 100% => TOTAL RANGE.

$\boxed{IQR = 75^{th}\ perc. - 25^{th}\ perc.}$

- AVERAGE

HOW far are the values from the average?

ex.: 1 ; 3 ; 5 ; 7 ; 10 average = 5,2

↳ How far are the values away from AVG (5,2) on average?

1 ← 5,2 = -4,2
3 ← 5,2 = -2,2
5 ← 5,2 = -0,2
7 ← 5,2 = +1,8
10 ← 5,2 = +4,8

$$\frac{(-4,2)+(-2,2)+(-0,2)+(1,8)+(4,8)}{5} = \frac{0}{5} = 0$$

on average, the values are average!

16.09.13

⚠ if we take the **absolute values**:

$$\frac{|-4{,}2| + |-2{,}2| + |-0{,}2| + 1{,}8 + 4{,}8}{5}$$

|values|

average deviation $= 2{,}64$

how far the values are from 5,2, no matter left or right they are (- or +) on average.

$$\boxed{\text{absolute value} = {}^{+}\sqrt{x^2}}$$

$$\sqrt{\frac{(-4{,}2)^2+(-2{,}2)^2+(-0{,}2)^2+(1{,}8)^2+(4{,}8)^2}{5}}$$

standard deviation (SD) $= 3{,}12$ **BIGGER.**

the big numbers have a bigger weight.

RMS: (root mean square)

SD = RMS of the { Distances / deviations } from average.

AVG = average. $= \bar{x}$

SD =
1). calculate AVG.
2). calculate Distances from AVG (from middle) → if intervals!
3). square Distances.
VARIANCE: 4). take mean (AVG) of squared Distances.
5). take root of the mean.

ex.: $\boxed{1 ; 3 ; 5 ; 7 ; 10}$

1). AVG = 5,2
2). (-4,2) ; (-2,2) ; (-0,2) ; (1,8) ; (4,8)
3). 17,64 ; 4,84 ; 0,4 ; 3,24 ; 23,04
$= (-4{,}2)^2$

4). $$\frac{17{,}64 + 4{,}84 + 0{,}4 + 3{,}24 + 23{,}04}{5} = 9{,}832$$ = **VARIANCE**.

5). $\sqrt{9{,}832} = 3{,}14$ = **SD**
$= \sqrt{\text{variance}}$

FORMULA of these steps:

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

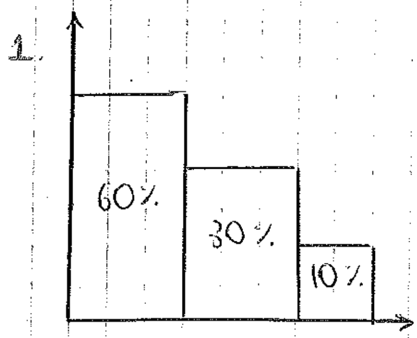
1. 2 3 4 5

STAT 101- Quantitative Methods. 23.09.13

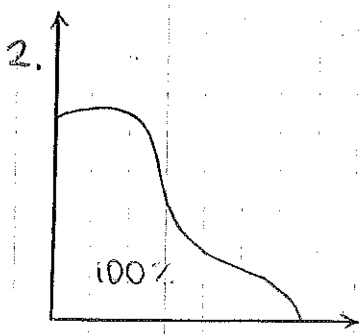
MIDTERM: WEDNESDAY 09.10

HISTOGRAM

1.

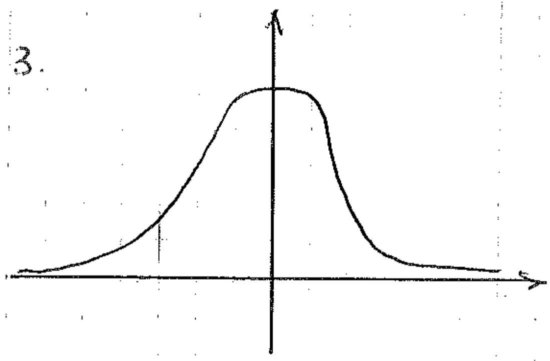


2.



the area is always representing 100% !

3.



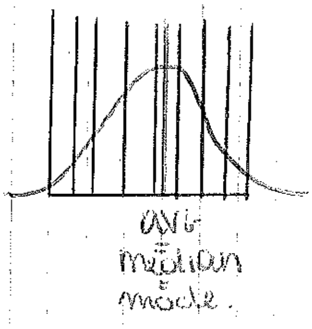
→ NORMAL CURVE HISTOGRAM.

always 50% that the ball goes left or right.

GALTON BOARD.

If we do this experiment with lot of balls, we will obtain a normal curve as the result of wher the balls fall.

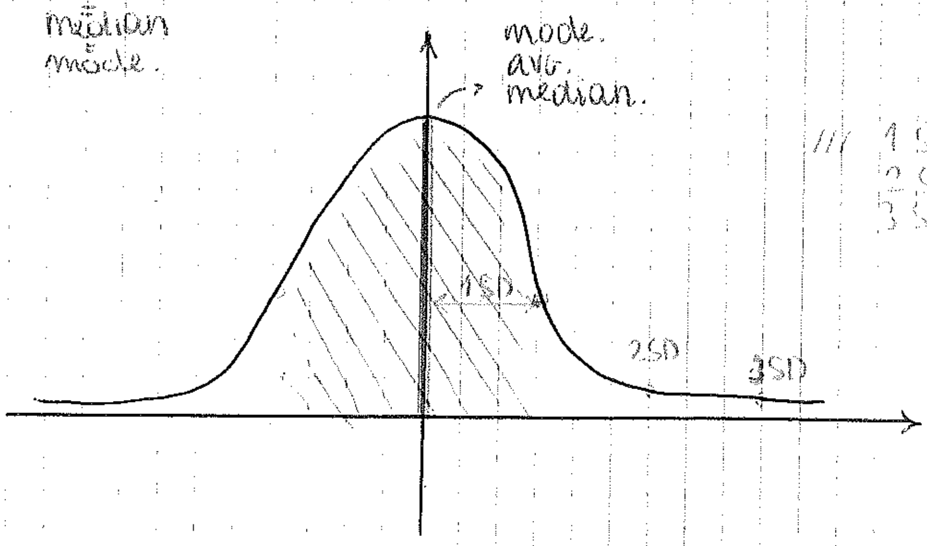
⇓



Final result with more balls: normal curve.

Because it's symetric, the average = the median

Because it's the highest point, it's also the mode.

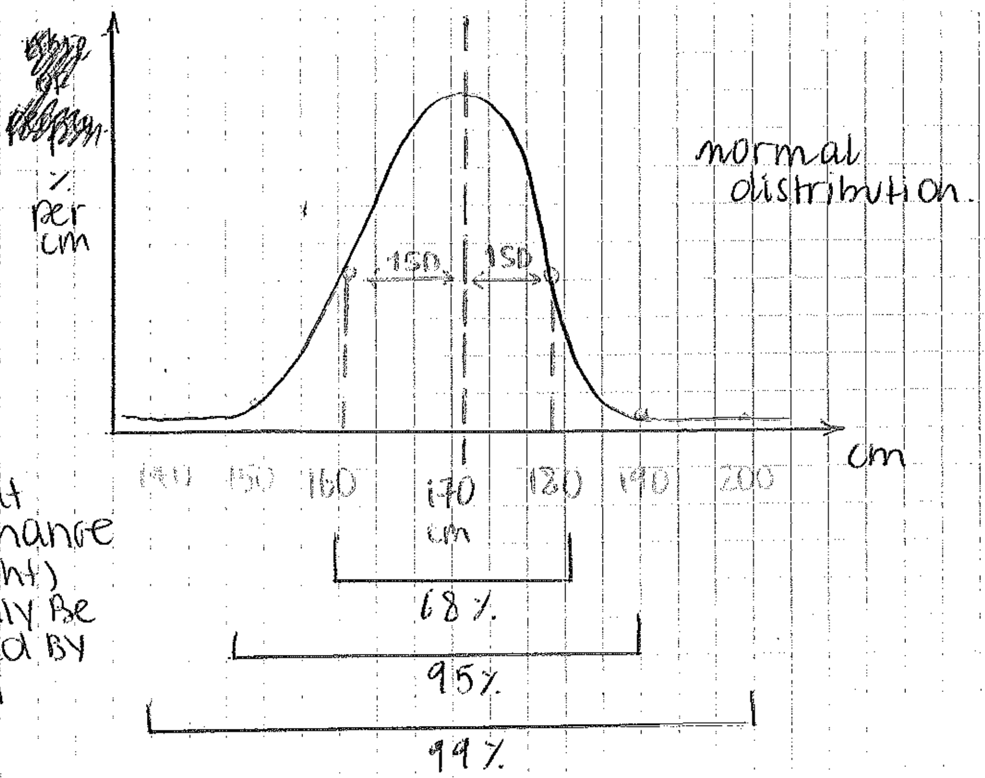


/// 1SD = 68%
2SD = 95%
3SD = 99%

KAMPLE:

:age
ht
-0 cm.

cm.



ngs that
nnot change
ke height)
an usually be
presented by
normal
curve.

N THE CALCULATOR:

[2nd] [Distr]

(.. ; ..)
×100%

= normal cdf.

- normal cdf = (-1 ; +1) × 100%
= 68,27 %

- normal cdf = (-2 ; +2) × 100%
= 95,45 %

- normal cdf = (-3 ; +3) × 100%
= 99,7 %

To calculate a part, for example, between 140 cm and 180 cm.

$\boxed{\text{normal cdf (lower bound ; upper bound ; avg ; SD)}}$ ×100%

normal cdf (140 ; 180 ; 170 ; 10) = 83,9 %
×100%

or: normal cdf (180 ; ∞ ; 170 ; 10) = 15,8 %
10^{99}
×100%

23.09.13

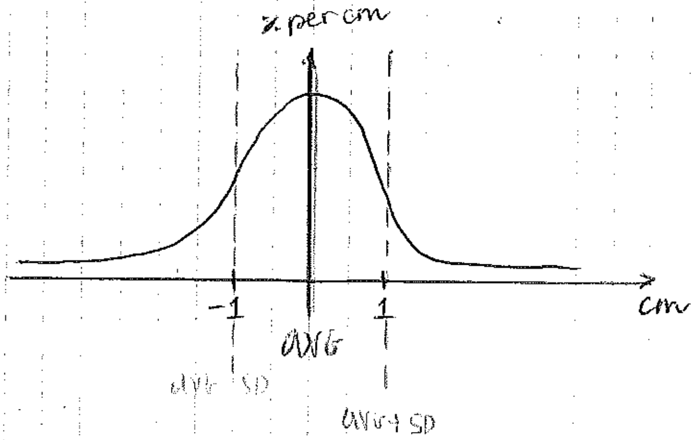
Standard unit:

To compare things that are ≠ from each other.

definition.

AVG → 0
AVG + SD → 1
AVG − SD → −1

$$\boxed{SU = \frac{\text{value} - \text{AVG}}{SD}}$$



$$\Rightarrow \frac{180\text{ cm} - 170\text{ cm}}{10\text{ cm}}$$

$$= \frac{10}{10} = 1$$

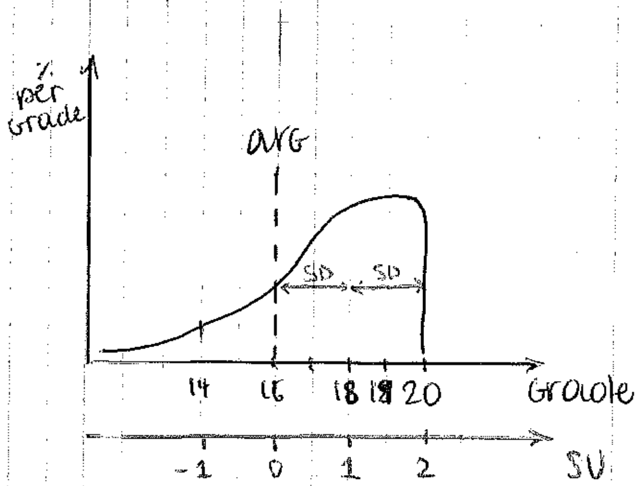
How many SDs is a value away from average.

180 cm is 1SD (10 cm) away from AVG (170 cm).

$$\boxed{\begin{array}{l} SU \times SD = \text{value} - \text{AVG} \\ \text{value} = SU \times SD + \text{AVG} \end{array}}$$

$$\Rightarrow \quad 2 \times 10 + 170 = 190$$

EXAMPLE:



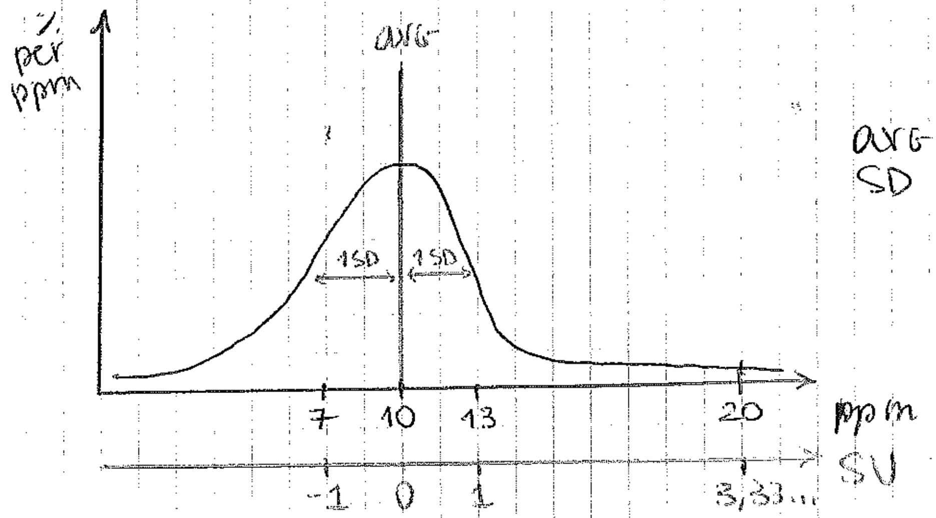
How many SDs is a student who has 14 away from AVG?

→ Translate in SU.

$$SU = \frac{14 - 16}{2} = \frac{-2}{2} = -1$$

EXAMPLE:

calculating air quality in ppm (particules per million).



avg = 10
SD = 3

HOW much is 20 ppm in SU?

↳ $\frac{20-10}{3} = \frac{10}{3} = 3,33...$

HOW much is 0 ppm is SU?

↳ $\frac{0-10}{3} = \frac{-10}{3} = -3,33...$

---

are	count	%	$x_i - \bar{x}$
.10	6	24	-1,25
.10	0		-0,25
.10	19	76	0,75
	25	100%	

$(x_i - \bar{x})^2$

$6 \times 1,5625$
$0 \times 0,0625$
$19 \times 0,15625$

↳ count $\times (x_i - \bar{x})^2$
9,375
0
10,6875

$\text{avg} = \frac{0 \times 7 + 9 \times 19}{25} = 8,25$

$SD = \sqrt{\text{variance}} = \sqrt{\frac{1}{25}\sum_{i=1}^{25}(x_i - \bar{x})^2}$

$= \sqrt{0,8025} = 0,8958$

$\Rightarrow 9,375 + 0 + 10,6875 = 20,0625$ $\Big]\ \sum_{i=1}^{25}$

divided by the total count (25): $\Big]\ \frac{1}{25}$

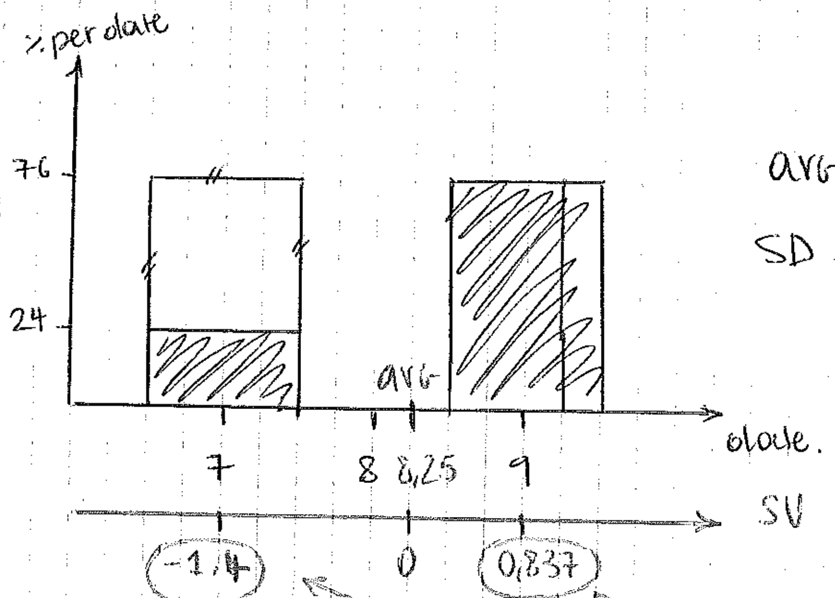
$20,0625 / 25 = 0,8025$ = VARIANCE.

H INTERVAL, IKE THE MIDDLE ALUE OF THE NTERVALS.

23.09.13

date	count	%	width	height
7	6	24	1	24
8	0	0	1	0
9	19	76	1	76
	25			

HISTOGRAM:



avg = 8,25

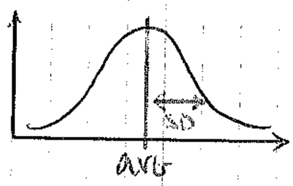
SD = 0,8958

How many SDs is 9 away from avg?

Translate 9 in SU: $\frac{9 - 8,25}{0,896} = 0,837$

$\frac{7 - 8,25}{0,896} = -1,4$

Normal curve:



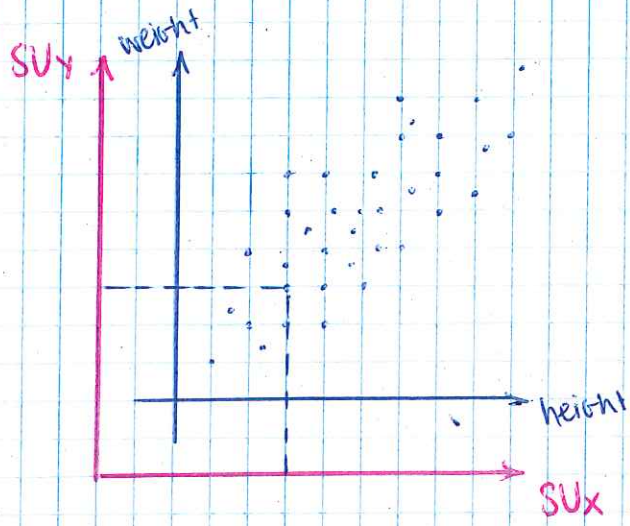
ie) -1SD & 1SD = 68%
-2SD 2SD = 95%
-3SD 3SD = 99%

$$SU = \frac{\text{value} - \text{avg}}{SD} = \frac{x_i - \bar{x}}{SD}$$

$$\text{value} = (SU \times SD) + \text{avg} = (SU \times SD) + \bar{x}$$

STAT 101 - Quantitative methods. 30.09.13

XIII. Correlations.

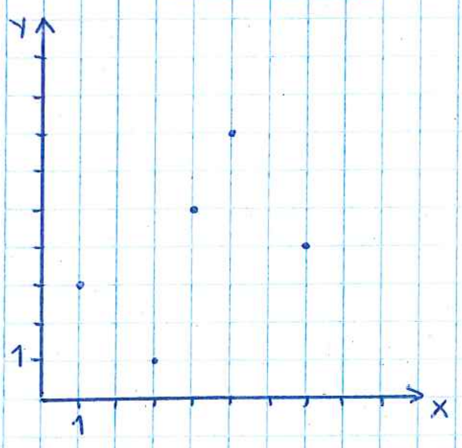


- 2 different variables (heigh & weight) cannot be compared.
- you can only compare 2 different variable from SUx and SUy.
- the degree of dependency between 2 variables is called the correlation.

To calculate the correlation coefficient (r):

1. Convert x values in SUx $\left(\frac{x_i - \bar{x}}{SD_x}\right)$.
   Convert y values in SUy $\left(\frac{y_i - \bar{y}}{SD_y}\right)$.
2. Multiply all SUx and SUy.
3. Average of all products is r.
   the correlation is always between -1 and 1
   → r = -1 : perfect negative correlation.
   → r = 0 : no correlation.
   → r = 1 : perfect correlation.

EXAMPLE:



x	y	SUx	SUy	SUx.SUy
1	3	-1,5	-0,5	0,75
3	1	-0,5	-1,5	0,75
4	5	0	0,5	0
5	7	0,5	1,5	0,75
7	4	1,5	0	0
				2,25

$\bar{x} = 4$ $\bar{y} = 4$
$SD_x = 2$ $SD_y = 2$

$\frac{2,25}{5} = 0,45 = r.$

↳ how strong the correlation is between x and y.

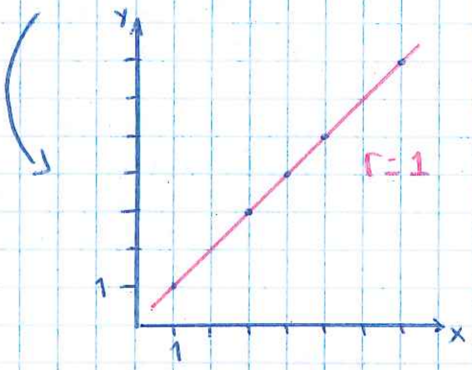
EXAMPLE:

X	Y	SU_x	SU_y	$SU_x . SU_y$
1	1	1,5	1,5	2,25
3	3	-0,5	-0,5	0,25
4	4	0	0	0
5	5	0,5	0,5	0,25
7	7	1,5	1,5	2,25
				5

$\bar{x} = 4 \qquad \bar{y} = 4$
$SD_x = 2 \qquad SD_y = 2$

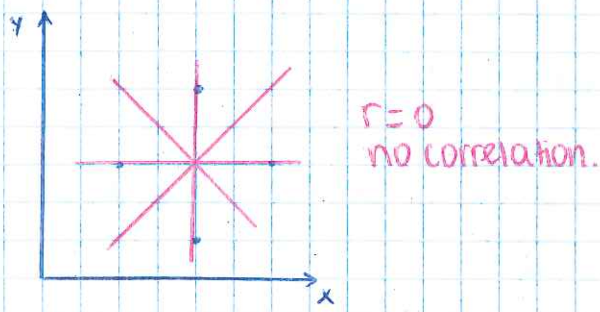
$r = \frac{5}{5} = 1$

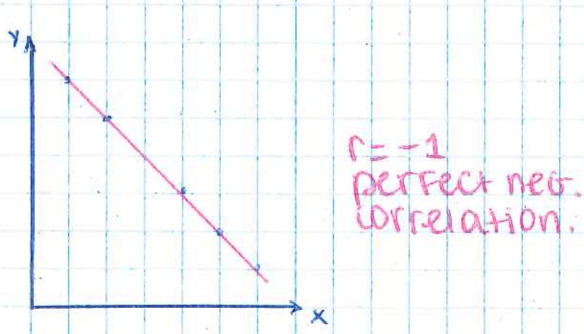
↳ perfect correlation

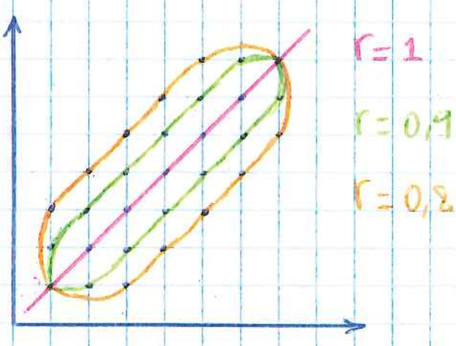


⚠ the line in a perfect correlation does not always cross (0;0).

OTHER EXAMPLES:







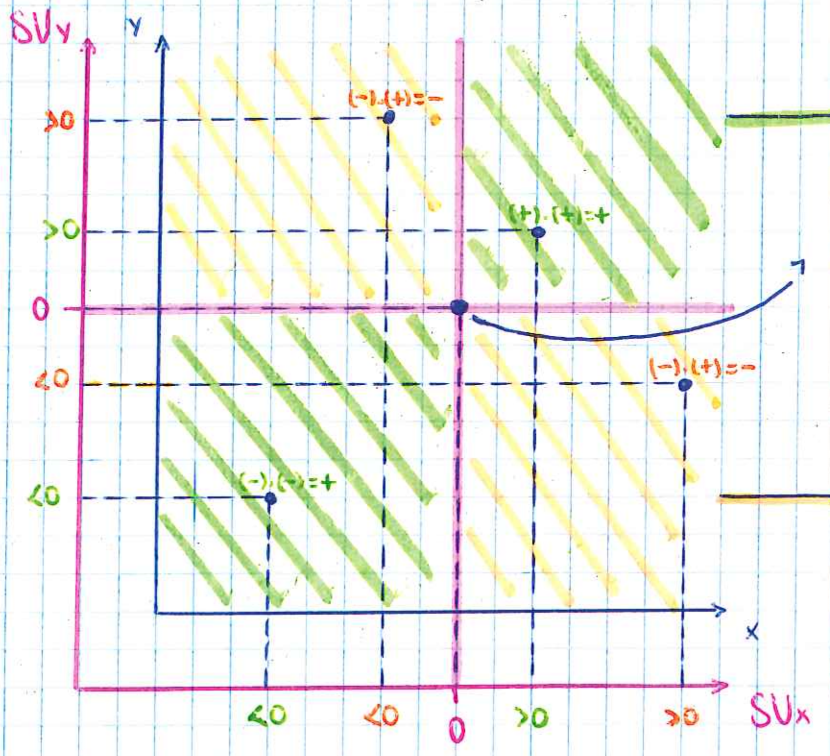
→ Objectives of calculating the correlation:

- Determine how strong is a correlation between 2 variables.
- Predict y values from x values (only in a perfect correlation).

2. To determine how strong a correlation is:

1. convert x and y values into SU_x and SU_y $\frac{x_i - \bar{x}}{SD_x}$
2. Multiply all SU_x and SU_y.
3. r = avg of all products.

$$\boxed{r = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i - \bar{x}}{SD_x} \cdot \frac{y_i - \bar{y}}{SD_y}}$$



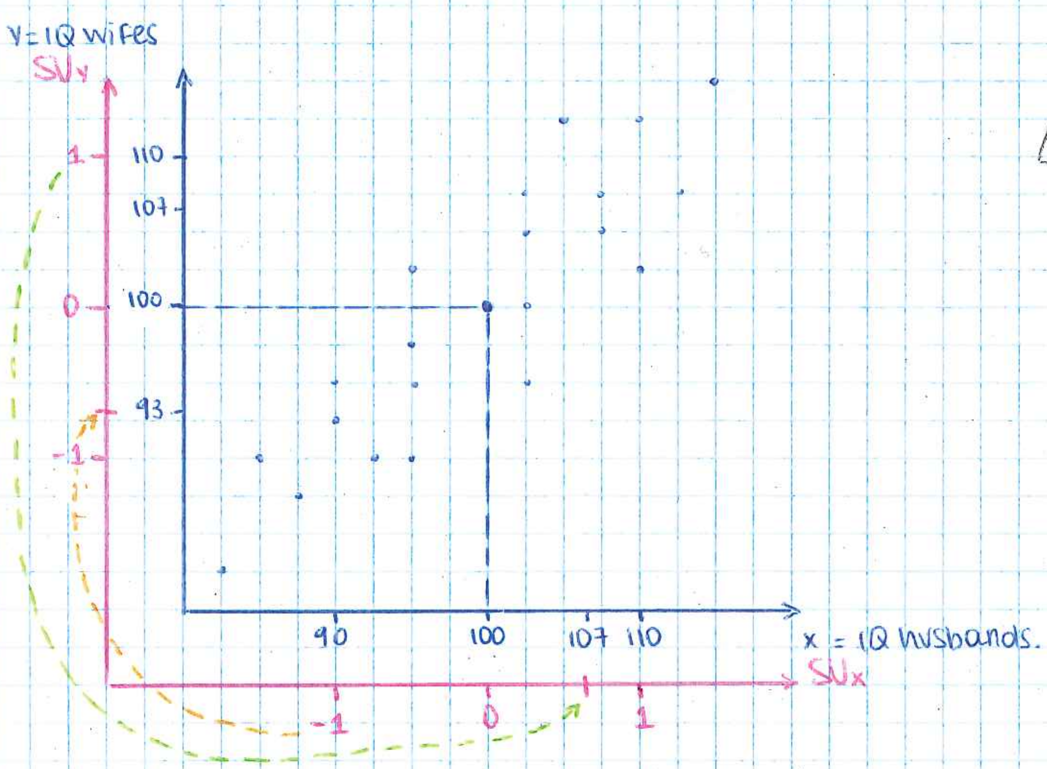
3. To predict y values from some x values:

1. convert x into SUx $\left(\frac{x - \bar{x}}{SD_x} = SU_x\right)$
2. Multiply by r to obtain SU_y $(SU_x \cdot r = SU_y)$
3. convert SU_y into y $(SU_y \cdot SD_y + \bar{y} = y)$.

$$\boxed{y_i = \left(\frac{x_i - \bar{x}}{SD_x} \cdot r\right) \cdot SD_y + \bar{y}}$$

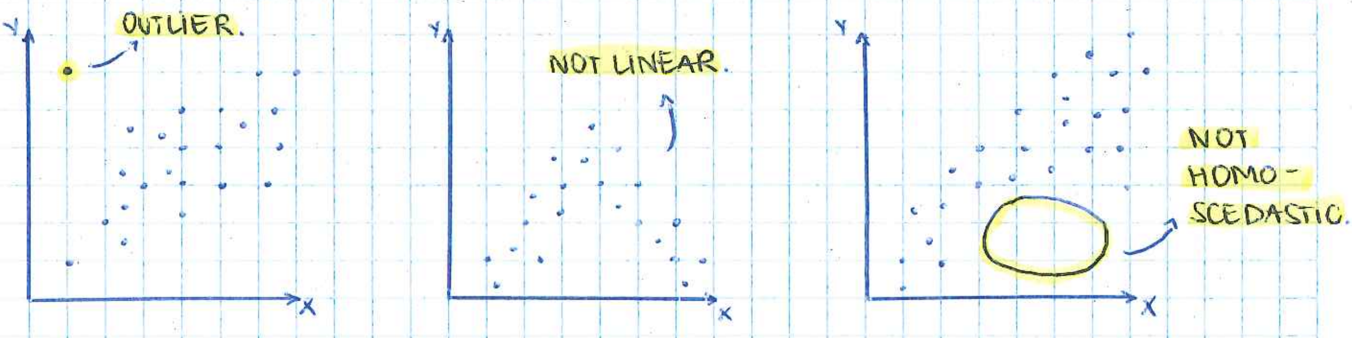
→ The "Regression effect":

When we predict values from the other variable, the result will always be closer and closer to the average (SD=0).



⚠ regression effect works better when the points form a football shape!

→ How to recognize a normal curve?



a normal curve should:

- have no outlier.
- Being not linear.
- Being homoscedastic.

4. COMPLETE EXAMPLE:

x cost of machine in €	y times before it breaks in years	cost in SU_x	time in SU_y	$SU_x . SU_y$
500	1	-1,34	-1,52	2,0368
1000	4	-0,45	0,506	-0,2277
1500	3	0,48	-0,17	-0,0816
2000	5	1,34	1,18	1,5812

1. Averages:

$$\text{avg } x = \frac{500 + 1000 + 1500 + 2000}{4} = 1250 = \bar{x}$$

$$\text{avg } y = \frac{1 + 4 + 3 + 5}{4} = 3,25 = \bar{y}$$

2. SD's:

1. $(x_i - \bar{x})^2$
2. $\sqrt{\frac{\text{added together}}{\text{nbre of values}}}$

SD_x:

$500 - 1250 = -750 \xrightarrow{(\ )^2} 562\,500$
$1000 - 1250 = -250 \rightarrow 62\,500$
$1500 - 1250 = 250 \rightarrow 62\,500$
$2000 - 1250 = 750 \rightarrow 562\,500$

$\sqrt{\frac{\text{added together}}{4}} = 559$

SD_y:

$1 - 3,25 = -2,25 \xrightarrow{(\ )^2} 5,0625$
$4 - 3,25 = 0,75 \rightarrow 0,5625$
$3 - 3,25 = -0,25 \rightarrow 0,0625$
$5 - 3,25 = 1,75 \rightarrow 3,0625$

$\sqrt{\frac{\text{added together}}{4}} = 1,48$

→ How strong is the correlation?

3. Convert x and y values into SU_x and SU_y $\left(\frac{x_i - \bar{x}}{SD_x}\right)$

x:
$500 - 1250 / 559 = -1,34$
$1000 - 1250 / 559 = -0,45$
$1500 - 1250 / 559 = 0,48$
$2000 - 1250 / 559 = 1,34$

y:
$1 - 3,25 / 1,48 = -1,52$
$4 - 3,25 / 1,48 = 0,506$
$3 - 3,25 / 1,48 = -0,17$
$5 - 3,25 / 1,48 = 1,18$

4. Multiply all SU_x and SU_y

5. Average of all products:

$$\frac{2,0368 + (-0,2277) + (-0,0816) + 1,5812}{4} = \frac{3,3087}{4}$$

$$= 0,83 = r.$$

→ How long do we expect a machine of 3000 € to last?

(prediction of y value)

AVG x = 1250
SD x = 559

AVG y = 3,25
SD y = 1,48

6. $$y_i = \left(\frac{x_i - \bar{x}}{SD_x} \cdot r\right) \cdot SD_y + \bar{y}$$

$$\hookrightarrow y_{3000} = \left(\frac{3000 - 1250}{559} \cdot 0{,}83\right) \cdot 1{,}48 + 3{,}25$$

$$= 7 \text{ years.}$$

→ What price for a machine that would last 6 years?

(prediction of x value)

AVG x = 1250
SD x = 559

AVG y = 3,25
SD y = 1,48

7. $$x_i = \left(\frac{y_i - \bar{y}}{SD_y} \cdot r\right) \cdot SD_x + \bar{x}$$

$$\hookrightarrow x_6 = \left(\frac{6 - 3{,}25}{1{,}48} \cdot 0{,}83\right) \cdot 559 + 1250$$

$$= 2112 \text{ €.}$$

5. To find the line on the computer with Excel.

x years	y number of members
1990	5
1995	20
2000	35
2005	30
2010	40

→ To find the line to predict y values:

- create a graph with all values.
- select all points in the graph.
- click on "trend line" in the presentation section.

$y = 1{,}6x - 3174$

AVG x = 2000
SD x = 7,071068

AVG y = 26
SD y = 12,40967

r = 0,91

6. To find the line (the function) manually.

→ To predict y from x:

- $\boxed{\text{slope} = \frac{SD_y}{SD_x} \cdot r}$ → $\text{slope} = \frac{12{,}4}{7{,}07} \cdot 0{,}91 = 1{,}6$
- $\boxed{\text{intercept} = \left(\frac{0 - \bar{x}}{SD_x} \cdot r\right) \cdot SD_y + \bar{y}}$ formula to find y values from x values!

→ $\text{intercept} = \left(\frac{0 - 2000}{7{,}07} \cdot 0{,}91\right) \cdot 12{,}4 + 26 = -3166$

- function: $y = 1{,}6x - 3166$

→ To predict x from y:

- $\boxed{\text{slope} = \frac{SD_x}{SD_y} \cdot r}$ → $\text{slope} = \frac{7{,}07}{12{,}4} \cdot 0{,}91 = 0{,}52$
- $\boxed{\text{intercept} = \left(\frac{0 - \bar{y}}{SD_y} \cdot r\right) \cdot SD_x + \bar{x}}$

→ $\text{intercept} = \left(\frac{0 - 26}{12{,}4} \cdot 0{,}91\right) \cdot 7{,}07 + 2000 = 1986{,}5$ (june 1986)

- function: $x = 0{,}52y + 1986{,}5$
  $y = 1{,}92x - 3820$

7. To calculate the prediction error (RMS error).

→ When predicting y from x : RMS error $= \sqrt{1-r^2} \cdot SD_y$

→ When predicting x from y : RMS error $= \sqrt{1-r^2} \cdot SD_x$

If r is 1 or -1, all the points are on the line, there is no prediction error.

If r is 0, the prediction error is equal to SD_x or SD_y (the biggest possible error).

Ex: When predicting y from x:

RMS error $= \sqrt{1-0{,}91^2} \cdot 12{,}4 = 5{,}14$

In 2015, we predicted 50 members with an RMS error of 5,14. The result is then between 44,86 and 55,14 members in 2015.

HOMEWORK

X	Y
year	price in €
2000	134
2003	120
2004	125
2008	107
2010	95

1). Find correlation between x & Y.

2). What price do we expect in 2013? + RMS error.

3). Equation of best line through points.

XIV. Introduction to logic and probabilities.

- **Logic**: "the light in the class is on" — proposition (T/F).
  - ↳ True : 1.
  - False : 0.
- **Probabilities**: "it will rain tomorrow"
  - ↳ I'm sure for 80% : 0,8.
  - I'm not sure for 20% : 0,2.
- "Is the light on?"
  "Turn the light on!" — sentences which are not propositions.
- "There is life on a planet around the star Sirius"
  - ↳ it is not a proposition because we cannot check if it is true or not.

EXAMPLE PROPOSITIONS:

a. $1+1=2$ ⟹ true, 1.
b. $3+4=7$ ⟹ true, 1.

a + b. $\underbrace{1+1=2}_{1}$ and $\underbrace{3+4=7}_{1}$ ⟹ true, 1.

c. $\sqrt{4}=87$ ⟹ false, 0.

a + c. $\underbrace{1+1=2}_{1}$ and $\underbrace{\sqrt{4}=87}_{0}$ ⟹ false, 0.

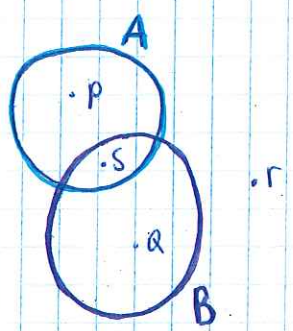
1. To make a truth table.

0 = False.
1 = True.

A	B	A ∧ B (and)
0	0	0
0	1	0
1	0	0
1	1	1

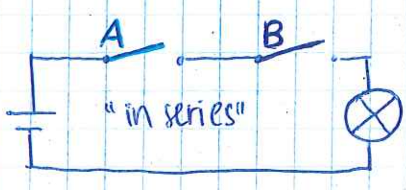
↳ we can replace logic calculation by multiplication. ×

//



A ∩ B the point needs to be inside both circles.

//

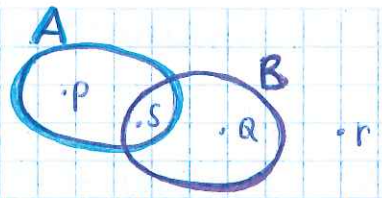


A ∧ B both interrupters need to be closed.

$+$ like an addition.

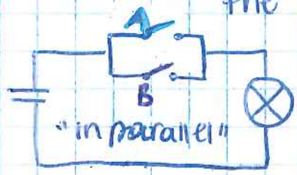
A	B	inclusive or $A \lor B$
0	0	0
0	1	1
1	0	1
1	1	1

//



$A \cup B$ one of the points needs to be inside the circle.

//



A	B	exclusive or $A \oplus B$
0	0	0
0	1	1
1	0	1
1	1	0

"either A or B".

will not be used in quantitative methods class.

A	$\neg A$ ($\sim A$) (not A)	// $1-A$
0	1	$1-0=1$
1	0	$1-1=0$

A	B	and $A \land B$	or $A \lor B$	ex or $A \otimes B$	not $\neg A$	not $\neg B$	$A + B$	if A, then B $A \supset B$
0	0	0	0	0	1	1	1	1
0	1	0	1	1	1	0	0	1
1	0	0	1	1	0	1	0	0
1	1	1	1	0	0	0	1	1
operator name		conjunct°	disjunct°	negat° of equivalence	negation		equivalence	implicat°

STAT 101- Quantitative methods. | 21.10.13

2. Combinations.

A	B	$A \wedge B$	$\neg(A \wedge B)$	$\neg A$	$\neg B$	$\neg A \vee \neg B$
			not (A and B)			(not A) or (not B)
0	0	0	1	1	1	1
0	1	0	1	1	0	1
1	0	0	1	0	1	1
1	1	1	0	0	0	0

SAME RESULT

$\Rightarrow \boxed{\neg(A \wedge B) = \neg A \vee \neg B}$

not (A and B) = (not A) or (not B).

A	B	C
0 0	0 0	0 1
0 0	1 1	0 1
1 1	0 0	0 1
1 1	1 1	0 1

the number of lines in a truth table is $(\text{number})^2$ of propositions.

$\rightarrow$ A, B = 2 propositions.
$= 2^2 = 4$ lines.

$\rightarrow$ A, B, C = 3 propositions.
$= 3^2 = 8$ lines.

A	B	$A \vee B$	$\neg(A \vee B)$	$\neg A$	$\neg B$	$\neg A \wedge \neg B$	$\neg(A \vee B) = \neg A \wedge \neg B$
			not (A or B)			(not A) or (not B)	not (A or B) = (not A) & (not B)
0	0	0	1	1	1	1	1
0	1	1	0	1	0	0	1
1	0	1	0	0	1	0	1
1	1	1	0	0	0	0	1

TAUTOLOGY = ALWAYS TRUE.

SAME RESULT

$\Rightarrow \boxed{\neg(A \vee B) = \neg A \wedge \neg B}$

not (A or B) = (not A) & (not B).

3. Use in Probabilities.

0 = False (certain).
1 = true (certain).

↳ Between the two is uncertainty.

0,001% ↳ 0,00001
99,9999% = 0,99999
50% 0,5
... ...

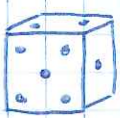
$$\boxed{\begin{array}{c} 0 \leq \text{PROBABILITY} \leq 1. \\ 0\% \leq \underset{P(A)}{\text{PROBABILITY}} \leq 100\%. \end{array}}$$

the chance that something Don't happen

$$\boxed{\begin{array}{l} P(\text{not } A) = P(\neg A) = 1 - P(A) \\ P(A) = 1 - P(\text{not } A) = 1 - P(\neg A) \end{array}}$$

→ there is an infinite of true values in probabilities.

→ How to calculate a probability?



What is the $\underbrace{\text{chance}}_{P}$ to $\underbrace{\text{throw 5 on top}}_{A}$?

- $P(A) = P(⚄)$

$= \dfrac{\#\text{ valid possibilities}}{\#\text{ total possibilities}} = \dfrac{1}{6}$

⚀ ⚁ ⚂ ⚃ ⚄ ⚅ — 1 GOOD possibility

6 possibilities.

- $P(A \overset{or}{\vee} B) = P(⚄ \text{ or } ⚅) = P(A) + P(B)$ ⚠

$= \dfrac{1}{6} + \dfrac{1}{6} = \dfrac{2}{6}$

because A∨B is like an addition (+).

you can use an addition only if A & B are mutually exclusive!
⇒ 1 DICE.

- $P(A \overset{and}{\wedge} B) = P(A) \times P(B)$ ⇒ 2 DIE mutually independent

↳ I have 2 Die. 1 yellow and 1 orange.
What is the chance to throw 5 on the yellow one, and six on the orange one?

STAT 101- Quantitative Methods. 23.10.13

4. To find the number of possible outcomes: possibilities

Die:

6 possibilities. 6 possibilities.

36 possibilities.

> the number of possibilities of outcomes $= b^n$
>
> b = base: number of possible outcomes for 1 variable ("thing")
>
> n = number of variables ("things").

Coins:

2 possible outcomes 2 possible outcomes

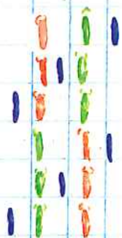
if I have 1 coins: 2 possible outcomes.
2 coins: $2^2 = 4$ possible outcomes.
3 coins: $2^3 = 8$ possible outcomes.
...

5. To find the number of possible orders: permutations:

I have 3 pencils: 

In how many different orders can we put them?
order n things

If I have 2 pencils:  2 possible orders.

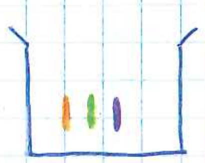
If I have 3 pencils:  6 possibilities (2×3).

If I have 4 pencils: $2 \times 3 \times 4 = 24$ possibilities.

If I have 5 pencils: $2 \times 3 \times 4 \times 5 = 120$ possibilities.

> $n! = 1 \times 2 \times 3 \times \dots = (n-1) \times n$
>
> Ex.: $5! = 1 \times 2 \times 3 \times 4 \times 5 = (5-1) \times 5 = 120$

6. To find in how many ways we can take n things out of a "Box" with m things in it: combinations:



In the Box, befor we take something out, the things can be ordered differently (in m! different ways / factonal ways).

What I take out of the box can also be ordered in n! factorial ways.

What remains in the box can also be ordered in (m−n)! factorial ways.

The total number of combinations is:

$$\boxed{C_m^n = \frac{m!}{n!\,(m-n)!}}$$

If I have a box in which there are 7 things inside, and I want to take 3 things out:

$$C_7^3 = \frac{7!}{3!\,(7-3)!} = \frac{\cancel{1\times2\times3\times4}\times5\times6\times7}{1\times2\times3\times\cancel{1\times2\times3\times4}} = 30.$$

$$C_7^3 = C_7^4$$

Ex.: In a group of 25 people, how many different groups of 5 people can I make?

$$C_{25}^5 = \frac{25!}{5!\,(25-5)!} = 53\,130.$$

$$C_{25}^{20} = \frac{25!}{20!\,(25-20)!} = 53\,130.$$

XV. Probabilities 1. Basic formulas.

$$0 \leq P(A) \leq 1$$

$$P(\text{not } A) = 1 - P(A)$$

$$P(A \text{ or } B) = P(A) + P(B)$$

$$P(A \text{ and } B) = P(A) \times P(B)$$

$$P(A) = \frac{\#\text{ of valid outcomes}}{\#\text{ of possible outcomes}}.$$

- Possibilities:

If I have 6 possible outcomes for 1 six-sided dice, and I have now 3 die, then the total number of possible outcomes is:

$\boxed{6^n} \rightarrow 6^3 = 216.$

- Permutations:

In how many different ways can I order n things?

$\boxed{n! = n \text{ factorial} = (n-1) \times n.}$

- Combinations:

In how many different ways can I take n things out of a box with m things in it?

$$\boxed{C_m^n = \frac{m!}{n!\,(m-n)!}}.$$

2. Application of probabilities in games:

1. I take a coin. If it's head, I win 1 €.
   If it's tails, I lose 1 €. 1

   We can replace this situation by a box in which we have to take 1 ticket between 2: "ticket + 1 €" or "ticket − 1 €".

   $$\left\lfloor \boxed{+1€}\ \boxed{-1€} \right\rfloor$$

2. If I get 6 on top of a dice, then I win 6 €. Anything else, I lose 1 €.

   We can replace this situation by a box in which we have to take 1 ticket between 6.

   $$\left\lfloor \boxed{1}\ \boxed{2}\ \boxed{3}\ \boxed{4}\ \boxed{5}\ \boxed{6} \right\rfloor = \left\lfloor \boxed{-1}\ \boxed{-1}\ \boxed{-1}\ \boxed{-1}\ \boxed{-1}\ \boxed{+6} \right\rfloor$$

3. Roulette game:

   $$\left\lfloor \boxed{1}\ \boxed{2}\ \boxed{3}\ \dots\ \boxed{35}\ \boxed{36}\ \boxed{0}\ \boxed{00} \right\rfloor$$

   Black
   Red
   Green

   $$= \left\lfloor \boxed{+1}\ \boxed{-1}\ \boxed{+1}\ \dots\ \boxed{-1}\ \boxed{+1}\ \boxed{-1}\ \boxed{-1} \right\rfloor = \left\lfloor 18\ \boxed{+1}\ \ 20\ \boxed{-1} \right\rfloor$$

   All casino games offer more chances to lose than to win.

⇒ The "box model" is used to compare different situation.

STAT 101 - Quantitative methods. 04.11.13

- **Coin game:** ① $= \boxed{+1} \boxed{-1}$
  - Average of the box: $\frac{+1-1}{2} = 0$

    If I play 10 times: $10 \cdot \frac{+1-1}{2} = 0$
  - Probability to lose 10 times in a row: $2^{10} = 1024$

    $P(A) = \frac{1}{1024}$

- **Roulette game:**
  - Avg box: $\frac{18 \times 1 + 20(-1)}{38} = \frac{-2}{38} = -0{,}0526$.
  - If I play 100 times: $100 \times (-0{,}0526) = -5{,}26$

    I expect to lose 5,26 €.

SD:
1). find avg.
2). Distances from avg.
3). Square the distances.
4). Avg of products.
5). $\sqrt{\ }$

↳ **Coin game:**

avg = 0
SD = 1

SD: $1 - 0 = 1 \xrightarrow{(\ )^2} 1$
$-1 - 0 = -1 \rightarrow 1$

$SD = \sqrt{\frac{1+1}{2}} = \sqrt{1} = 1$.

↳ **Roulette game:**

avg = $-0{,}0526$
SD = $0{,}999$

SD: $18 \times 1 - (-0{,}0526) = 1{,}0526 \xrightarrow{(\ )^2} 1{,}108$
$20 \times -1 - (-0{,}0526) = -0{,}9474 \rightarrow 0{,}8975$

$SD = \sqrt{\frac{20(0{,}8975) + 18(1{,}108)}{38}}$

$= \sqrt{0{,}997} = 0{,}999$

- Cards Game:

If I get an Ace, I get 10€.
Any other card, I pay 1€.

48 $\boxed{-1}$ 4 $\boxed{+10}$

Avg Box: $\dfrac{48(-1)+4(10)}{52} = \dfrac{-48+40}{52} = -0,1538$

On avg, I will lose 0,1538€ per time.
If I play 10 times, I will lose 15,38€ on avg.

SD:
48 ×: $-1-(-0,1538) = -0,8462 \xrightarrow{(\ )^2} 0,716$
4 ×: $10-(-0,1538) = 10,1538 \longrightarrow 103,1$

$$SD = \sqrt{\frac{48(0,716)+4(103,1)}{52}} = \sqrt{8,59} = 2,93.$$

To find SD faster, when you only have 2 kinds of tickets:

CARDS
$$|\text{Big value} - \text{small value}| \times \sqrt{\text{Fraction Big val} \times \text{Fraction small val}}$$

$$\rightarrow |10-(-1)| \times \sqrt{\frac{4}{52} \times \frac{48}{52}} = 11.\sqrt{0,071} = 2,93.$$

ROULETTE
$$|1-(-1)| \times \sqrt{\frac{18}{48} \cdot \frac{20}{48}} = 2\sqrt{0,2493} = 0,999$$ same.

3. Exercises.

- We have 2 die. What are the chances that we will have 2 "six" on top?

$$\frac{1}{6} \times \frac{1}{6} = \frac{1}{36} = 0,028 = 2,8\%$$

"x" because – independance from each other
– ⚅ AND ⚅

- What are the chances to have no "six" on top?

$$\frac{5}{6} \times \frac{5}{6} = \frac{25}{36} = 0,694 = 69,4\%$$

- What are the chances to have only 1 "six" on top? That is to say one is "six" OR the other is "six"?

We cannot use "+" for "or" because they are not mutually exclusive: they are independant.

So we could say: "what are the chances to have not "six" OR not "six"?"

the opposite of what we want.
→ not (⚅ OR ⚅)?
→ not(⚅) AND not(⚅)?

$$\underbrace{P(⚅ \text{ or } ⚅)}_{\text{what we want}} = 1 - \underbrace{P(\text{not } (⚅ \text{ or } ⚅))}_{\text{the opposite.}}$$

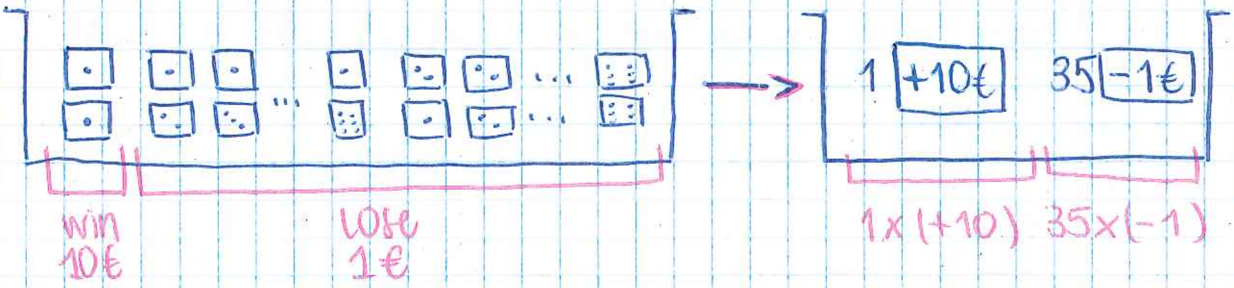
to inverse the result.

$$= 1 - \frac{25}{36} = \frac{36}{36} - \frac{25}{36} = \frac{11}{36}$$

$$= 0,30556 = 30,556\%$$

$$\boxed{\begin{aligned} P(A) &= 1 - P(\text{not } A) \\ P(\text{not } A) &= 1 - P(A) \end{aligned}}$$

- We still have 2 Die. We play the "snake eyes" game. We win if we have 2 times "one" on top. In other cases, we lose. Let's say that we win 10 € and we lose 1€.



$$\text{AVG BOX} = \frac{(1\times 10)+(35\times(-1))}{36} = \frac{10-35}{36} = \frac{-25}{36} = -0{,}694$$

$$\text{SD BOX} = |10-(-1)| \times \sqrt{\frac{1}{36}\times\frac{35}{36}} = 11\sqrt{0{,}027}$$
$$= 11 \cdot 0{,}164$$
$$= 1{,}8$$

We can use the quick formula because we only have 2 kinds of "tickets".

- If we play 10 times, what do we expect to lose or win, on average?

$$EV = \text{expected value} = \text{nbre of games} \times \text{AVG BOX}$$
$$= 10 \times (-0{,}694)$$
$$= -6{,}94.$$

How to calculate the standard error (SE)?

$$\boxed{SE = \sqrt{\text{nbr of games}} \times \text{SD BOX}}$$

If we play 10 times: $SE = \sqrt{10} \times 1{,}8 = 5{,}72$

↳ we expect to lose $-6{,}94$ € if we play 10 times, $\pm$ 5,72.

- We play the roulette game. I play "Black" for 1€.

$$\left[\,18\ \boxed{\text{Black}}\quad 20\ \boxed{\text{not Black}}\,\right] = \left[\,18\ \boxed{+1€}\quad 20\ \boxed{-1€}\,\right]$$

$$\text{AVG BOX} = \frac{18(1) + 20(-1)}{38} = \frac{-2}{38} = -0{,}0526$$

$$\text{SD BOX} = |1-(-1)|\sqrt{\frac{18}{38} \times \frac{20}{38}} = 2\sqrt{0{,}2493}$$
$$= 0{,}999$$

The casino plays 100 Roulette games per evening for one year. What will they lose or win?
We only play red or Black.

The AVG Box $-0{,}0526$ is for the player. We have to inverse it for the casino: $0{,}0526$.

$$EV = 100 \cdot 365 \cdot 0{,}0526 = 1919{,}9\ €.$$

$$SE = \sqrt{100 \cdot 365} \cdot 0{,}999 = 190{,}9\ €.$$

→ They will win $1919{,}9€ \pm 190{,}9€$.

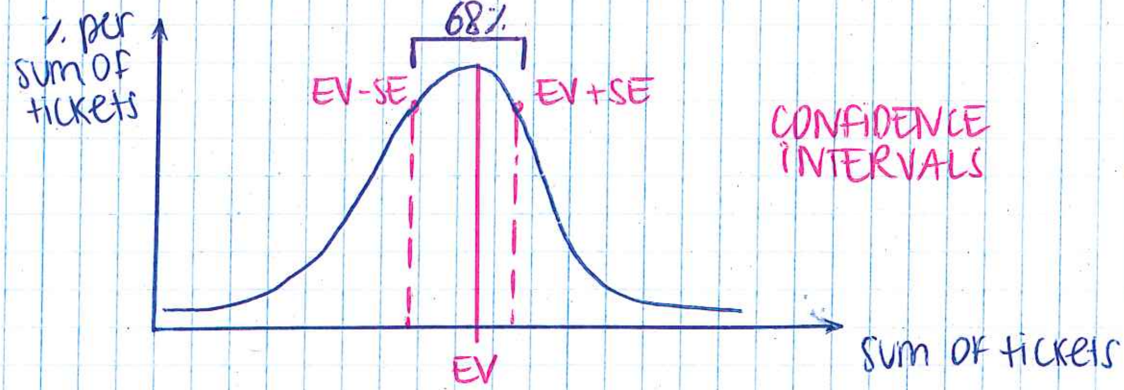
The casino will win between 1729€ and 2111€ in 1 year on a player who plays "Black" or "red" 100 times per evening.

STAT 101- Quantitative Methods. 13.11.13

4. Central limit theorem (CLT).

If we take many tickets out of a box, and we repeat this many times, the sum of the tickets will be a normal curve, with the top at the EV and the "shoulder points" at the EV − SE and SE + SE.

We have to take the tickets 50 times or more, and take 50 or mor tickets each time.



CONFIDENCE INTERVALS

Any random process repeated many times will give a normal curve around the expected value (EV).

EX.

- Box model: [5 $\boxed{1}$ 5 $\boxed{3}$]
- Histogram of box:

value	count	width	area(%)	height
1	5	1	50	50
3	5	1	50	50
	10			

- Avg box: $\frac{(5\times1)+(5\times3)}{10} = \frac{5+15}{10} = 2$
- SD box: $|3-1| \cdot \sqrt{\frac{5}{10}\times\frac{5}{10}} = 2\sqrt{0{,}25} = 1$
- We take 100 tickets: $EV = 100 \cdot 2 = 200$
  $SE = \sqrt{100} \cdot 1 = 10$

→ If I take 100 times 100 tickets, I expect that in 68% of the experiment, the sum of the tickets will be between EV − 1SE and EV + 1SE

Ex..

- roulette game: we play "Black" or "red"

$$\left[18 \boxed{+1} \; 20 \boxed{-1} \right]$$

$$\text{AVG BOX} = \frac{18 \cdot 1 + 20 \cdot -1}{38} = -0{,}0526$$

$$\text{SD BOX} = |1-(-1)| \sqrt{\frac{18}{38} \times \frac{20}{38}}$$

$$= 2\sqrt{0{,}249} = 0{,}999$$

- we play 100 times.

$EV = 100 \cdot -0{,}0526 = -5{,}26$ we expect to lose 5,26 €, if we play 100 times.

$SE = \sqrt{100} \cdot 0{,}999 = 9{,}99$

→ I am 68 % confident that I'll win between:

- ★ $-5{,}26 - (9{,}99) = -15{,}25$ € = EV − SE
- ★ $-5{,}26 - (9{,}99) = 4{,}73$ € = EV − SE

→ I am 95% confident that I'll win between:

- ★ $-5{,}26 - 2(9{,}99) = -25{,}24$ € = EV − 2SE
- ★ $-5{,}26 + 2(9{,}99) = 14{,}72$ € = EV − 2SE

→ I am 99,7% confident that I'll win between:

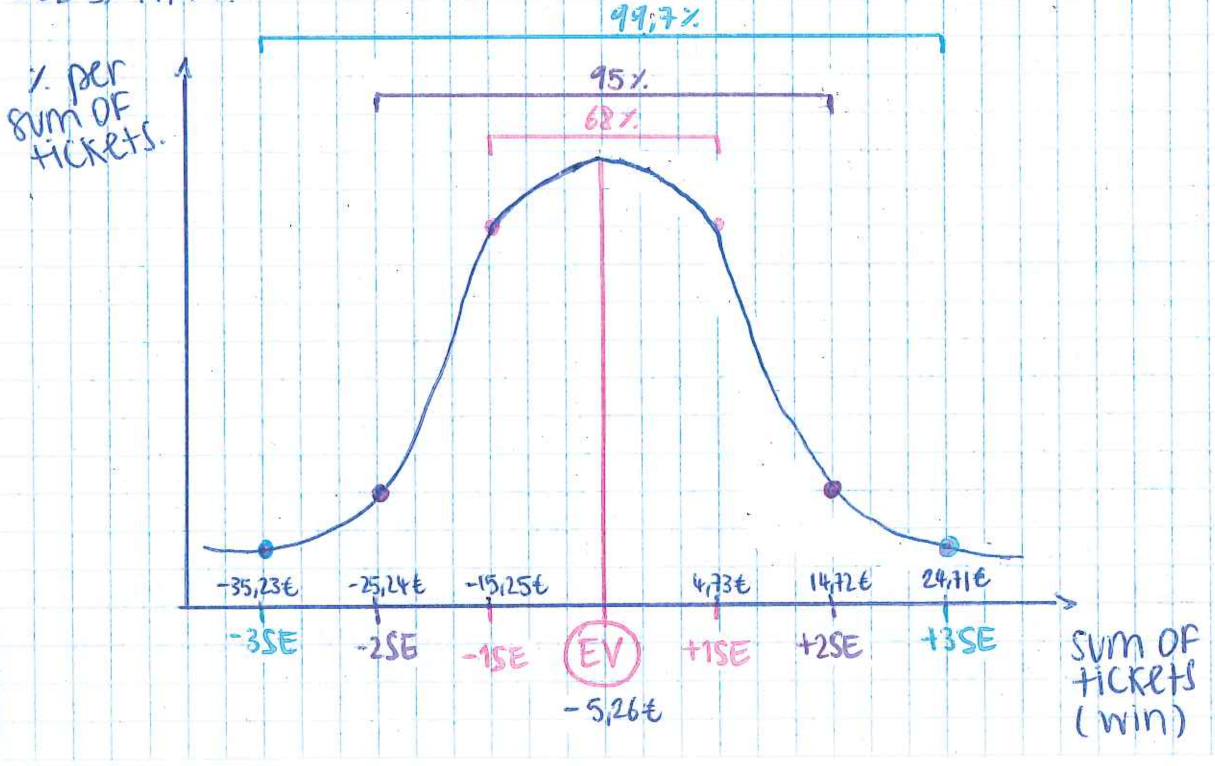
- ★ $-5{,}26 - 3(9{,}99) = -35{,}23$ € = EV − 3SE
- ★ $-5{,}26 + 3(9{,}99) = 24{,}71$ € = EV + 3SE

Percentage: see normal table

1SD ⇒ 68%

2SD ⇒ 95%

3SD ⇒ 99,7%



- To calculate a % that is not on the normal table:

5 SD => ?

↳ normal cdf between -5 and 5 = 0,999999

0,999999 × 100 = 99,9999 %

→ 5 SD => 99,9999 %

⚠ "SD" can be replaced by "σ" (sigma).

5. The law of big numbers / law of averages.

Ex.: Roulette game, playing "Black" or "Red"

| 18 $\boxed{+1}$ 20 $\boxed{-1}$ |

AVG BOX: -0,0526
SD BOX: 0,999

- we play 10 times: EV = 10 . -0,0526 = -0,526
  SE = $\sqrt{10}$. 0,999 = 3,16 |SE|>|EV|
- we play 100 times: EV = 100 . -0,0526 = -5,26
  SE = $\sqrt{100}$. 0,999 = 9,99 |SE|>|EV|
- we play 1000 times: EV = -52,6
  SE = 31,59 |SE|<|EV|
- we play 1000000 times: EV = -52600
  SE = 999 |SE|<|EV|

↳ the casino is 68% confident to make 52600 € ± 999 per 1 million games.

↳ the casino is 99,99 % confident (4 σ) to make 52600 € ± 4(999), that is to say between:
* 48 604 €
* 56 596 €

⇒ the SE goes up slower than the EV, that is why the casino is certain to win when there is a huge number of games. For big numbers, the SE becomes smaller relative to the EV.

EX.: we play "head" or "tails" with a coin.

$\boxed{-1}$ $\boxed{+1}$

AVG BOX : 0.
SD BOX : 1.

we play 10 times : EV = 0
SE = 3,16.

100 times : EV = 0
SE = 10

1000 times : EV = 0
SE = 31,6

1000 000 times : EV = 0
SE = 1000

SE relative to the number of times we played:

$\frac{SE}{\# games}$ → 10 times : $\frac{3,16}{10} = 0,316$

100 times : $\frac{10}{100} = 0,1$

1000 times : $\frac{31,6}{1000} = 0,0316$

1000 000 times : $\frac{1000}{1000\,000} = 0,001$

6. Binomials.

→ If I pick a card from a deck (52 cards), what is the chance that it is an ace or a diamond?

Ace: $\frac{4}{52}$ chances. **Diamond**: $\frac{13}{52}$ chances.

- **We know that**:

$$P(A \text{ or } B) = P(A) + P(B) \text{ if mutually exclusive}$$
$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$
if not mutually exclusive.

So, if A = ace and B = diamond:

$$P(\boxed{1} \text{ or } \Diamond) = \frac{4}{52} + \frac{13}{52} - \left(\frac{4}{52} \times \frac{13}{52}\right)$$
$$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52}$$

- **And we know that**:

$$P(A \text{ or } B) = 1 - P(\text{not } (A \text{ or } B))$$
$$= 1 - (P(\text{not } A) \text{ and } P(\text{not } B))$$

So, if A = ace and B = diamond:

$$P(\boxed{1} \text{ or } \Diamond) = 1 - \left(\frac{48}{52} \times \frac{39}{52}\right)$$
$$= 1 - \frac{1872}{2704} = \frac{2704}{2704} - \frac{1872}{2704} = \frac{832}{2704}$$
$$= \frac{16}{52}$$

→ IF it rains, what are the chances that I become wet?

conditional chances:

$$\text{Chances that A happens if B happens} = P(A/B) = \frac{P(A \text{ and } B)}{P(B)}$$

TO Find $P(A \text{ and } B)$:

we know that $P(B) = 0{,}8$ and $P(A|B) = 0{,}5$.

$$\hookrightarrow \quad P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

$$0{,}5 = \frac{P(A \text{ and } B)}{0{,}8}$$

$$P(A \text{ and } B) = 0{,}8 \times 0{,}5 = 0{,}4.$$

→ IF I try something "n" times, the chances that something happens if I try 1 time is equal to "p".

Then what are the chances that it happens exactly "K" times?

$n!$
$= n$ Factorial
$= (n-1) \times n$

$$\frac{n!}{K!(n-K)!} \times p^K \times (1-p)^{n-K} = C_n^K \cdot p^K \cdot (1-p)^{n-K}$$

n = number of times I play.

K = number of times we want something to happen.

p = number of times it happens if we play once.

ON THE CALCULATOR: TO DO $n!$

menu → Run → OPTN → ▷ → PROB → x!

CASIO GRAPH 35+.

Ex.: I toss a coin 10 times. What are the chances to have exactly 5 times "head"?

$n = 10$
$k = 5$
$p = 1/2$

$\hookrightarrow \dfrac{10!}{5!(10-5)!} \cdot \left(\dfrac{1}{2}\right)^5 \cdot \left(1-\dfrac{1}{2}\right)^{10-5}$

$= \dfrac{10!}{5!\,5!} \cdot \left(\dfrac{1}{2}\right)^5 \cdot \left(\dfrac{1}{2}\right)^5 = 0{,}24609 = 24{,}609\,\%$

Ex.: I toss a coin 10 times, what are the chances to have 10 times "head"?

$n = 10 \quad k = 10 \quad p = 1/2$

$\hookrightarrow \dfrac{\cancel{10!}}{\cancel{10!}\underbrace{(10-10)!}_{=1}} \cdot \left(\dfrac{1}{2}\right)^{10} \cdot \left(1-\dfrac{1}{2}\right)^{10-10}$

$x^0 = 1$

$= 1 \cdot \left(\dfrac{1}{2}\right)^{10} \cdot 1 = 0{,}000976 = 0{,}0976\,\%$

OR Binomial pdf $(\overset{n}{10}; \overset{p}{\tfrac{1}{2}}; \overset{k}{10}) = 0{,}000976.$

ON THE CALCULATOR : TO DO BINOMIAL PDF

menu → STAT → DIST (F5) → BINM (F5) → Bpd (F1)

→ Data : variable.
x : k
numtrial : n.
p : p.

CASIO GRAPH 35+

9,7656E-04 = 0,00097656

Ex.: I throw a 20 sided Dice $\overset{n}{15}$ times. what are the chances to have at least $\underset{k}{13}$ times, "15" or more, on the Dice?

over 15 total times.

$$\text{Binomial pDF}\left(\overset{n}{15};\overset{p}{\frac{6}{20}};\overset{k}{13}\right)$$
$$+\ \text{Binomial pDF}\left(15;\frac{6}{20};14\right)$$
$$+\ \text{Binomial pDF}\left(15;\frac{6}{20};15\right)$$

$= 0,0008719\ \%$

Ex.: I throw a 6 sided Dice 5 times. what are the chances to have to have "6" 5 times?

$$\text{Binomial pDF}\left(5;\frac{1}{6};5\right) = 0,0001286.$$
$$= \left(\frac{1}{6}\right)^5$$

Ex.: I throw a 6 sided Dice 5 times. what are the chances to have at least 1 time "6"?

over 5 total times.

$$\text{Binomial pDF}\left(5;\frac{1}{6};1\right)$$
$$+\ \text{Binomial pDF}\left(5;\frac{1}{6};2\right)$$
$$+\ \text{''}\ \left(5;\frac{1}{6};3\right)$$
$$+\ \text{''}\ \left(5;\frac{1}{6};4\right)$$
$$+\ \text{''}\ \left(5;\frac{1}{6};5\right)$$

$= 59,81\ \%$

Binomial

$P(A \text{ or } B) = P(A) + P(B)$ → if mutually exclusive

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

If I pick a card from a deck (52 cards) what is the chance that it is an ace or a diamond?

Ace = $\frac{4}{52}$ diamonds = $\frac{13}{52}$

$P(A \text{ or } B) = P(A) + P(B) - (P(A \text{ and } B))$

$= \frac{4}{52} + \frac{13}{52} - \frac{1}{52}$ → ace + diamond

$= \frac{16}{52}$

Chance that card is ace or diamond ⇒ Law of Morgan

$P(A \text{ or } \diamond) = 1 - P[\text{NOT } (\text{Ace or } \diamond)]$

$= 1 - [P(\text{not } A) \text{ and } P(\text{not } \diamond)]$

$= 1 - \left[\frac{48}{52} \times \frac{39}{52}\right]$

$= 1 - \frac{1872}{2704} = \frac{2704}{2704} - \frac{1872}{2704}$

$= \frac{832}{2704} = \frac{16}{52}$

Chance that A happens if B happens

$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$

conditional chances

$P(\text{rain}) = 0.8$

$P(\text{if it rains I become wet}) = 0.5$

$P(\text{wet}|\text{rain})$

$P(\text{wet}|\text{rain}) = \frac{P(\text{wet and rain})}{P(\text{rain})}$

$0.5 = \frac{P(\text{wet and rain})}{0.8}$

$P(\text{wet and rain}) = 0.5 \times 0.8 = 0.4$

Binomial Formula, Binomium of Newton

- if I try something n times
- the chance that something happens if I try 1 time is p
- then what the chance that this happens exactly k times?

$$= \frac{n!}{k!\,(n-k)!}\, p^k (1-p)^{n-k}$$

$$= C_n^k\, p^k (1-p)^{n-k}$$

(1) Ex: I toss a coin 10 times. What is the chance to have exactly 5 heads

$$= \frac{10!}{5!\,(10-5)!} \left(\frac{1}{2}\right)^5 \left(1-\frac{1}{2}\right)^{10-5}$$

$$= \frac{10!}{5!\,5!} \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^5$$

$$= 0.24609$$

$$\boxed{\text{Binompdf}\left(10, \tfrac{1}{2}, 5\right) = 0.246}$$

(2) Ex: I toss a coin 10 times, head 10 times?

$$\left(\frac{1}{2}\right)^{10} = \frac{1}{1024} = 0.000976 = 0.097656\,\%$$

$$\text{Binompdf}\left(10, \underset{p}{\tfrac{1}{2}}, 10\right) = 0.000976$$

(3) Ex = head 9 times?

$$\text{Binompdf}\left(10, \tfrac{1}{2}, 9\right) = 0.00976 \Rightarrow 0.976\,\%$$

Ex = 9 or 10 times? ⇒ at least 9 times

$$0.9765\,\% + 0.09765\,\% = \boxed{\text{chance of } 1.07\,\%}$$

↓ mutually exclusive

# of heads	% chance
1	0.9765
2	4.394
3	11.718
4	20.507
5	24.6
6	~~[illegible]~~ 20.507
7	~~20.507~~ 11.718
8	4.394
9	0.9765
10	0.09765 $\left(\frac{1}{2^{10}}\right)$

$$\text{Binompdf}\left(10, \tfrac{1}{2}, x\right)$$

x = # of heads

Ex: I throw a 20 sided dice 15x. What is the chance to have atleast 13x ~~on~~ $\underbrace{\text{15 or more}}_{\frac{6}{20}}$ on the dice?

(15, 16, 17, 18, 19, 20)

$$\text{Binompdf}\left(15, \tfrac{6}{20}, 13\right) + \text{Binompdf}\left(15, \tfrac{6}{20}, 14\right) + \text{Binompdf}\left(15, \tfrac{6}{20}, 15\right)$$

$= 0.000\,8719\ \%$

Ex: I throw 6 sided dice 5 times
What is chance to have 5 times ?

1/6

$$\left(\tfrac{1}{6}\right)^5 = \text{Binompdf}\left(5, \tfrac{1}{6}, 5\right)$$

What is chance to have atleast 1 ?

1/6

either:

1) Binompdf $(5\ \tfrac{1}{6}, 1)$ + Binompdf $(5, \tfrac{1}{6}, 2)$ + Binompdf $(5, \tfrac{1}{6}, 3)$ + Binompdf $(5, \tfrac{1}{6}, 4)$ + Binompdf $(5, \tfrac{1}{6}, 5)$ = 59.81 %

or

~~What is chance to have no~~ ?

2) = 1 − chance to have no
= 1 − Binompdf $(5\ \tfrac{1}{6}, 0)$
= 59.81 %

or

3) (inverse)
$1 - \left(\tfrac{5}{6}\right)^5 = 59.81\ \%$

Sampling

10,000,000 belgians (population)
How many are Flemish / Walloon / German?

Flemish	Walloon	German
5200000	4200000	600000

5200000 4200000 600000

[1] [2] [3]

10000,000 Belgian population

(1, 2, 3 ⇒ just random tickets)

5200 4200 600

[1] [2] [3]

10000 sample

What is the chance that in random sample 10000 german speaking

$$\left(\frac{600,000}{10000,000}\right)^{10000} = (0.6)^{10000}$$

$$= 0$$ ⇒ chance is so small, can't be calculated

flemish)

$$= \left(\frac{5,200,000}{10,000,000}\right)^{10000}$$

$$= (0.52)^{10000}$$
$$= 0$$

What if I take 100 people

52 42 6

[ ] [ ] [ ]

100

$$\left(\frac{5200000}{10000000}\right)^{100} = 3.98 \times 10^{-29}$$

7. Sampling.

- census : ask everybody the same question.
- sample : take a part of "everybody" and ask the same question.
  You should select randomly the people of your sample.

EX. : TOTAL : 10 000 000 people. $\boxed{1}$ Dutch speaking.
sample : 10 000 people. $\boxed{2}$ french speaking.

1. Take random sample & ask question (Be as neutral as posible).
2. Make a box model, and give numbers to your answers. Quantisize your answers (tickets with numbers).
3. Put the answers in percentages ($\frac{\text{count}}{\text{total c.}} \times 100\%$).
4. Extrapolate your percentages to the total population.

6000 $\boxed{1}$	4000 $\boxed{2}$

60 % $\boxed{1}$ 40 % $\boxed{2}$

→ We can assume that 60% of the 10 000 000 people speak Dutch, & 40 %. speak french.

5. Calculate AVG Box & SD Box

AVG BOX : $\frac{60 . 1 + 40 . 2}{100} = 1,4$

or $\frac{6\,000\,000 . 1 + 4\,000\,000 . 2}{10\,000\,000} = 1,4$

SD BOX : $|2-1| . \sqrt{\frac{40}{100} \times \frac{60}{100}} = 1 . \sqrt{0,24} = 0,4899$

6. Calculate the Standard error (SE).

$SE\,\% = \frac{0,4899 \times 100\,\%}{10\,000} = 0,4899\,\%$

$$SE = \sqrt{\text{sample size}} \times SD$$

$$SE\% = \frac{\sqrt{\text{sample size}} \times SD \times 100\%}{\text{sample size}}$$

$$= \frac{SD \times 100\%}{\sqrt{\text{sample size}}}$$

⚠ $\frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}}$

=> We can assume that 6 000 000 Belgians speak Dutch as 1st language, with a standard error of 48 990 (0,4899 % of 10 000 000).

=> We are 68 % confident that 6000000 Belgians speak Dutch as 1st language, ± 48 990.

=> We are 95 % confident that 6 000 000 Belgians speak Dutch as 1st language, ± 2 . 48 990.

=> We are 99 % confident that 6 000 000 Belgians speak Dutch as 1st language, ± 3 . 48 990.

If we would have taken a sample of only 100 people, the number of the box model would have been the same :

60 $\boxed{1}$ 40 $\boxed{2}$

60% $\boxed{1}$
40% $\boxed{2}$

But the SE % would have been bigger:

$$SE\% = \frac{0{,}4899 \times 100\%}{\sqrt{100}} = 4{,}899\%$$

→ We are more confident & more precise if the sample is big.

The size of the total population does not matter in the calculation of the SE %.

⚠ If there are more than 2 types of "tickets", the long complete formula should be used to calculate the SD !

EX: (with more than 2 kinds of QUALITATIVE tickets)

- total population: 6 000 000.
- sample size: 1000
- types of answers: $\boxed{0}$ people who didn't vote for this party.
  $\boxed{1}$ people who vote for this party.

a. Open VLD - 15 %
b. SPa - 15 %
c. CD&V - 20 %
d. NVA - 30 %
e. Groen - 10 %
f. Other - 10 %

d. NVA:

70 $\boxed{0}$ 30 $\boxed{1}$

- $\text{AVG BOX} = \dfrac{70.0 + 30.1}{100} = 0{,}3$
- $\text{SD BOX} = |1-0|\sqrt{\dfrac{30}{100} \times \dfrac{70}{100}} = 0{,}458$
- $\text{SE\%} = \dfrac{45{,}8\%}{\sqrt{1000}} = 1{,}45\%$

→ 1,45 % of 6 000 000:
$= \dfrac{1{,}45 \cancel{\%} \times 6000000}{100 \cancel{\%}}$
$= 86\,948$

e. Groen:

90 $\boxed{0}$ 10 $\boxed{1}$

- $\text{AVG BOX} = \dfrac{90.0 + 10.1}{100} = 0{,}1$
- $\text{SD BOX} = |1-0|\sqrt{\dfrac{90}{100} \times \dfrac{10}{100}} = 0{,}09$
- $\text{SE\%} = \dfrac{9\%}{\sqrt{1000}} = 0{,}285\%$

→ 0,285 % of 6 000 000:
$= \dfrac{0{,}285 \cancel{\%} \times 6000000}{100 \cancel{\%}}$
$= 17\,076$

⇒ We are 68 % confident that:

- NVA will obtain 1800 000 ± 86 948 votes.
- Groen will obtain 600 000 ± 17 076 votes.
- ...