

Distribuția normală standard

Statistica z

Intervalul de încredere

Tendința generală + dispersia → media valorilor + deviația standard

Dacă dorim să știm cum arată evoluția în timp a unei anumite valori, trebuie determinată *probabilitatea de apariție a acelei valori printre toate valorile* ce caracterizează populația de interes.

Dacă setul de date al esanționului este *normal distribuit* (sau aproape normal distribuit), datele vor determina *probabilități foarte particulare*, care pot fi exploatare în combinație cu *media aritmetică și deviația standard* pentru a face *predicții* despre probabilitatea de apariția a unei valori particulare.

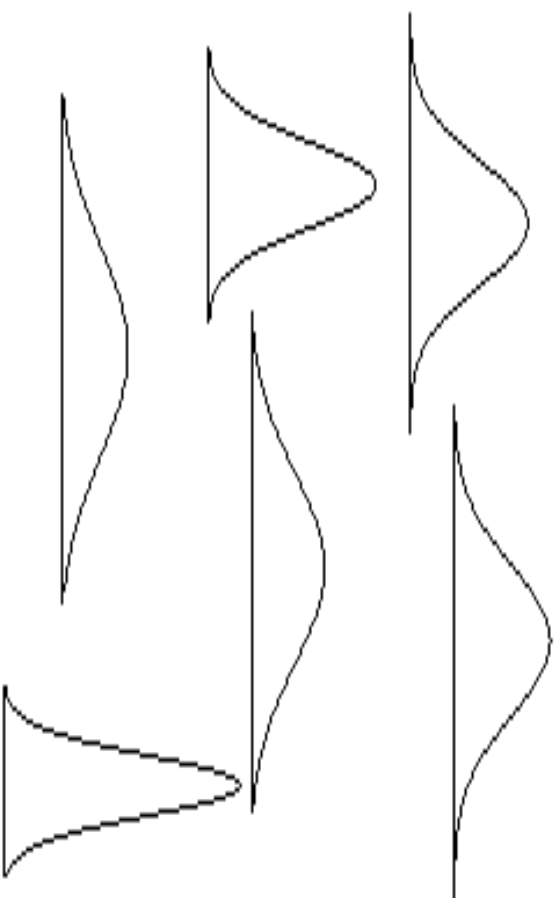
Statistica inferentiala permite enunțarea unor:

- afirmații cantitative probabilistice
- afirmații cantitative predictive

Distribuția normală

Un set de date are o **distribuție normală** dacă media aritmetică, mediana și modul vor avea *valori identice* (sau foarte apropiate).

Graficul unei distribuții normale este *simetric* (sau aproape simetric) față de valoarea mediană.



Înălțimea și lățimea unei distribuții normale pot fi exprimate matematic în funcție de doi parametri: *media aritmetică* și *deviația standard*.

Carl Friedrich Gauss (1777-1855)



O distributie de date este normala daca:

- graficul distributiei are forma de clopot, atingand inaltimea maxima pentru media aritmetica (μ)
- este simetrica fata de media aritmetica (μ)
- este unimodala
- este o distributie continua
- se apropie de axa orizontala, dar nu o atinge niciodata

Caracteristica cea mai folositoare:

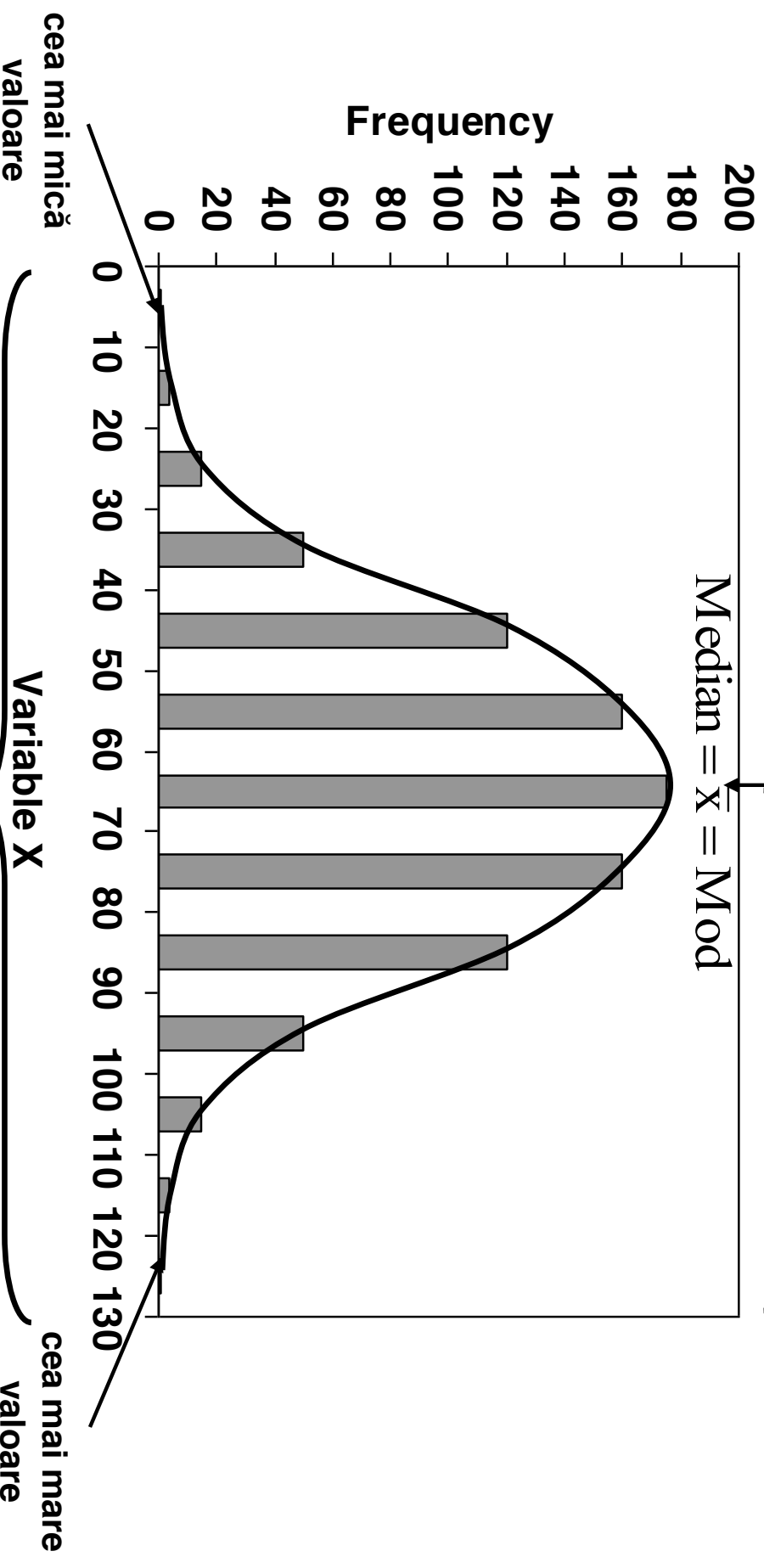
aria de sub grafic – se converteste in probabilitate

**Distribuția normală este simetrică
de ambele părți ale mediei aritmetice**

**50% din date sunt mai mici
decât media aritmetică**

**50% din date sunt mai mari
decât media aritmetică**

Median = \bar{x} = Mod

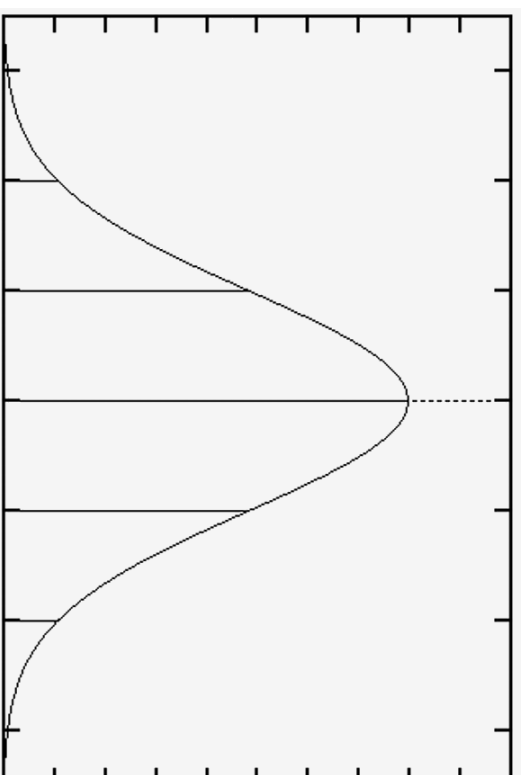


**100% probabilitatea ca toate valorile din
setul de date să se găsească între extreme**

probabilitatea de 100% = probabilitate egală cu 1,0

Afirmatii *logice*:

- probabilitatea ca toate măsurătorile din setul de date să se afle între cele două extreme este 100%
- probabilitatea ca orice dată din setul de date considerat să fie mai mică decât valoarea medie, este 50%
- probabilitatea ca orice dată din setul de date considerat să fie mai mare decât valoarea medie, este 50%
- cele mai multe valori din setul de date se află în apropierea mediei
- este mai probabil ca o valoare individuală să fie mai aproape de valoarea medie decât de valorile extreme
- în apropierea extremităților se afla mai puține valori decât în apropierea mediei



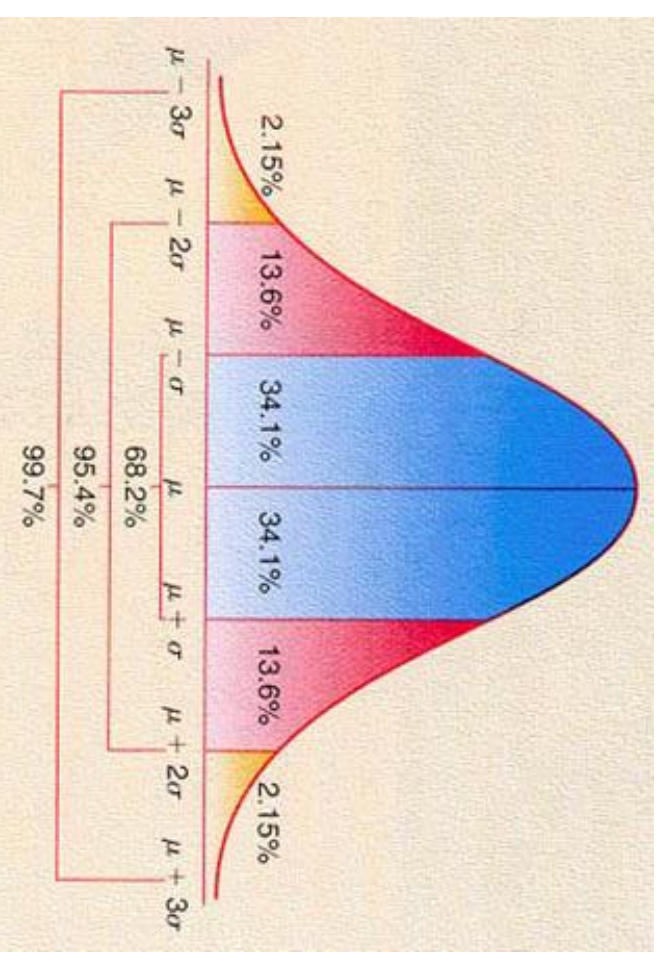
Afirmatii deduse matematic:

- există o probabilitate de ~68% ca orice valoare dintr-un set de date normal distribuit să se afle în intervalul definit de punctele care se află la o deviație standard deasupra și sub medie:

$$\bar{X} - S, \bar{X} + S$$

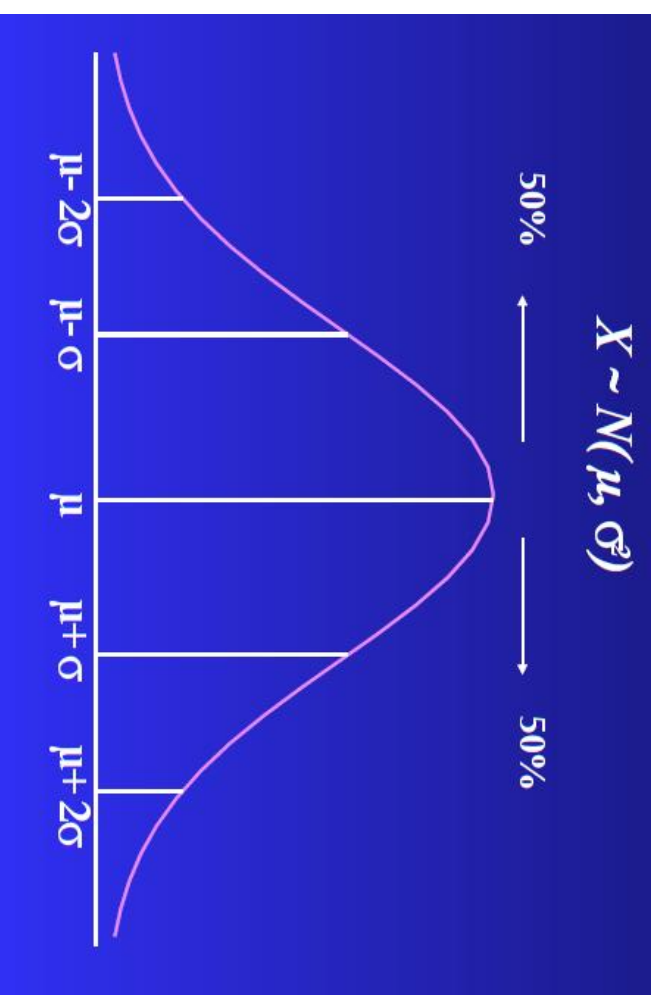
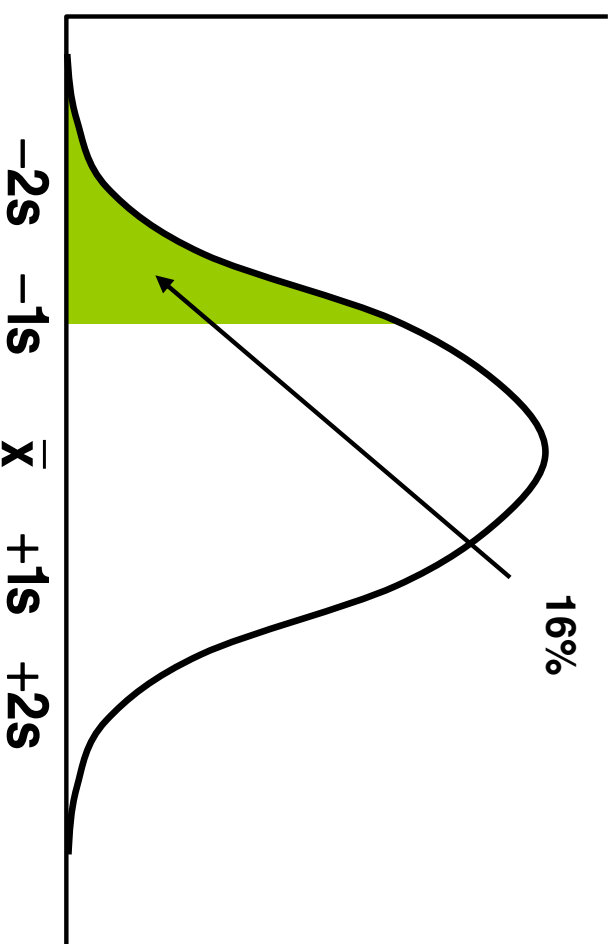
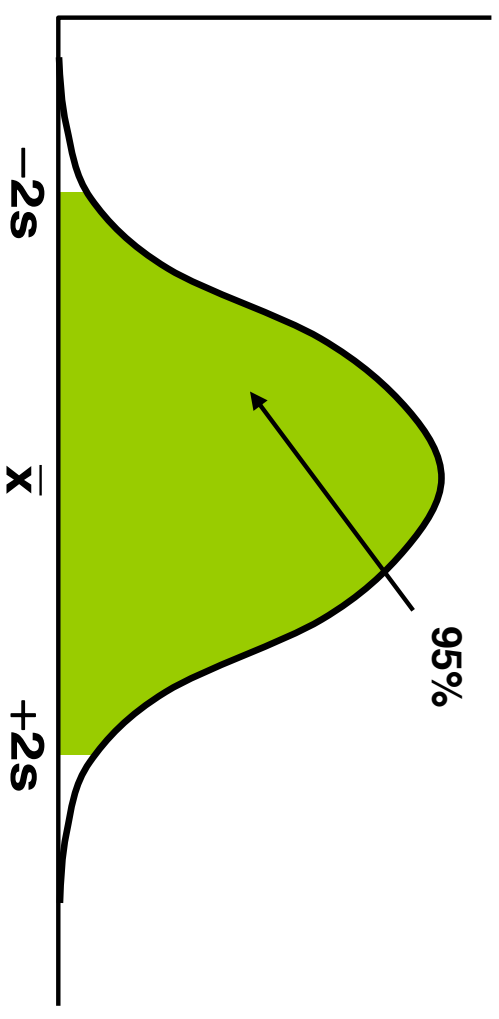
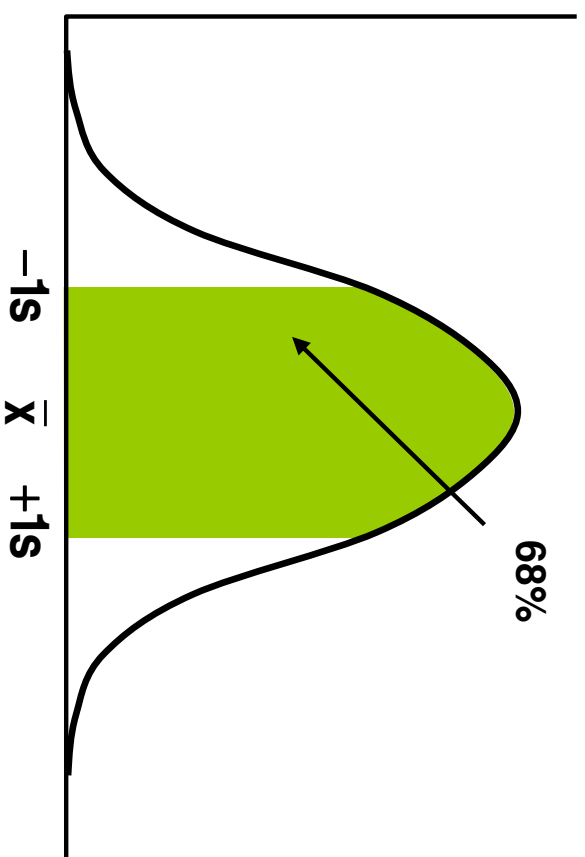
- există o probabilitate de ~95% ca orice valoare dintr-o distribuție normală de date să se afle la cel mult două deviații standard de medie:

$$\bar{X} - 2S, \bar{X} + 2S$$



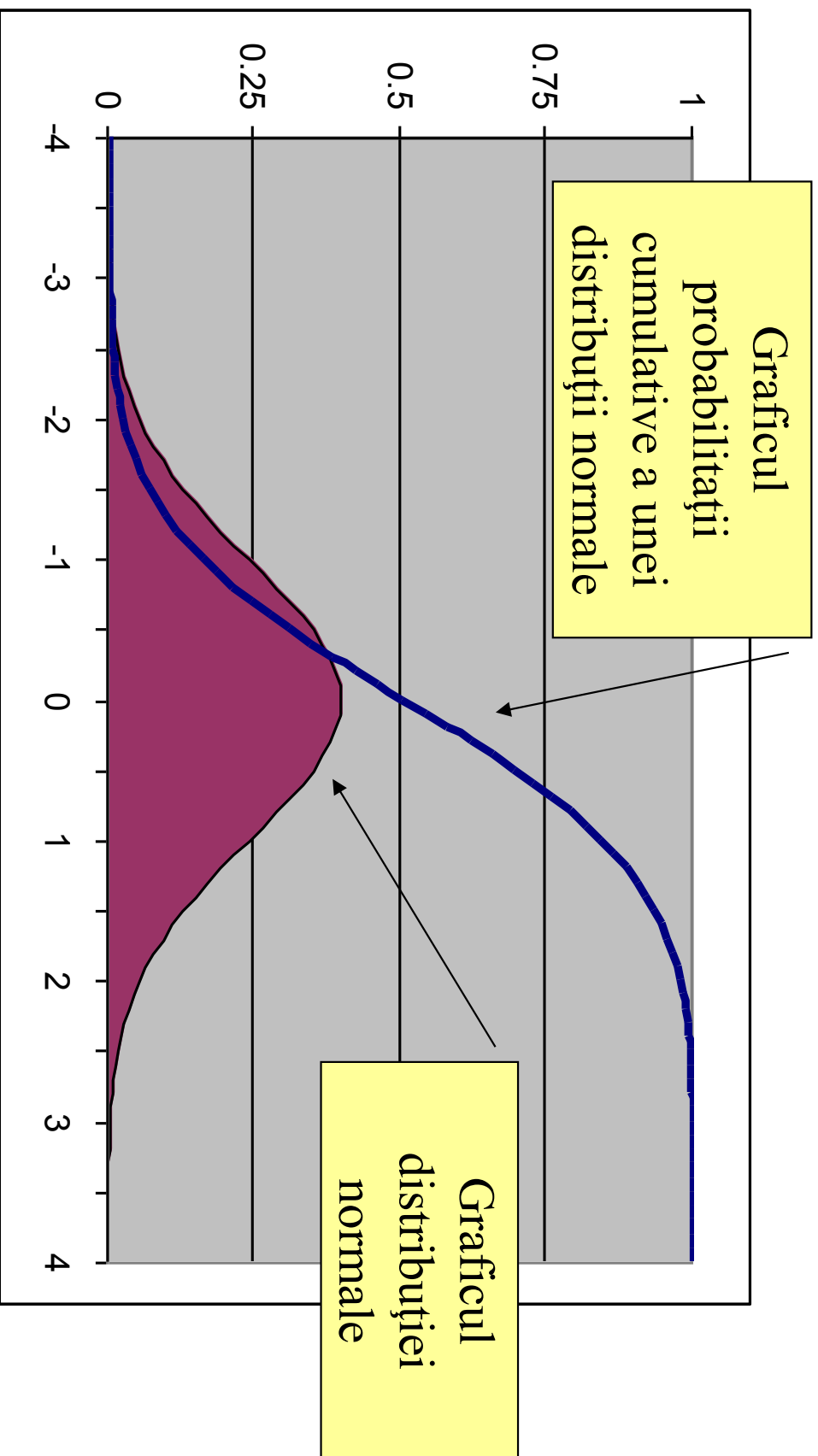
- există o probabilitate de ~16% ca orice valoare dintr-o distribuție normală de date să fie mai mică decât: $\bar{X} - S$

probabilitate calculată folosind prima afirmatie: $(100-68)/2 = 16$



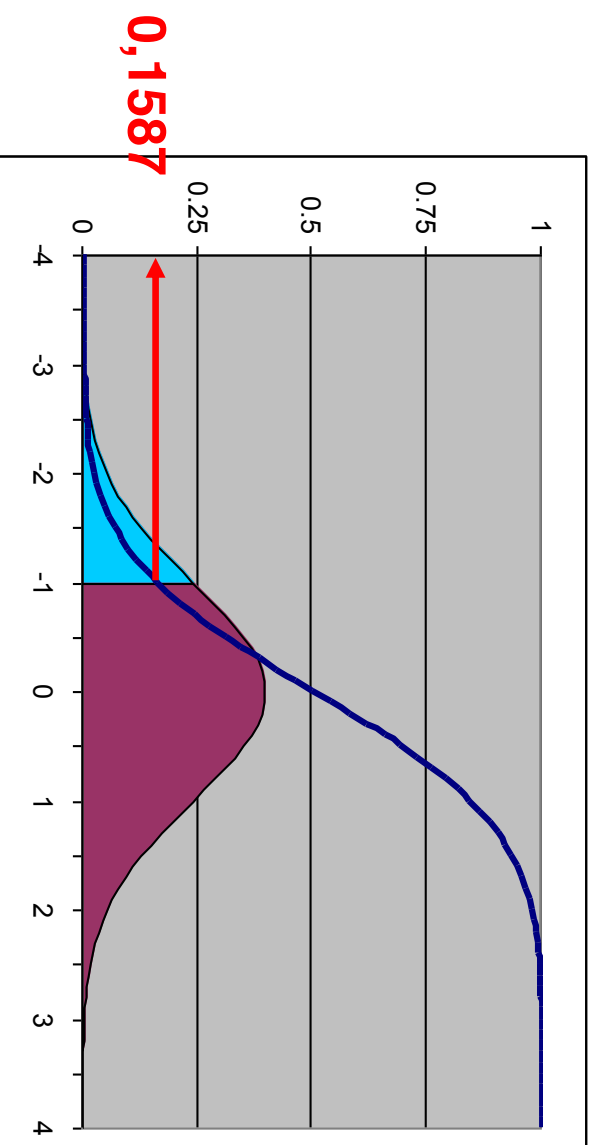
Probabilitatea cumulativă

Probabilitatea cumulativă: probabilitatea ca o variabilă aleatoare dintr-un set de date să aibă valoarea într-un anumit domeniu.

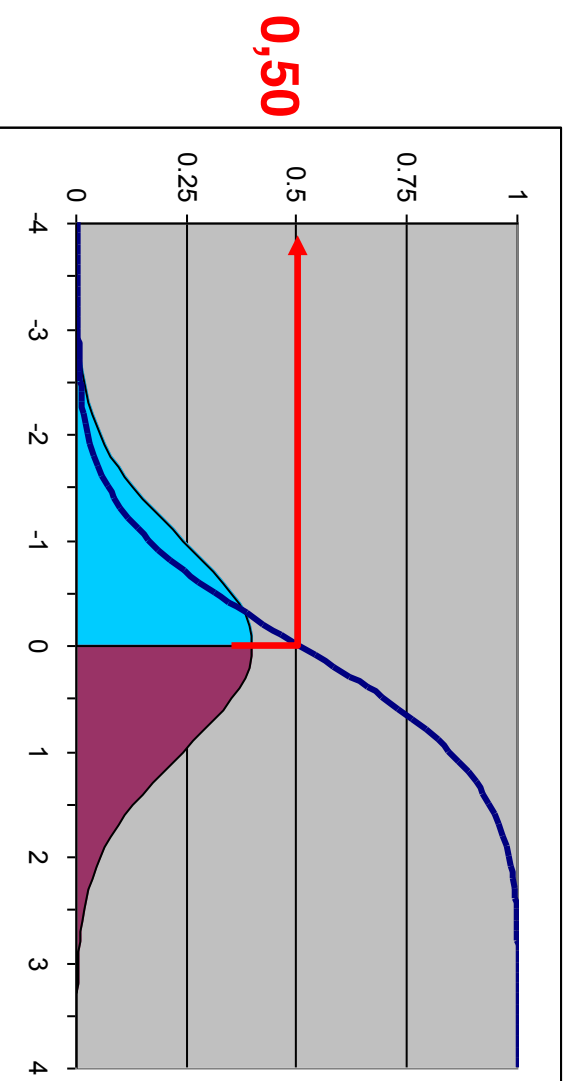


Probabilitatea cumulativă este probabilitatea ca o variabilă aleatoare să fie mai mică sau egală cu o valoare dată.

- probabilitatea ca o valoare aleatoare sa se afle sub o deviatie standard fata de media aritmetica este 15,87 %



- probabilitatea ca o valoare aleatoare sa se afle sub media aritmetica este 50 %:



Excel: functia **NORMDIST**: determina probabilitatea cumulativa asociata unei valori (distributie normala):

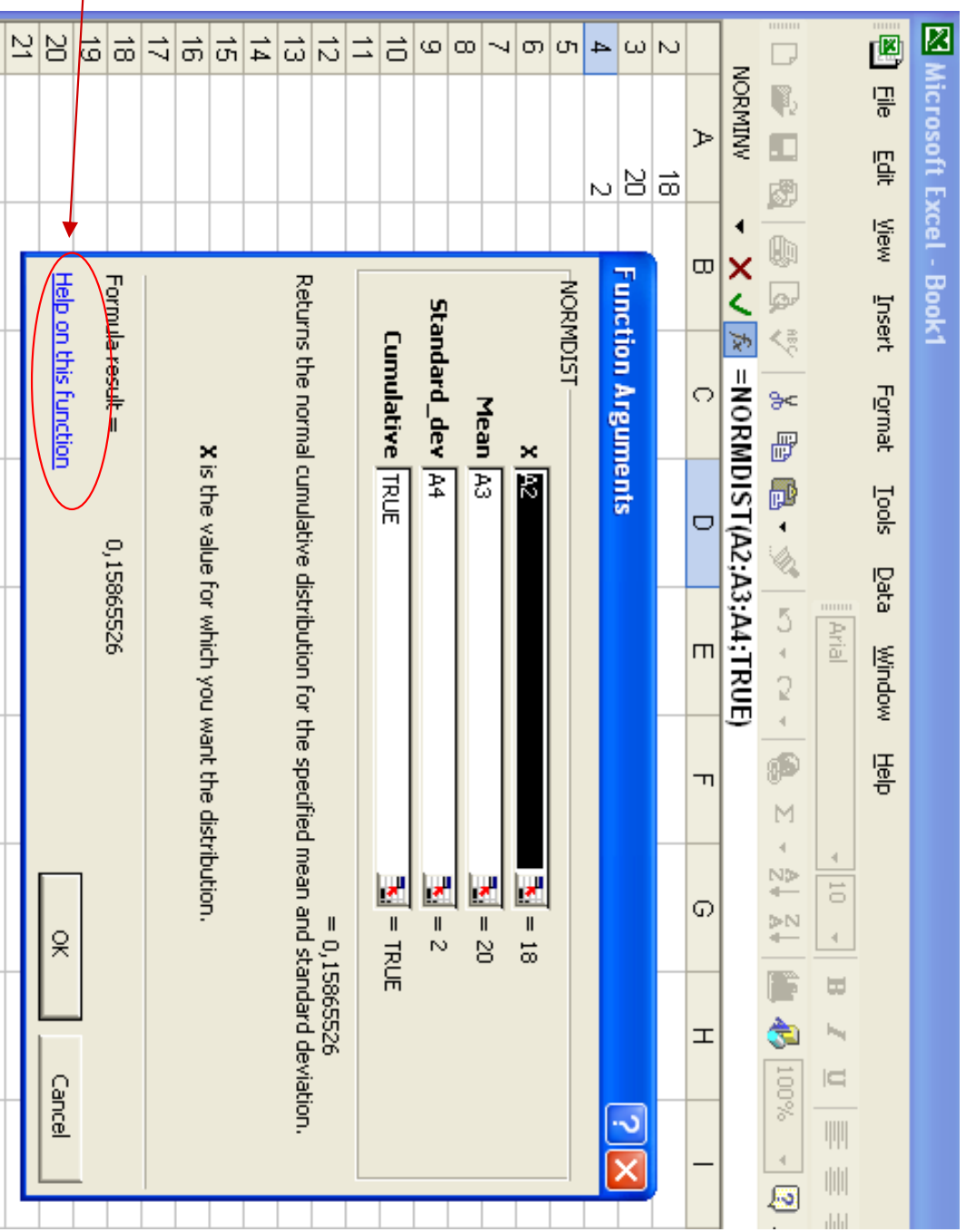
- probabilitatea ca o valoare din setul de date sa fie mai mica sau egala cu valoarea indicata (aria de la $-\infty$ la valoarea indicata)

Cumulative = TRUE

Se indica:

- valoarea
- media
- deviatia standard

descrierea functiei!



Excel: functia **NORMINV**: determina valoarea asociata unei probabilitati cumulative (distributie normala)

Se indica:

- probabilitatea de aparitie
- media
- deviatia standard

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

NORMINV \times \checkmark fx =NORMINV(D4:A3:A4)

	A	B	C	D	E	F	G	H
2	18							
3	20							
4	2			0,158655				

Function Arguments

NORMINV

Probability D4 = 0,15865526

Mean A3 = 20

Standard_dev A4 = 2

Returns the inverse of the normal cumulative distribution for the specified mean and standard deviation.

Standard_dev is the standard deviation of the distribution, a positive number.

Formula result = 18

[Help on this function](#)

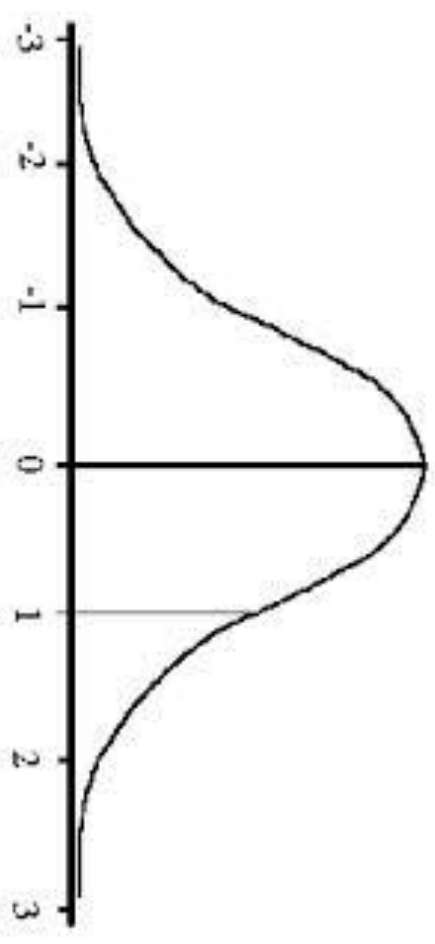
OK Cancel

Distribuția normală standard

O distribuție normală standard are:

media aritmetică = 0


deviația standard = 1



Distribuția normală standard este denumită *distribuția z*

- Orice distribuție normală poate fi transformată într-o distribuție normală standard
- Aria de sub curba unei distribuții normale standard este egală cu 1!
- Aria de sub graficul distribuției standard normale = Probabilitatea de a avea date între valorile extreme (100%!)

Relatia de transformare a unei distributii normale (x_i) intr-o **distributie normala standard**:

punctajul z 
$$Z_i = \frac{(X_i - \bar{X})}{S}$$

Z_i reprezinta numărul deviațiilor standard la care se află valoarea x_i față de media aritmetică: deasupra (> 0), sub (< 0).

O valoare egală cu media aritmetică se transformă în 0!

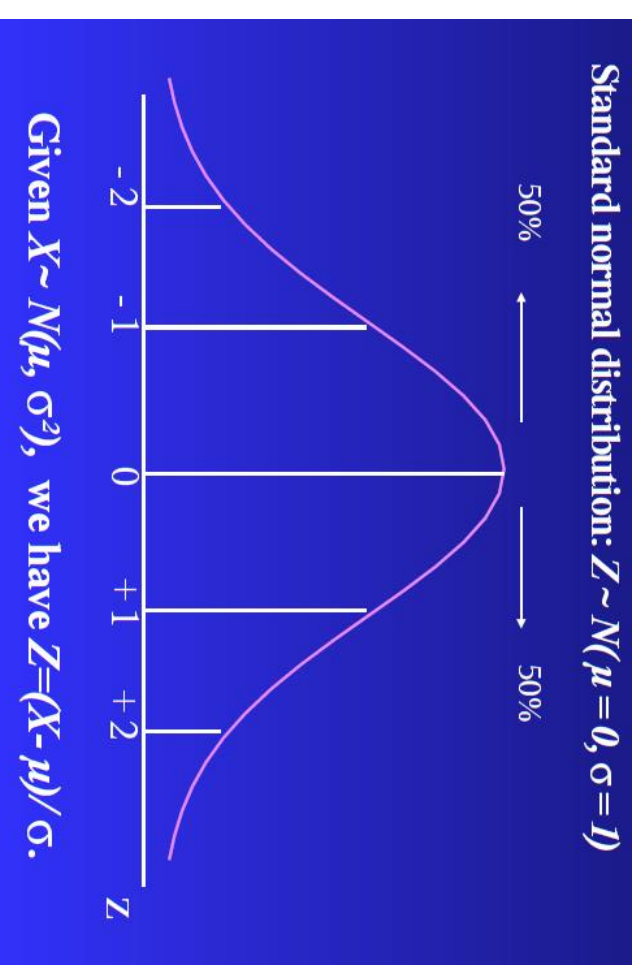
O valoare mai mare cu o deviație standard decât media aritmetică se transformă în +1.

O valoare mai mica cu doua deviații standard decât media aritmetică se transformă în -2.

Punctajul z : valoarea x corespunzatoare se afla la z deviații standard față de media aritmetică.

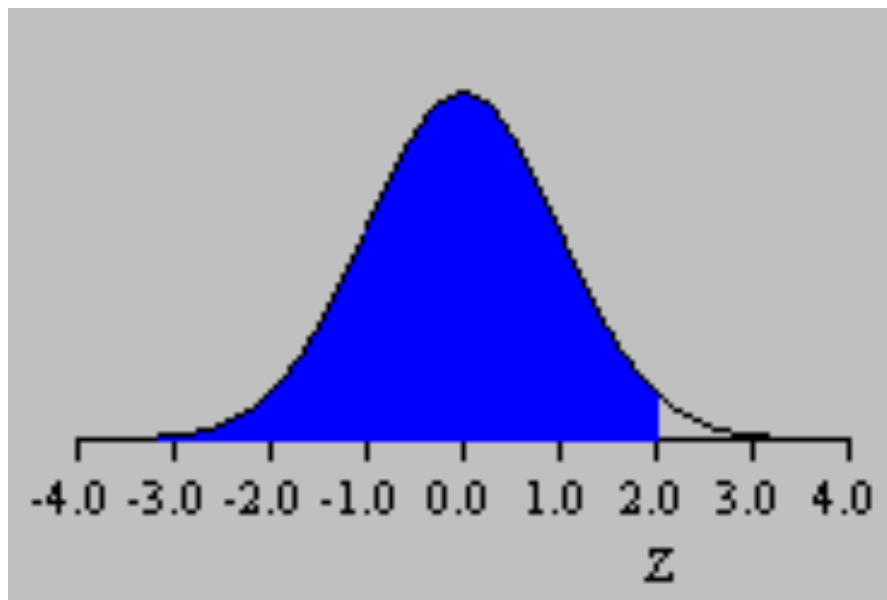
Tabelul probabilităților pentru statistica Z

Probabilitatea asociată unei valori ce este mai mare (sau mai mica) decât media aritmetica, dar nu la un număr întreg de deviații standard poate fi determinata folosind **tabelul probabilităților normale** (sau folosind functii Excel dedicate).



Valorile din acest tabel indică **aria** ce se află sub graficul distribuției standard normale la **valori mai mici** decât punctajul z corespunzător.

Aria de sub grafic cuprinsă între orice două valori este egală cu probabilitatea de a avea date între aceste două valori!



Probabilitatea unei distribuții normale standard

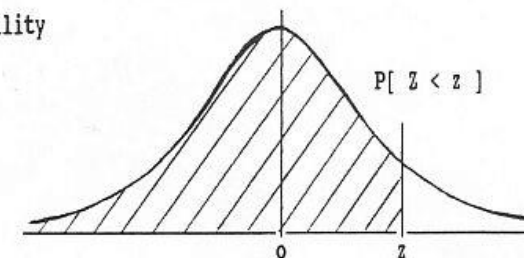
Tabelul indică probabilitatea (p) ca o variabilă standard normală, să aibă o valoare mai mică sau egală cu z (zona hașurată din diagramă).

STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

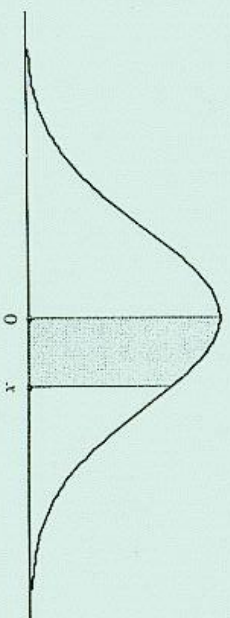
The table gives the cumulative probability up to the standardised normal value z i.e.

$$P[Z < z] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

Appendix 4 – Values of the Standard Normal distribution

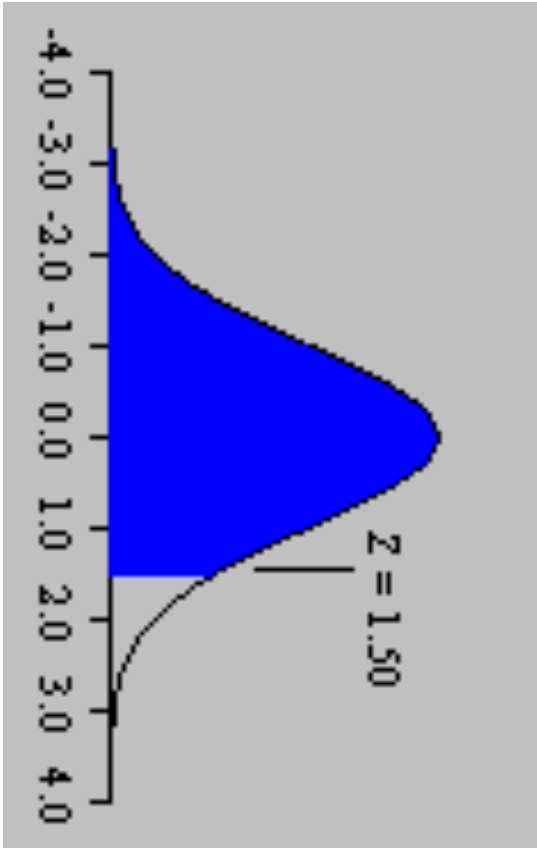


The table gives the probability that a standard Normal variable lies between 0 and x (which is equivalent to the shaded area on the figure).

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0754
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2258	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.2	0.2580	0.2612	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.3	0.2881	0.2910	0.2939	0.2967	0.2996	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4990	0.4993	0.4995	0.4997	0.4998	0.4998	0.4999	0.4999	0.5000

P1. Care este probabilitatea ca o valoare, dintr-un set de date cu distribuție normală, să se afle la cel mult 1,5 deviații standard deasupra mediei? ($p=?$ dacă $z \leq 1,5$)

z	p	z	p	z	p
0.00	0.5000	0.5	0.6915	1.00	0.8413
0.01	0.5040	0.51	0.6950	1.01	0.8438
0.02	0.5080	0.52	0.6985	1.02	0.8461
0.03	0.5120	0.53	0.7019	1.03	0.8485
0.04	0.5160	0.54	0.7054	1.04	0.8508
0.05	0.5199	0.55	0.7088	1.05	0.8531
↓	↓	↓	↓	↓	↓
↓	↓	↓	↓	↓	↓
↓	↓	↓	↓	↓	↓
0.45	0.6736	0.95	0.8289	1.45	0.9265
0.46	0.6772	0.96	0.8315	1.46	0.9279
0.47	0.6808	0.97	0.8340	1.47	0.9292
0.48	0.6844	0.98	0.8365	1.48	0.9306
0.49	0.6879	0.99	0.8389	1.49	0.9319
0.50	0.6915	1.00	0.8413	1.50	0.9332



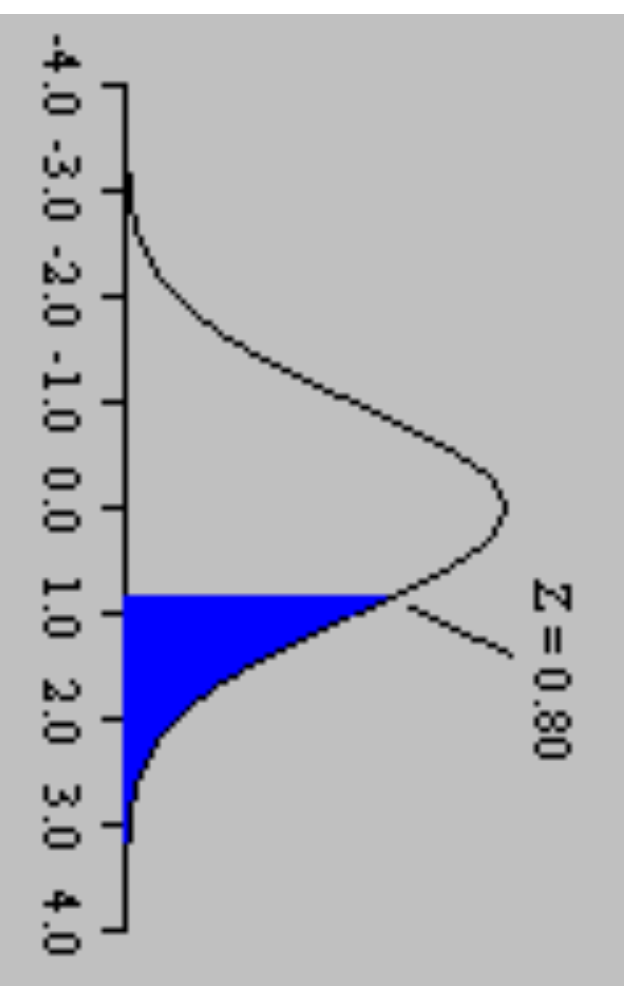
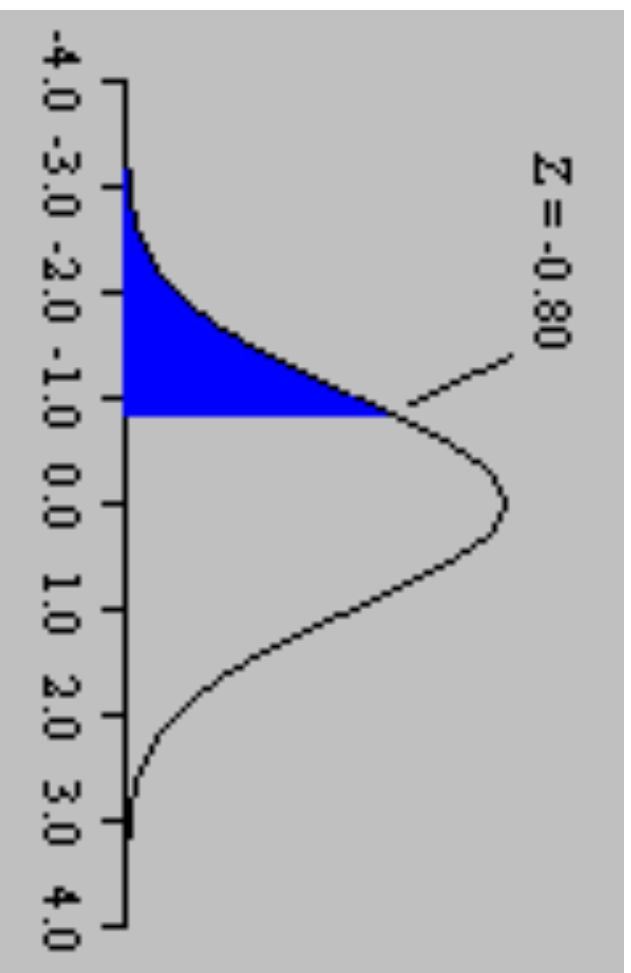
$z = 1,5$

$p = 0,933 \quad (93,3\%)$

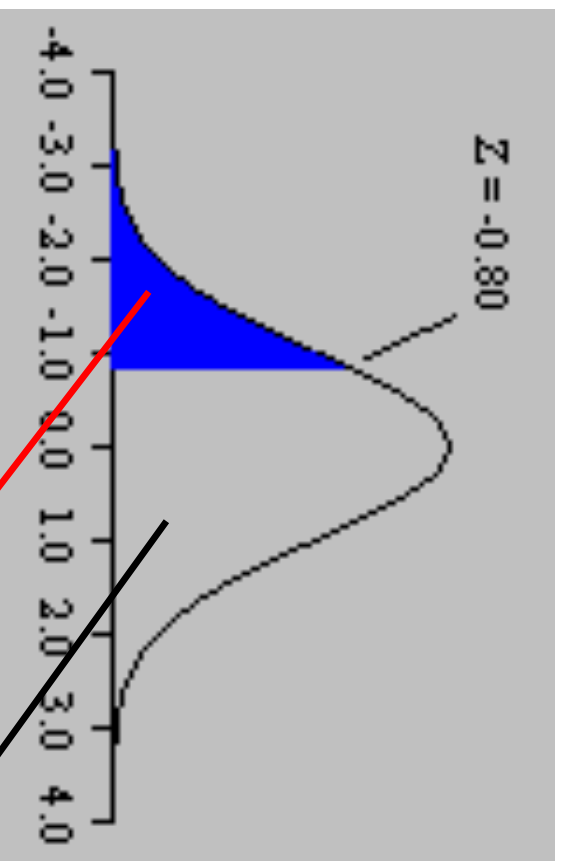
TOTDEAUNA trebuie folosită o diagramă!

Atentie!

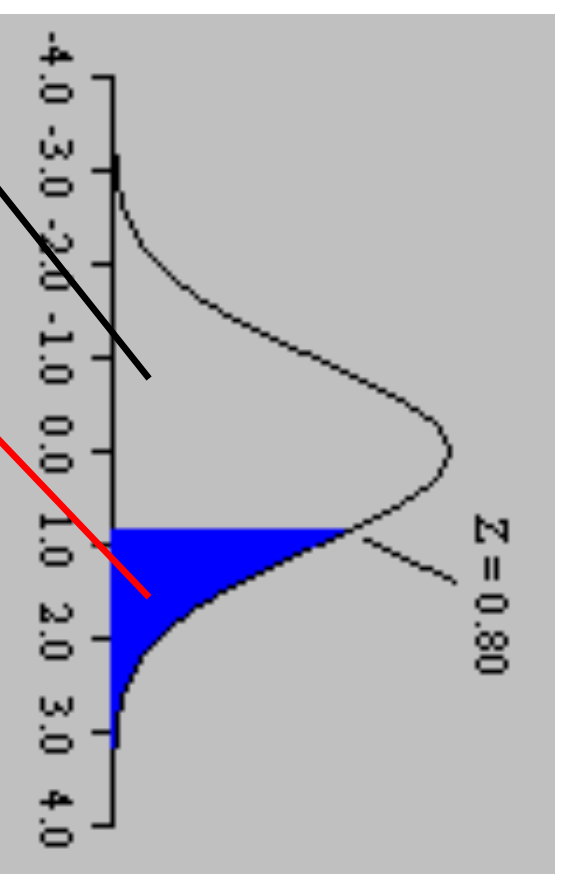
- în tabelul statistici z toate valorile z sunt pozitive!
- distribuția normală este simetrică



P2. Care este probabilitatea ca o valoare, dintr-un set de date cu distribuție normală, să se fie mai mică decât valoarea situată la 0,8 deviații standard sub medie? ($p=?$ dacă $z \leq -0,8$).



$p = 0,7881$

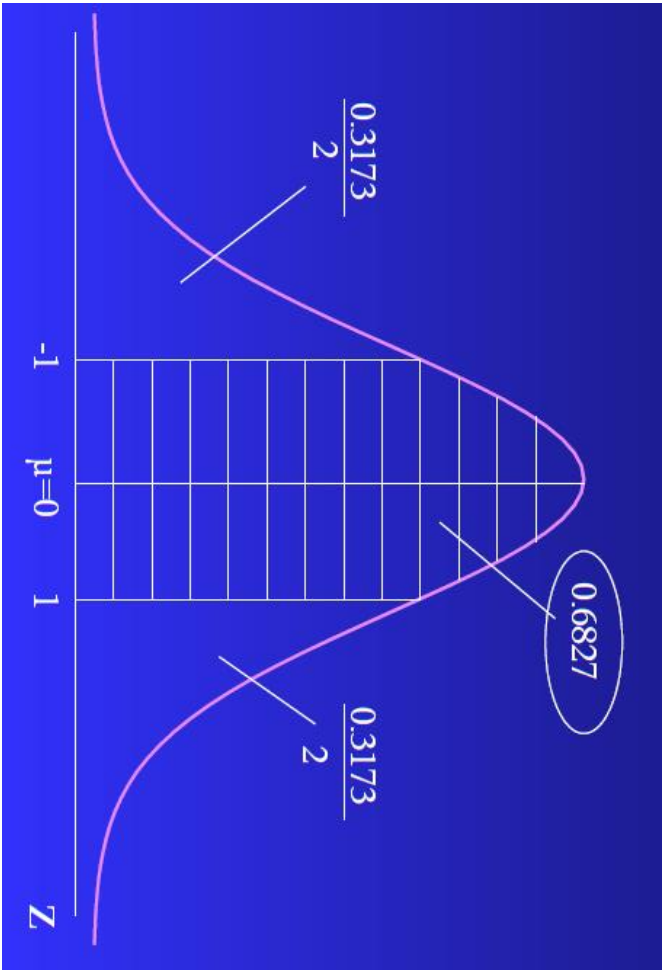


$$p = 1 - 0,7881 = 0,2119$$

valoarea pentru $z = 0,80 \rightarrow p = 0,7881$
(simetrie fata de medie, aria totala =1!)

(Two-Sided Tail Probabilities of the Normal Curve)

Z	0.00	0.01	0.02	.	.	0.09
.0	1.000	.9920	.9840	.	.	.
.
.
.
.9	.3681	.3628	.3576	.	.	.
1.0	.3173	.3125	.3077	.	.	.
1.1	.2713	.2670	.2627	.	.	.
.
.

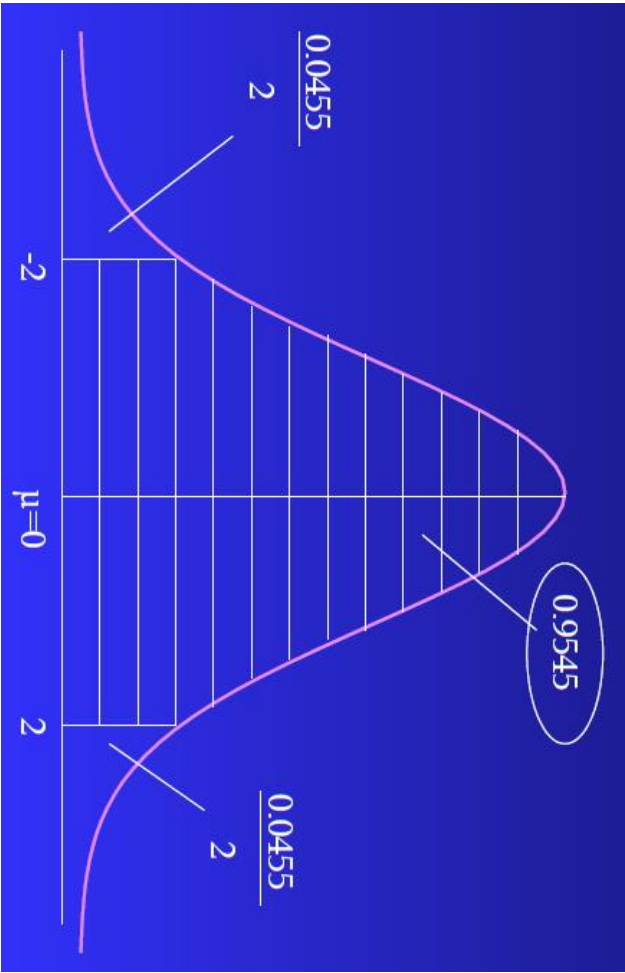


tabele "two -tail"

limitate la ambele capete

(Two-Sided Tail Probabilities of the Normal Curve)

Z	0.00	0.01	0.02	.	.	0.09
.0	1.000	.9920	.9840	.	.	.
.
.
.
1.9	.0574	.0561	.0549	.	.	.
2.0	.0455	.0444	.0434	.	.	.
2.1	.0357	.0349	.0340	.	.	.
.
.

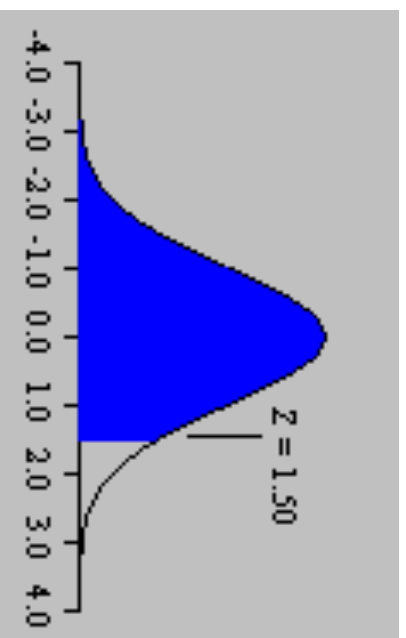


Excel: functia **NORMSDIST** determina probabilitatea ca o valoare sa se afle la cel mult z deviatii standard fata de medie.

P1. Calculati

probabilitatea ca o valoare, dintr-un set de date cu distributie normală, să se afle la cel mult 1,5 deviații standard deasupra mediei?

($p=?$ daca $z \leq 1,5$)



Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

Font: Arial, Size: 10

Formulas: NORMSDIST X fx $=\text{NORMSDIST}(A4)$

	A	B	C	D	E	F	G	H
2								
3	z							
4	1,5							
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								

Function Arguments

NORMSDIST

z $A4$ $= 1,5$

Returns the standard normal cumulative distribution (has a mean of zero and a standard deviation of one).

z is the value for which you want the distribution.

Formula result = 0,933192771

[Help on this function](#)

OK Cancel

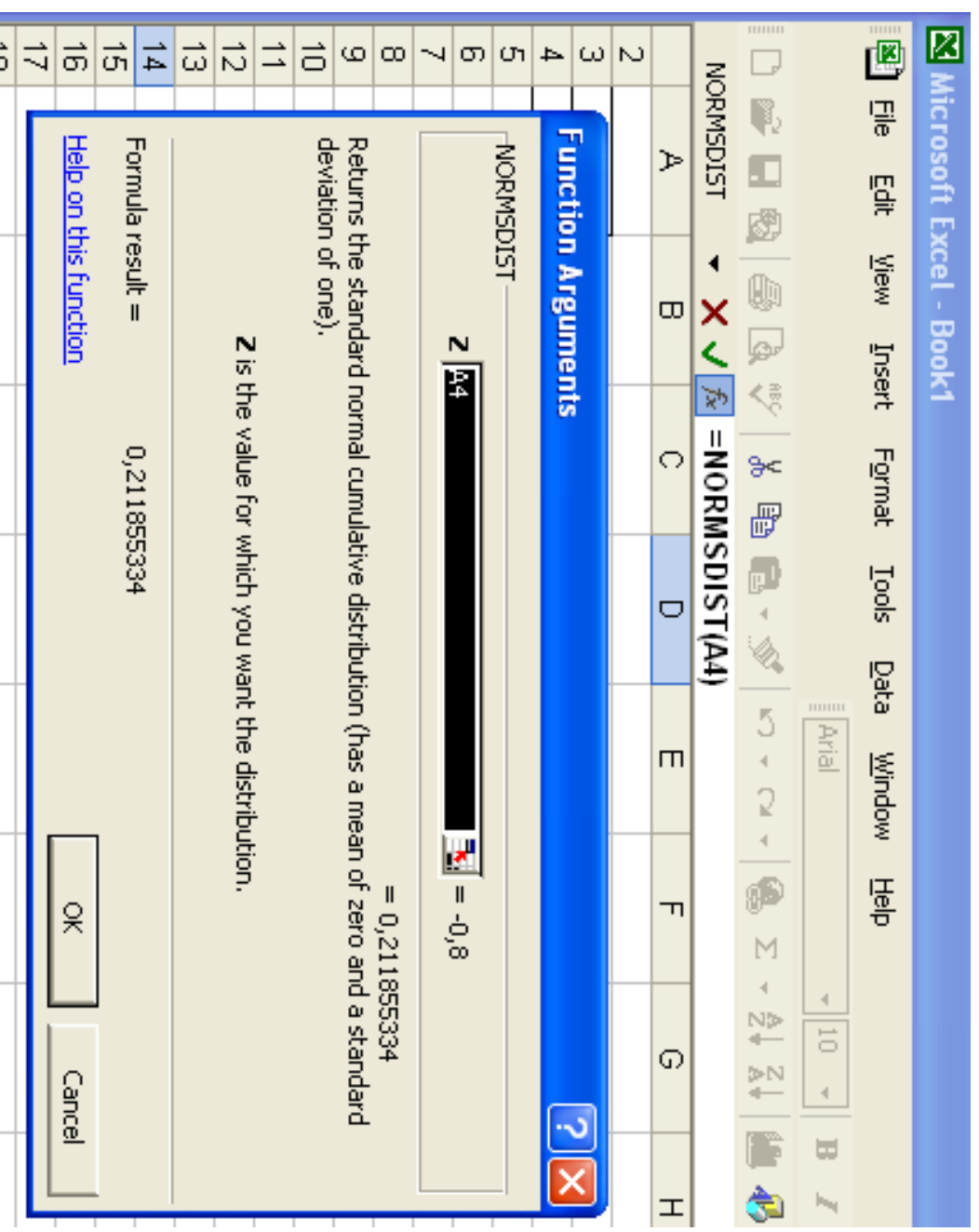
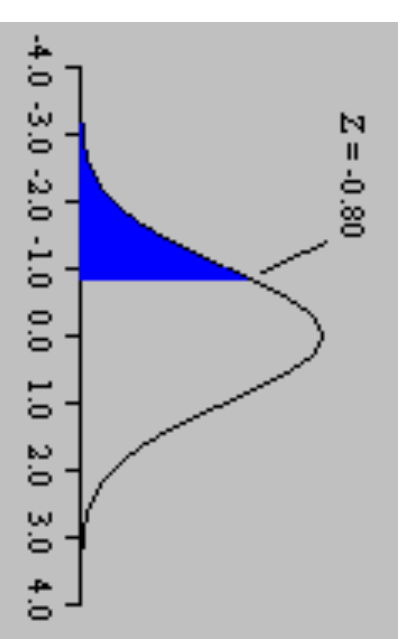
Valoarea punctajului z poate sa fie si negativa!

P2. Care este

probabilitatea ca o valoare, dintr-un set de date cu distribuție normală, să se fie mai mică decât valoarea situată la 0,8 deviații standard **sub medie**? (p=? daca $z \leq -0,8$).

$Z = -0,8$

$p = 0,21$



Excel: functia **NORMSINV** calculeaza punctajul z asociat unei probabilitati.

Ex: la câte deviații standard sub media aritmetică trebuie să fie o măsurătoare dintr-o distribuție normală pentru a avea 21% probabilitate de apariție?

$$p = 0,21$$

$$Z = -0,8$$

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

Arial 10 B I

NORMSINV \times \checkmark f_x =NORMSINV(A2)

	A	B	C	D	E	F	G
1							
2		0,21					
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							

Function Arguments

NORMSINV

Probability A2 = 0,21

Returns the inverse of the standard normal cumulative distribution (has a mean of zero and a standard deviation of one).

Probability is a probability corresponding to the normal distribution, a number between 0 and 1 inclusive.

Formula result = -0,806421028

[Help on this function](#)

OK Cancel

P3. Se considera ca valoarea nivelului seric al 1,25 dihidroxivitamina D in sangele adolescentelor este normal distribuita, avand o medie de 65 pg/ml si o deviatie standard de 12,5 pg/ml.

a) ce procentaj din totalul adolescentelor au nivelul seric al dihidroxivitaminei D mai mare (mic) de 65 pg/ml?

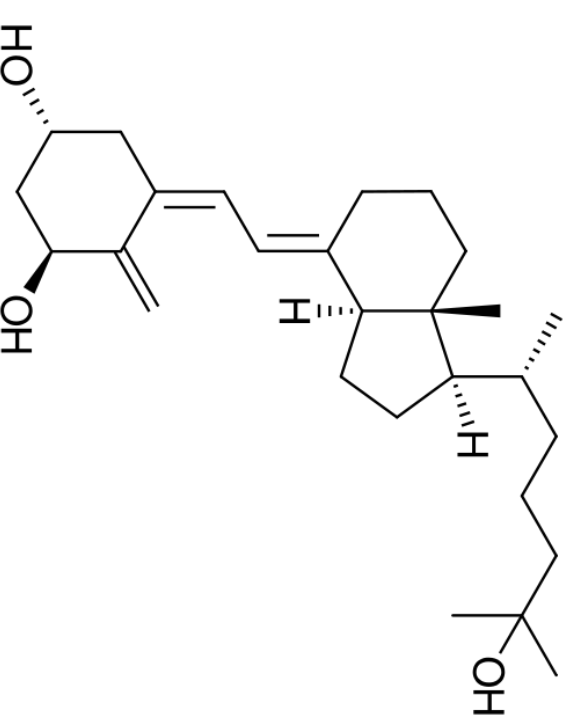
b) cat la suta din totalul adolescentelor au nivelul seric al dihidroxivitaminei D situat intre 40pg/ml si 90 pg/ml?

1,25-dihydroxivitamin D3 (*Calcitriol*) = forma hormonal activa a vitaminei D (are 3 grupuri hidroxil: 1,25-(OH)₂D₃).

Calcitriolul creste nivelul de calciu (Ca²⁺) din organism prin:

(1) cresterea absorbtiei de calciu din intestin in sange

(2) posibila crestere a eliberarii in sange a calciului din oase.



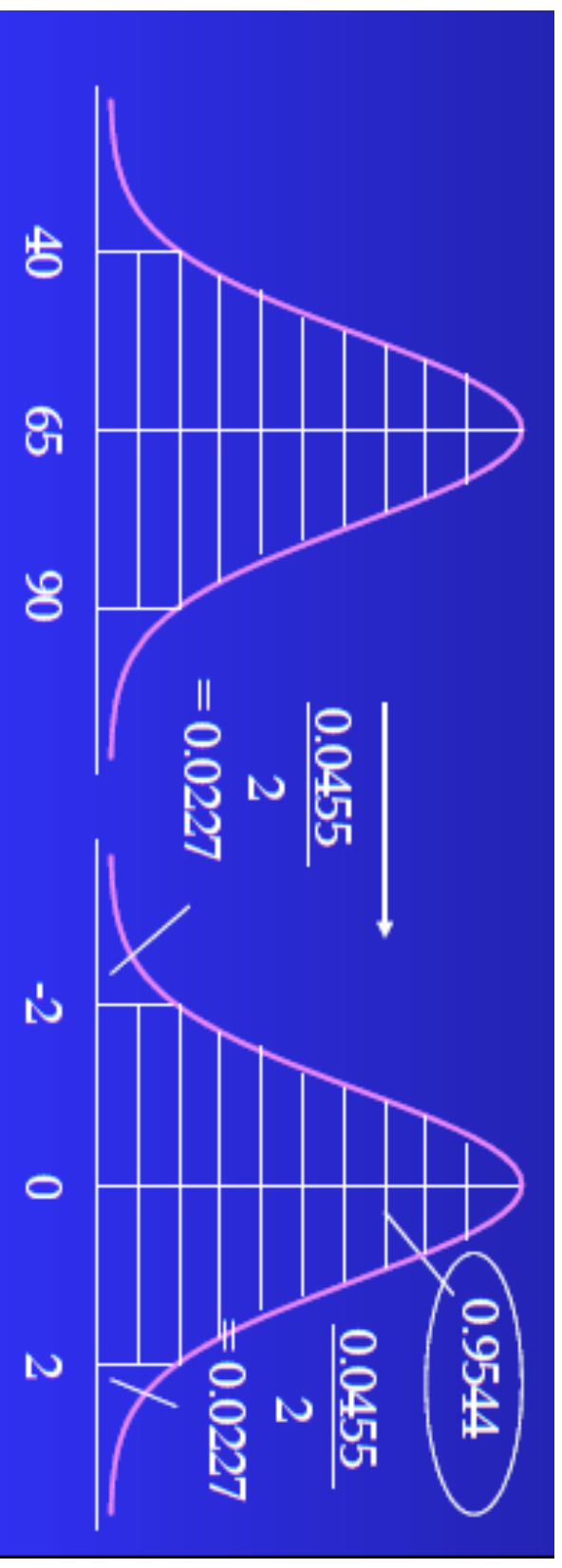
$$a) P(z > 65) = P(z < 65) = 50\%$$

$$b) P(45 < z < 90) = ?$$

$$z_1 = \frac{40 - 65}{12,5} = -2 \quad z_2 = \frac{90 - 65}{12,5} = 2$$

$$P_{(45 < z < 90)} = 1 - \frac{0,0455}{2} - \frac{0,0455}{2} = 0,9545$$

(Two-Sided Tail Probabilities)			
Z	0.00	0.01	0.02
.0	1.000	.9920	.9840
.	.	.	.
.	.	.	.
1.9	.0574	.0561	.0549
2.0	.0455	.0444	.0434
2.1	.0357	.0349	.0340



Teorema limitei centrale

Dacă mărimea eșantionului este mare (>30), distribuția *mediei aritmetice a eșantioanelor* va fi o distribuție normală.

Dacă mărimea eșantionului este mică (<30), distribuția *mediei aritmetice a eșantioanelor* va fi o distribuție normală *numai dacă* populația din care provin eșantioanele are o distribuție normală

Media aritmetică a **distribuției mediei eșantioanelor** este egală cu media aritmetică a populației (μ):

$$\mu_{\bar{x}} = \mu$$

Eroarea standard a **mediei** este *deviația standard a distribuției mediei eșantioanelor*:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

σ = deviația standard a întregii populații

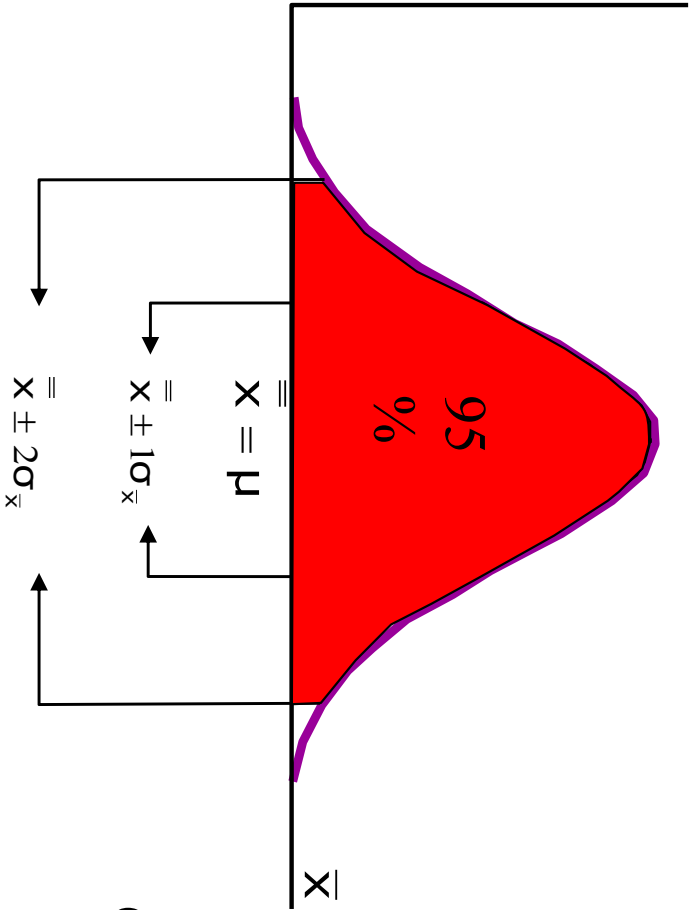
$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

s = deviația standard a eșantionului

Distributia valorilor individuale ale unei populatii



Distributia valorilor medii ale tuturor esantioanelor unei populatii



distributia mediei esantioanelor

Intervalul de încredere pentru media aritmetică

Intervalul de încredere = Confidence Interval = CI

Intervalul de încredere (CI) în jurul mediei aritmetice indică *acuratețea* cu care media aritmetică a eșantionului estimează *media aritmetică a întregii populații*.

Ex: Dimensiunea medie a calculilor biliari este $2,5 \text{ cm} \pm 0,20 \text{ cm}$ pentru un nivel de încredere de 95%, folosind un eșantion de 100 calculi.

În formularea de mai sus se afirmă că: suntem 95% siguri că media populației în discuție se află între 2,3 cm și 2,7 cm.

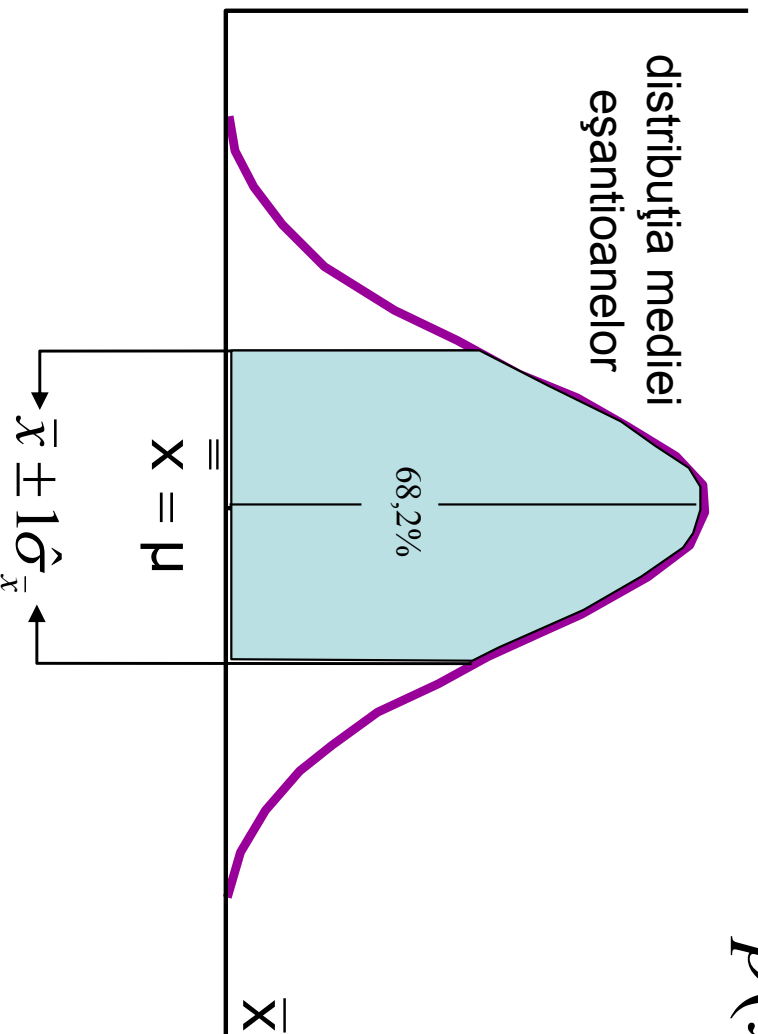
Intervalul de încredere = regiunea care conține valoarea reală (parametrul de interes), data cu o probabilitate specificată.

Există o probabilitate de 95,4% ca media oricărui eșantion să fie la distanță de 2 erori standard față de media întregii populații.

Există o probabilitate de 68,2% ca media oricărui eșantion să fie la distanță de 1 eroare standard față de media întregii populații.

$$p(\bar{x} - 1\hat{\sigma}_{\bar{x}} < \mu < \bar{x} + 1\hat{\sigma}_{\bar{x}}) = 0,68$$

$$p(\bar{x} - 2\hat{\sigma}_{\bar{x}} < \mu < \bar{x} + 2\hat{\sigma}_{\bar{x}}) = 0,95$$



$$CI = \bar{x} \pm z\hat{\sigma}_{\bar{x}}$$

$$p(\bar{x} - z\hat{\sigma}_{\bar{x}} < \mu < \bar{x} + z\hat{\sigma}_{\bar{x}}) = \text{din tabelul probabilitatilor}$$

$$CI = \bar{x} \pm z \sigma_{\bar{x}}$$

CI = Intervalul de încredere
(confidence interval)

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

σ_x - eroarea standard a mediei
s - deviația standard a eșantionului

Valoarea z se calculeaza în funcție de *nivelul de siguranță* (probabilitatea)!

Nivel de siguranță (p) – cât de siguri dorim să fim că media aritmetică a populației se află în intervalul de încredere

(90%; 95%; 99% → **p** = 0,90; 0,95; 0,99)

Nivel de semnificație (a sau α) – cât de puțin vrem să greșim

(10%; 5%; 1% → a = 0,10; 0,05; 0,01)

Relatia dintre nivelul de siguranță (p) si nivelul de semnificatie (a):

$$a = 1 - p$$

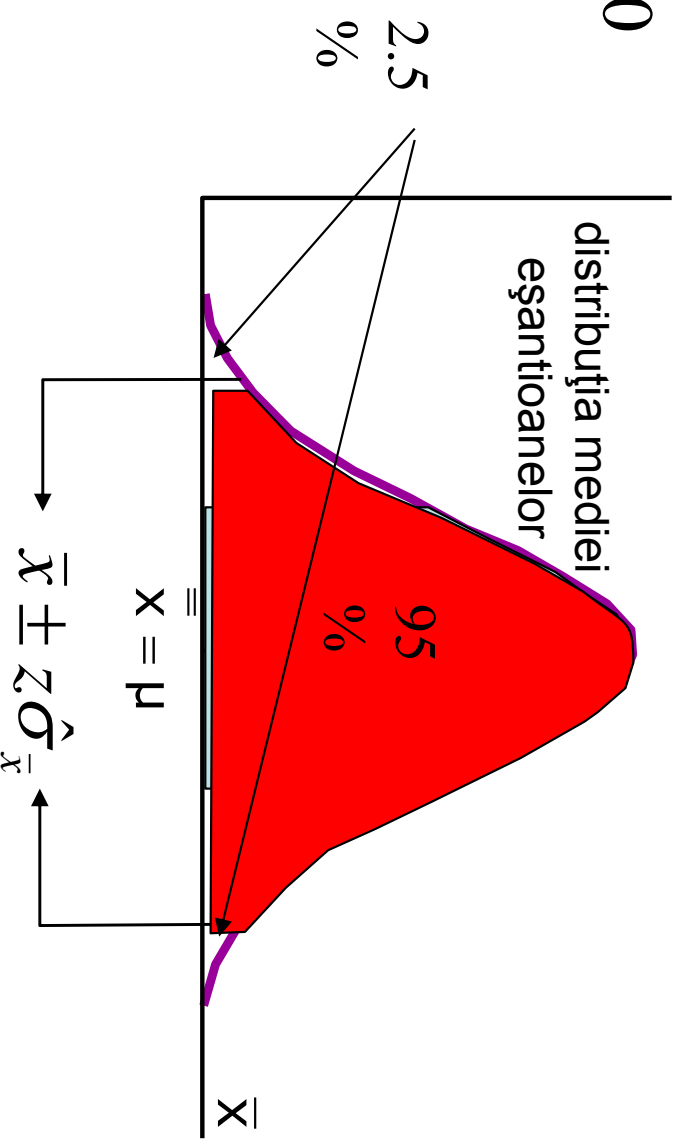
P4. Calculați intervalul de încredere în jurul mediei, corespunzător unei probabilități de 95%, cunoscând următoarele: media aritmetica = 2,5; deviația standard = 0,2; mărimea esanționului = 100

$$\bar{x} = 2,5 \quad s = 0,2 \quad n = 100$$

$$CI = \bar{x} \pm z \cdot \sigma_{\bar{x}}$$

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{0,2}{\sqrt{100}} = 0,02$$

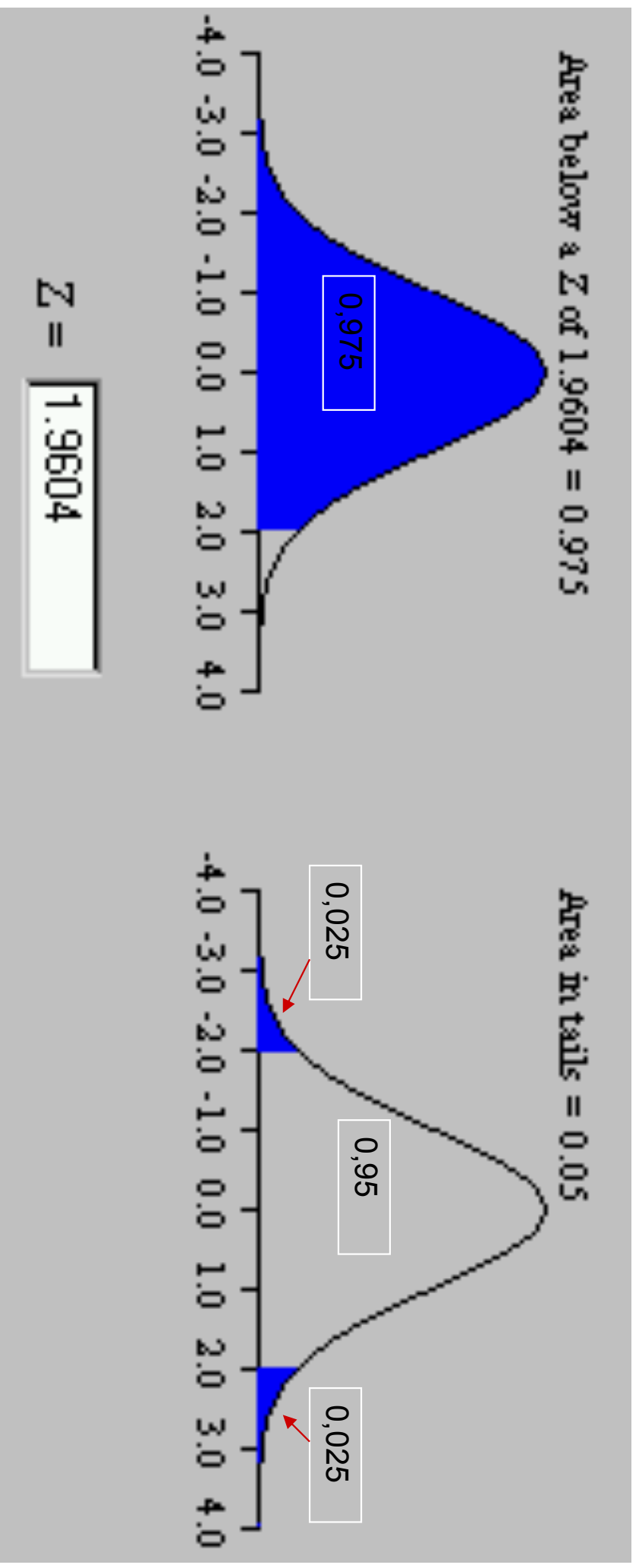
$$CI = 2,5 \pm z \cdot 0,02$$



nivelul de sigurantă (se consideră în jurul mediei!): $p = 0,95$

Valoarea z va fi asociată unei valori p' diferită de 0,95!!!

Aria unui singur interval este $[(1-0,95)/2] = 0,025$, deci trebuie determinată **valoarea lui z asociată unei valori $p' = 0,975$**



z	p	z	p	z	p
1.00	0.8413	1.50	0.9332	2.00	0.97725
1.01	0.8438	1.51	0.9345	2.01	0.97778
1.02	0.8461	1.52	0.9357	2.02	0.97831
1.46	0.9279	1.96	0.9750	2.46	0.99305
1.47	0.9292	1.97	0.9756	2.47	0.99324
1.48	0.9306	1.98	0.9761	2.48	0.99343
1.49	0.9319	1.99	0.9767	2.49	0.99361
1.50	0.9332	2.00	0.9772	2.50	0.99379

$p' = 0,975$
 $z = 1,96$

$CI = 2,5 \pm 1,96 \cdot 0,02$

$CI = 2,5 \pm 0,04$

nivelul de siguranta	p'	z
90%	0,90+0,05	1,65
95%	0,95+0,025	1,96
99%	0,99+0,005	2,58

Atentie: Pentru un tabel al probabilitatilor limitat la un singur capat (tip "one tail") la deducerea punctajului z trebuie folosita probabilitatea p'

Excel: functia **CONFIDENCE**: determina intervalul de incredere $Z \cdot \sigma_{\bar{x}}$

Obs.: rezultatul obtinut se adauga/scade la media esantionului!

$$CI = \bar{x} \pm Z \cdot \sigma_{\bar{x}}$$

P4. Calculați intervalul de încredere în jurul mediei, corespunzător unei probabilități de 95%, cunoscand: media aritmetica = 2,5; s = 0,2; n = 100

The screenshot shows the Microsoft Excel interface with the **CONFIDENCE** function being used. The formula bar displays **=CONFIDENCE(C6;C3;C4)**. The spreadsheet data is as follows:

	A	B	C	D	E	F	G	H	I
1									
2		x med	2,5						
3		s	2						
4		n	100						
5		p	0,95						
6		a (1-p)	0,05						
7									
8									
9									
10									

The **Function Arguments** dialog box for the **CONFIDENCE** function is open, showing the following values:

- Alpha:** C6 = 0,05
- Standard_dev:** C3 = 2
- Size:** C4 = 100

The dialog box also displays the formula result as **0,391992557** and includes a link to [Help on this function](#).

Raspuns: Media populatiei din care a fost extras esantionul va fi situata in domeniul: $2,5 \pm 0,39199$

P5. Timpul de spitalizare dupa interventia chirurgicala in cazul protezei totale de genunchi a fost inregistrat pentru 90 pacienti ai spitalului Z.

Media esantionului este 4,2 zile, iar deviatia standard este 1,05 zile.

Determinati cu o probabilitate de 90% intervalul de incredere pentru media populatiei: zile de spitalizare in cazul protezarii totale a genunchiului.

$n = 90 \rightarrow$ esantion
mare,

media = 4,2

$s = 1,05$

90 % $\rightarrow z = 1,65$

$$CI = \bar{x} \pm 1,65 \frac{s}{\sqrt{n}}$$

$$CI = 4,2 \pm 1,65 \frac{1,05}{\sqrt{90}}$$

$$CI = 4,2 \pm 0,1826$$

Excel: $z \cdot \sigma = 0,18205$

The image shows an Excel spreadsheet with data for a confidence interval calculation. The data is as follows:

	A	B	C	D	E	F	G	H
1								
2		x med	4,2					
3		s	1,05					
4		n	90					
5		p	0,9					
6		a (1-p)	0,1					
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								

Below the spreadsheet is the 'Function Arguments' dialog box for the CONFIDENCE function. The arguments are:

- Alpha: C6 = 0,1
- Standard_dev: C3 = 1,05
- Size: C4 = 90

The result of the function is displayed as 0,182051919.

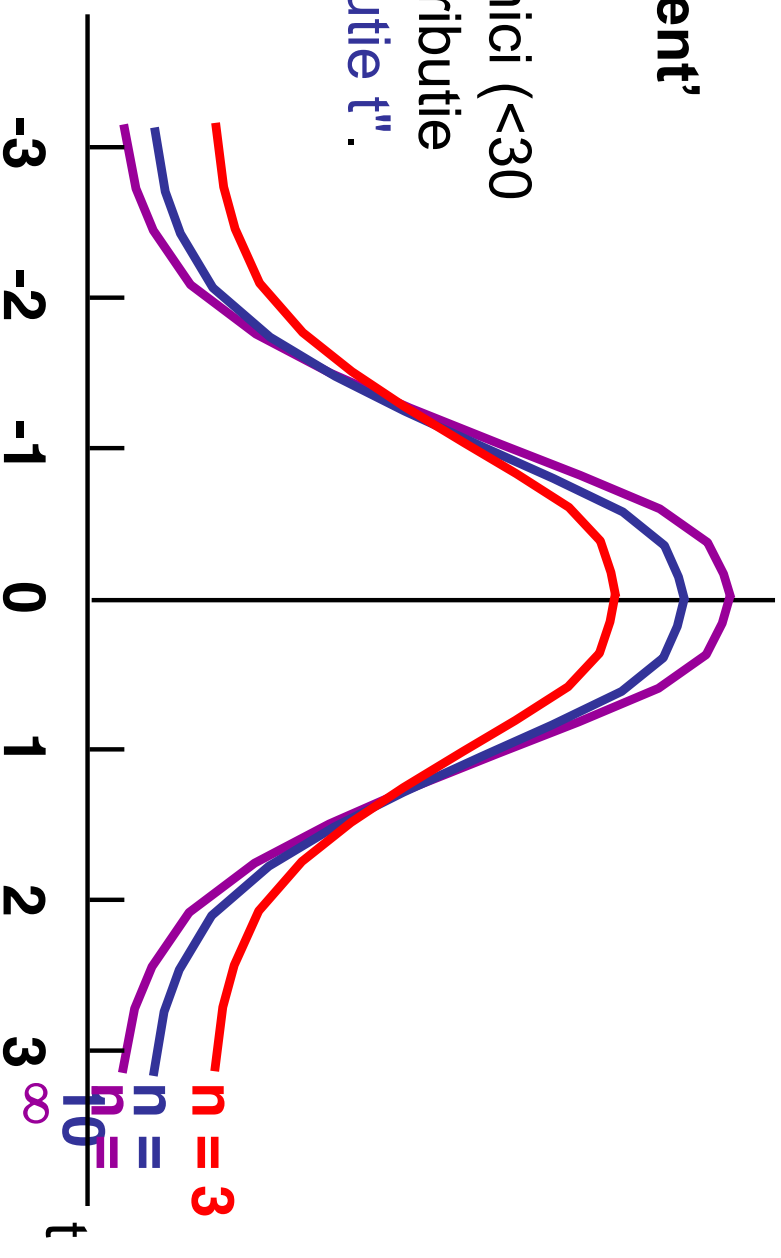
Formula result = 0,182051919

Help on this function

OK Cancel

Distribuția 't-student'

În cazul eșantioanelor mici (<30 valori) trebuie folosită o distribuție ușor diferită, numită "distribuție t".



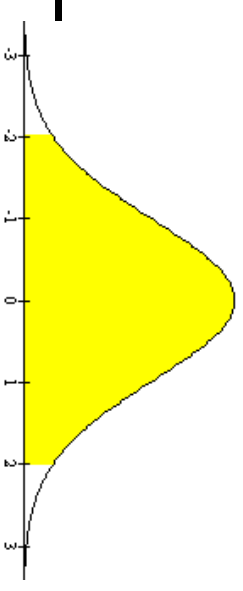
Distribuția t descrie o familie de distribuții dependente de mărimea eșantioanelor.

Pentru un eșantion ce conține mai mult de 30 măsurători, distribuția t devine identică cu o distribuția z.

$$CI = \bar{X} \pm t \cdot \hat{\sigma}_{\bar{X}}$$

Tablul distribuției "t"

(Student's t Distribution)



Degrees of Freedom	Significance Level					
	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.001$
1	3.078	6.314	12.706	25.452	63.656	636.578
2	1.886	2.920	4.303	6.205	9.925	31.600
3	1.638	2.353	3.182	4.177	5.841	12.924
18	1.330	1.734	2.101	2.445	2.878	3.922
19	1.328	1.729	2.093	2.433	2.861	3.883
20	1.325	1.725	2.086	2.423	2.845	3.850
21	1.323	1.721	2.080	2.414	2.831	3.819

a (α) este probabilitatea ca o valoare să fie mai extremă decât t

Obs.:

- tabelul distribuției t este tăiat la ambele capete!
- valorile sunt grupate în funcție de - *nivelul de semnificație* ($\alpha = 1-p$)
 - *gradul de libertate* ($df = n-1$).

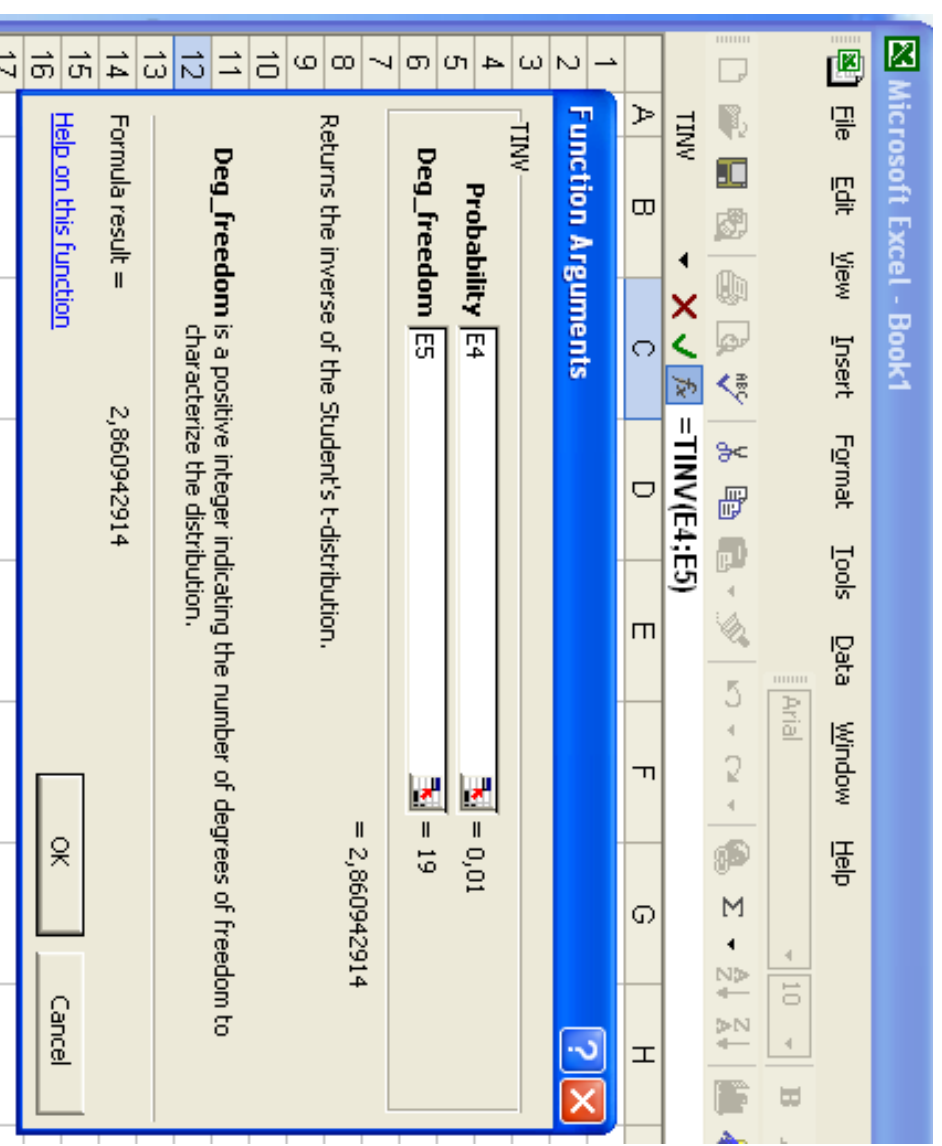
Excel: functia **TINV**: determina *valoarea distributiei t* daca se cunoaste probabilitatea ($\alpha = 1 - p$) si marimea esantionului ($df = n - 1$)

valoarea distributiei t = punctajul t corespunzator probabilitatii date

Distributia t este de tipul "two tail"!

Se indica:

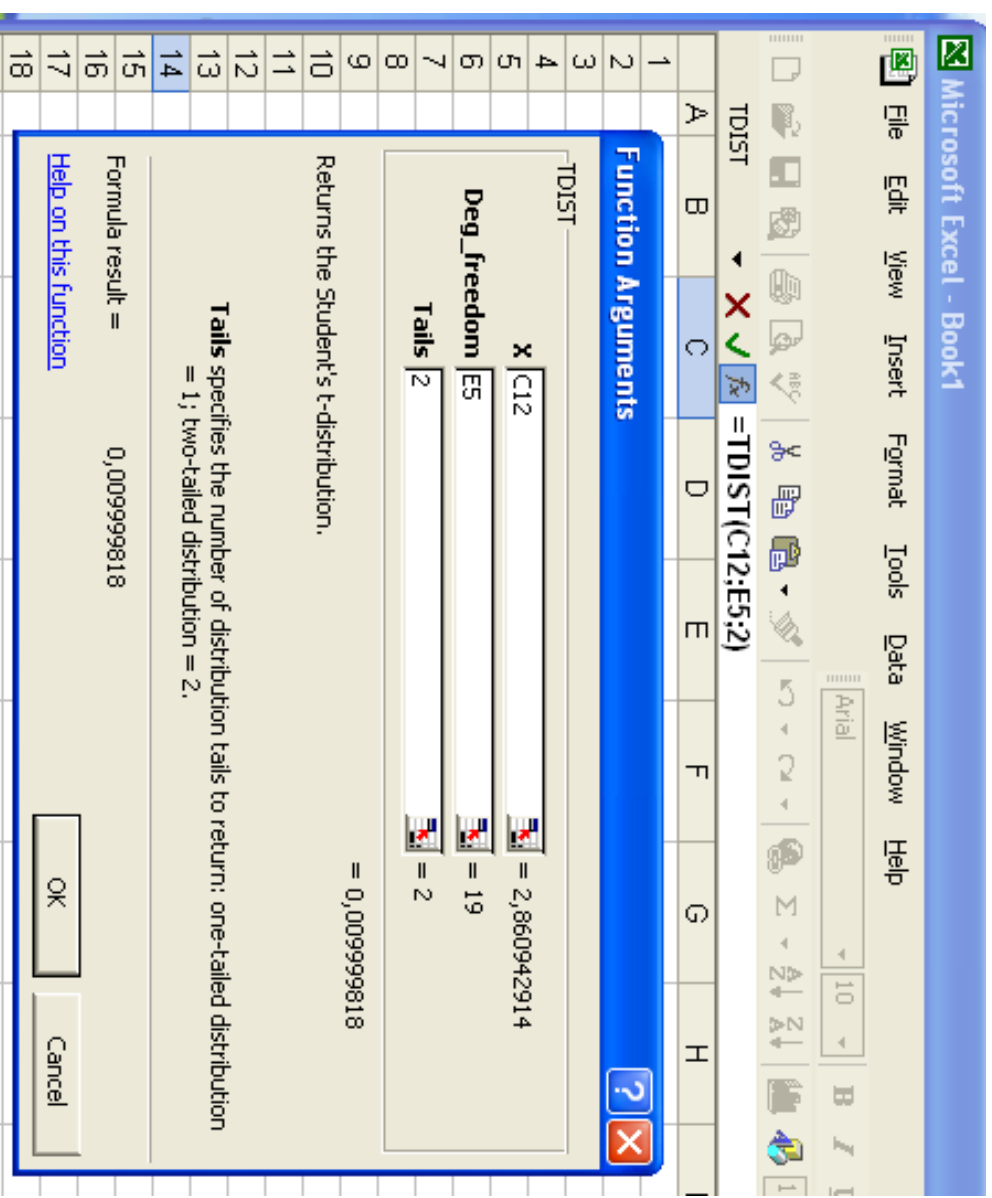
- nivelul de semnificatie (notat "Probability" in fereastra functiei TDIST!)
- gradul de libertate



Excel: functia **TDIST**: determina *nivelul de semnificație* ($\alpha = 1-p$)

Se indica:

- valoarea distribuției t (notată "x" în fereastra funcției TDIST!)
- gradul de libertate
- tipul de distribuție (one-tailed/two tailed)

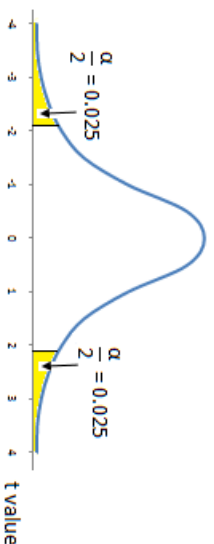


t – Table (tail probabilities of the t-distributions)

Degrees of Freedom	2Q (Q)	0.10 (0.05)	0.05 (0.025)	0.01 (0.005)	0.005 (0.0025)	0.001 (0.0005)
1		6.3138	12.706	63.657	127.32	636.62
2		2.9200	4.3026	9.9251	14.0911	31.6075
3		2.3534	3.1825	5.8408	7.4633	12.9258
4		2.1318	2.7764	4.6040	5.5980	8.6087
5		2.0151	2.5706	4.0323	4.7734	6.8701
6		1.9432	2.4469	3.7075	4.3169	5.9590
7		1.8946	2.3646	3.4995	4.0293	5.4088
8		1.8595	2.3060	3.3555	3.8326	5.0421
9		1.8331	2.2621	3.2498	3.6895	4.7805
10		1.8125	2.2281	3.1693	3.5814	4.5871
11		1.7969	2.2010	3.1057	3.4967	4.4374
12		1.7823	2.1788	3.0546	3.4285	4.3184
13		1.7709	2.1604	3.0122	3.3726	4.2215
14		1.7613	2.1448	2.9768	3.3258	4.1412
15		1.7530	2.1314	2.9467	3.2862	4.0735
16		1.7469	2.1199	2.9207	3.2521	4.0157
17		1.7396	2.1098	2.8982	3.2226	3.9659
18		1.7341	2.1009	2.8784	3.1967	3.9224
19		1.7291	2.0930	2.8609	3.1738	3.8841
20		1.7247	2.0860	2.8453	3.1536	3.8502
21		1.7207	2.0796	2.8313	3.1353	3.8200
22		1.7171	2.0739	2.8187	3.1189	3.7928
23		1.7139	2.0687	2.8073	3.1041	3.7683
24		1.7109	2.0639	2.7969	3.0906	3.7461
25		1.7081	2.0595	2.7874	3.0783	3.7258
26		1.7056	2.0555	2.7787	3.0670	3.7073
27		1.7033	2.0518	2.7707	3.0566	3.6903
28		1.7011	2.0484	2.7632	3.0470	3.6746
29		1.6991	2.0462	2.7564	3.0382	3.6601
30		1.6973	2.0443	2.7500	3.0299	3.6466

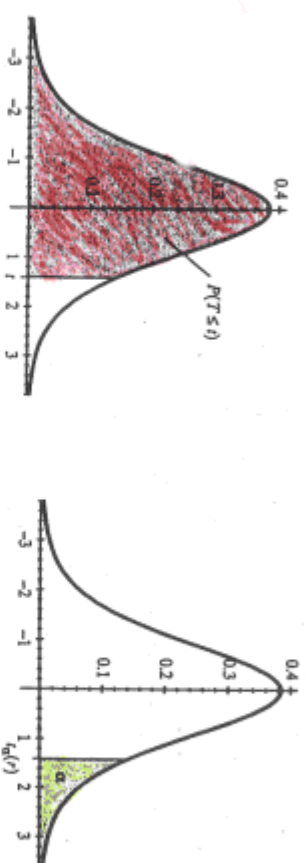
Student's t Distribution Table

For example, the t value for
18 degrees of freedom
is 2.101 for 95% confidence
interval (2-Tail $\alpha = 0.05$).



	90%	95%	97.5%	99%	99.5%	99.95%	1-Tail Confidence Level
	80%	90%	95%	98%	99%	99.9%	2-Tail Confidence Level
	0.100	0.050	0.025	0.010	0.005	0.0005	1-Tail Alpha
df	0.20	0.10	0.05	0.02	0.01	0.001	2-Tail Alpha

Table VI The t Distribution



$$P(T \leq t) = \int_{-\infty}^t \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r} \Gamma(r/2) (1 + w^2/r)^{(r+1)/2}} dw$$

$$P(T \leq -t) = 1 - P(T \leq t)$$

	$P(T \leq t)$							
r	0.60	0.75	0.90	0.95	0.975	0.99	0.995	
	$t_{0.40}(r)$	$t_{0.25}(r)$	$t_{0.10}(r)$	$t_{0.05}(r)$	$t_{0.025}(r)$	$t_{0.01}(r)$	$t_{0.005}(r)$	
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	

P6. Calculați intervalul de încredere în jurul mediei corespunzător unei probabilități de 95%

$$\bar{x} = 2,5 \quad s = 0,2 \quad n = 20$$

$n = 90 \rightarrow$ esantion mic

\rightarrow distribuția t, $df = 19$

95 %, $\alpha = 0,05 \rightarrow t = 2,093$

$$CI = \bar{x} \pm t \cdot \sigma_{\bar{x}}$$

$$CI = \bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$$

$$CI = 2,5 \pm 2,093 \cdot 0,0447$$

$$CI = 2,5 \pm 0.09$$

The screenshot shows a Microsoft Excel spreadsheet with the following data in cells B2 to B7:

1	A	B	C	D	E	F	G	H
2		x med	2,5					
3		s	2					
4		n	20					
5		p	0,95					
6		a (1-p)	0,05					
7		df	19					

The formula bar shows the active cell (C6) containing the formula: `=TINV(C6;C7)`.

The "Function Arguments" dialog box for the TINV function is open, showing the following arguments:

- Probability:** C6 (value: 0,05)
- Deg_freedom:** C7 (value: 19)

The dialog box also displays the result: `=2,093024705`.

Below the arguments, a description of the TINV function is provided: "Returns the inverse of the Student's t-distribution. **Probability** is the probability associated with the two-tailed Student's t-distribution, a number between 0 and 1 inclusive."

The "Formula result =" field shows the calculated value: 2,093024705.

Buttons for "OK" and "Cancel" are visible at the bottom of the dialog box.

Determinati intervalul de incredere pentru media populatiei tinta (nivelul de creatinina serica) cu un nivel de siguranta de 95%.

gradul de libertate: $df = n-1$

$t = 2, 1448$

$$CI = 0,94 \pm 0,08$$

The screenshot shows the Excel 'Function Arguments' dialog box for the TINV function. The formula bar at the top displays '=TINV(C23:C21)'. The dialog box has a blue header with the text 'Function Arguments' and a question mark icon. The 'Probability' field is set to C23, and the 'Deg_freedom' field is set to C21. The calculated result is 2,144788596. The dialog includes 'OK' and 'Cancel' buttons at the bottom.