

Κατανομές

1. Ομοιόμορφη κατανομή

Η κατανομή (**distribution**) μιας μεταβλητής (**variable**) φαίνεται από το σχήμα του ιστογράμματος (**histogram**).

Ας κάνουμε ένα απλό πείραμα (μπορεί να γίνει και στην τάξη).

Έστω ότι ρίχνουμε ένα ζάρι. Το αποτέλεσμα είναι 1, 2, 3, 4, 5 ή 6.

Έστω ότι ρίχνουμε ένα τέτοιο ζάρι 10 φορές. Ή ότι ρίχνουν ένα ζάρι 10 φοιτητές.

Θα μπορούσαμε να πάρουμε αποτελέσματα παρόμοια με τα παρακάτω:

αύξων αριθμός	αποτέλεσμα ρίψης ζαριού
1	5
2	5
3	1
4	4
5	5
6	5
7	6
8	5
9	6
10	6

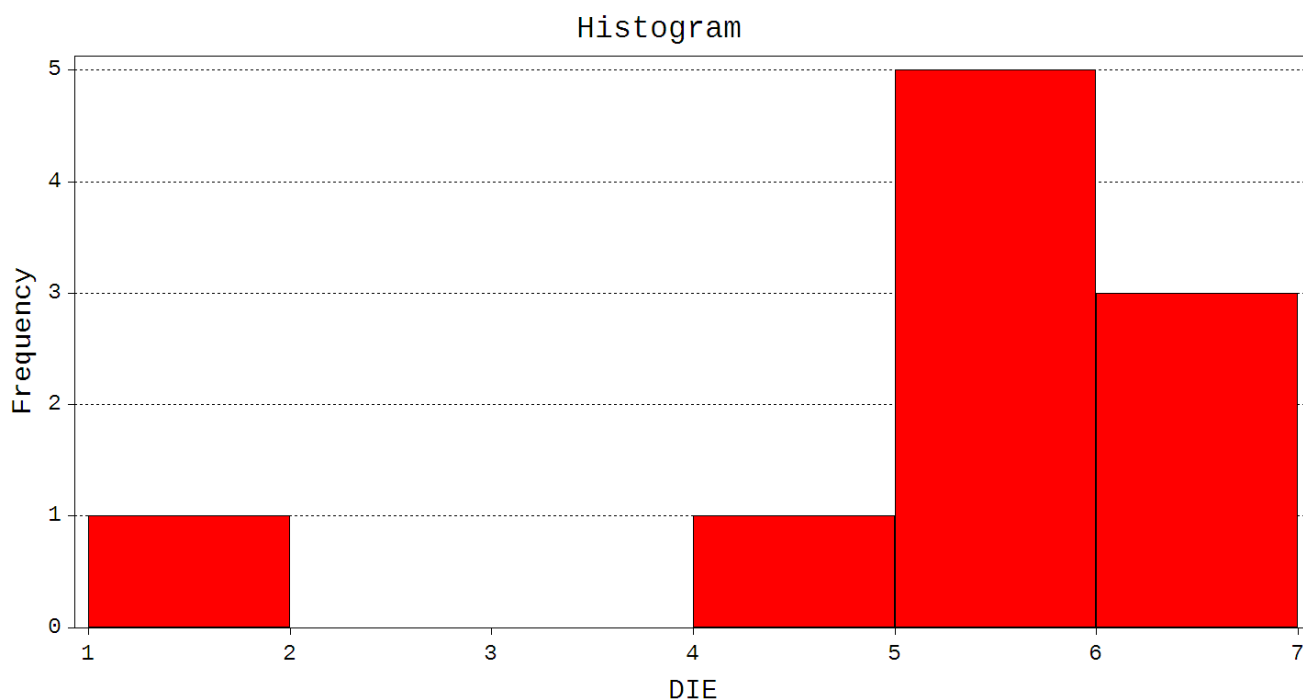
Η κατανομή συχνοτήτων των ανωτέρω αποτελεσμάτων ρίψης ζαριού είναι

Frequency Distribution of DIE

Value	Freq	Percent	Cumulative	
			Freq	Percent
1	1	10.0	1	10.0
4	1	10.0	2	20.0
5	5	50.0	7	70.0
6	3	30.0	10	100.0
Total	10	100.0		

Παρατηρούμε ότι στις 10 ρίψεις δεν εμφανίστηκαν όλες οι δυνατές τιμές που μπορεί να φέρει ένα ζάρι: είχαμε τον αριθμό 1 μια φορά, τον αριθμό 4 μια φορά, τον αριθμό 5 πέντε φορές και τον αριθμό 6 τρεις φορές. Οι τιμές 2 και 3 δεν εμφανίστηκαν ούτε μια φορά.

Εάν ο διδάσκων καταχωρήσει τις τιμές σε ένα στατιστικό πακέτο και σχεδιάσει ιστόγραμμα, θα πάρει κάτι σαν το παρακάτω σχήμα.



Η κλάση [1-2) περιέχει ένα **1**, η κλάση [4-5) περιέχει ένα **4**, η κλάση [5-6) περιέχει πέντε **5** και η κλάση [6-7) περιέχει τρία **6**.

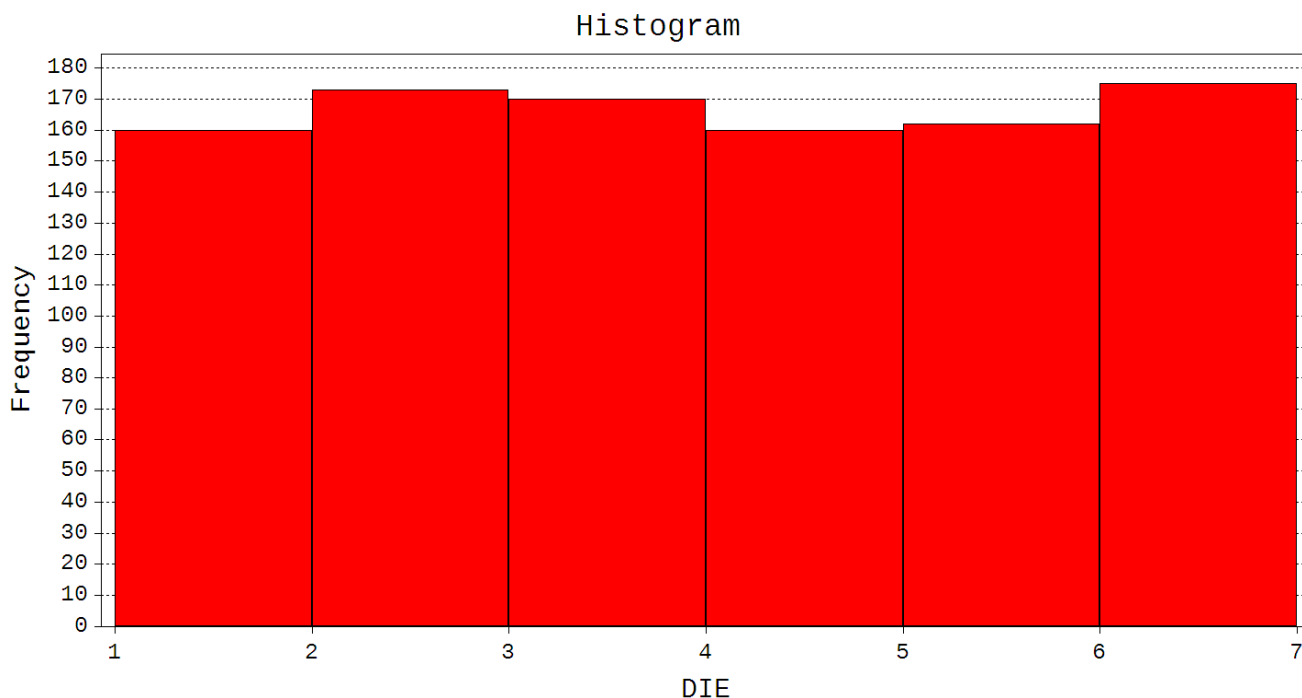
Με τόσο λίγες επαναλήψεις (10), δηλαδή με τόσο μικρό μέγεθος δείγματος δεν μπορούμε να διακρίνουμε το σχήμα της κατανομής.

Ας προσομοιάσουμε τώρα στην υπολογιστή (Excel) 1000 ρίψεις ζαριού. Η κατανομή συχνοτήτων αυτών των 1000 ρίψεων έχει ως ακολούθως:

Frequency Distribution of DIE

Value	Freq	Percent	Cumulative	
			Freq	Percent
1	160	16.0	160	16.0
2	173	17.3	333	33.3
3	170	17.0	503	50.3
4	160	16.0	663	66.3
5	162	16.2	825	82.5
6	175	17.5	1000	100.0
Total	1000	100.0		

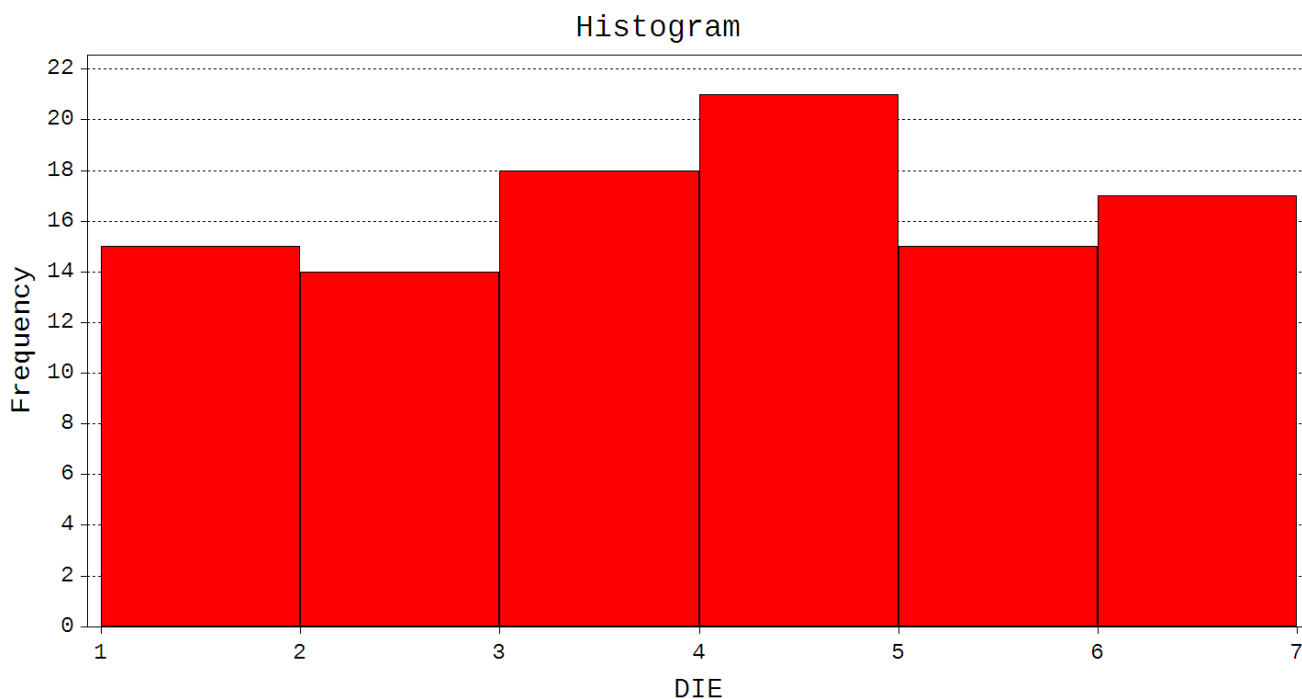
Το αντίστοιχο ιστόγραμμα είναι



Με τόσες πολλές επαναλήψεις του πειράματος ([experiment](#)) ρίψης του ζαριού, έχουν πλέον γεμίσει όλες οι τιμές και το ιστόγραμμα αποκτά την όψη της ομοιόμορφης κατανομής ([uniform distribution](#)).

Όταν μια μεταβλητή ακολουθεί την ομοιόμορφη κατανομή, οι τιμές της είναι κατανομημένες ομοιόμορφα σε όλο το εύρος ([range](#)) που καταλαμβάνουν. Επομένως, οι μπάρες του ιστογράμματος θα έχουν (περίπου) το ίδιο ύψος.

Γίνεται εύκολα αντιληπτό ότι για ενδιάμεσα μεγέθη δείγματος, θα είχαμε μια κατανομή που θα πλησίαζε το σχήμα της ομοιόμορφης αλλά όχι τόσο καλά όσο το μεγαλύτερο δείγμα. Για 100 ρίψεις, για παράδειγμα, το ιστόγραμμα είναι



2. Κανονική κατανομή

Εκτός από την ομοιόμορφη κατανομή ([uniform distribution](#)) είναι σημαντικό να εξετάσουμε την κανονική κατανομή ([normal distribution](#)).

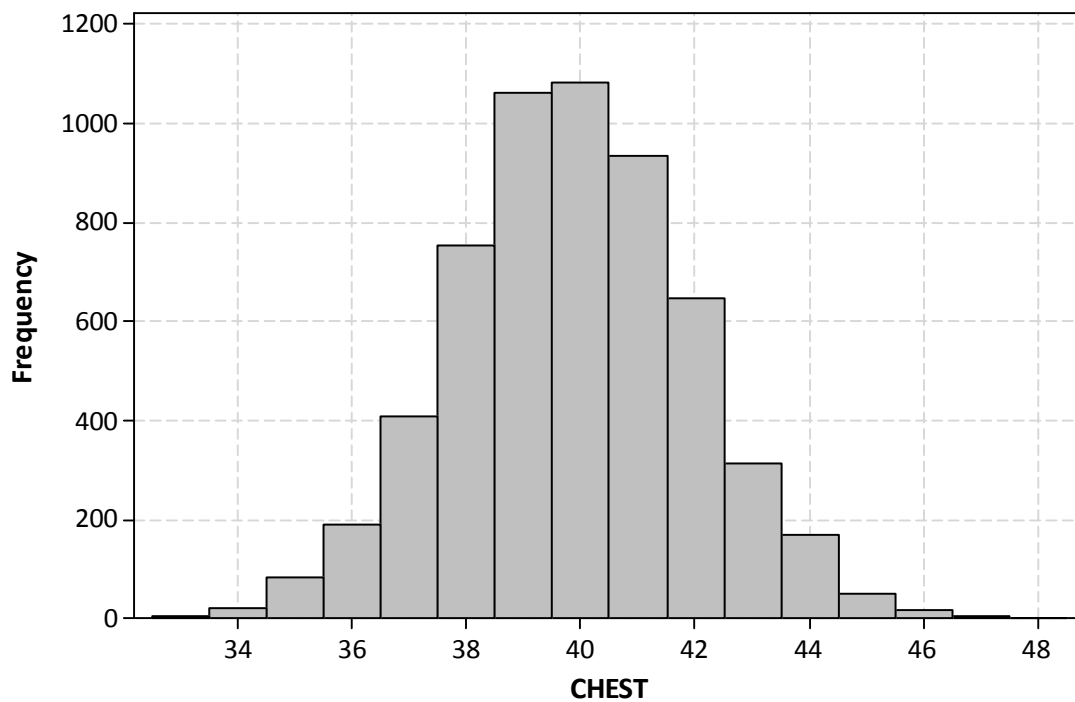
Ενώ μπορούμε να παράγουμε εύκολα αριθμούς κατανεμημένους ομοιόμορφα, π.χ. με τη ρίψη ενός ζαριού, δεν είναι το ίδιο εύκολο να παράγουμε αριθμούς κατανεμημένους κανονικά.

Πολλές φυσικές ιδιότητες όπως το ύψος, το βάρος, η αρτηριακή πίεση ή ο δείκτης ευφυΐας είναι κατανεμημένοι ομοιόμορφα. Για να μπορέσουμε όμως να δούμε το σχήμα της κατανομής πρέπει να μαζέψουμε αρκετές παρατηρήσεις, δηλαδή να έχουμε ένα δείγμα μεσαίου ή μεγάλου μεγέθους.

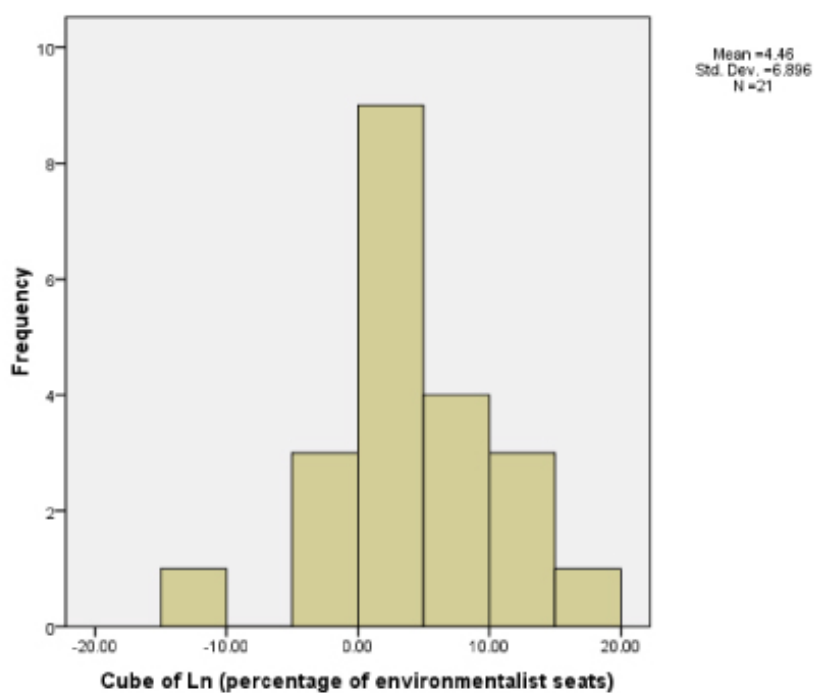
Για παράδειγμα, ο κατωτέρω πίνακας δείχνει την περιφέρεια θώρακα Σκωτσέζων στρατιωτών του έτους 1817:

CHEST	FREQ
33	3
34	19
35	81
36	189
37	409
38	753
39	1062
40	1082
41	935
42	646
43	313
44	168
45	50
46	18
47	3
48	1

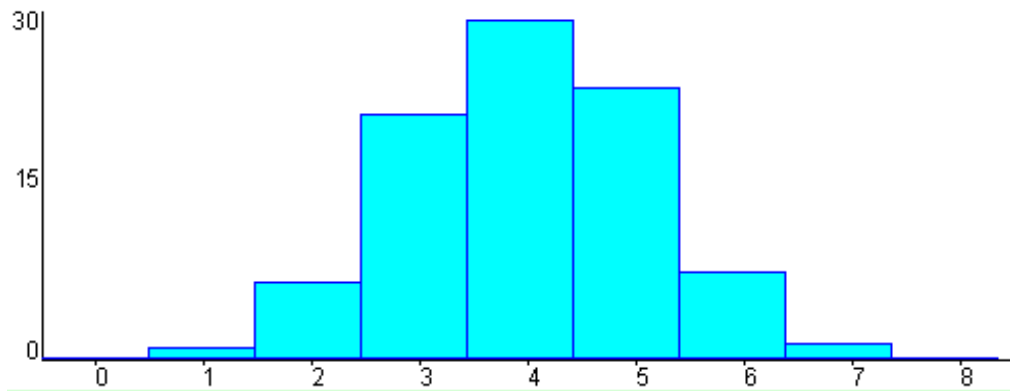
Το αντίστοιχο ιστόγραμμα είναι



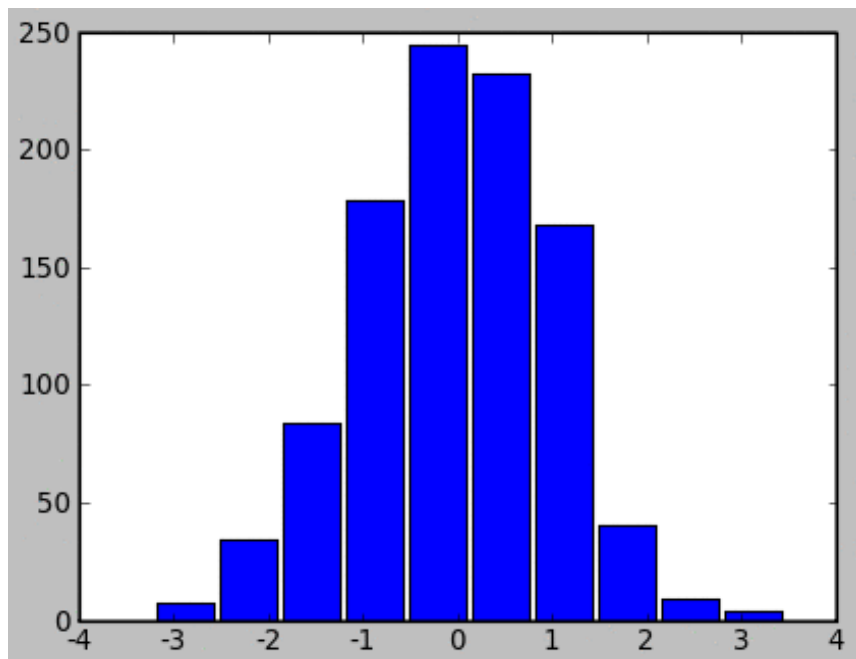
Όταν το δείγμα είναι μικρό, δεν θα μπορούμε να είμαστε σίγουροι από την όψη του ιστογράμματος ότι το δείγμα είναι κατανεμημένο κανονικά.

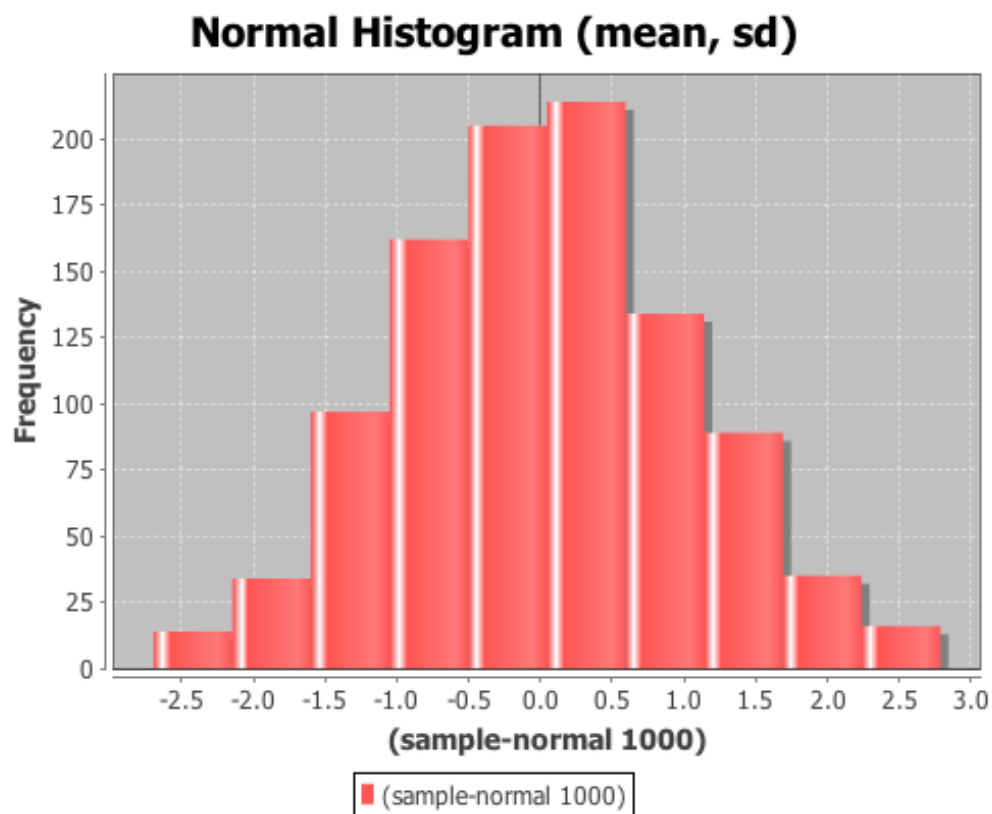


Να ένα ιστόγραμμα δείγματος μεσαίου μεγέθους, που πλησιάζει περισσότερο το σχήμα της κανονικής κατανομής:

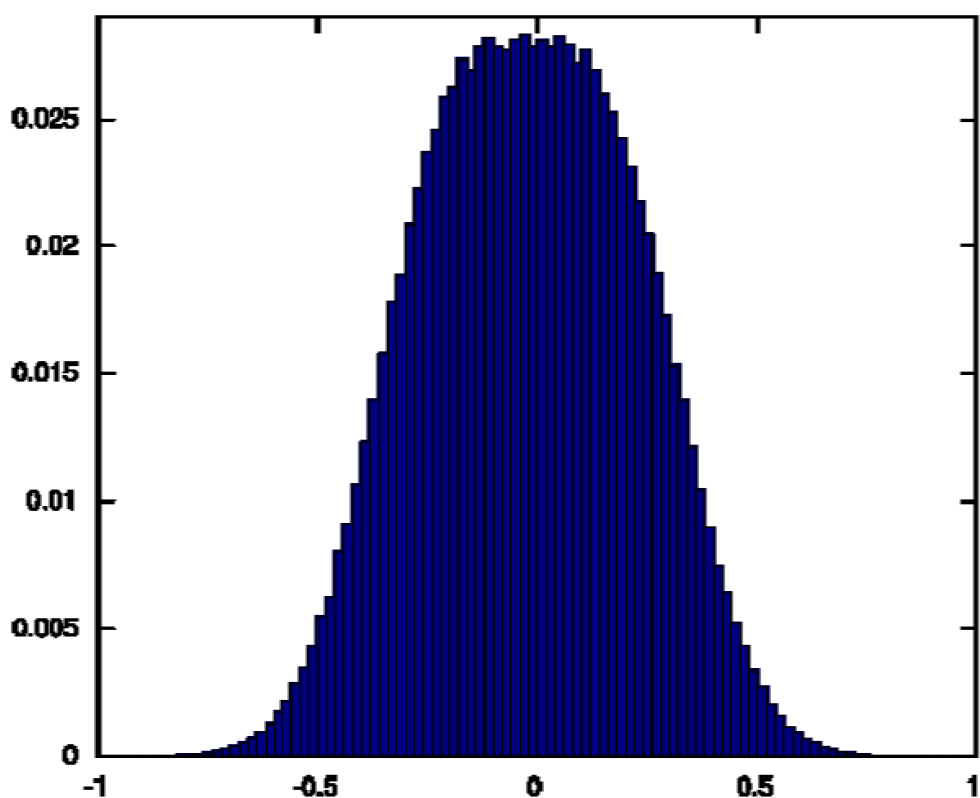


Να και δυο ιστογράμματα μεγαλύτερων δειγμάτων, τα οποία είναι εμφανώς κατανεμημένα κανονικά:



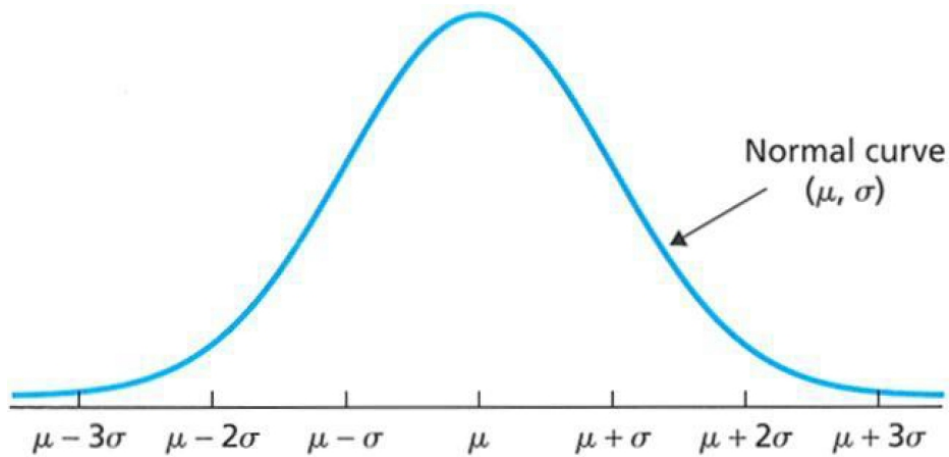


Τέλος, να και ένα ιστόγραμμα που είναι κατασκευασμένο από πάρα πολλές παρατηρήσεις και απεικονίζει ένα σχεδόν τέλειο σχήμα κανονικής κατανομής:



2.1. Τυπική κανονική κατανομή

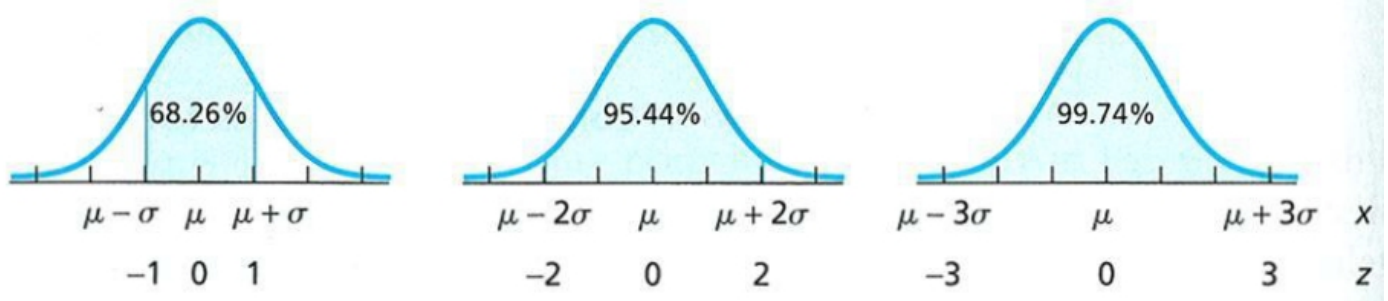
Η κανονική κατανομή έχει το κλασικό σχήμα της καμπάνας. Η μαθηματική συνάρτηση που την εκφράζει περιέχει δυο παραμέτρους: την μέση τιμή (μ) και την τυπική απόκλιση (σ).



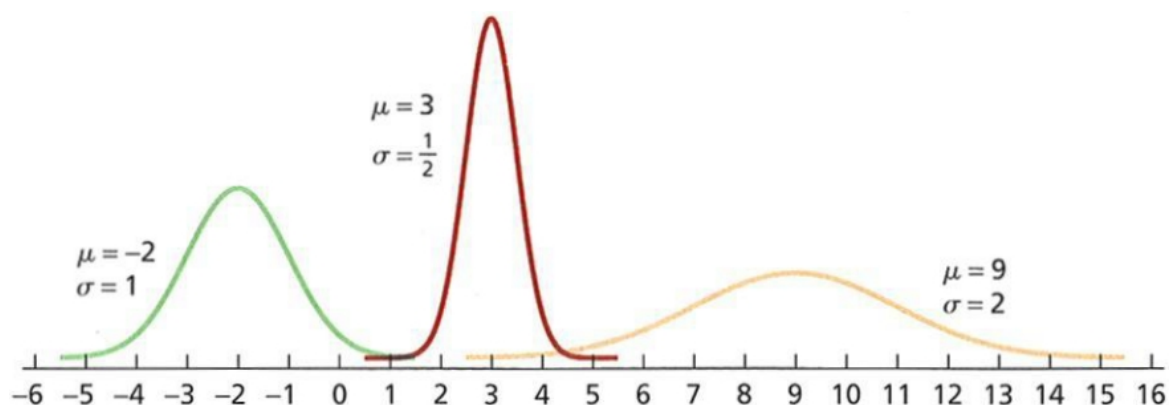
Αν και η ελάχιστη τιμή της κανονικής καμπύλης είναι το μείον άπειρο ($-\infty$) και η μέγιστη το συν άπειρο ($+\infty$), το σύνολο σχεδόν του εμβαδού ανάμεσα στην καμπύλη και τον οριζόντιο άξονα βρίσκεται από το $\mu - 3\sigma$ έως το $\mu + 3\sigma$, δηλαδή στο διάστημα από τη μέση τιμή μείον 3 τυπικές αποκλίσεις έως τη μέση τιμή συν 3 τυπικές αποκλίσεις.

Άλλα χαρακτηριστικά σημεία που έχουν ιδιαίτερο ενδιαφέρον είναι τα $\mu \pm \sigma$, $\mu \pm 2\sigma$ και $\mu \pm 3\sigma$.

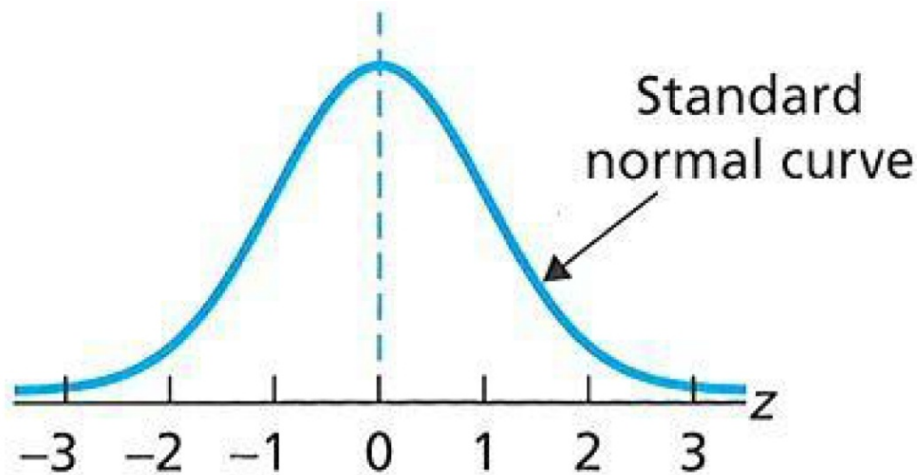
Το εμβαδόν ανάμεσα σε αυτά τα σημεία ορίζεται από τον λεγόμενο εμπειρικό κανόνα ([empirical rule](#)):



Ανάλογα με τις τιμές της μέσης τιμής (μ) και της τυπικής απόκλισης (σ), η κανονική καμπύλη μπορεί να μετακινηθεί αριστερά ή δεξιά καθώς και να γίνει πιο στενή ή πιο φαρδιά.



Έτσι, για να μπορούμε να κάνουμε χρήση της κανονικής καμπύλης σε πρακτικά προβλήματα, αναφερόμαστε πάντα σε μια συγκεκριμένη κανονική καμπύλη που έχει μέση τιμή μηδέν ($\mu = 0$) και τυπική απόκλιση ένα ($\sigma = 1$) και ονομάζεται τυπική (ή μοναδιαία) κανονική καμπύλη ([standard normal curve](#)).



Οι τιμές της κατανομής αυτής απεικονίζονται σε πίνακες όπως αυτόν στο τέλος αυτού του φυλλαδίου.

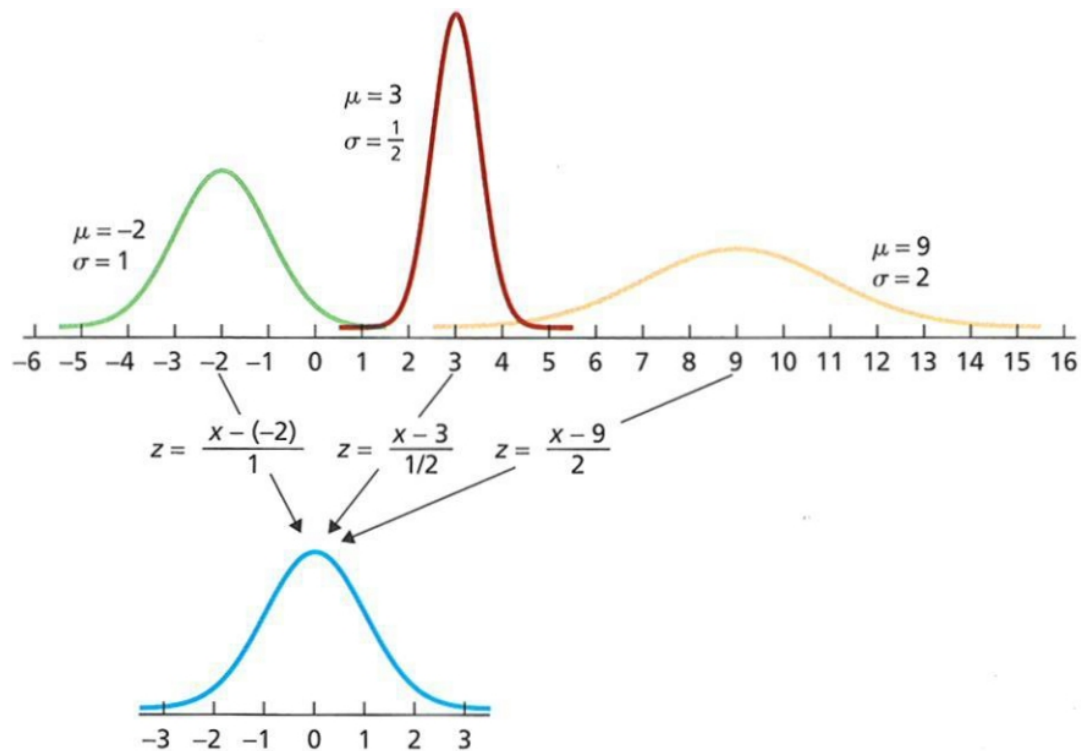
Για να μετατρέψουμε τις πρωτογενείς τιμές ενός συνόλου δεδομένων σε αριθμούς συγκρίσιμους αυτούς του πίνακα της τυπικής κανονικής κατανομής, κάνουμε τον ακόλουθο μετασχηματισμό:

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

όπου z είναι η συντεταγμένη του οριζόντιου άξονα, x είναι το πρωτογενές δεδομένο (που θέλουμε να φέρουμε στην κλίμακα του πίνακα), μ είναι η μέση τιμή και σ είναι η τυπική απόκλιση των δεδομένων.

Φυσικά, η μετατροπή κάποιων δεδομένων σε τιμές συγκρίσιμες με την τυπική κανονική κατανομή έχει νόημα μόνο εάν τα δεδομένα είναι κατανεμημένα κανονικά.

Στο επόμενο σχήμα φαίνονται τρία παραδείγματα μετατροπής κανονικών καμπυλών (με διάφορες τιμές μ και σ) στην τυπική κανονική κατανομή.



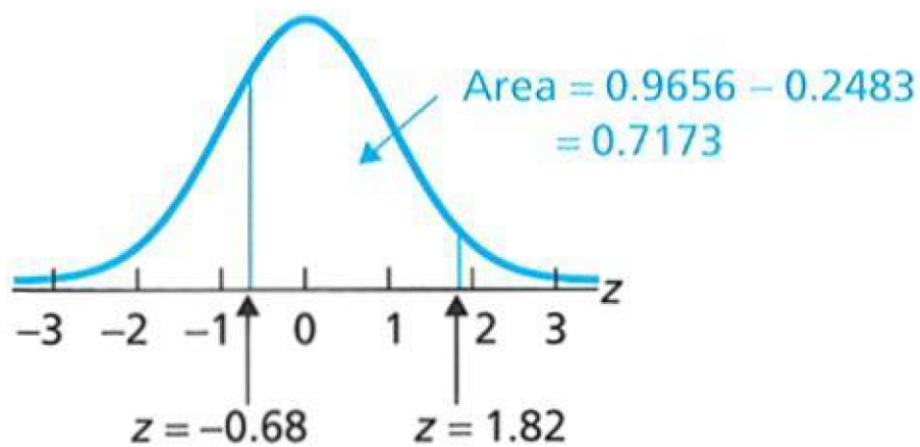
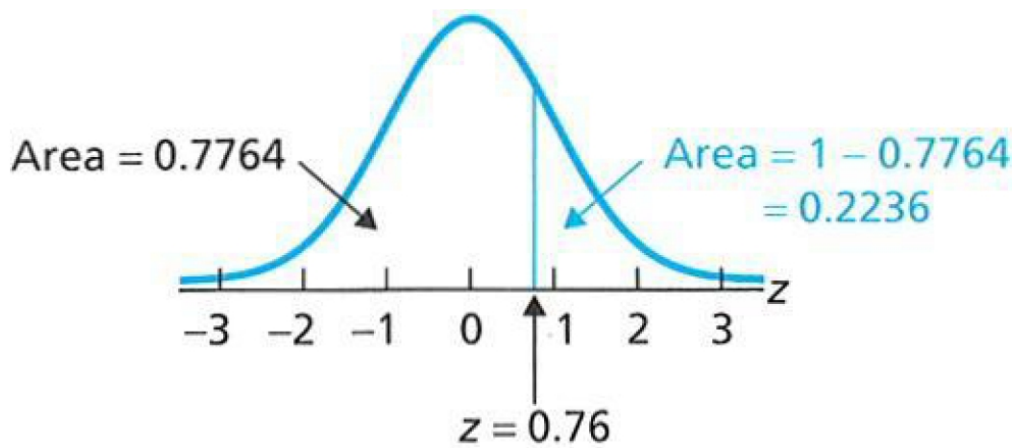
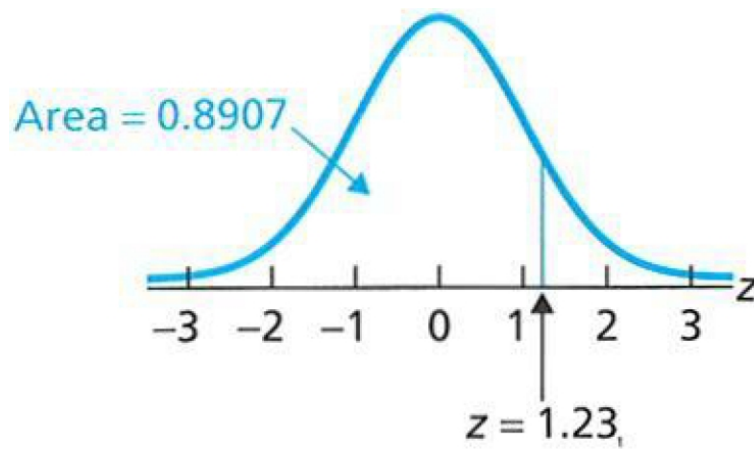
Οι πίνακες της τυπικής κανονικής κατανομής δίνουν το εμβαδόν μεταξύ της καμπύλης και του οριζόντιου άξονα που ευρίσκεται π.χ. αριστερά της τιμής του z εκφρασμένης σε 2 δεκαδικά ψηφία. Το συνολικό εμβαδόν από το $-\infty$ έως το $+\infty$ ισούται με 1.

Διαφορετικοί πίνακες μπορεί να δίνουν διαφορετικά εμβαδά σε σχέση με τη θέση του z .

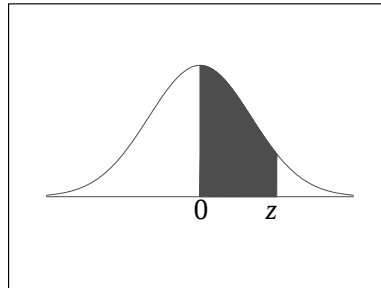
Στο παρακάτω παράδειγμα, στην τιμή $z = -1.75$ αντιστοιχεί τιμή εμβαδού 0.0401. Δηλαδή από το $z = -\infty$ έως το $z = -1.75$ υπάρχει εμβαδόν ίσο με 0.0401, δηλαδή 4.01% του συνόλου.

Second decimal place in z										z
0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.
0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287	-1.9
0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359	-1.8
0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446	-1.7
0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548	-1.6
0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668	-1.5

Στα επόμενα σχήματα φαίνονται παραδείγματα υπολογισμού διάφορων εμβαδών από τιμές του z και το ανάποδο.



Standard Normal Distribution Table



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998