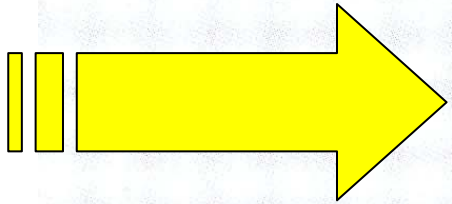
The normal distribution

The normal distribution, which may already be familiar to someone as the curve with the bell shape, is sometimes associated with the name of **Pierre Laplace** and **Carl Gauss**, who figured prominently in its historical development.

Gauss derived the normal distribution mathematically as the probability distribution of the error of measurements, which he called the **"normal laws of errors"**.

Subsequently many researchers in a wide variety of fields found that **their histograms** exhibited the common feature of first rising gradually in height to a maximum and then decreasing in a symmetric manner.

A normal distribution has a bell-shaped density



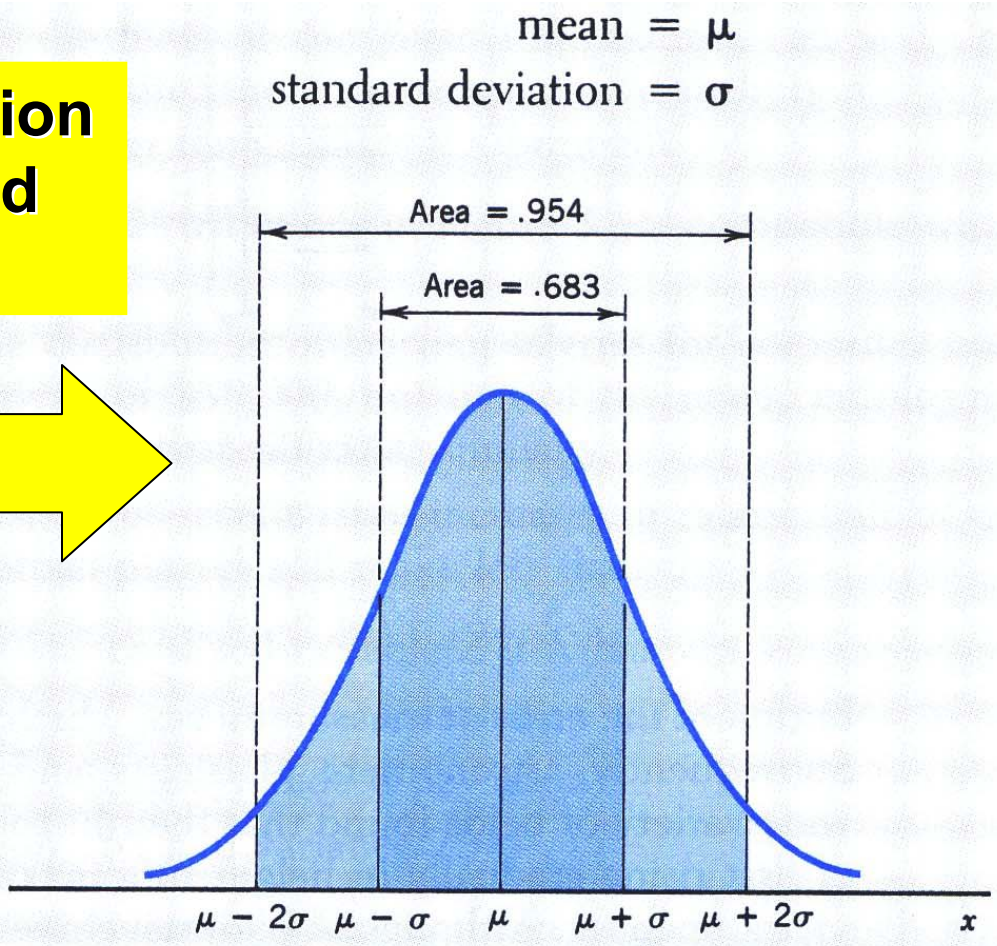


Figure 5 Normal distribution.

The probability of the interval extending

one sd on each side of the mean: $P[\mu - \sigma < X < \mu + \sigma] = .683$

two sd on each side of the mean: $P[\mu - 2\sigma < X < \mu + 2\sigma] = .954$

three sd on each side of the mean: $P[\mu - 3\sigma < X < \mu + 3\sigma] = .997$

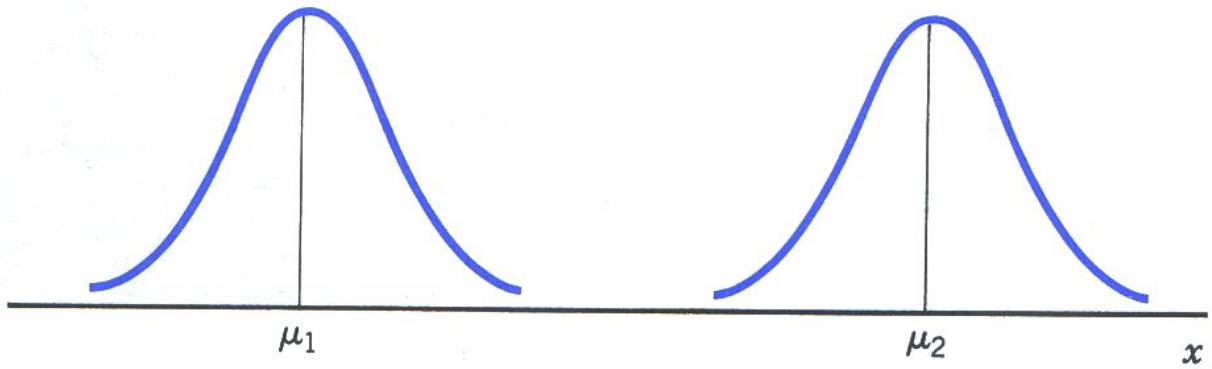
The formula, which need not concern us, is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{for} \qquad -\infty < x < \infty$$

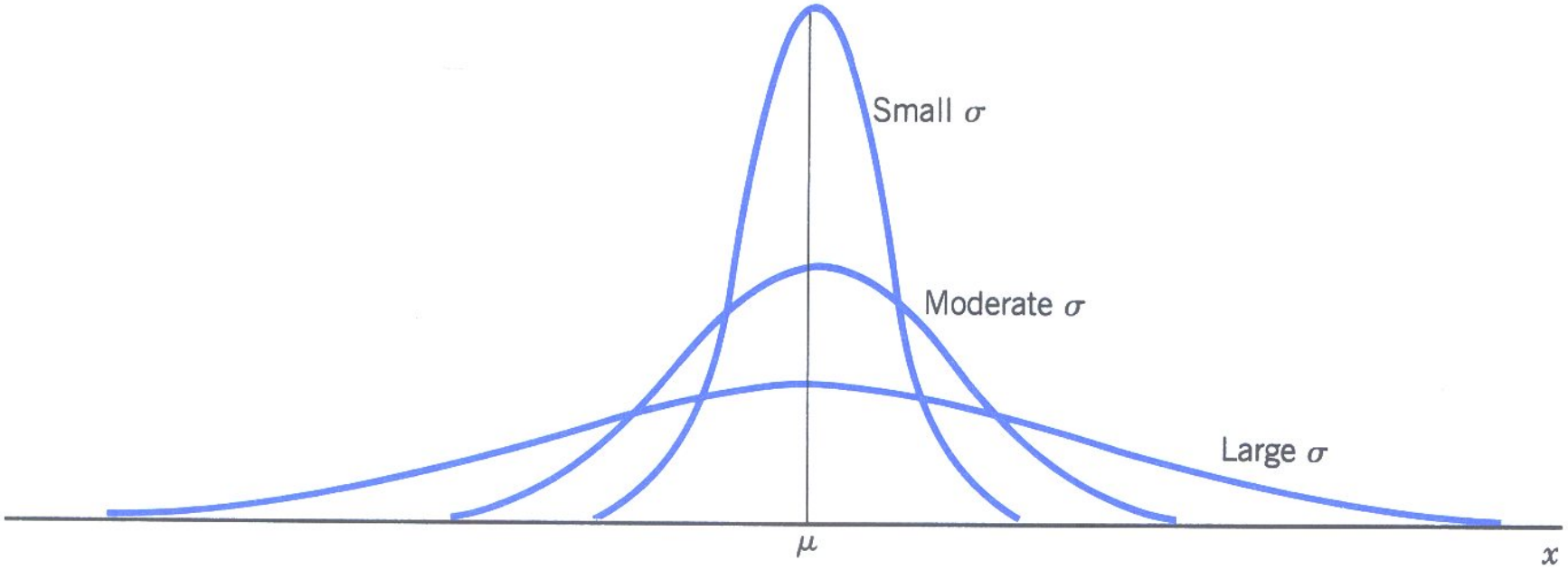
where π is the area of a circle having unit radius, or approximately 3.1416, and e is approximately 2.7183.

Notation

The normal distribution with a mean of μ and a standard deviation of σ is denoted by $N(\mu, \sigma)$.



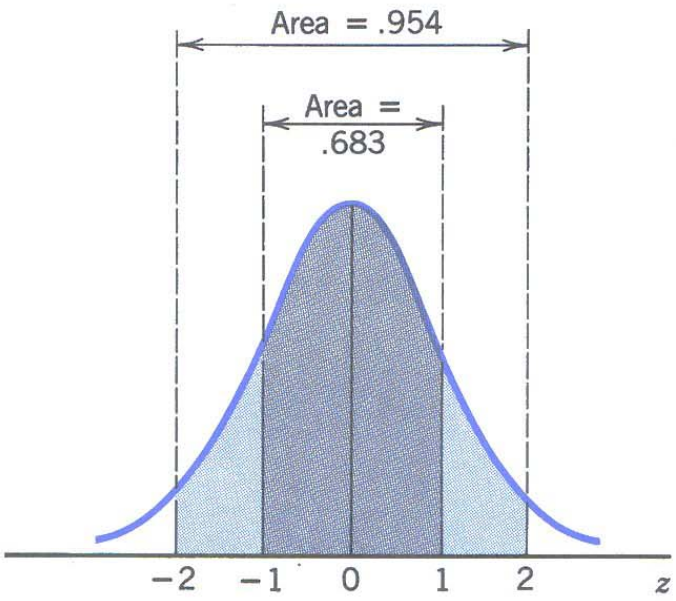
Two normal distributions with different means, but the same standard deviation.



Decreasing σ increases the maximum height and the concentration of probability about μ.

However, around one, two or three standard deviation on each side of the mean we have the **same probability** to find data.

The standard normal distribution Z



The standard normal curve.

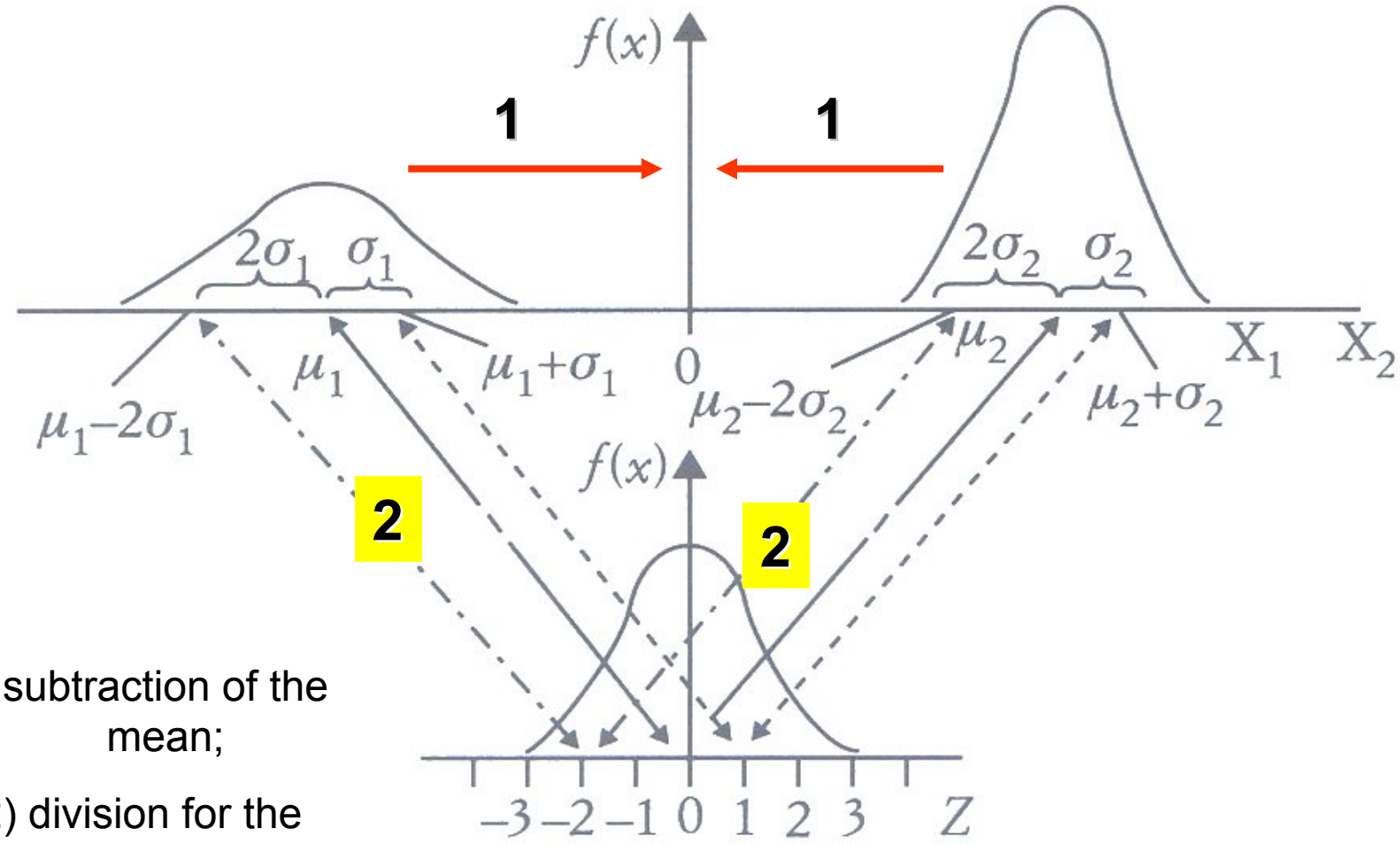
The **standard normal distribution** has a bell-shaped density with

$$\text{mean } \mu = 0$$

$$\text{standard deviation } \sigma = 1$$

The standard normal distribution is denoted by $N(0, 1)$.

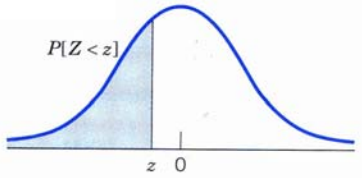
The process of standardization



(1) subtraction of the mean;

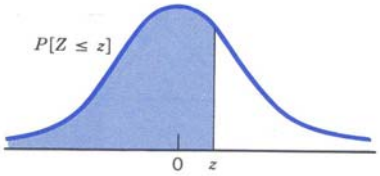
(2) division for the standard deviation as the natural unit of measurement.

Standard normal probabilities table



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−.7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
−.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Standard normal probabilities table (continued)



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

Use of the standard Normal table

The standard normal table gives the area to the left of a specified value of z as:

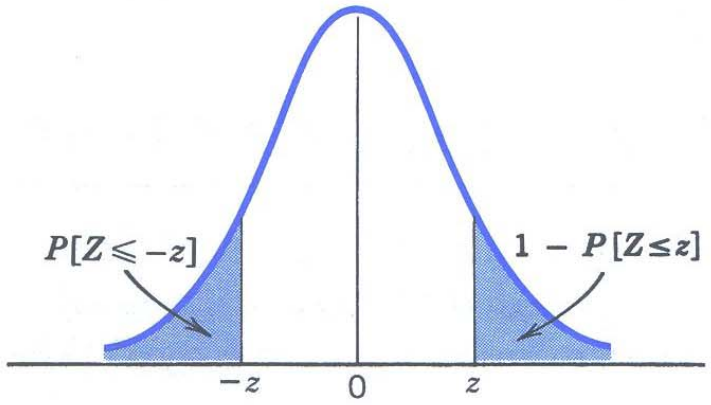
$$P[Z \leq z] = \text{Area under curve to the left of } z$$

For the probability of an interval $[a, b]$,

$$P[a \leq Z \leq b] = [\text{Area to left of } b] - [\text{Area to left of } a]$$

The following properties can be observed from the symmetry of the standard normal curve about 0 as exhibited in Figure

1. $P[Z \leq 0] = .5$
2. $P[Z \leq -z] = 1 - P[Z \leq z] = P[Z \geq z]$



Equal normal tail probabilities.

Example Find $P[Z \leq 1.37]$ and $P[Z > 1.37]$.

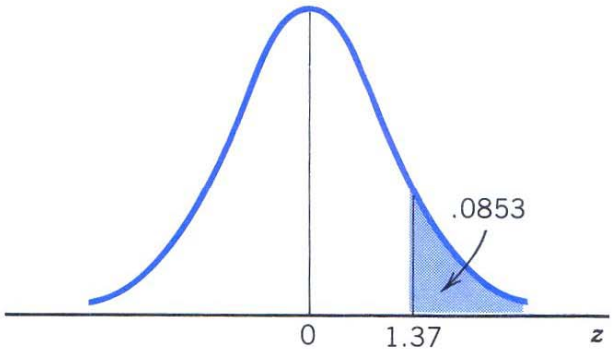
SOLUTION From the normal table, we see that the probability or area to the left of 1.37 is .9147. (See Table) Consequently, $P[Z \leq 1.37] = .9147$. Moreover, because $[Z > 1.37]$ is the complement of $[Z \leq 1.37]$,

$$P[Z > 1.37] = 1 - P[Z \leq 1.37] = 1 - .9147 = .0853$$

as we can see in Figure . An alternative method is to use symmetry to show that $P[Z > 1.37] = P[Z < -1.37]$, which can be obtained directly from the normal table.

TABLE 1 How to Read

z	.00	⋯	.07	⋯
.0				
.				
.				
.				
1.3	----------		.9147	
.				
.				
.				

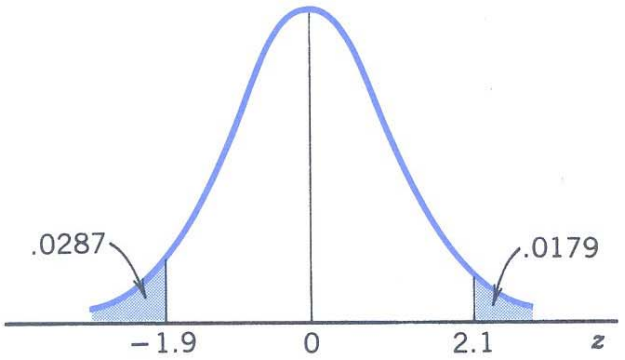


An upper tail normal probability.

Example Find $P[Z < -1.9 \text{ or } Z > 2.1]$.

SOLUTION The two events $[Z < -1.9]$ and $[Z > 2.1]$ are incompatible, so we add their probabilities:

$$P[Z < -1.9 \text{ or } Z > 2.1] = P[Z < -1.9] + P[Z > 2.1]$$



Normal probabilities

As indicated in Figure , $P[Z > 2.1]$ is the area to the right of 2.1, which is $1 - [\text{Area to left of } 2.1] = 1 - .9821 = .0179$. The normal table gives $P[Z < -1.9] = .0287$ directly. Adding these two quantities, we get

$$P[Z < -1.9 \text{ or } Z > 2.1] = .0287 + .0179 = .0466$$

PROBABILITY CALCULATIONS WITH NORMAL DISTRIBUTIONS

Fortunately, no new tables are required for probability calculations regarding the general normal distribution. Any normal distribution can be set in correspondence to the standard normal by the following relation.

> If X is distributed as $N(\mu, \sigma)$, then the standardized variable
>
> $$Z = \frac{X - \mu}{\sigma}$$
>
> has the standard normal distribution.

This property of the normal distribution allows us to cast a probability problem concerning X into one concerning Z. To find the probability that X lies in a given interval, convert the interval to the z-scale and then calculate the probability by using the standard normal table

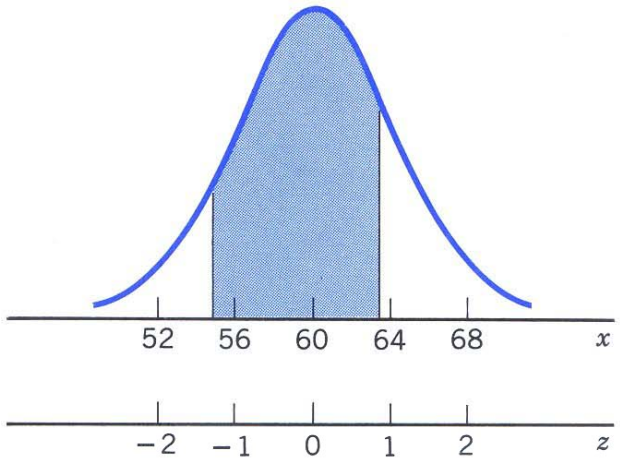
Example 6 Given that X has the normal distribution $N(60,4)$, find $P[55 \le X \le 63]$.

SOLUTIO

Here, the standardized variable is $Z = \dfrac{X - 60}{4}$.

$x = 55$ gives $z = \dfrac{55 - 60}{4} = -1.25$

$x = 63$ gives $z = \dfrac{63 - 60}{4} = .75$



Converting to the z-scale.

Therefore,

$$P[55 \le X \le 63] = P[-1.25 \le Z \le .75]$$

Using the normal table, we find $P[Z \le .75] = .7734$ and $P[Z \le -1.25] = .1056$ so the required probability is $.7734 - .1056 = .6678$.

In general:

If X is distributed as $N(\mu, \sigma)$, then

$$P[a \le X \le b] = P\left[\frac{a - \mu}{\sigma} \le Z \le \frac{b - \mu}{\sigma}\right]$$

where Z has the standard normal distribution.

Diffusion and dispersion of pollutants

When a **pollutant** is released into the environment, many diverse, unrelated forces act on it simultaneously. Because of the complexity of these processes, it is difficult to construct a single model for the movement, transformation, and fate of a pollutant.

To gain insight into these phenomena, it is preferable to consider several very simple models of one or more of the important processes at work in the environment.

Such models are idealized, but their purpose is to illustrate how the **statistical properties** of observed environmental concentrations come about.

diffusion

It is one of the most important processes that acts upon a pollutant released into the environment. In its simplest form, diffusion occurs when a molecule changes place with an adjacent molecule.

In the absence of outside mechanical forces, such movement will take place naturally due to the constant motion of the molecules of the pollutant and the material comprising the carrier medium.

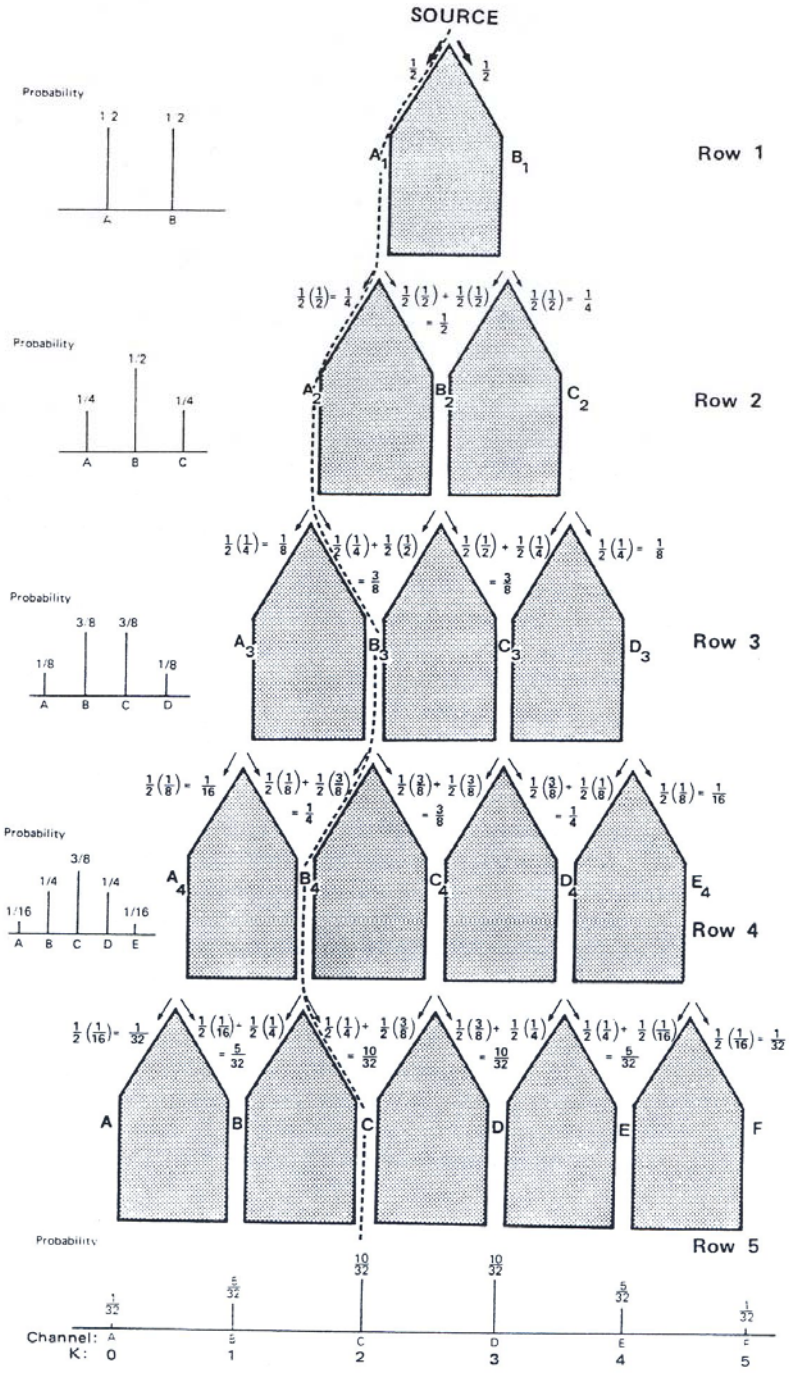
If the carrier medium moves as the diffusion occurs, causing the molecules of the pollutant to exhibit a predominant motion in a particular direction, the process is called **diffusion with drift**.

As diffusion processes become more complex, incorporating the effects of many real factors (winds, temperature changes, turbulence,), they sometimes are called **dispersion processes** in one, two or three dimensions.

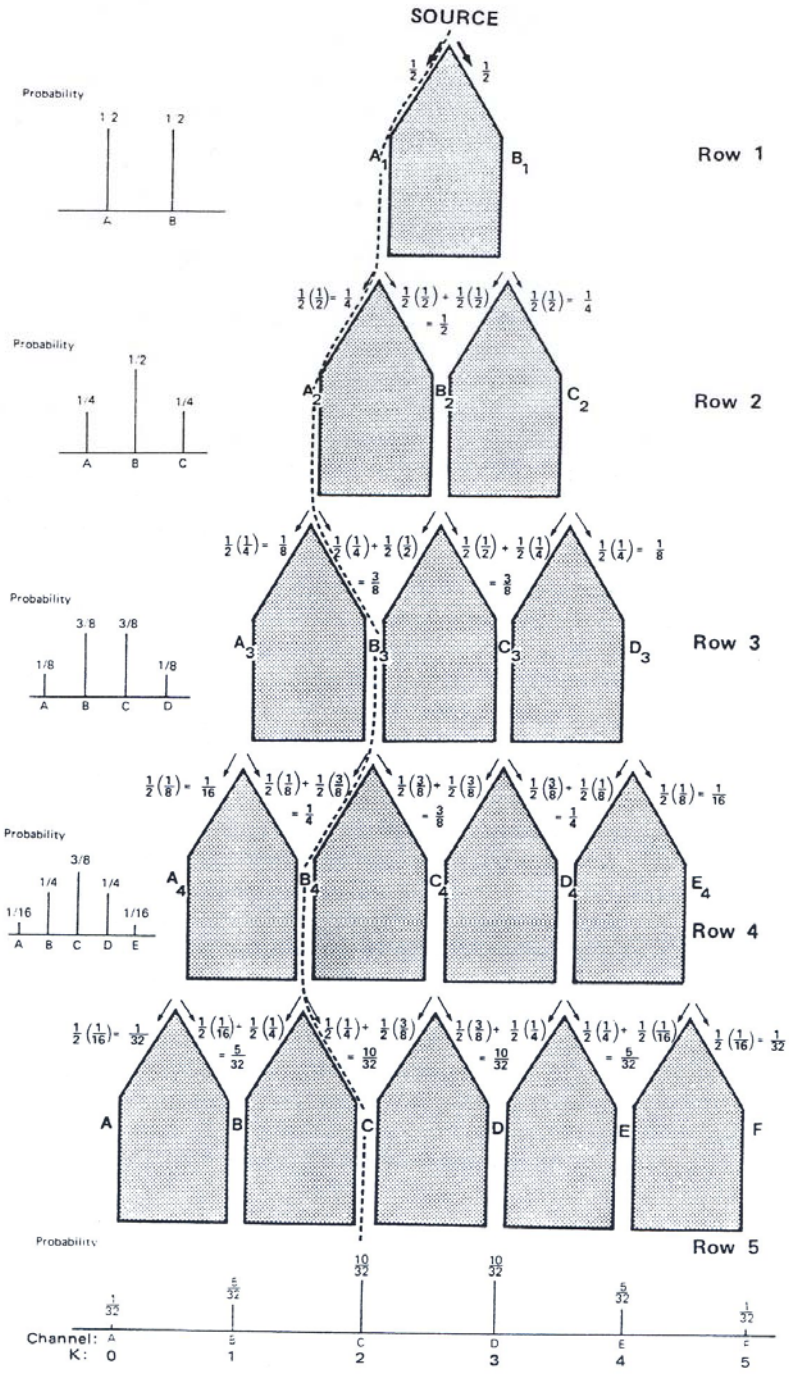
Consider a physical system consisting of an array of wedges in uniform rows. The wedges might be constructed of pieces of wood. Suppose that many small particles say, grains of sand, can be released from a “source” at the very top of the structure and that they eventually fall by gravity through the array to the bottom of the structure.

As we shall see, the movement of particles downward through the machine is a mechanical analogue of the **dispersion** and **diffusion** of pollutants in the environment.

Which type of probability distribution with respect to space will arise naturally from this mechanical analogue?



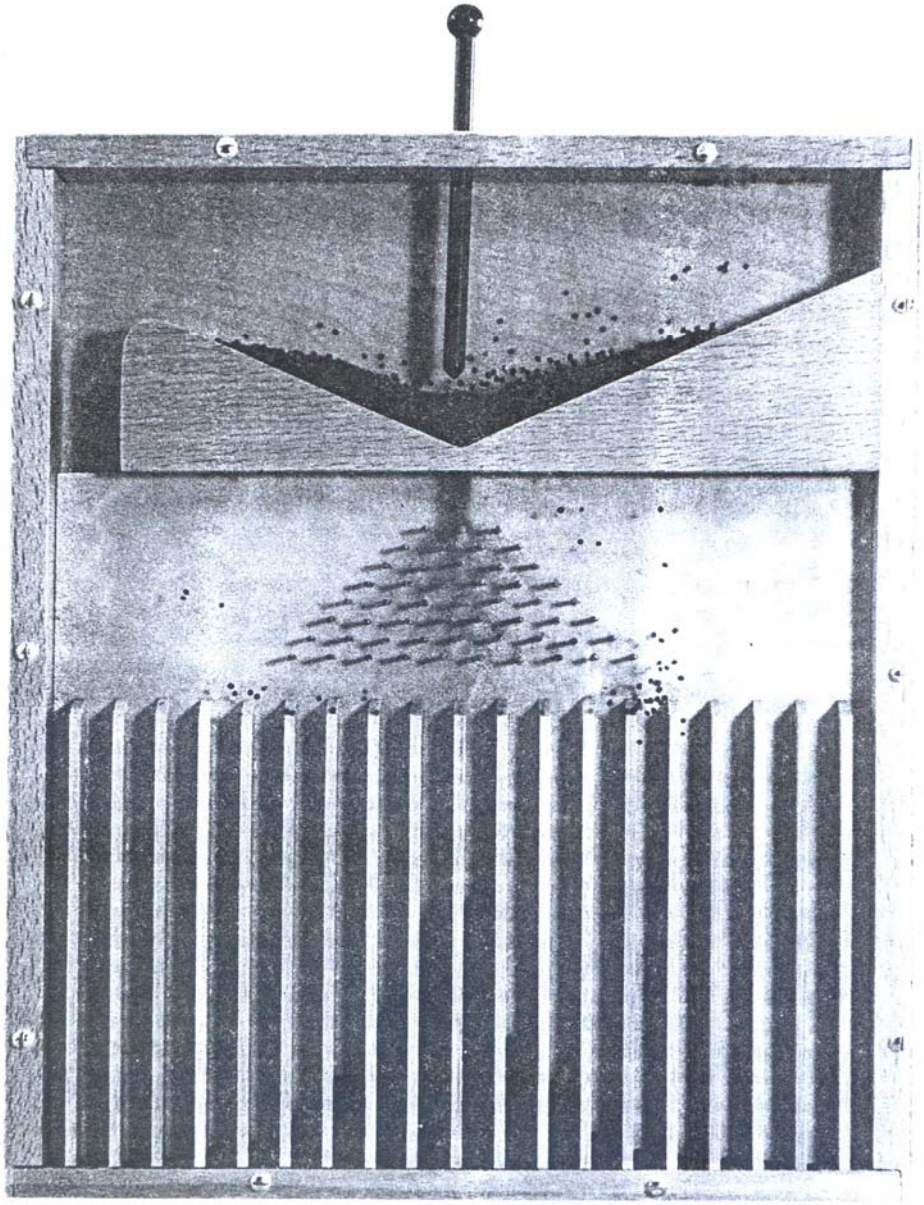
If one considers the probability distribution of particle arrivals with respect to space (distance from the midpoint of any row), the result is a **symmetrical binomial distribution** (p = ½), causing the expected number of particles arrivals to be symmetrical also.



The Normal Approximation to the Binomial

When np and $n(1 - p)$ are both large, say, greater than 15, the binomial distribution is well approximated by the normal distribution having mean $= np$ and sd $= \sqrt{np(1 - p)}$. That is,

$$Z = \frac{X - np}{\sqrt{np(1 - p)}} \text{ is approximately } N(0, 1)$$

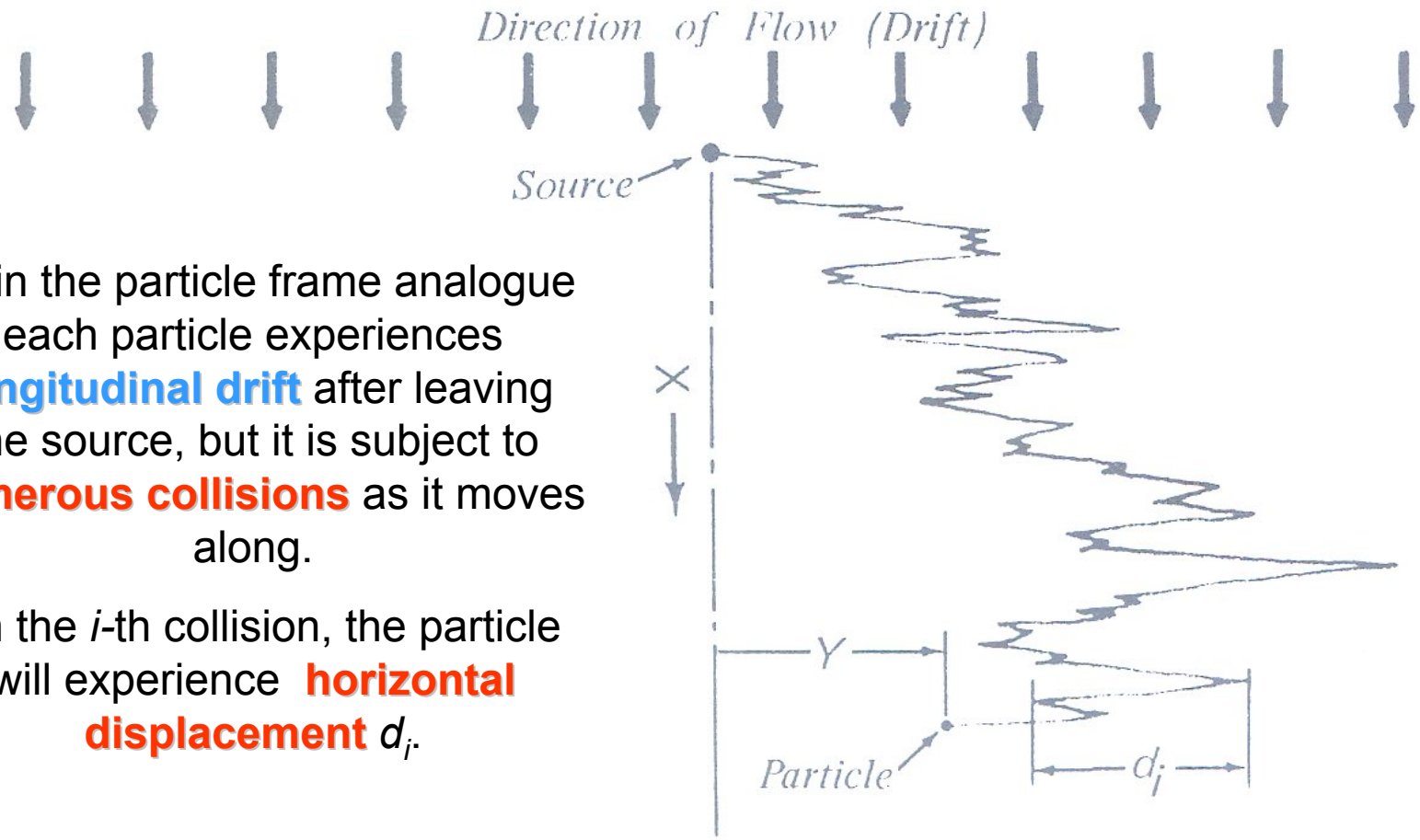


Particle frame analog of diffusion process showing reservoir with plunger released (top), particles spreading out as they fall through pin array (middle), and equally spaced collection columns with bell-shaped distribution emerging (bottom).

Suppose that the particles (or molecules) of a pollutant are released from a point source as a continuous stream into a moving carrier medium.

Instead of gravity which was responsible for the **drift** in the wedge machine and particle frame machine, assume that the drift now is caused by the **predominant motion of the carrier medium** at a constant speed and direction.

What will happen to the particles released into this moving carrier medium simplify the problem in two dimensions?

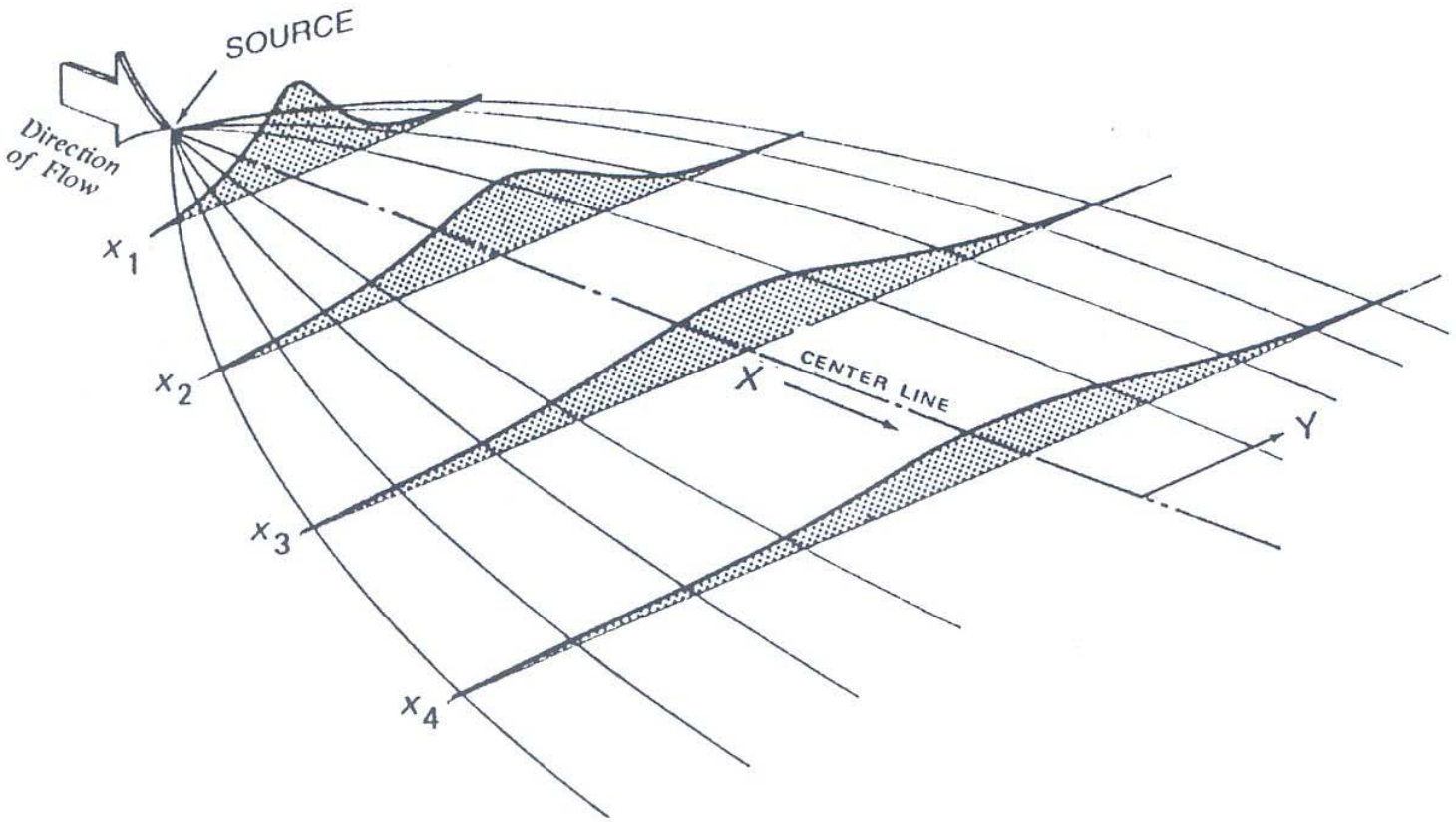


As in the particle frame analogue each particle experiences **longitudinal drift** after leaving the source, but it is subject to **numerous collisions** as it moves along.

On the *i*-th collision, the particle will experience **horizontal displacement** d_i.

The final position Y of the particle (or of many similar particles) after ***m*** collisions will asymptotically approach a normal distribution whose mean is at the center line and variance is proportional to ***m***.

If many particles were released at once from the source, then their "expected arrival density" (expected number of particles per unit length) will follow similar normal distributions, except that the quantities will be multiplied by the number of particles released. The expected number of particles arriving in a given segment of the *Y*-axis will be given by the area under the normal curve .



Probability distributions of the horizontal position of a particle subject to Brownian motion and drift. The vertical axis denotes probability, and the normal distributions occur at four different distances (and times) from the source.

Normal processes: are those that give rise to normally distributed random variables.

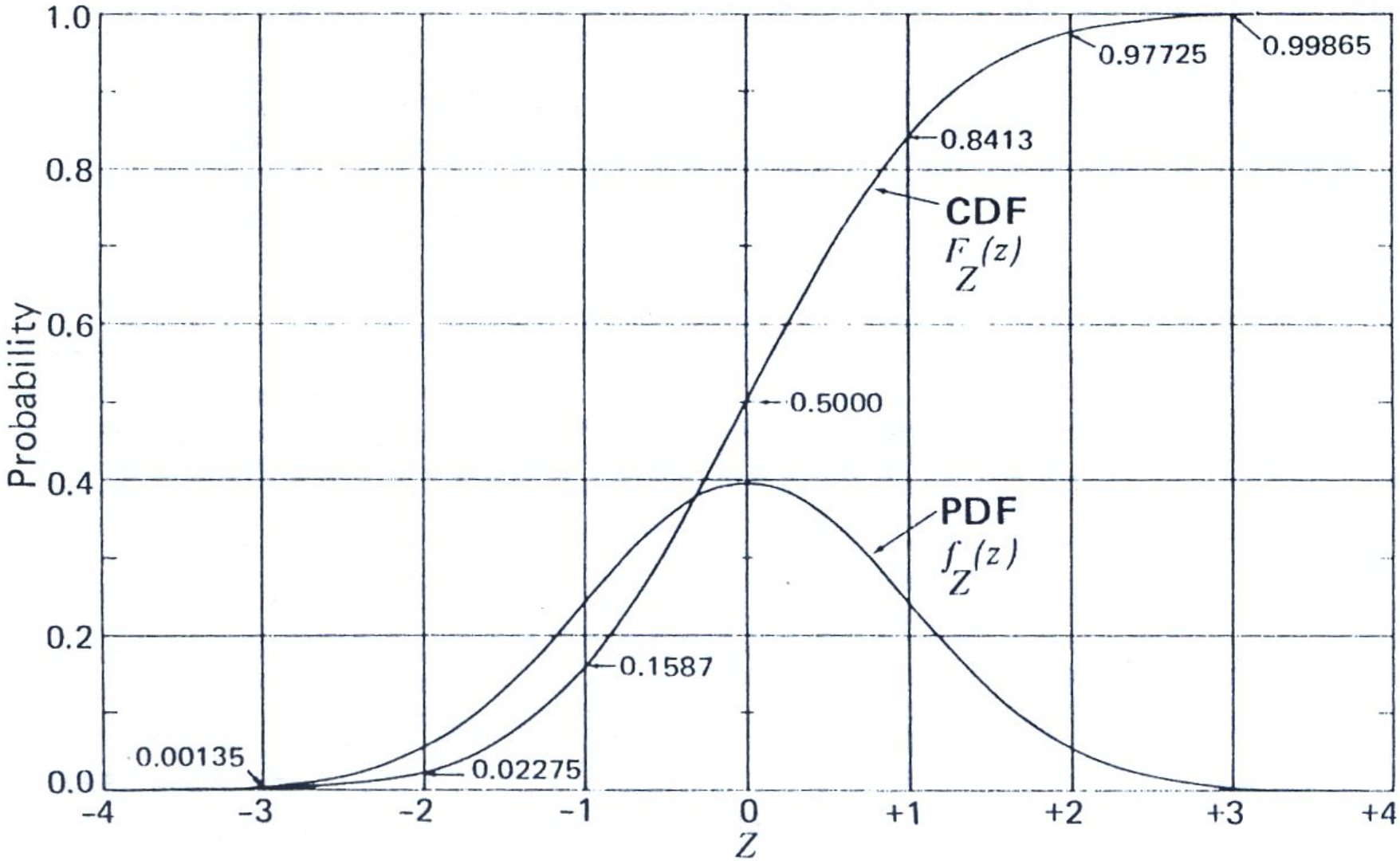
Normally distributed random variables tend to arise naturally when **many** continuous, **independent** random variables are **added** together.

Conditions for normal processes: many variables found in nature result form the summing of numerous unrelated components. When the individual components are sufficiently **unrelated** and **complex**, then the resulting sum tends toward normality as the number of components comprising the sum becomes increasingly large.

Two important conditions for normal processes are:

1) summation of many continuous random variables,
2) independence of these random variables.

In summary → a normal process (o **random-sum process**) result when a number of **unrelated**, **continuous random variables** are **added** together.



Probability distribution function (**PD**F) and cumulative distribution function (**CDF**) of the standardised normal distribution.

Checking the plausibility of a normal model

Does a normal model serve as a **reasonable model** for the population that produced the sample?

One reason for our interest in this question is that many commonly used statistical (inferential) procedures **require** the population to be nearly normal.

An effective way to check the plausibility of a normal model is to construct a special graph, called a **normal-scores plot** of the sample data.

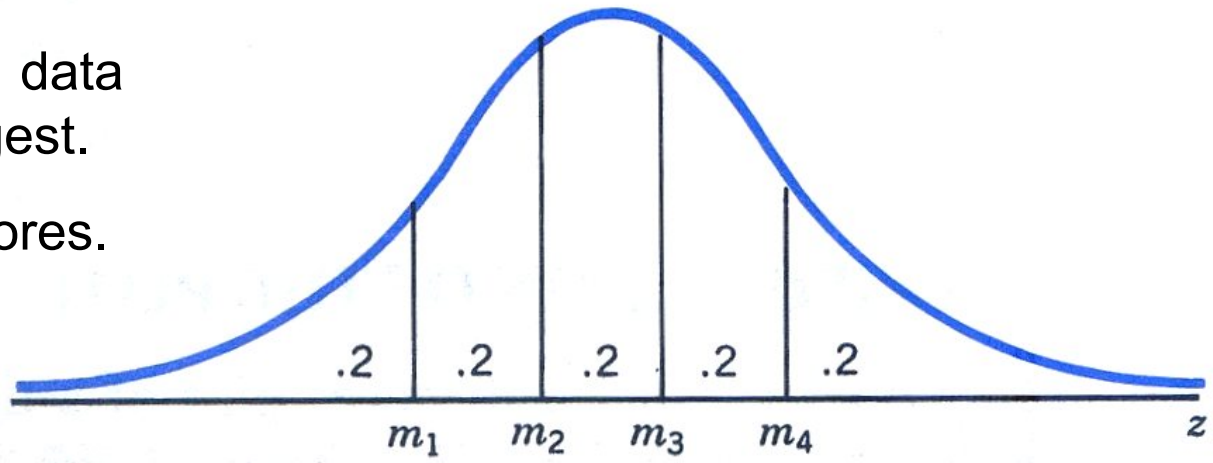
For an easy explanation of the ideas, we work with a small sample size. In practical applications, at least **15** or **20** observations are needed to detect a meaningful pattern in the plot.

The term normal scores refers to an idelised sample from the standard normal distribution, namely the z values that divide the standard normal distribution into equal-probability intervals.

Suppose the sample size is n = 4. The figure below shows the standard normal distribution where four points are located on the z-axis so the distribution is divided into five segments of equal probability 1/5 = 0.2. The four points are denoted m_1, m_2, m_3 and m_4. Thus:

1) Order the sample data from the smallest to largest.

2) Obtain the normal scores.



The $N(0, 1)$ distribution and the normal scores for $n = 4$.

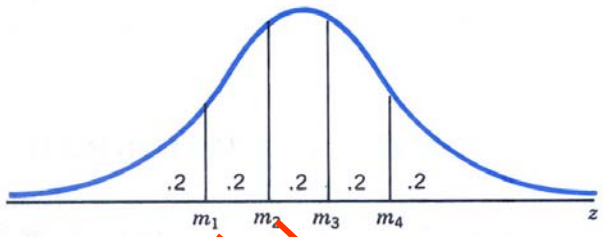


Figure 18 The $N(0, 1)$ distribution and the normal scores for $n = 4$.

case 1 0.2

case 2 0.4

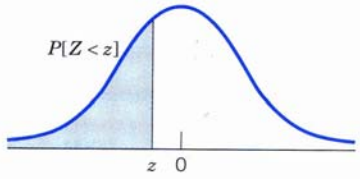
case 3 0.6

Case 4 0.8

cumulative probability

Find the probability **inside** the table and take the correspondent z-score value

TABLE 3 Standard Normal Probabilities



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−.7	.2420	.2389	.2358	.2327	.2297	.2266	.2236	.2206	.2177	.2148
−.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

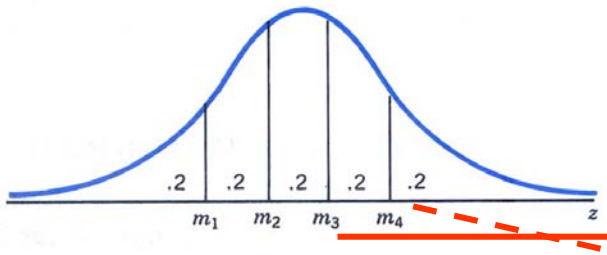
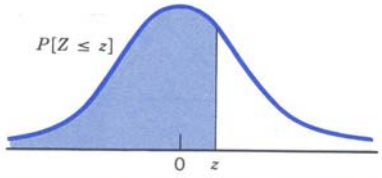


Figure 18 The $N(0, 1)$ distribution and the normal scores for $n = 4$.

When the normal scores are obtained, plot the pairs ***observation – correspondent normal score*** in a plot thus checking for **linearity**.

A straight line pattern in a normal-scores plot support the plausibility of a normal model. A curve appearance indicates a **departure** from normality.



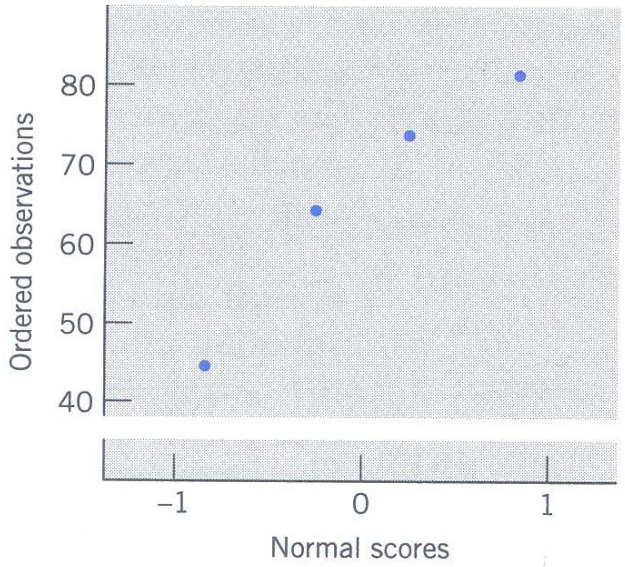
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998

Example Suppose a random sample of size 4 has produced the observations 68, 82, 44, and 75. Construct a normal-scores plot.

SOLUTION The ordered observations and the normal scores are shown in Table and the normal-scores plot of the data is given in Figure .

TABLE Normal Scores

Normal Scores	Ordered Sample
$m_1 = -.84$	44
$m_2 = -.25$	68
$m_3 = .25$	75
$m_4 = .84$	82



Normal-scores plot of Table

Transforming observations to obtain normal distribution

There is no rule for determining the best transformation in a given situation. For any data set that does not have a symmetric histogram, we consider a variety of transformations.

Some Useful Transformations

Make large values larger:	Make large values smaller:
x^3, x^2	$\sqrt{x}$, $\sqrt[4]{x}$, $\log_e x$, $\frac{1}{x}$

You may recall that $\log_e x$ is the natural logarithm. Fortunately, computers easily calculate and order the transformed values, so that several transformations in a list can be quickly tested. Note, however, that the observations must be positive if we intend to use $\sqrt{x}$, $\sqrt[4]{x}$, and $\log_e x$.

The selection of a good transformation is largely a matter of trial and error. If the data set contains a few numbers that appear to be detached far to the right, $\sqrt{x}$, $\sqrt[4]{x}$, $\log_e x$, or negative powers that would pull these stragglers closer to the other data points should be considered.

HOW MUCH TIMBER IS IN THIS FOREST?
The volume of timber available for making lumber can only be estimated by sampling the number of trees in randomly selected plots within the forest. The distribution of tree size must also be taken into account.

Example

A forester records the volume of timber, measured in cords, for 49 plots selected in a large forest. The data are given in Table ' and the corresponding histogram appears in Figure The histogram exhibits a long tail to the right, so it is reasonable to consider the transformations $\sqrt{x}$, $\sqrt[4]{x}$, $\log_e x$, and $1/x$. Transform the data to near normality.

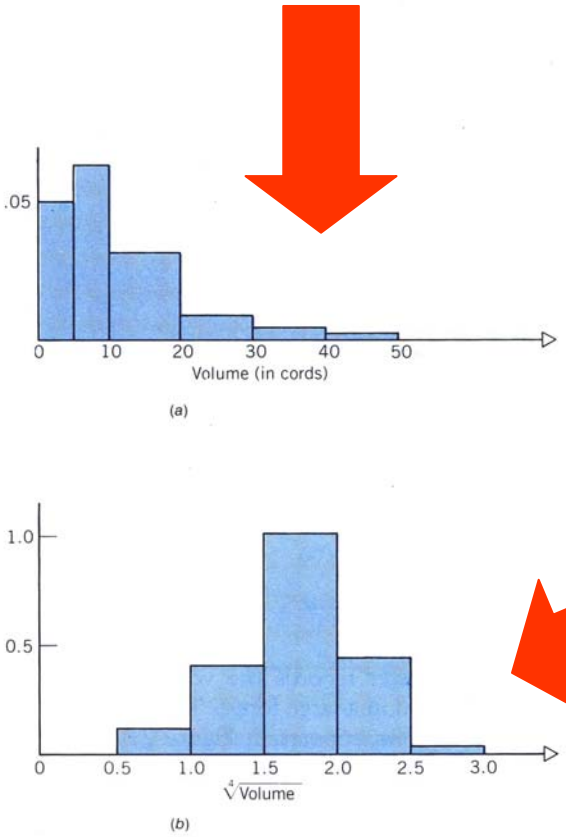
SOLUTION The most satisfactory result, obtained with

$$\text{Transformed data} = \sqrt[4]{\text{Volume}}$$

Transforming observations to obtain normality

TABLE Volume of Timber in Cords

39.3	14.8	6.3	.9	6.5
3.5	8.3	10.0	1.3	7.1
6.0	17.1	16.8	.7	7.9
2.7	26.2	24.3	17.7	3.2
7.4	6.6	5.2	8.3	5.9
3.5	8.3	44.8	8.3	13.4
19.4	19.0	14.1	1.9	12.0
19.7	10.3	3.4	16.7	4.3
1.0	7.6	28.3	26.2	31.7
8.7	18.9	3.4	10.0	



An illustration of the transformation technique. (*a*) Histogram of timber volume. (*b*) Histogram of $\sqrt[4]{\text{volume}}$.

TABLE The Transformed Data $\sqrt[4]{\text{Volume}}$

2.50	1.96	1.58	.97	1.60
1.37	1.70	1.78	1.07	1.63
1.57	2.03	2.02	.91	1.68
1.28	2.26	2.22	2.05	1.34
1.64	1.60	1.51	1.70	1.56
1.37	1.70	2.59	1.70	1.91
2.10	2.09	1.94	1.17	1.86
2.11	1.79	1.36	2.02	1.44
1.00	1.66	2.31	2.26	2.37
1.72	2.09	1.36	1.78	