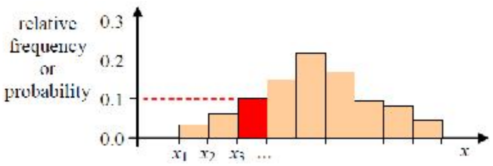
Graphical Resresentaiton of Experimental Data: Gaussian Distribution

Histograms

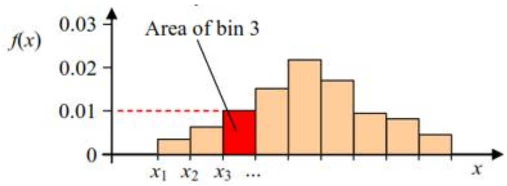
Probability Histogram

$$probability_j = \frac{\text{number of measurements in range } x_j \le x < x_{j+1}}{n}$$

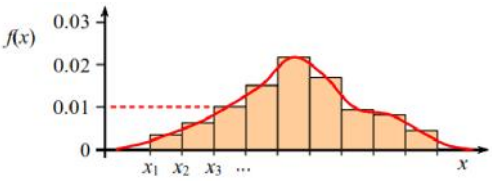


Verically Normalized Histogram

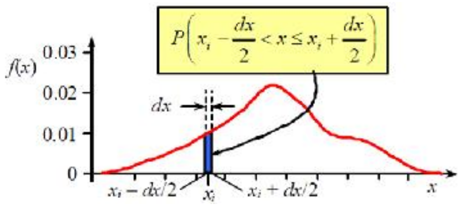
$$\frac{probability_j}{\Delta x} = \frac{\text{number of measurements in range } x_j \le x < x_{j+1}}{n\Delta x}$$



Probability Density Functions

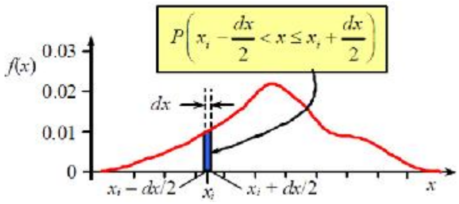


Infinite Number of Experiments



$$P(a < x < b) = \int_a^b F(x)dx \qquad P(\forall x) = \int_{-\infty}^{+\infty} F(x)dx$$

Probability Density Functions



$$P(a < x < b) = \int_a^b F(x)dx$$

$$P(\forall x) = \int_{-\infty}^{+\infty} F(x)dx = 1$$

Expected Value (or mean of population) and standard devation

Expected value is defined in terms of the probability density function as the mean of all possible x values in the continuous system.

$$\mu = E(x) = \int_{-\infty}^{+\infty} xF(x)dx$$

Standard deviation: In an ideal situation in which f(x) exactly represents the population, σ is the standard deviation of the entire population. It is therefore also called the population standard deviation.

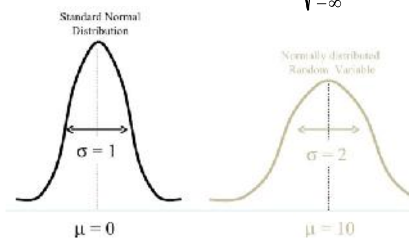
$$\sigma = \sqrt{\int_{-\infty}^{+\infty} (x-\mu)^2 F(x)dx}$$

Normalized probability density function

A normalized probability density function is constructed by transforming both the abscissa (horizontal axis) and ordinate (vertical axis) of the PDF plot as follows:

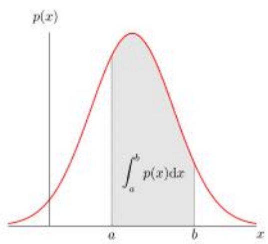
$$z = \frac{x-\mu}{\sigma}$$

$$f(z) = \sigma f(x)$$



Probability Density Functions

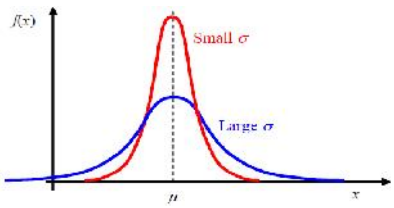
A probability density function (PDF), or density of a continuous random variable, is a function that describes the relative likelihood for this random variable to take on a given value.



The probability of the random variable falling within a particular range of values is given by the integral of this variable's density over that range

The Normal (Or Gaussian) Distribution

The normal (or Gaussian) distribution is a very common continuous probability distribution.
Normal distributions are important in statistics and are often used in the natural and social sciences to represent real-valued random variables whose distributions are not known.
The name Gaussian comes from Gaussian function that describes pure randamodness



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

«: mean or expectation of the distribution
±: standard deviation
$\pm^2$: Variance

- Bell shaped
- Symetrical mean=mode=median
- Area under curve is 1.

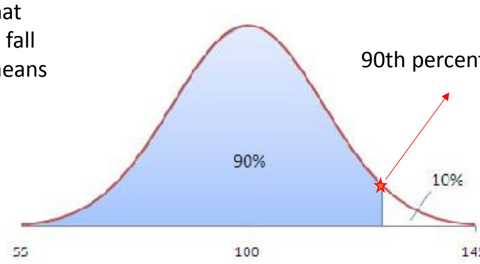
Standart Normal Density Function

Concept of percentile

The most common definition of a percentile is a number where a certain percentage of scores fall below that percentile.

You might know that you scored 119 out of 150 on a test. But that figure has no real meaning unless you know what percentile you fall into. If you know that your score is in the 90th percentile, that means you scored better than 90% of people who took the test

90th percentile is 119

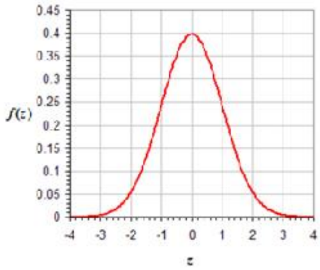


Standard Normal Density Function

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

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

«=0
±=1



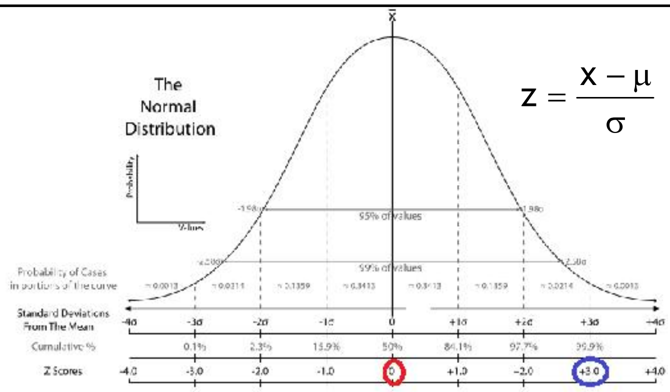
The standard density function is an important tool to calculate propability and statistics of distributions with normal probability density function.
Instead calculating integrals under curve , z score tables are used to do calculations.

Standart Normal Density Function

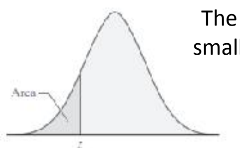
Z Scores and tables

A z-score is a measure of how many standard deviations below or above the population mean a raw score is.

- Z-scores mostly range from -3.45 std devs (which would fall to the far left of the normal distribution curve) to +3.45 std devs (which would fall to the far right of the normal distribution curve)..



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



The probaibility of a any value smaller than a specific z score is given in tables

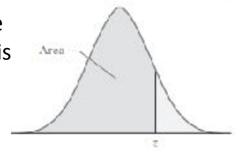


Table IV

Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004

Table IV (continued)

Standard Normal Distribution

z	.00	.01	.02	.03	.04	.05	.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636

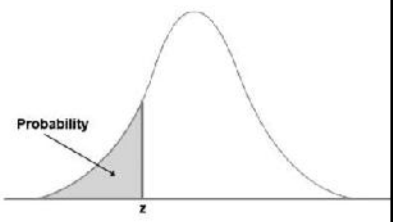
In order to use a z-score, you need to know the mean μ and the population standard deviation σ. Simply put, a z-score is the number of standard deviations from the mean a data point is

Standart Normal Density Function

Calculating Probabilities

The measured resistances of light bulbs on a prodcution line has
a mean of 550 Ohm and standard deviation of 140 Ohm

- the percentage of resistances which has a value below 655 Ohm
- the percentage of resistances which has a value above 746 Ohm
- the percentage of resistances which has a value between 655 and 746 Ohm

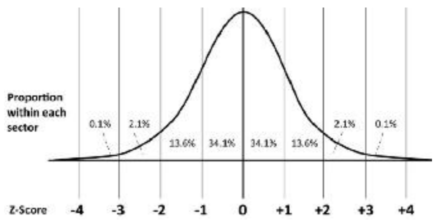


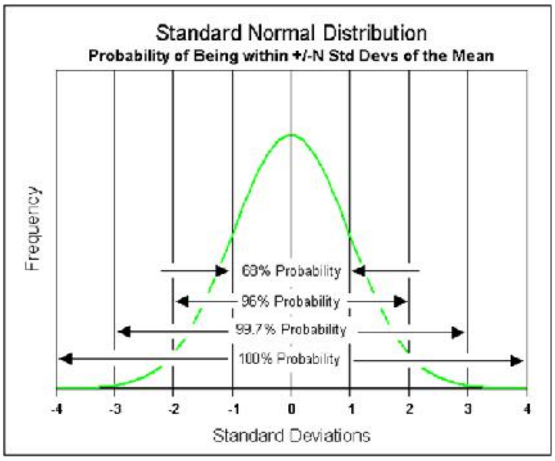
z	.00	.01	.02	.03	.04	.05	.06	.07
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418

(a) Z=0.75 P(x<655)=0.7734 answer 77.34%
(b) Z=1.4 P(x<746)=0.9192 answer 8.08%
(c) P(655x<6746)=0.9192-0.7734=0.1458 answer 14.58%

Probability Density Functions

The multiples of standard deviation or integer Z scores





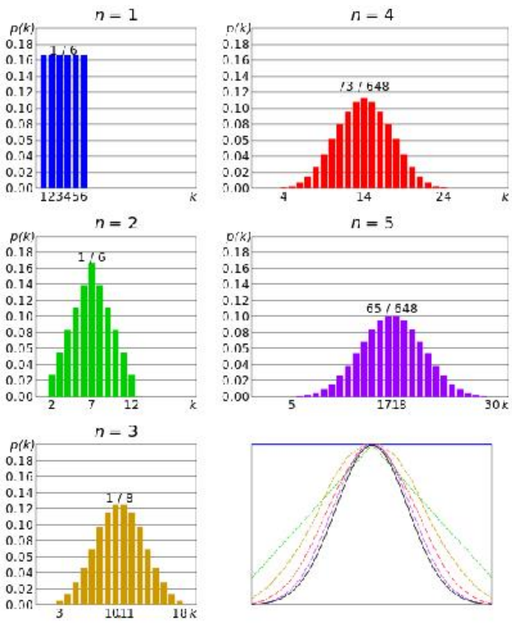
Probability Density Functions

Confidence Level

Confidence level, ***c***, is defined as ***the probability that a random variable lies within a specified range of values***. The range of values itself is called the ***confidence interval***.
For example, as discussed above, we are 95.44% confident that a purely random variable lies within +/- two standard deviations from the mean.
We state this as a confidence level of *c* = 95.44%, which we usually round off to 95% for practical engineering statistical analysis.

Level of significance, α, is defined as ***the probability that a random variable lies outside of a specified range of values***.

In the above example, we are 100 – 95.44 = 4.56% confident that a purely random variable lies either *below* or *above* two standard deviations from the mean.



The normal distribution is useful because of the central limit theorem. In its most general form, under some conditions (which include finite variance), it states that averages of random variables independently drawn from independent distributions converge in distribution to the normal, that is, become normally distributed when the number of random variables is sufficiently large. Physical quantities that are expected to be the sum of many independent processes (such as measurement errors) often have distributions that are nearly normal