Data collection, Analysis and Inference

Subject Code: CPE-RPE,

May 2021,
SRM Univeristy-AP, Andhrapradesh

Lecture- 3: Basic Statistical Distributions and their applications: Normal Distribution, Weibull Distribution.

Aim: To be able recognize systems of continuous variable nature and their probability distributions.

- Recall..
  - ► We defined sample space, events
  - ► Looked at three definitions of probability
  - ► Defined random variables and started to see events in terms of random variables
  - ► Narrowed down our focus to “discrete” random variables and defined expectation of a (discrete) random variable.
- Looked at four types of random variables - Bernoulli, binomial, Poisson, and geometric.

- “Discrete” random variables - random variables which takes on finite or countably many values
- We can even have random variables which can take uncountably many values
- Typical examples -

▸various times like service time, installation time, download time, failure time, and

▸ physical measurements like weight, height, distance, velocity, temperature, and connection speed etc.

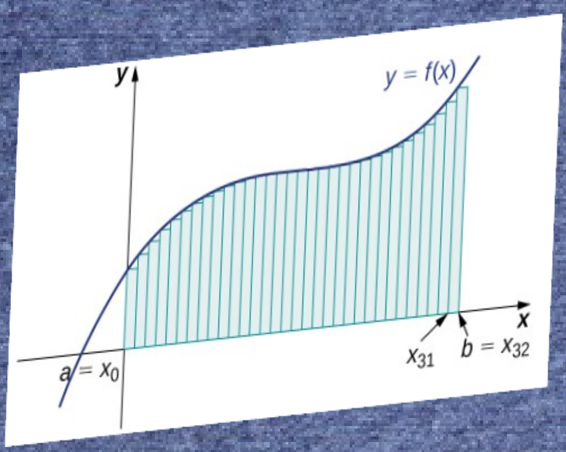
Probability density function

If X is a continuous random variable, then there exists a non-negative function f , called **probability density function**, defined for all real $x \in (-\infty, \infty)$, having the property that for any set B of real numbers,

$$P\{X \in B\} = \int_B f(x)\,\mathrm{d}x$$

$$\int_a^b f(x)\,\mathrm{d}x$$

Area under the graph of f(x) over the interval [a, b]



Total probability must be 1

$$\Longrightarrow P\{-\infty < X < \infty\} = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$$

If we let $B = [a, b]$, then

$$P(a \le X \le b) = \int_{a}^{b} f(x)\,\mathrm{d}x$$

If we let $a = b$ in the above equation,

$$P\{X = a\} = \int_{a}^{a} f(x)\,\mathrm{d}x = 0$$

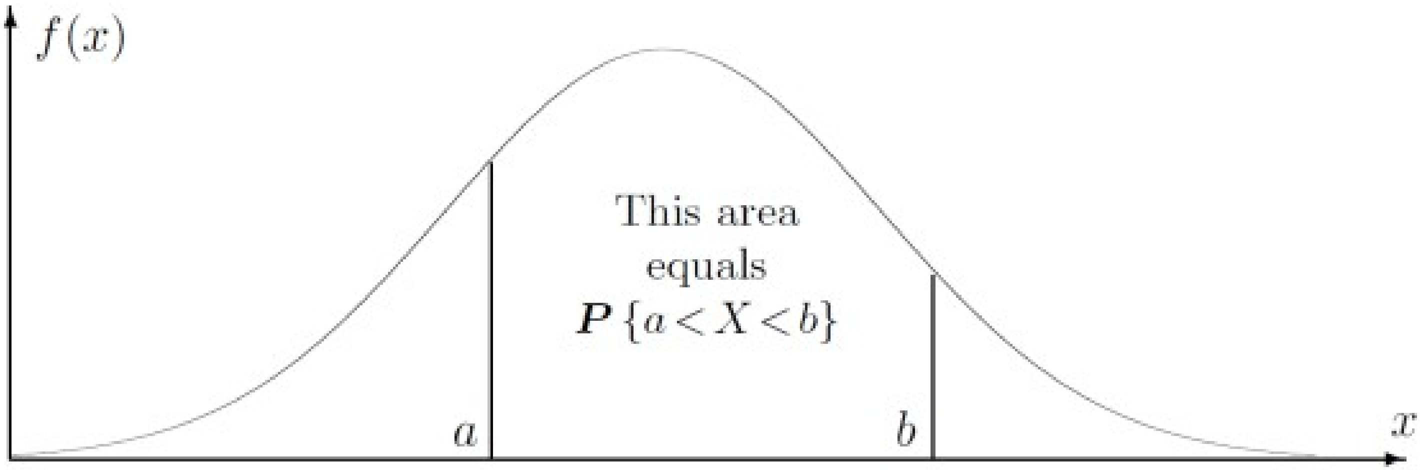
- That is, p(a) = P{X = a} = 0 for every real number a

Hence the probability mass function **does not carry any information** in the case of continuous random variables!

Basis for Normal, t-dist tables

By the **Fundamental Theorem of Calculus**,

$$\int_a^b f(x)dx = F(b) - F(a) = P\{a \le x \le b\}$$



Summary

Distribution	Discrete	Continuous
We use	p.m.f $p(x) = P\{X = x\}$	p.d.f $f(x)$ (p.d.f)
Computing probabilities	$P\{X \in A\} = \sum_{x \in A} p(x)$	$P\{X \in A\} = \int_A f(x)dx$
Cumulative distribution function	$F(a) = \sum_{x \leq a} p(x)$	$F(a) = \int_{-\infty}^{a} f(x)dx$
Total probability	$\sum_x p(x) = 1$	$\int_{-\infty}^{\infty} f(x)dx = 1$

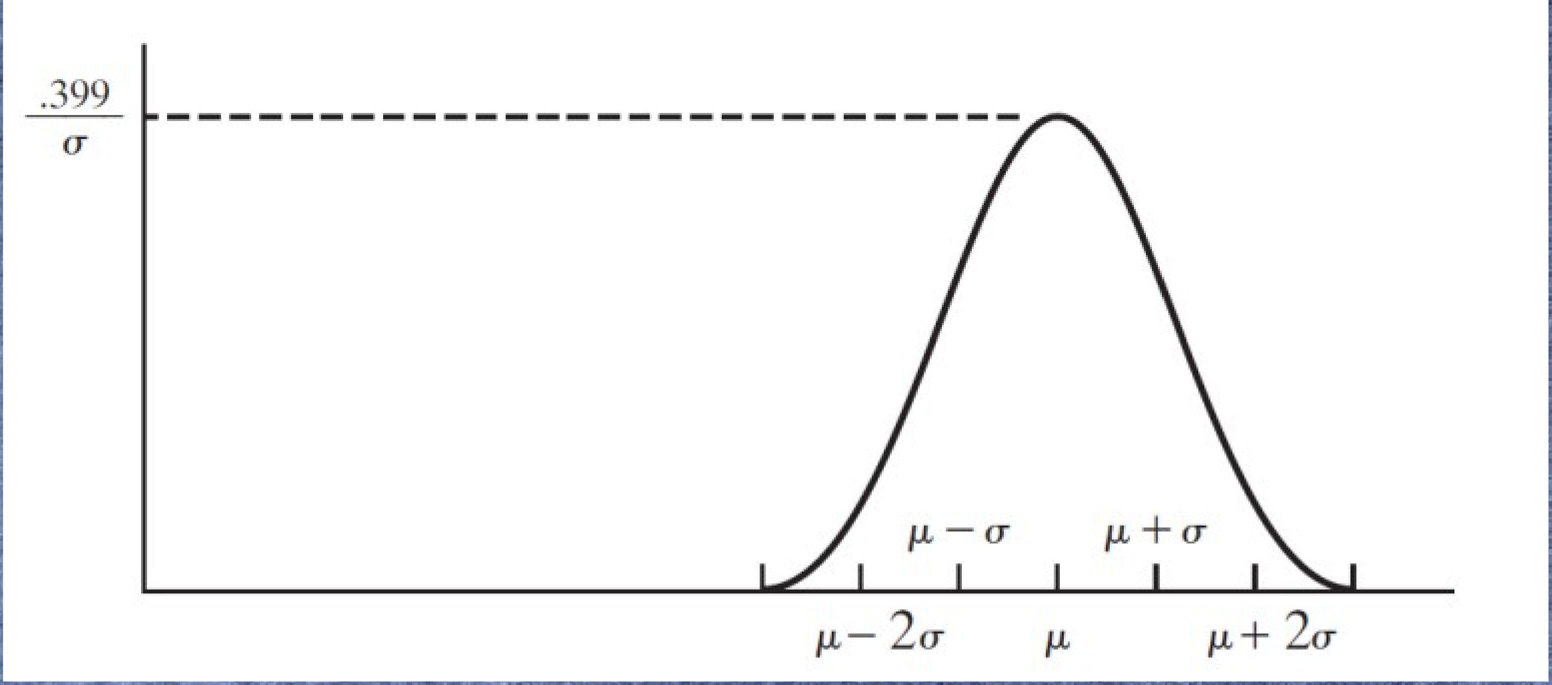
Exercise: Compute $\mathrm{Var}(X)$ if X has a density function given by

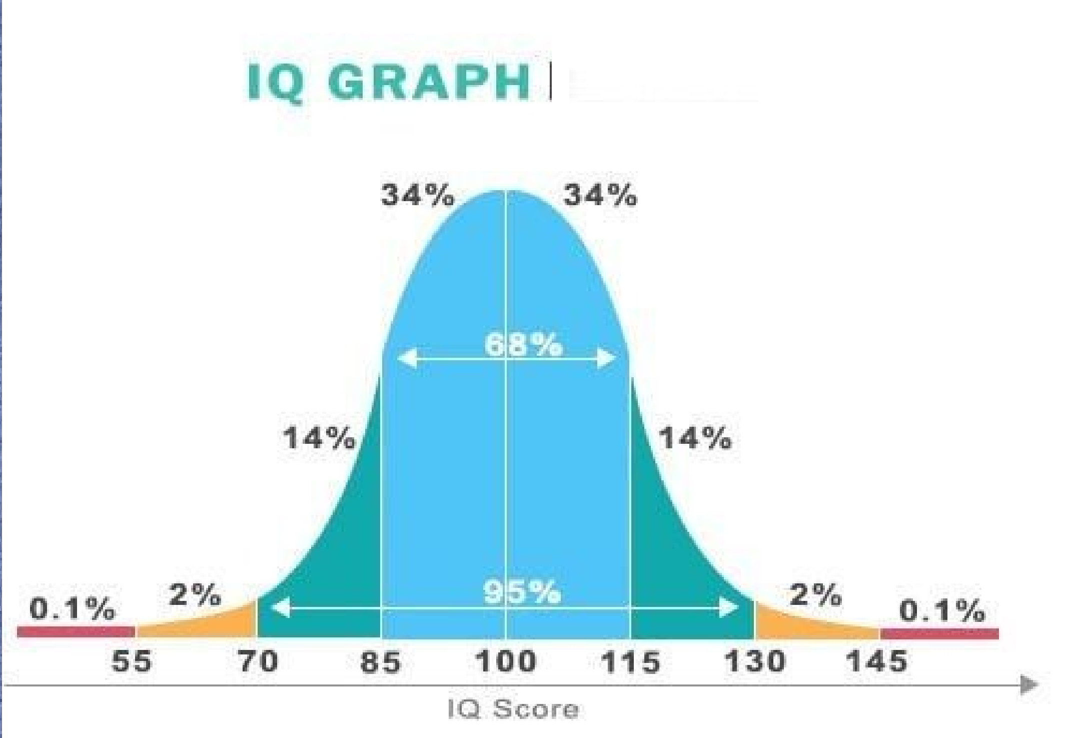
$$f(x) = \begin{cases} \frac{1}{4}xe^{-x/2}, & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

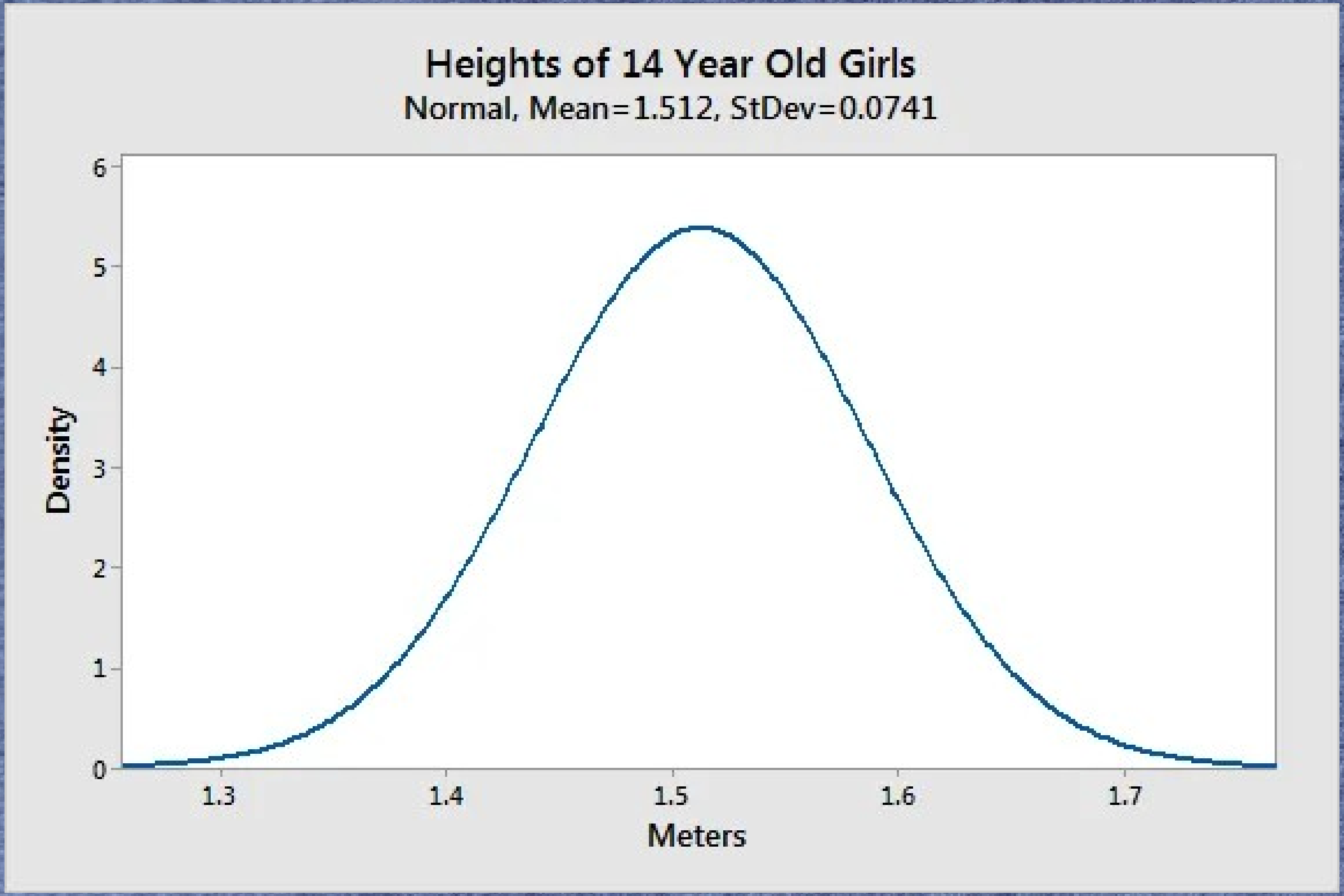
Normal random variable ★

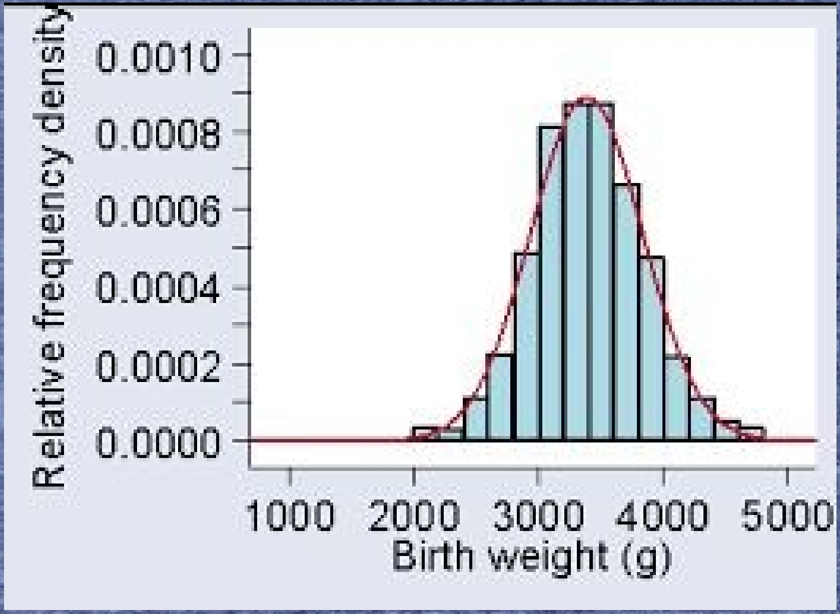
We say that X is a **normal random variable**, or simply that X is normally distributed, with parameters μ and σ^2 if the density of X is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$



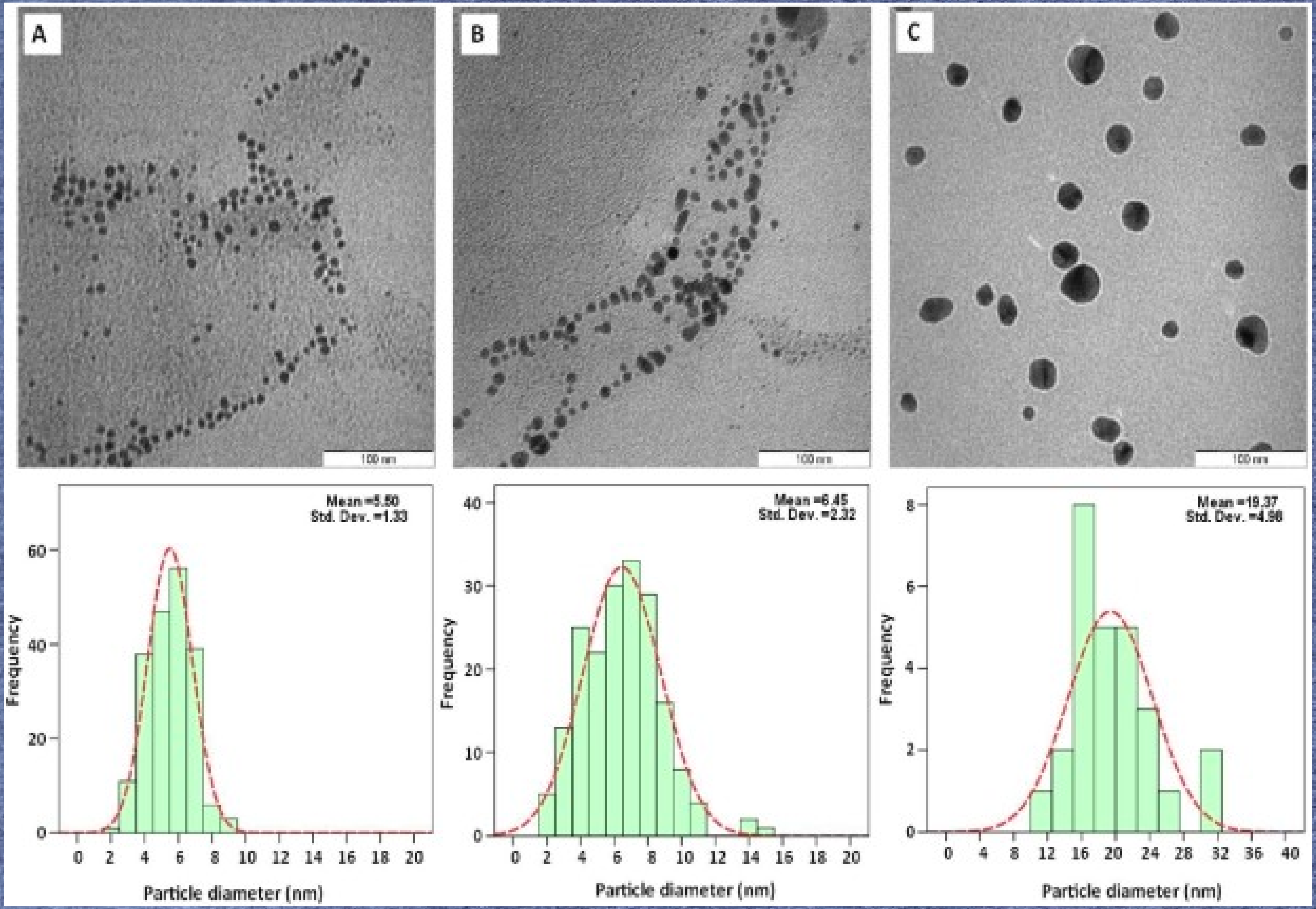

IQ GRAPH
34%
34%
68%
14%
14%
0.1%
2%
95%
2%
0.1%
55
70
85
100
115
130
145
IQ Score


Heights of 14 Year Old Girls
Normal, Mean=1.512, StDev=0.0741
Density
0
1
2
3
4
5
6
1.3
1.4
1.5
1.6
1.7
Meters

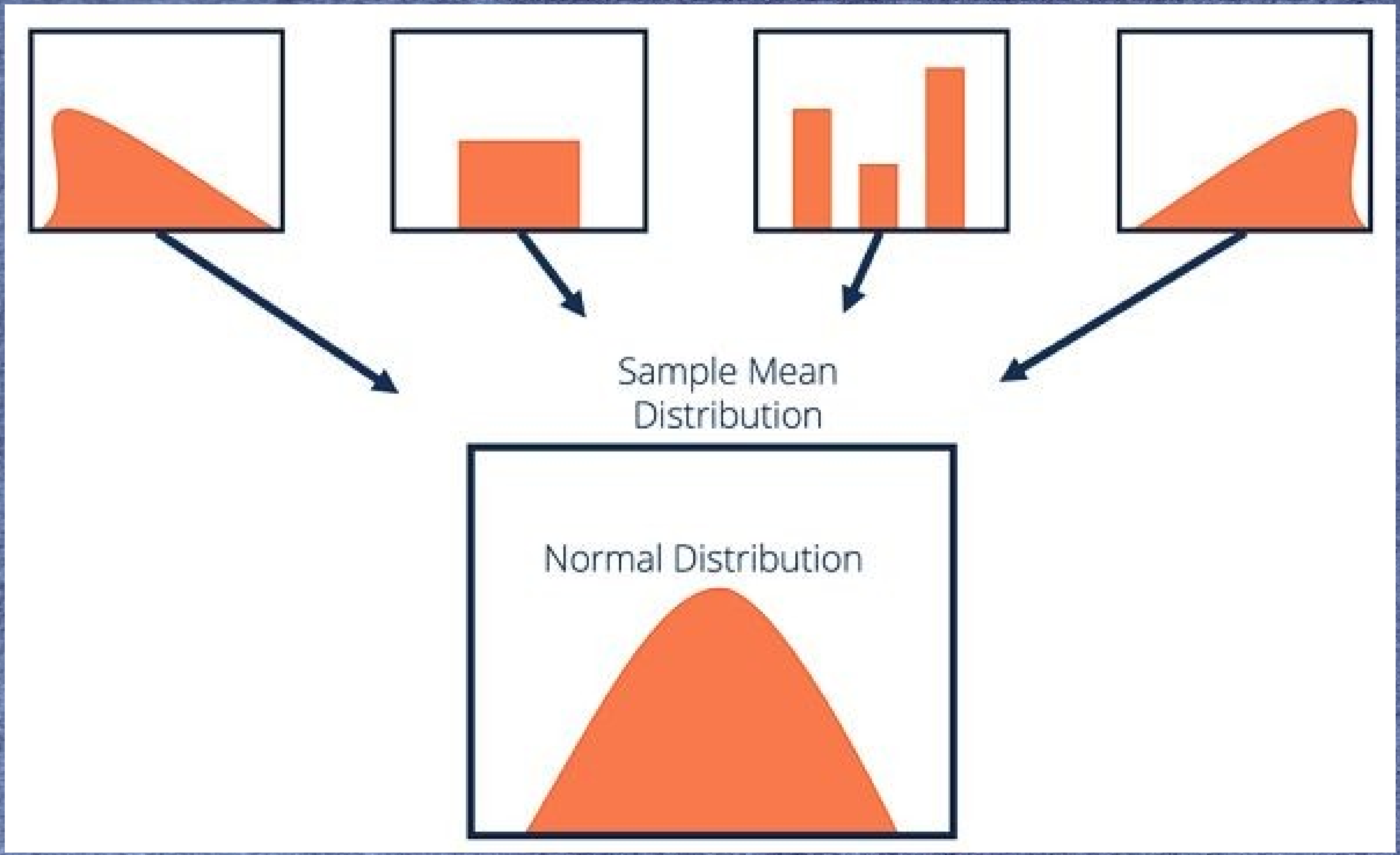


Relative frequency density
0.0010
0.0008
0.0006
0.0004
0.0002
0.0000
1000
2000
3000
4000
5000
Birth weight (g)

Transmission electron microscopy image and the particle size distribution for Ag/Cts/PEG



Central Limit Theorem



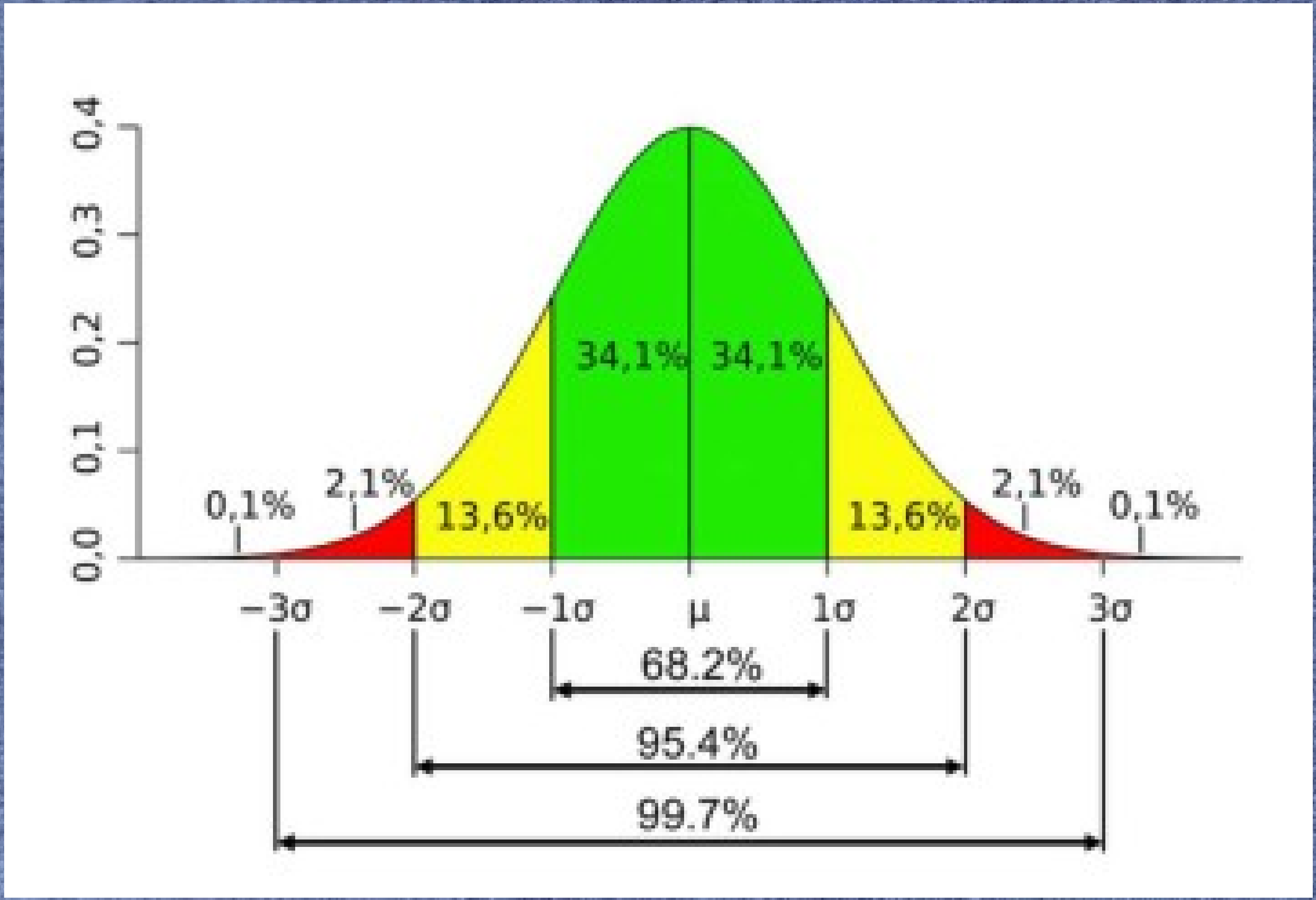
Standard normal random variable.

- If Z =(X-μ)/ σ , then Z is a normal random variable with parameters 0 and 1
- **A normal random variable with parameters 0 and 1 is called a standard normal random variable.**
- The alphabet Z is usually reserved for the standard normal random variable

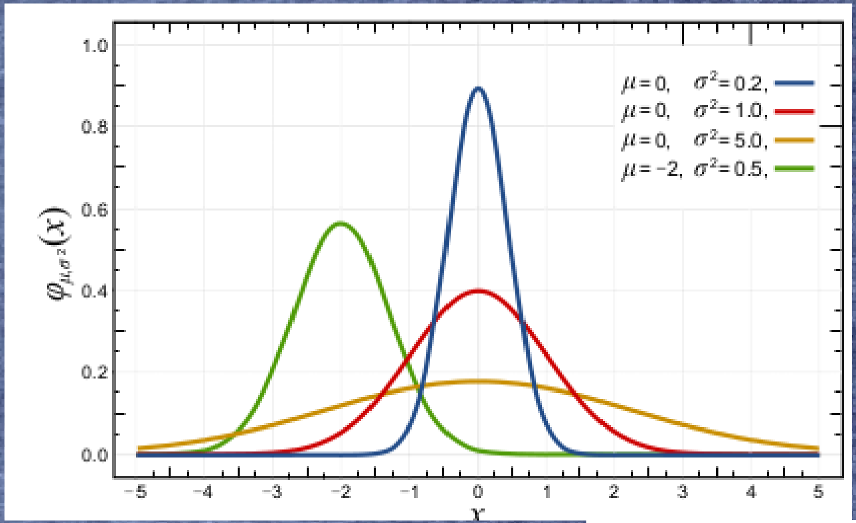
$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$$\phi_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

3σ limits

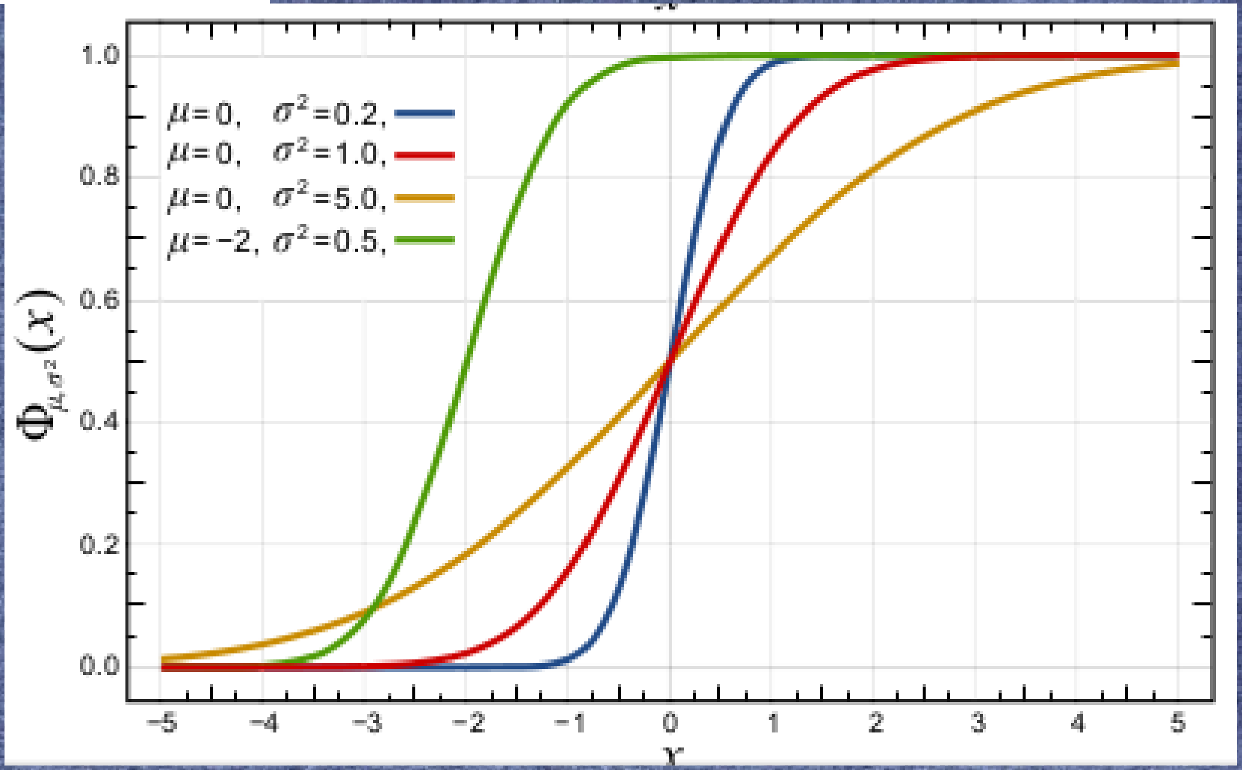

0,0
0,1
0,2
0,3
0,4
34,1%
34,1%
13,6%
13,6%
2,1%
2,1%
0,1%
0,1%
−3σ
−2σ
−1σ
μ
1σ
2σ
3σ
68.2%
95.4%
99.7%

Source for Normal Tables



$\phi_Z(z)$

$\Phi(z)$



Normal vs Standard Normal

$$P\{\mu - \sigma < X < \mu + \sigma\} = 0.682$$

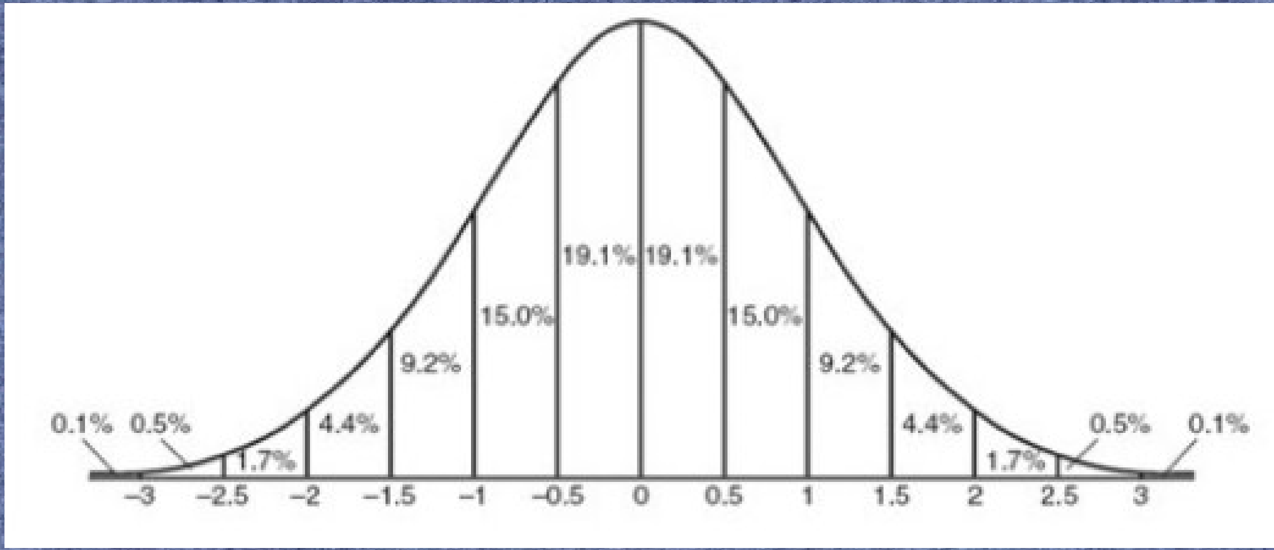
$$P\{\mu - 2\sigma < X < \mu + 2\sigma\} = 0.954$$

$$P\{\mu - 3\sigma < X < \mu + 3\sigma\} = 0.997$$

$$P\{-1 < Z < 1\} = 0.682$$

$$P\{-2 < Z < 2\} = 0.954$$

$$P\{-3 < Z < 3\} = 0.997$$



- Question: How do we compute P {a < Z < b} for any a and b?

  For instance, P {−1.75 < Z < 0.62} =?

- We use the distribution function Φ

  $P\{-1.75 < Z < 0.62\} = \Phi(0.62) - \Phi(-1.75)$

  Where,

$$\Phi(0.62) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0.62} e^{-x^2/2}\,dx$$

$$\Phi(-1.75) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-1.75} e^{-x^2/2}\,dx$$

Here comes the use of Standard Normal tables.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.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

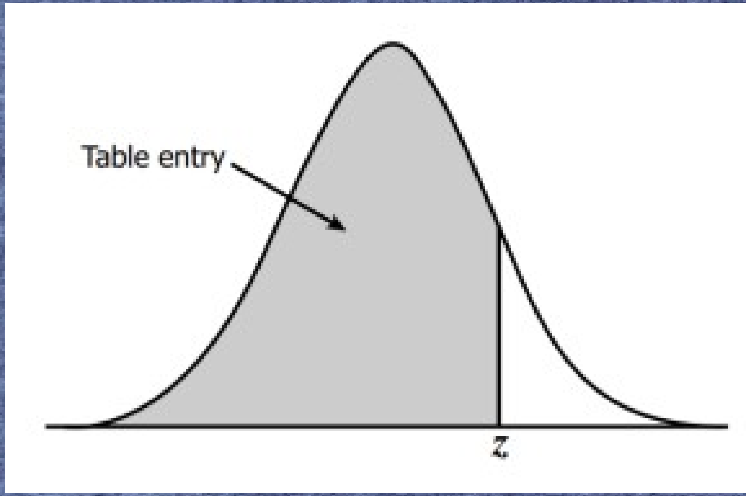


Table entries give Φ(z) for postive values of z.

$\Phi(0.62) = 0.7324$

z	.00	.01	.02	.03	.04	.05	.06	.07
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157
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
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790

- What about negative values of z?

  $\Phi(-1.75)$?

  Recall: $P\{Z < z\} = 1 - P(Z < -z)$ for any z.

  That is, $\Phi(z) = 1 - \Phi(-z)$.

  ►Thus, $\Phi(-1.75) = 1 - \Phi(1.75)$

  $=1 - 0.9599 = 0.0401$

For any normal variable X with parameters μ and σ

- How do we find $P\{c \leq X \leq d\}$?

  Trick: Transform it to standard form!

$$P\{c \leq X \leq d\} = P\left\{\frac{c-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{d-\mu}{\sigma}\right\} = P\{a \leq Z \leq b\}$$

where $a = \frac{c-\mu}{\sigma}$ and $b = \frac{d-\mu}{\sigma}$

Example: If X is a normal random variable with parameters $\mu = 3$ and $\sigma^2 = 9$, find

(a)$P\{2 < X < 5\}$, (b) $P\{X > 0\}$

Solution: $Z = \frac{X-3}{3}$

(a)

$$\begin{aligned}
P\{2 < X < 5\} &= P\left\{\frac{2-3}{3} < \frac{X-3}{3} < \frac{5-3}{3}\right\} \\
&= P\left\{-\frac{1}{3} < Z < \frac{2}{3}\right\} \\
&= \Phi\left(\frac{2}{3}\right) - \Phi\left(-\frac{1}{3}\right) \\
&= \Phi\left(\frac{2}{3}\right) - \left[1 - \Phi\left(\frac{1}{3}\right)\right] \quad (\text{since } \Phi(-x) = 1 - \Phi(x)) \\
&\approx 0.7486 - (1 - 0.6293) \quad \textbf{(standard normal table)} \\
&= 0.3779
\end{aligned}$$

Example: The systolic blood pressure in the population is usually modeled by a normal distribution with mean 120 mmHg (millimeters of mercury) and standard deviation 8 mmHg.

(a) Below which blood pressure do we find one third of the population?

(b) Above which blood pressure do we find 5% of the population?

Solution: Let X be the systolic blood pressure of a randomly selected individual.

Given that, X is a normal random variable with parameters 120 and 8^2

- Let $Z = (X - 120)/8$. Then Z is the standard normal random variable.

(a) Below which blood pressure do we find one third of the population?

We need to find c such that $P\{X < c\} = \frac{1}{3}$

$$\implies P\left\{Z < \frac{c-120}{8}\right\} = \frac{1}{3}$$

$$\implies \Phi\left(\frac{c-120}{8}\right) = \frac{1}{3}$$

$$\implies 1 - \Phi(-\frac{c-120}{8}) = \frac{1}{3} \qquad \textbf{(as } \Phi(x) = 1 - \Phi(-x)\textbf{)}$$

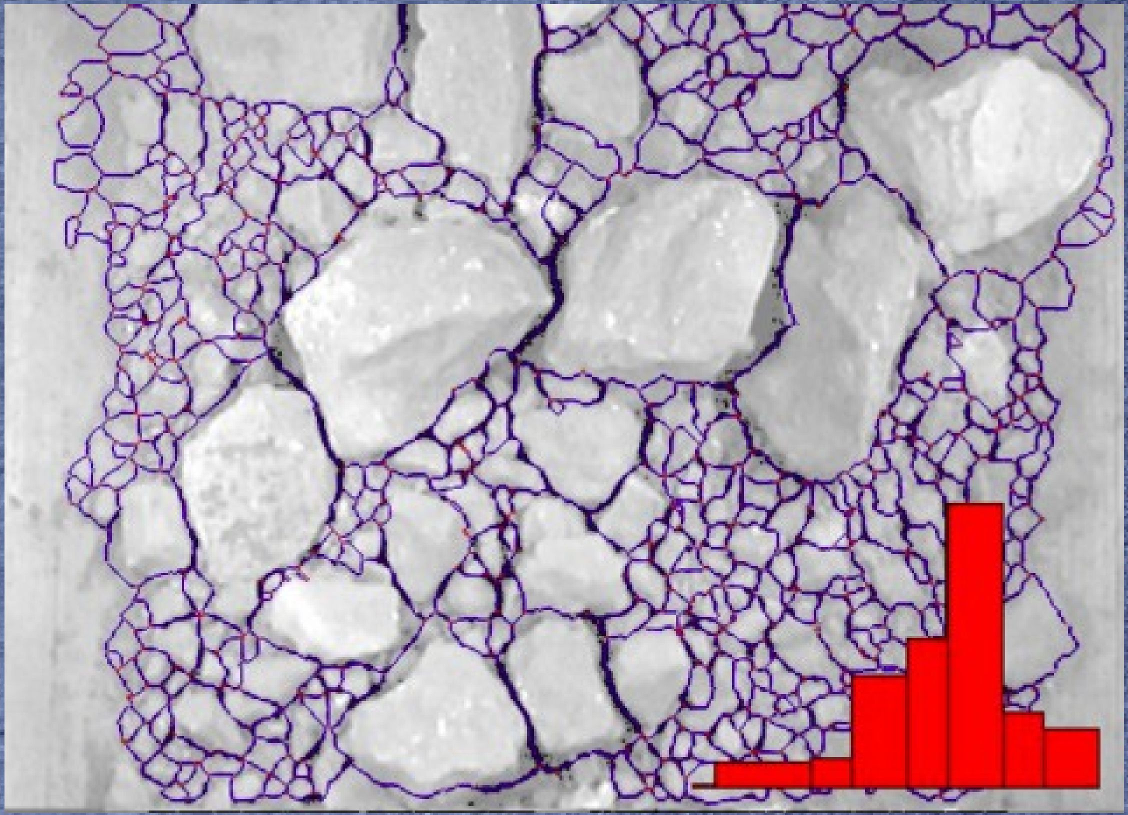
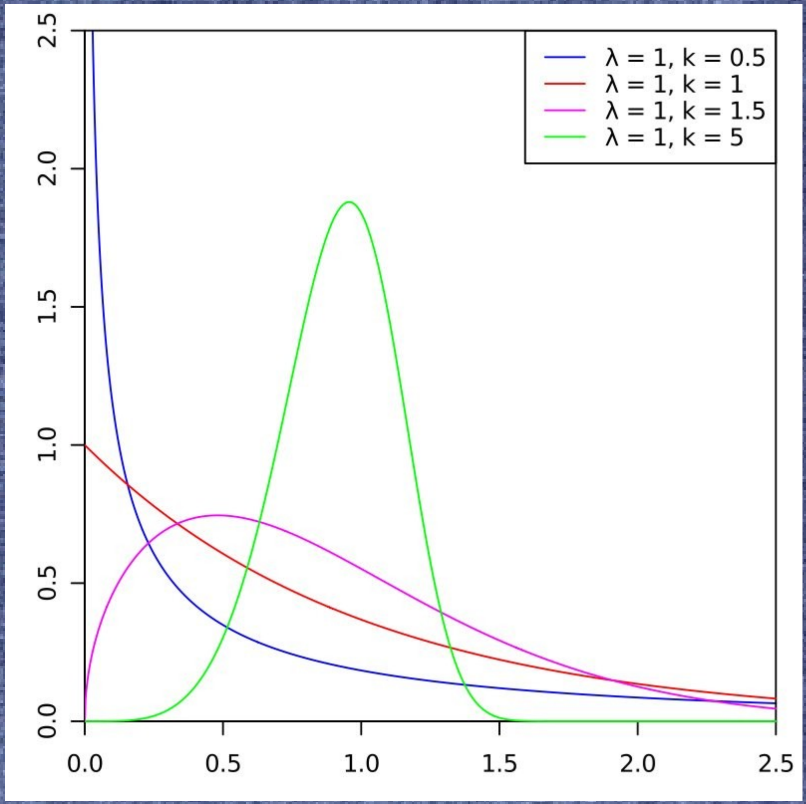
$$\implies \Phi(-\frac{c-120}{8}) = \frac{2}{3}$$

$$\implies -\frac{c-120}{8} \approx 0.43 \qquad \textbf{(from the standard normal table)}$$

$$\implies c \approx 116.56$$

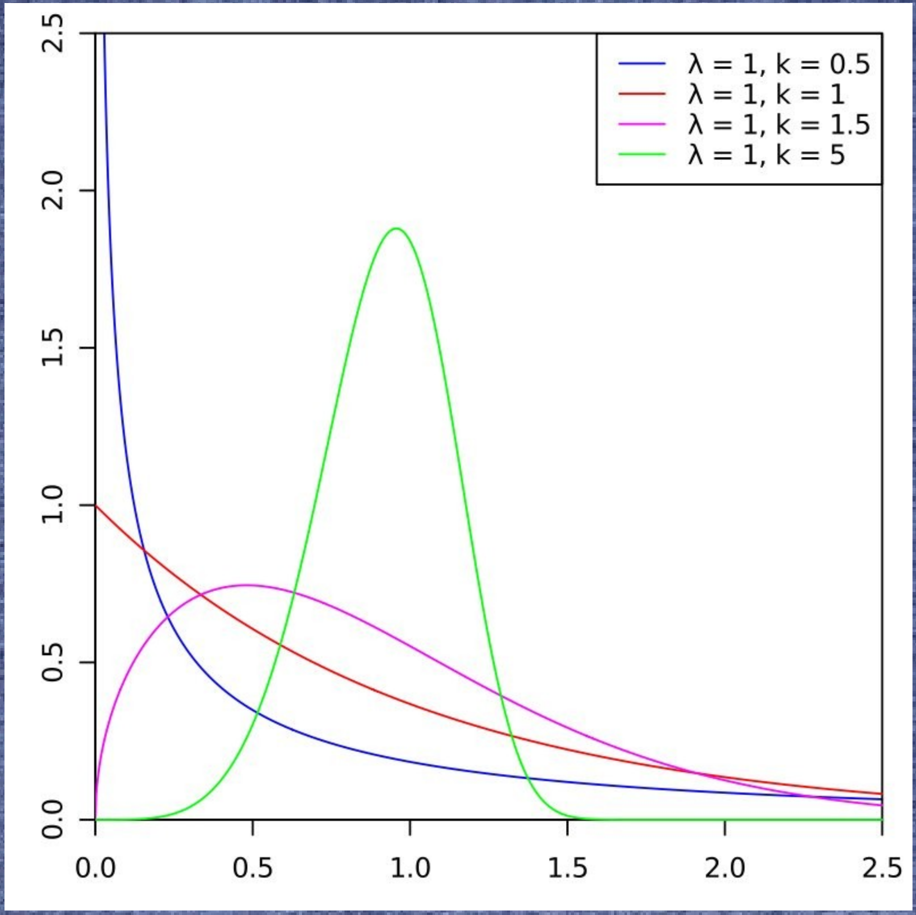
In words, one third of the population has a blood pressure below 116.56 mmHg.

Weibull distribution

- This is to describe a particle size distribution of a powder or granular material.
-  

- The probability density function of a Weibull random variable is:

$$f(x) = \begin{cases} \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

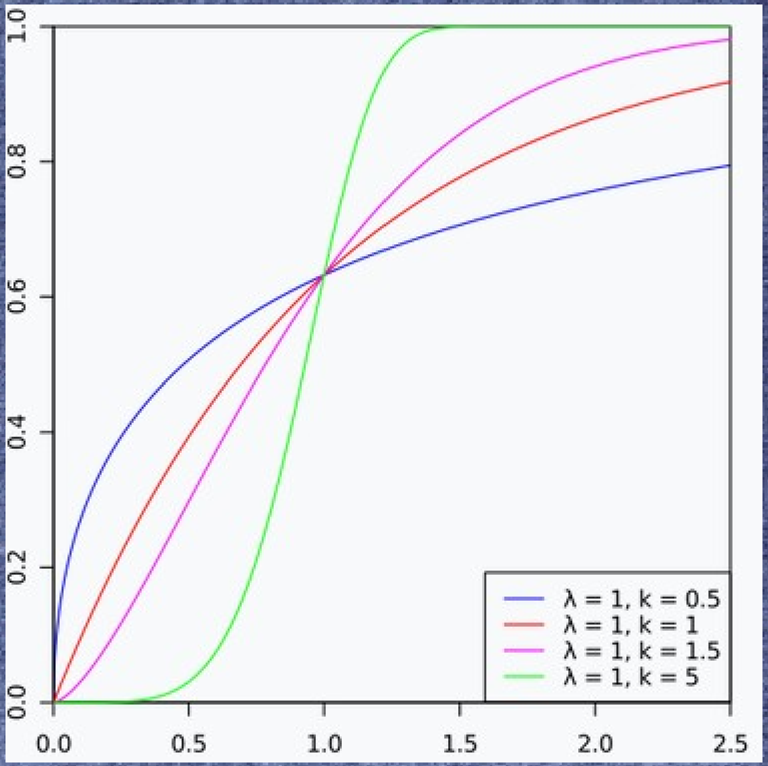


The cumulative distribution function is

$$F(x;k,\lambda) = 1 - e^{-(x/\lambda)^k}$$

for $x \geq 0$, and $F(x; k; \lambda) = 0$ for $x < 0$.

Mean	$\lambda\Gamma(1+1/k)$
Variance	$\lambda^2\left[\Gamma\left(1+\frac{2}{k}\right) - \left(\Gamma\left(1+\frac{1}{k}\right)\right)^2\right]$



- END