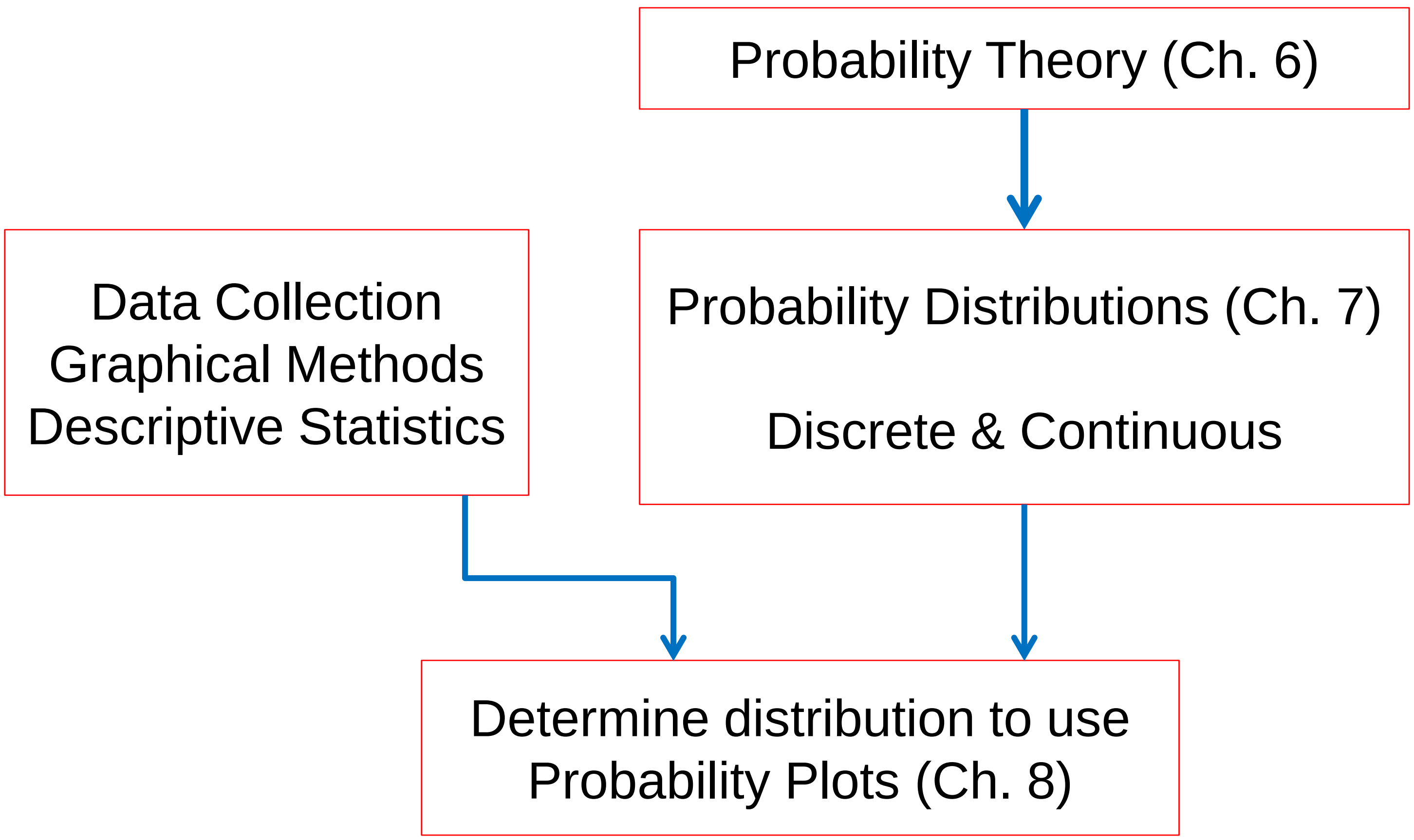
Normal and Lognormal Distributions

Chapter 7 Sections 7.1, 7.2, 7.3, 7.14

Note: Need Calculator & Z table Handout for lecture

Karen E. Schmahl Ph.D., P.E.

Relationship between Chapter Material



Probability Distributions

- Continuous
  - Normal,
  - Lognormal, Exponential, Weibull
  - Chi-Square, F, Student t
- Discrete
  - Binomial
  - Hypergeometric
  - Poisson

The Normal Distribution

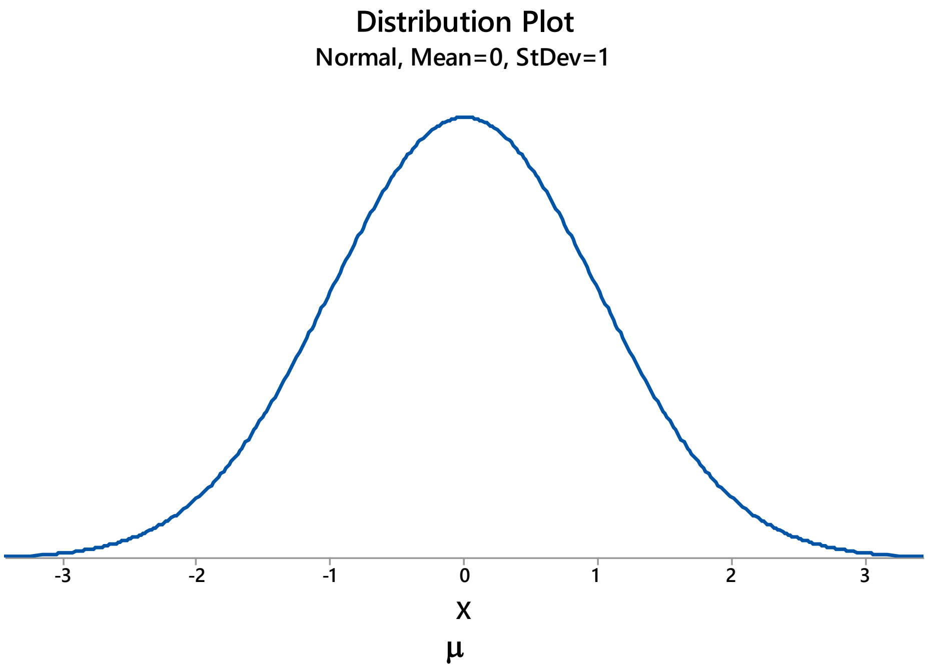
- The **normal distribution** is the most commonly used distribution in statistics.
- The normal distribution is continuous rather than discrete.
- The mean of a normal population may have any value, and the variance may have any positive value.
- It is also referred to as the Gaussian distribution.

The Normal Distribution

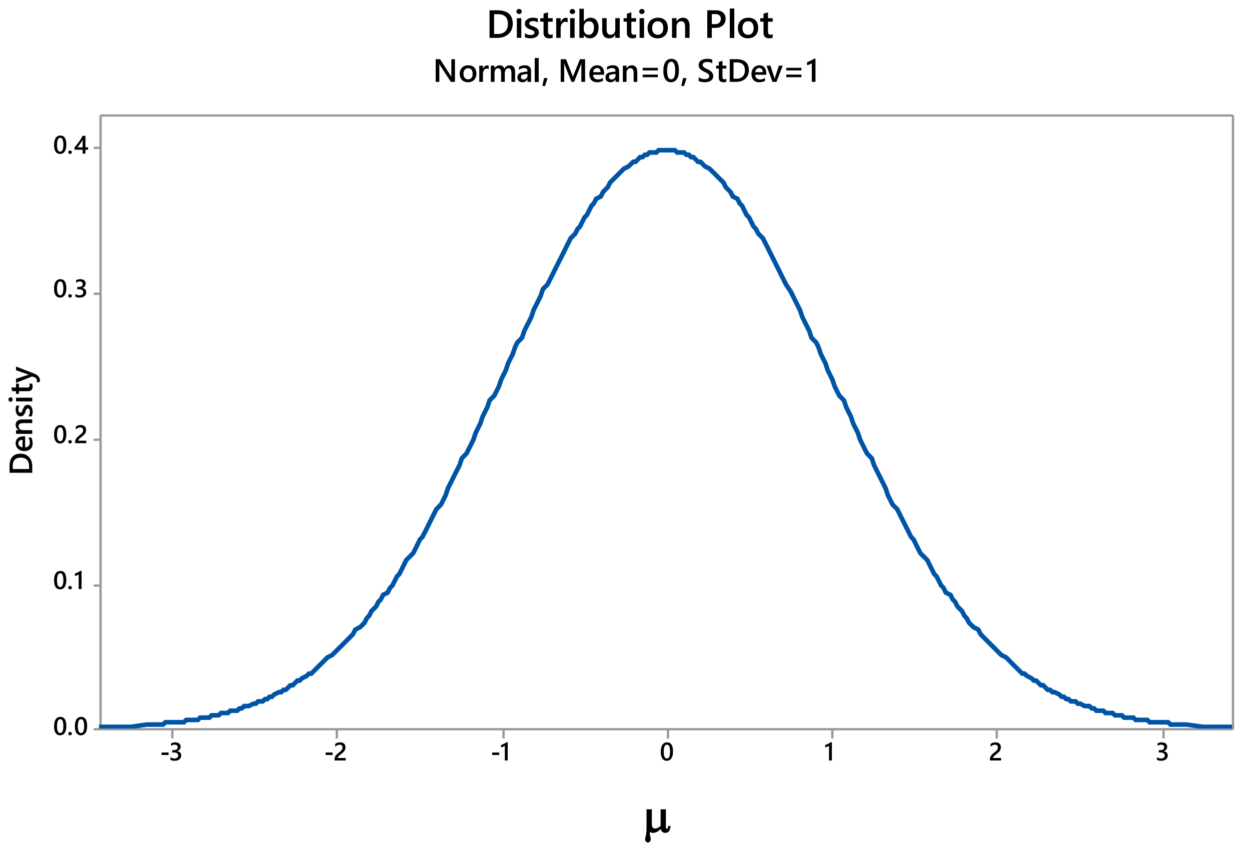
Normal distribution - Important **continuous** distribution
- Symmetric and bell-shaped.

For the population

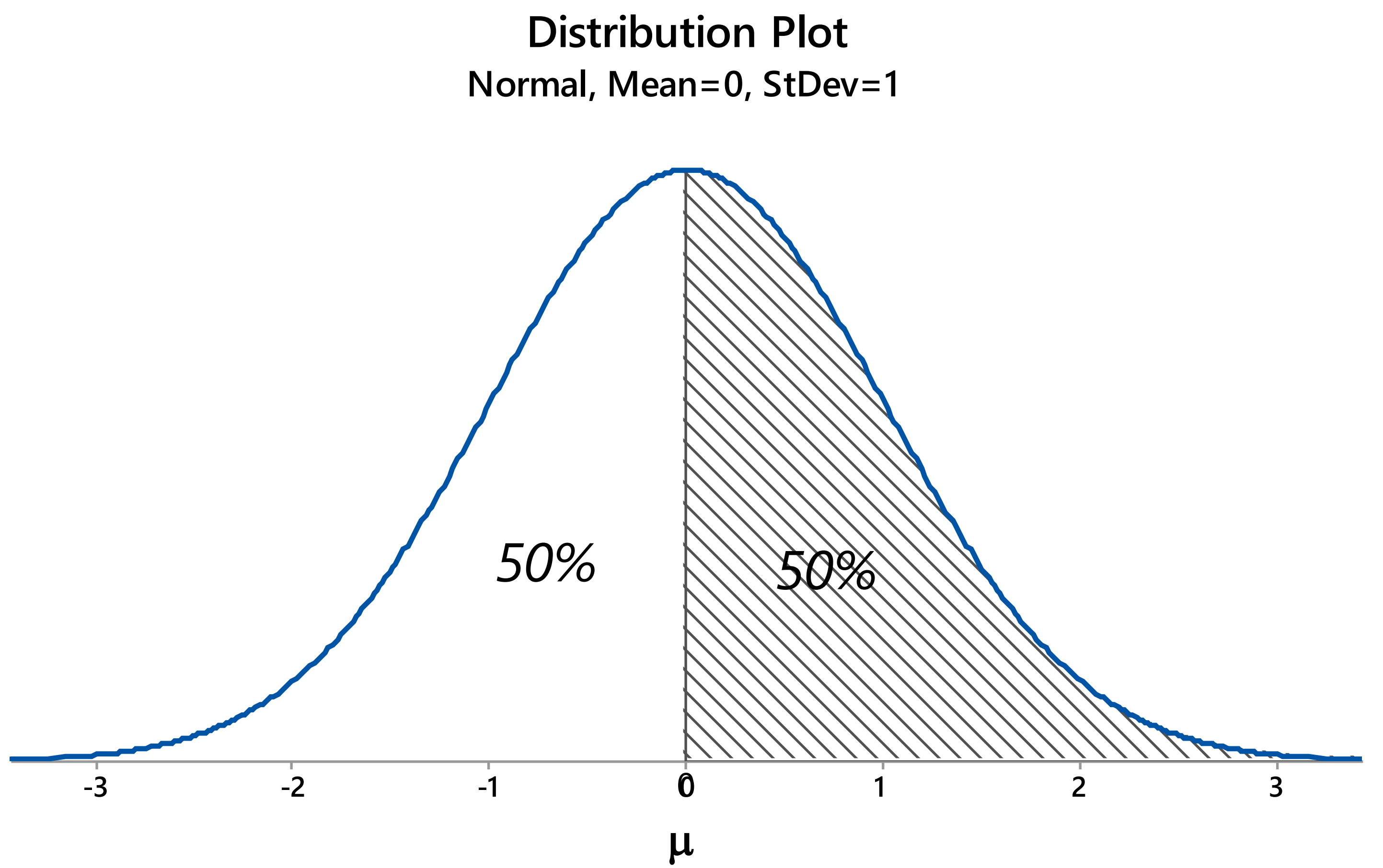
- Mean, μ
- Standard deviation, σ



Normal probability density function: $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-u)^2/2\sigma^2}$$



Area under the curve represents 100% of the population



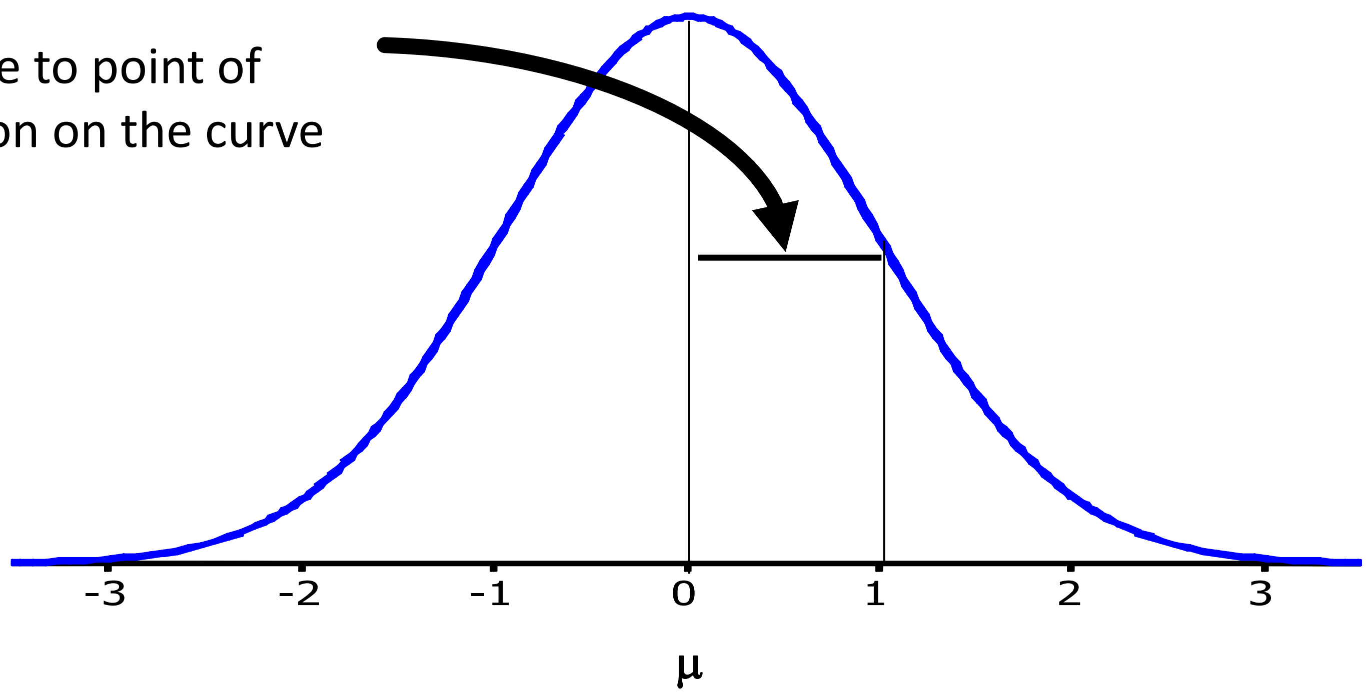
50% of observations from a normal distribution will be below the mean

50% of observations from a normal distribution will be above the mean

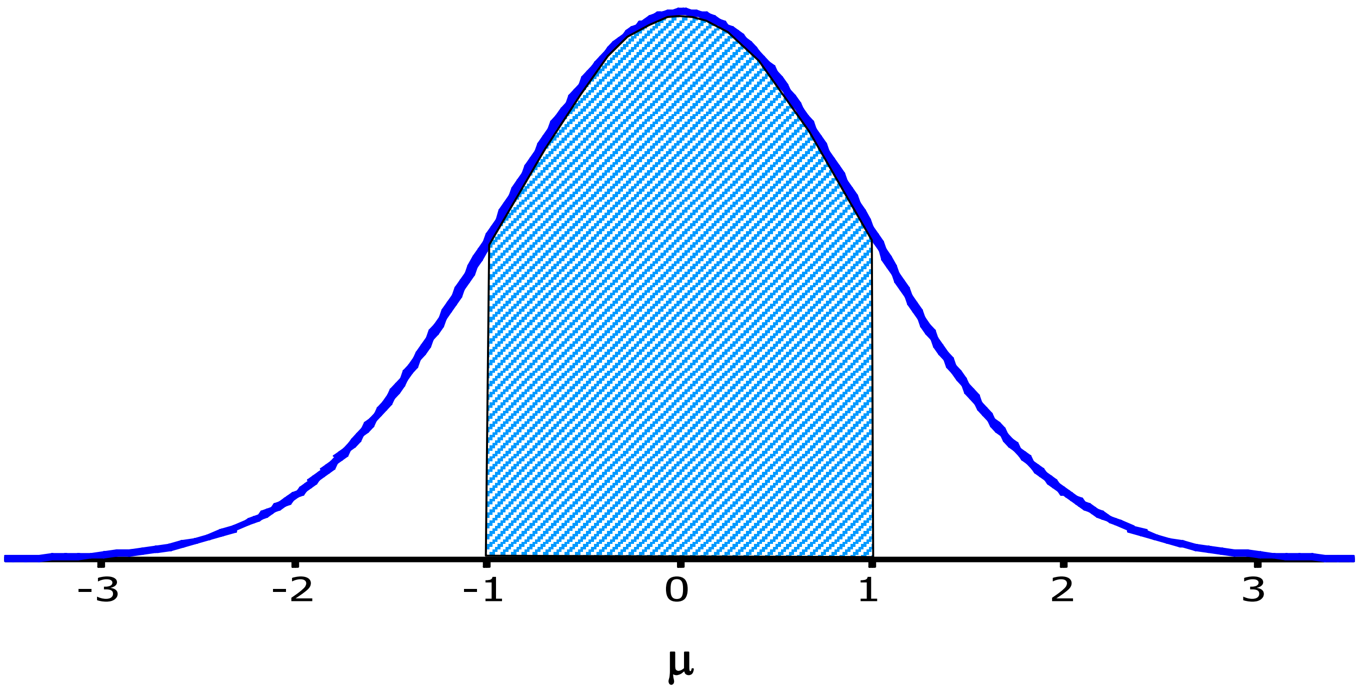
The Normal Distribution

Standard Deviation

Distance to point of inflection on the curve

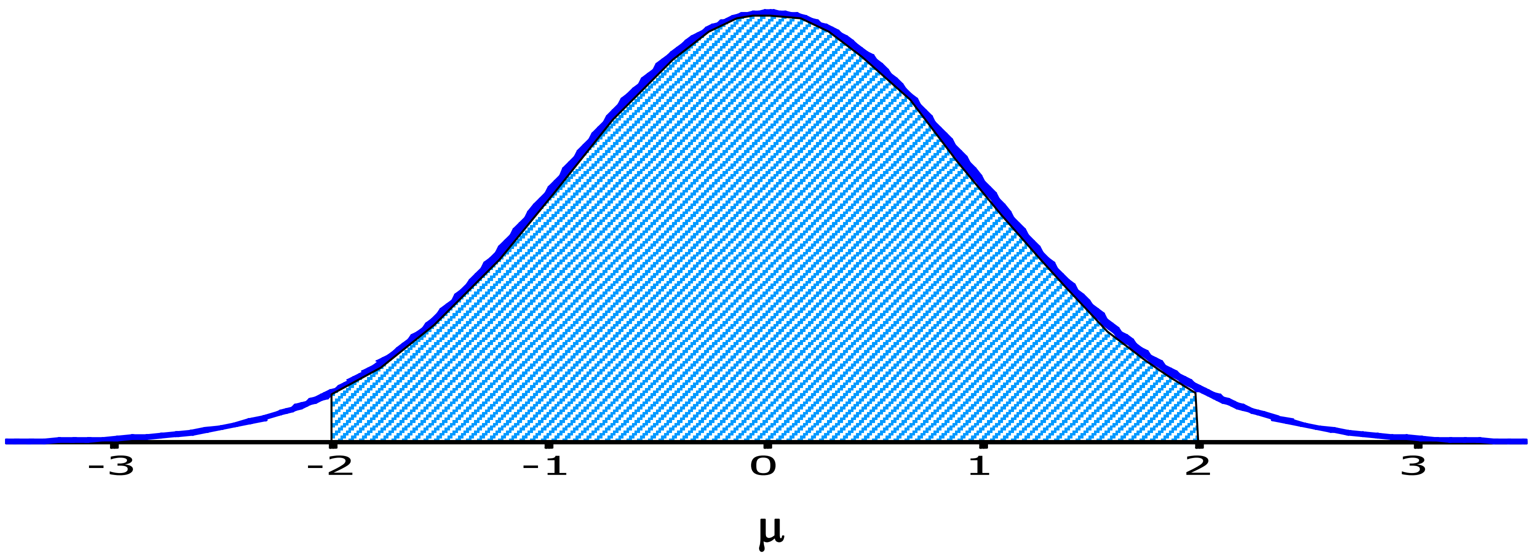


The Normal Distribution



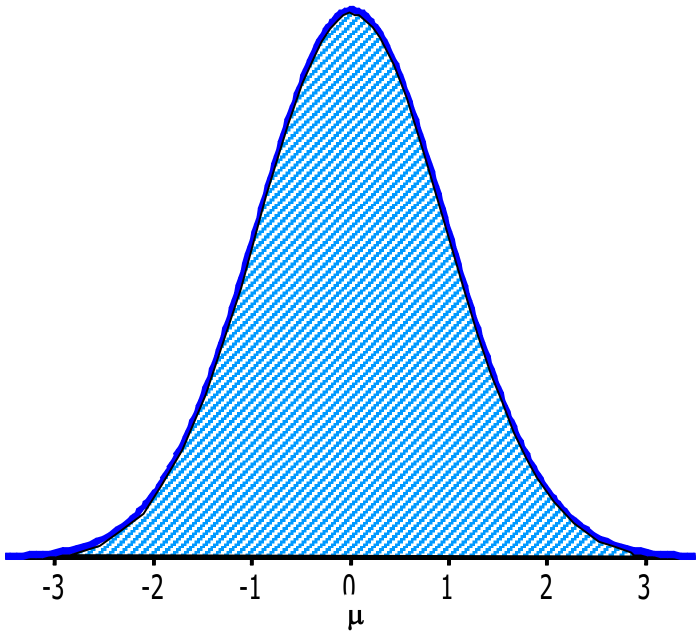
~ 68% of the data lie within 1 standard deviation of the mean.

The Normal Distribution



~ 95% of the data lie within 2 standard deviations of the mean.

The Normal Distribution



~ 99.7% of the data lie within 3 standard deviations of the mean.

+ / - 3 standard deviations is also called the natural process limits when referring to population data

Standard Normal Tables

- The proportion of a normal population that is within a given distance to the mean is the same for any normal population.
- Standard normal tables - summarize the probabilities of being less than Z units from the mean where $\mu = 0$ and $\sigma = 1$.
- To use the tables, we must convert from the units in which the population items were originally measured to the number of **standard units** denoted by Z.

The Z score

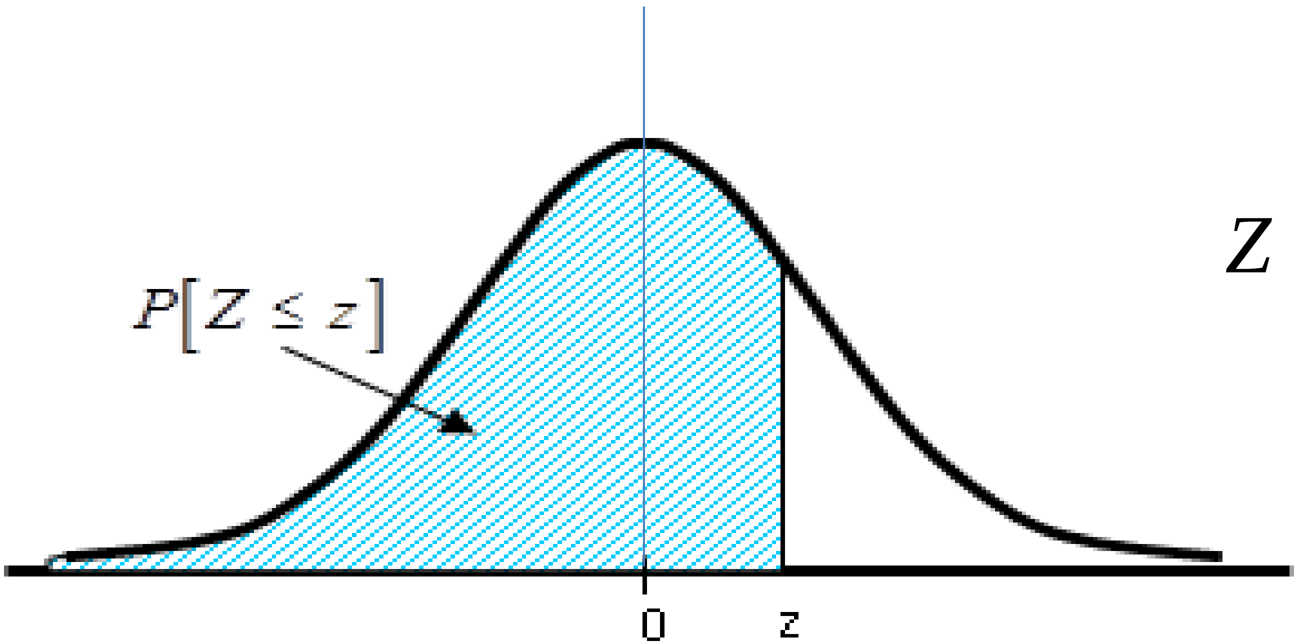
- Z - number of standard deviations an observation is from the mean

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

Standard normal distribution

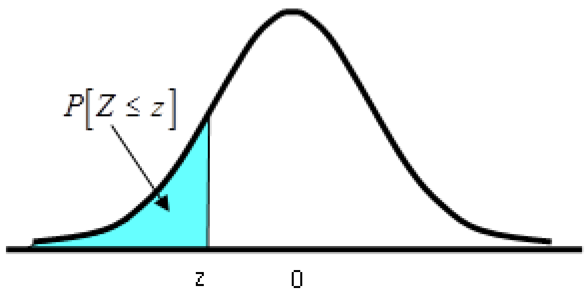
$\mu = 0$ and $\sigma = 1$

The *standardized* normal random variable, Z, is the number of standard deviations away from the mean, of a specified value, x.



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

Reading the Z table

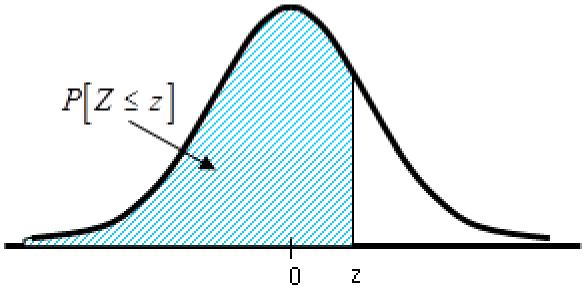


Standard Normal Probabilities

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

.Example: $P(Z<-3.17) = .0008$

Reading the Z table

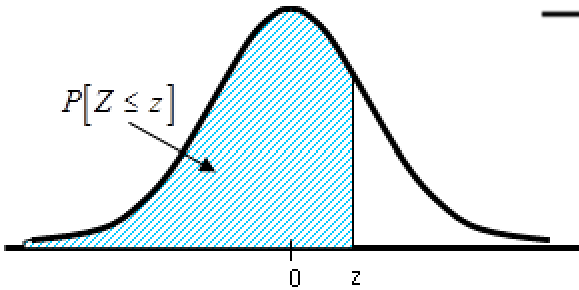


Standard Normal Probabilities

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	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

Example: $P(Z< +0.48) = 0.6844$

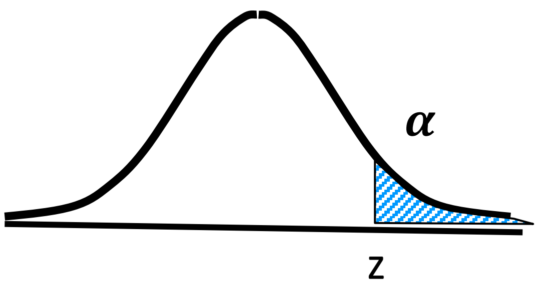
Reading the Z table



Standard Normal Probabilities

z	0	0.01	0.02	0.03	0.04	0.05	0.06
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
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123

Table A. page 1040



z	**0**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**
0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877

Includes only positive z values.
Gives area to the right of z. $\boldsymbol{\alpha} = \mathrm{p}(\mathrm{Z}>\mathrm{z})$
Since the normal curve is symmetric, $\boldsymbol{\alpha} = \mathrm{p}(\mathrm{Z}< -\mathrm{z})$

Normal Table Practice

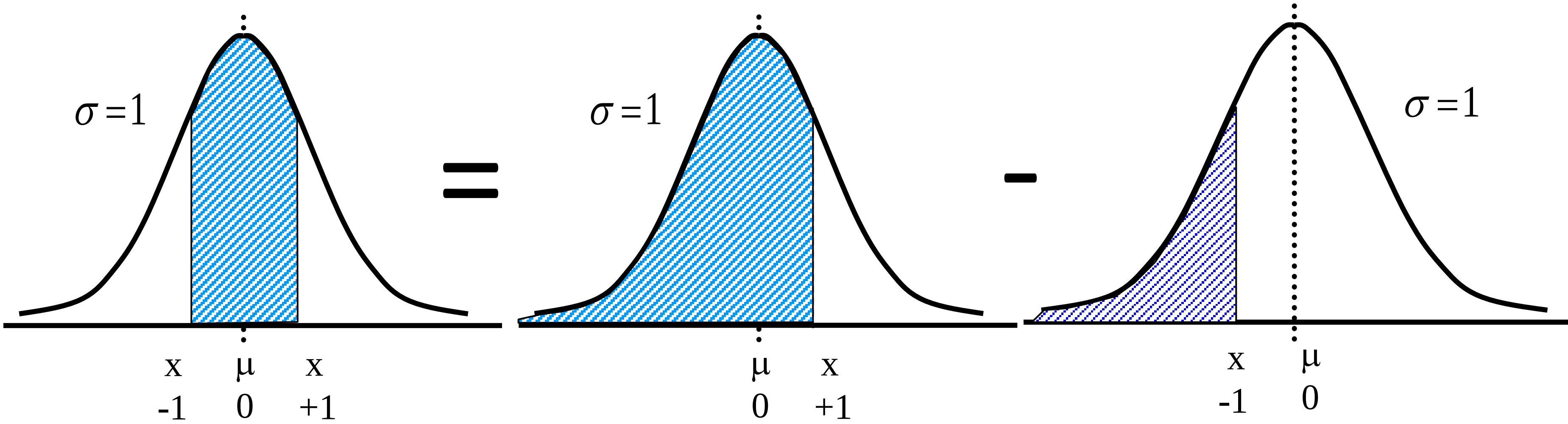
Find the area under curve to the left of $z = 0.48$.

A) 0.3156 B) 0.3446 C) 0.6554 D) 0.6844

Find the area under the curve to the right of $z = 1.35$.

A) 0.9099 B) 0.9115 C) 0.0885 D) 0.6844

Area between two values



$$Z = +1.00$$

$$Z = -1.00$$

P(Z < +1.00) = 0.8413

P(Z<-1.00) = .1587

P(-1< Z < +1) = P(Z<+1) – P(Z—1) = 0.8413 - .1587 = .6826 = 68.26%

Normal Table Practice

What is the area under the curve between $z = -2.20$ and $z = +1.35$?

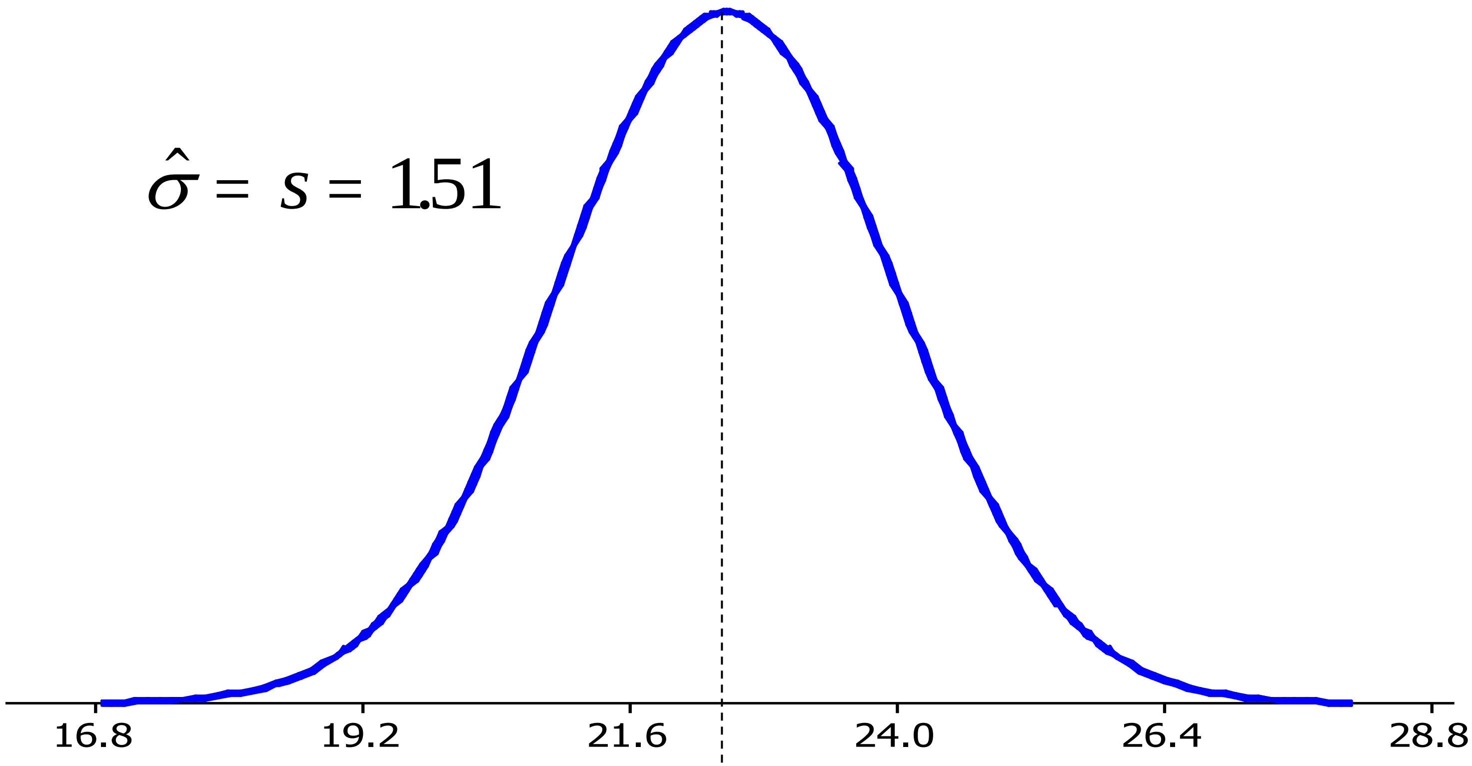
A) 0.9115 B) 0.8976 C) 0.7634 D) 0.5043

What *z*-score corresponds to the 75th percentile of a normal curve?

A) 0.675 B) -2.35 C) 0.750 D) -2.43

Distribution of gas mileage

$$\hat{\sigma} = s = 1.51$$

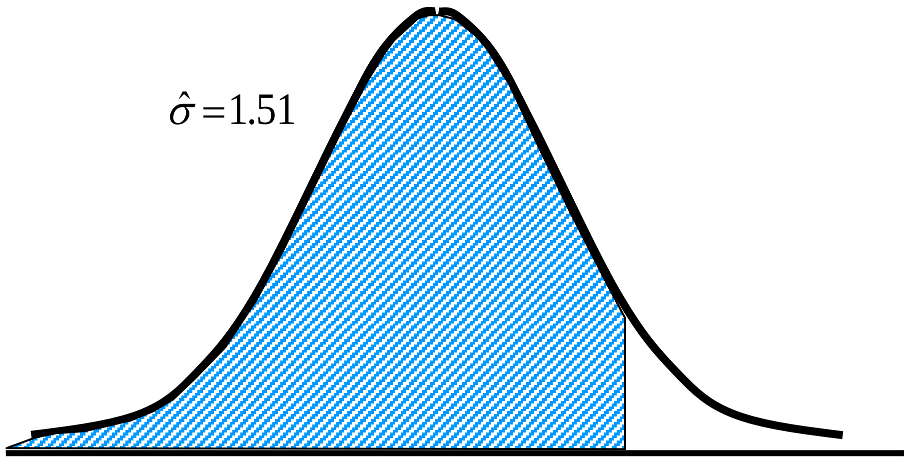


$$\hat{\mu} = 22.47$$

What is the probability that our mpg will be less than 25?

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

$$Z = \frac{25 - 22.47}{1.51} = 1.68$$



From the table:

P(Z<1.68) = 0.9535

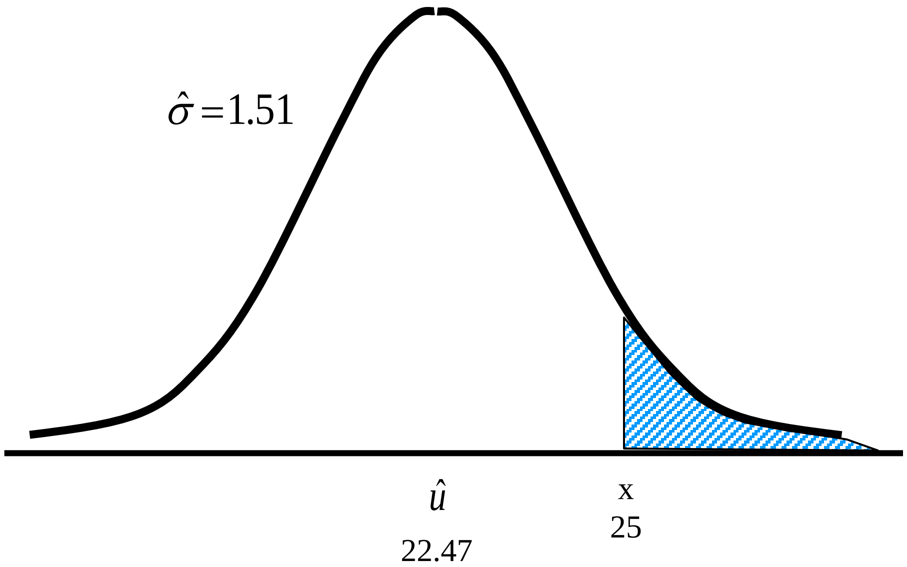
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	.08	.09
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545

The probability that the gas mileage will be less than 25 is 95.35%.

What is the probability that our mpg will be greater than 25?

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

$$Z = \frac{25 - 22.47}{1.51} = 1.68$$



From the table:

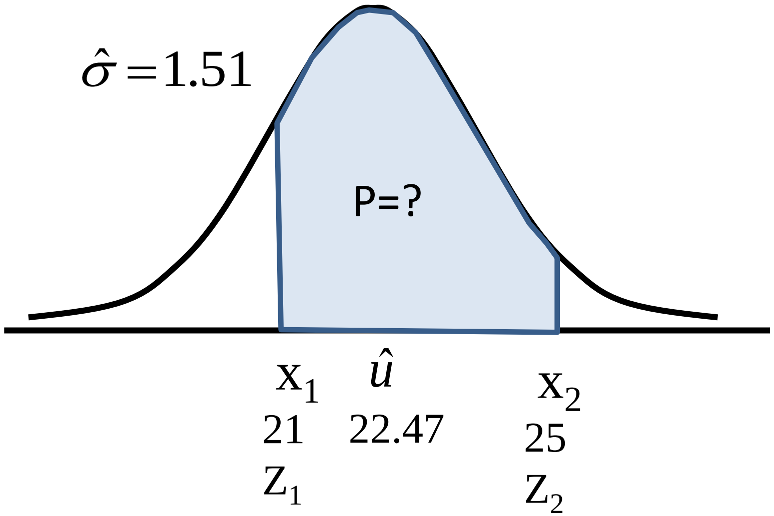
P(Z>1.68) = 1 - P(Z<1.68)

P(Z>1.68) = 1- 0.9535 = 0.0465

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	.08	.09
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

The probability that the gas mileage will be greater than 25 is 4.65%.

What is the probability that our MPG will be between 21 and 25?



$$Z_2 = \frac{25 - 22.47}{1.51} = +1.68$$

From Table: P(Z<1.68) = 0.9535

$$Z_1 = \frac{21 - 22.47}{1.51} = -0.97$$

From Table: P(Z< -0.97) = 0.1660

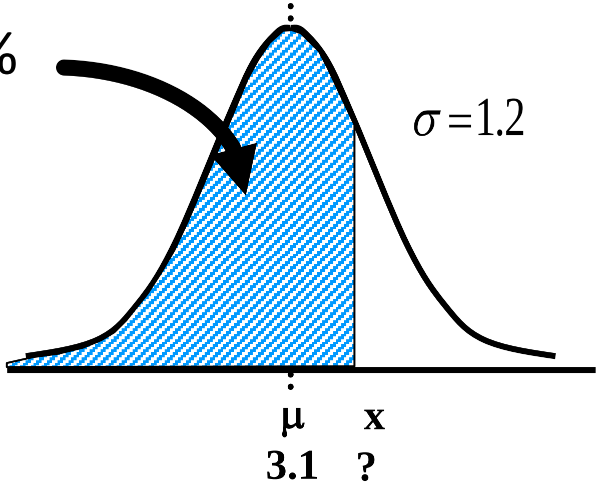
P(-.97<Z<+1.68) = P(Z<1.68) - P(Z< -0.97) = .9535-.1660=0.7875

The probability that the gas mileage will be between 21 and 25 is 78.75%.

Given a process mean of 3.1 and standard deviation of 1.2, below what value will 60% of the observations fall?

Area = 60%

What are we looking for?



What do we know?	P(Z < + z) = 0.600	$z = \dfrac{x - 3.1}{1.2}$

Look for probability in body of table, then over to find Z.

From Table: z = .25

Plug into equation and solve for x.

$$0.25 = \frac{x - 3.1}{1.2}$$

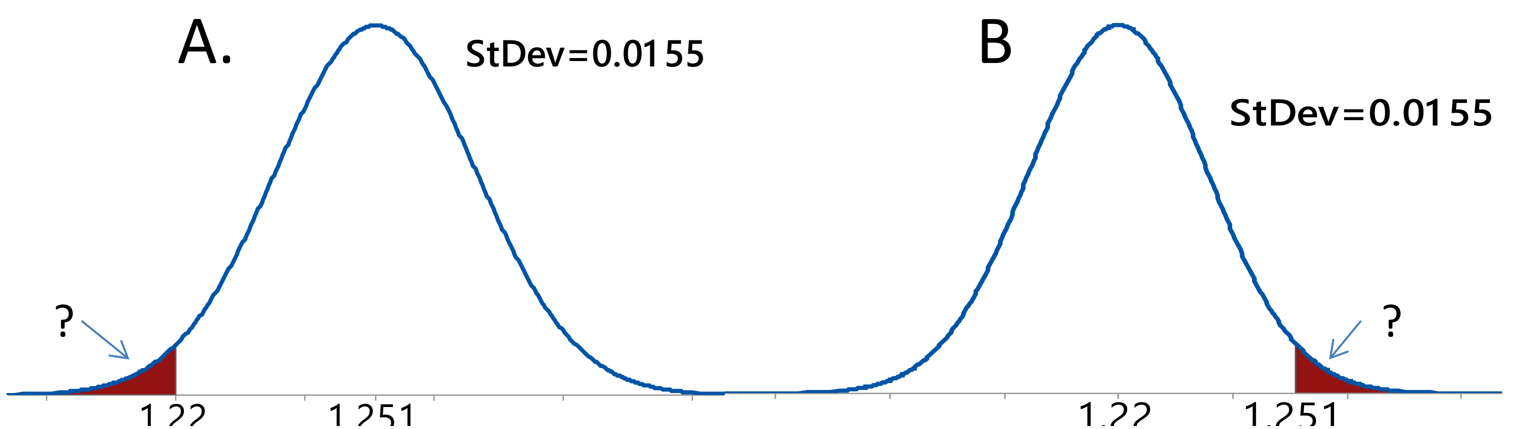
x=(.25*1.2)+3.1 = 3.4

Normal Distribution Problem

A part is machined to a critical dimension and the mean of the process is 1.251. The standard deviation of the process is .0155.

The lower specification limit for the dimension is 1.22 and all parts below the lower specification must be scrapped. What percent of the parts must be scrapped?

The diagram depicting the problem is



Normal Distribution Problem

A part is machined to a critical dimension and the mean of the process is 1.251. The standard deviation of the process is .0155.

The lower specification limit for the dimension is 1.22 and all parts below the lower specification must be scrapped. What percent of the parts must be scrapped?

Z used to work this problem is

A. +2.0 B. -2.0 C. -1.45 D. +1.45

The percent of the parts must be scrapped is

A. 1.5% B. 1.9% C. 3.5% D. 2.28%

Normal Distribution Problem

A part is machined to a critical dimension and the mean of the process is 1.251. The standard deviation of the process is .0155.

The upper specification limit for the dimension is 1.28 and all parts over the upper specification must be reworked. What percent of the parts must be reworked?

Z used to work this problem is

A. +2.15 B. -2.15 C. -1.87 D. +1.87

The percent of the parts must be reworked is

A. 1.5% B. 2.7% C. 3.1% D. 3.8%

Normal Distribution Problem

An improvement project focused on the process in the previous problem. The process mean is now 1.25 and is centered between the specification limits. The variation has been reduced, so the standard deviation is now .010.

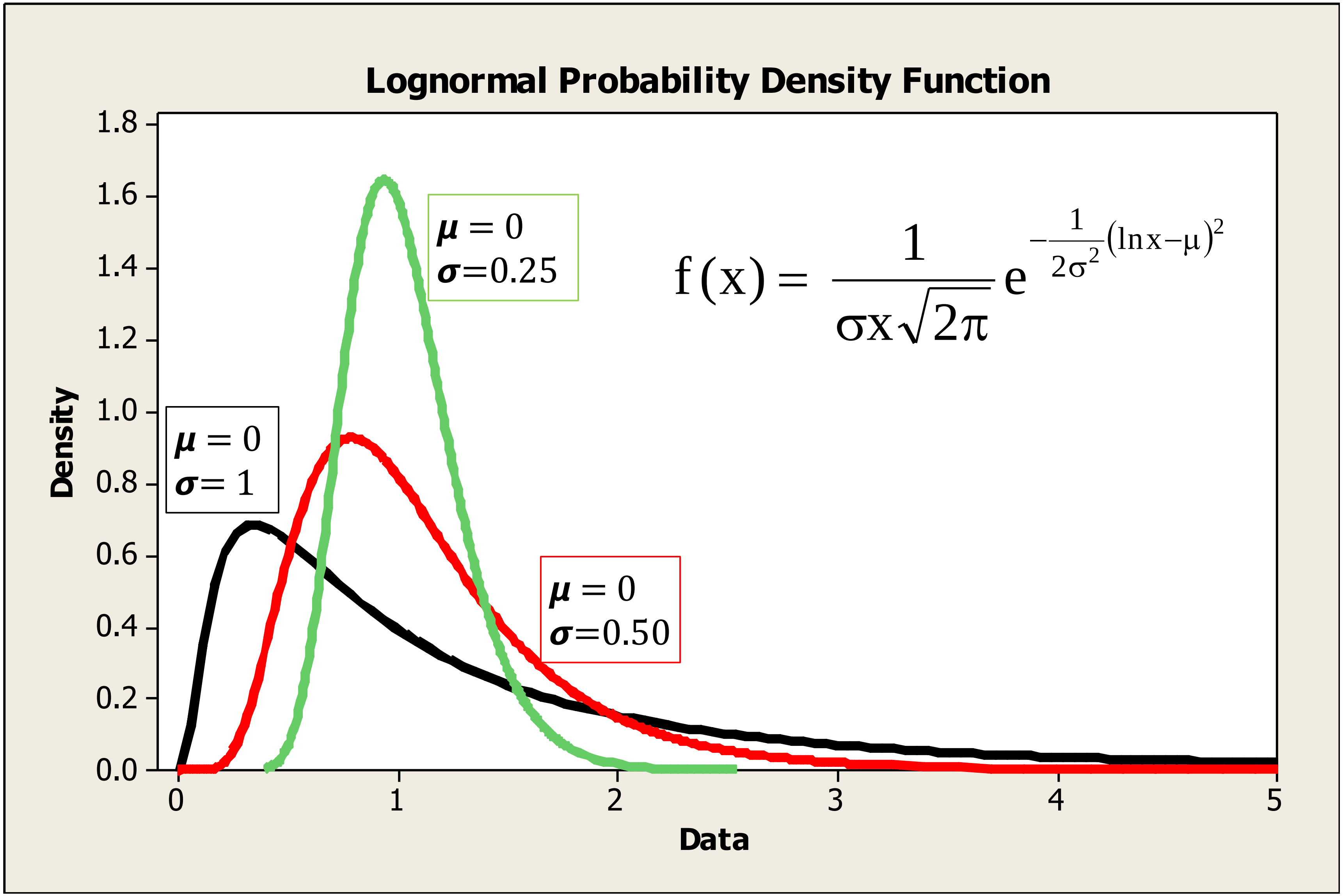
Given the specification limits of 1.22 and 1.28, what percent of the parts now meet the specification?

The percent of the parts now meeting specs is

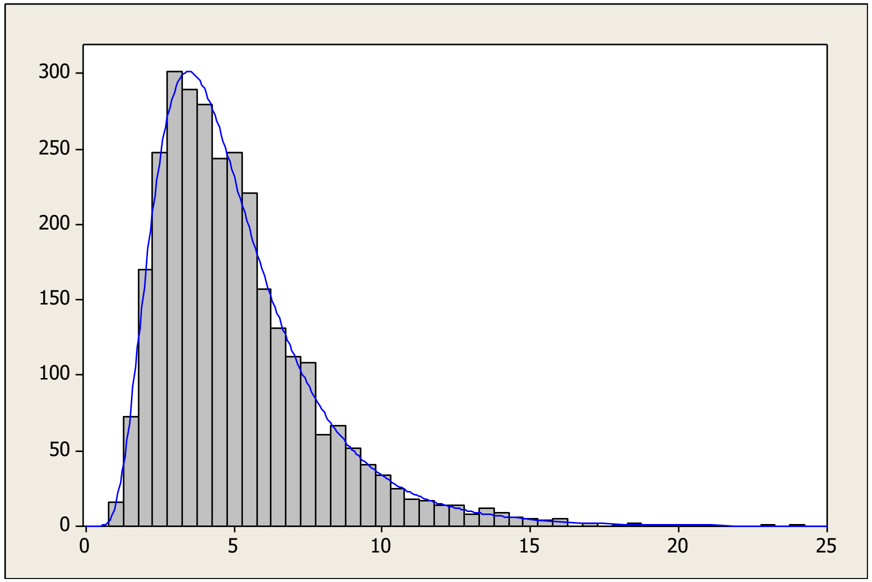
A. 98.32%   B. 99.74%   C. 96.50%   D. 97.32%

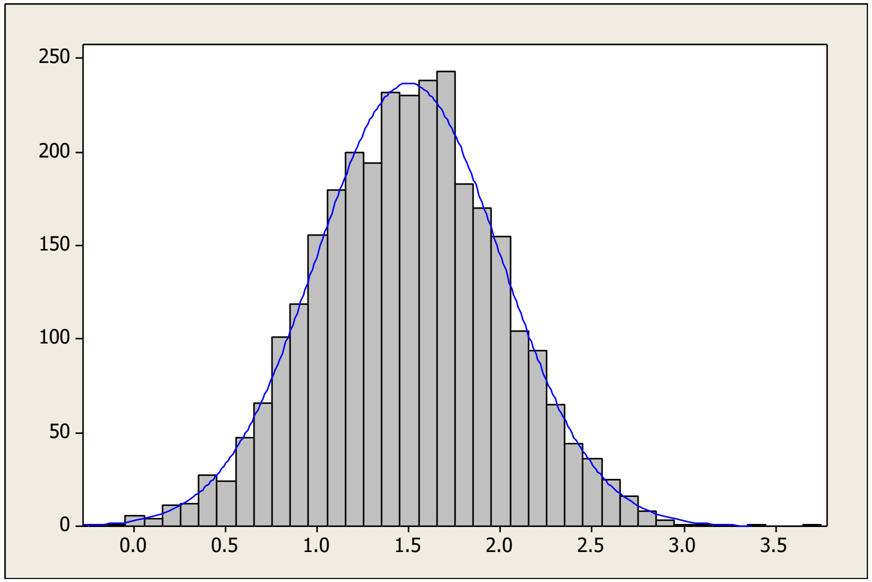
7.14 Lognormal Distribution

- For data that contain outliers, the normal distribution is generally not appropriate. The **lognormal distribution**, which is related to the normal distribution, is often a good choice for these data sets.


Lognormal Probability Density Function
Density
Data
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
$\mu = 0$
$\sigma = 0.25$
$\mu = 0$
$\sigma = 1$
$\mu = 0$
$\sigma = 0.50$
$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2}$$

Lognormal Distribution





3000 observations: **y**

Lognormal Distribution

Parameters $\boldsymbol{\mu} = 1.5 \quad \boldsymbol{\sigma} = 0.5$

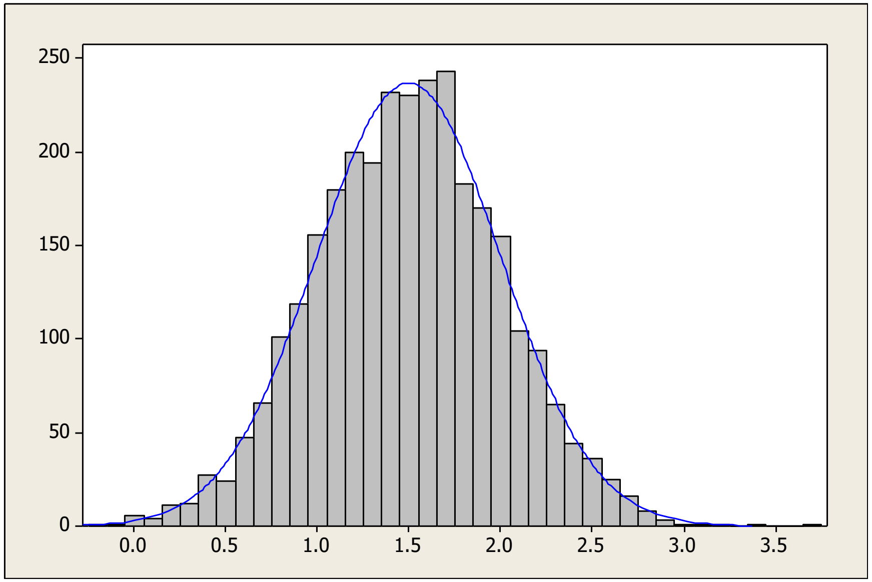
Same 3000 observations: x = **ln y**

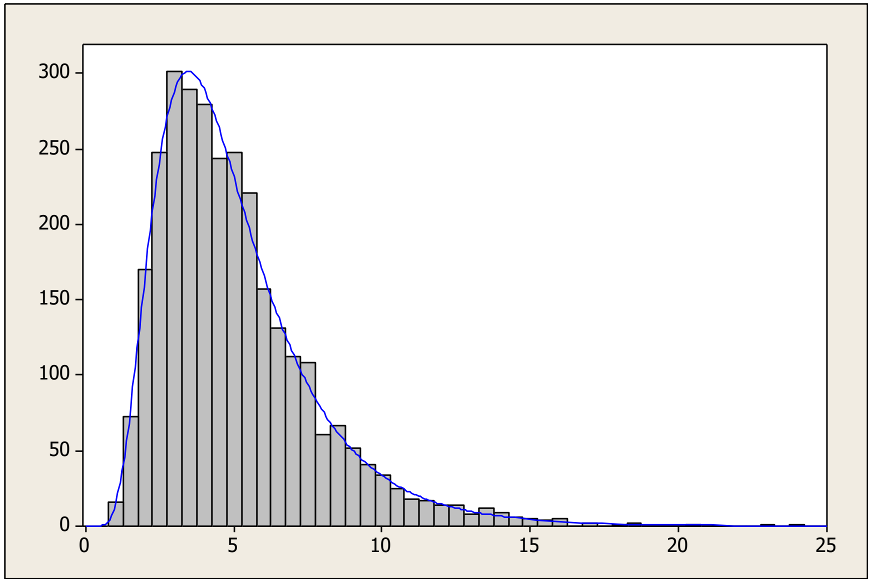
Normal Distribution

$\boldsymbol{\mu} = 1.5 \quad \boldsymbol{\sigma} = 0.5$

If Y has the lognormal distribution with parameters μ and σ^2, then the random variable $X = \ln Y$ has the $N(\mu, \sigma^2)$ distribution.

Lognormal Distribution





Same 3000 observations: x

Normal Distribution
$\boldsymbol{\mu} = 1.5 \quad \boldsymbol{\sigma} = 0.5$

3000 observations: $\mathbf{y} = \boldsymbol{e^x}$

Lognormal Distribution
Parameters $\boldsymbol{\mu} = 1.5 \quad \boldsymbol{\sigma} = 0.5$

If $X \sim N(\mu, \sigma^2)$, then the random variable $Y = e^X$ has the lognormal distribution with parameters μ and σ^2

Lognormal Problem

A process is know to follow a lognormal distribution with parameters $\boldsymbol{\mu} = 1.5$ and $\boldsymbol{\sigma} = 0.5$. What is the probability that an observed value will be less than 5?

P(y<5) = P(X< ln 5)

where X is normally distributed with $\boldsymbol{\mu} = 1.5$ and $\boldsymbol{\sigma} = 0.5$.

$\ln 5 =$ 1.609

To get P(x<1.609), calculate z and use the Z table.

$$z = \frac{x - \mu}{\sigma} = \frac{1.609 - 1.5}{0.5} = +.22$$

P(x<1.609) = P(Z<+.22) = .5871 *(from standard normal table)*

Lognormal Distribution

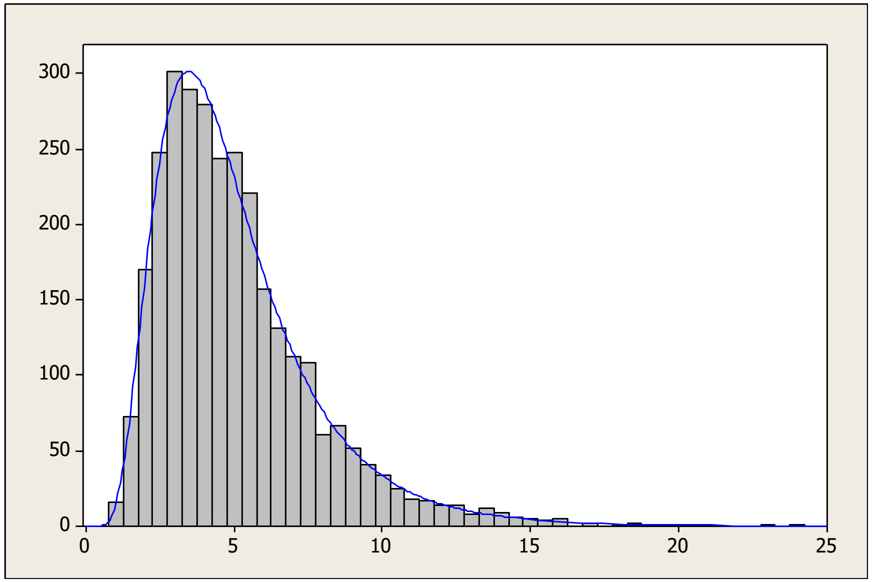
- Mean or Expected Value of Y

$$E(Y) = e^{\mu + \frac{1}{2}\sigma^2}$$

- Variance Y

$$V(Y) = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$$

Lognormal Distribution



$$E(Y) = e^{\mu + \frac{1}{2}\sigma^2}$$

$$= e^{1.5 + \frac{1}{2}(.5)^2} = 5.08$$

3000 observations - y

Lognormal Distribution
Parameters $\boldsymbol{\mu} = 1.5$ $\boldsymbol{\sigma} = 0.5$

Lognormal Problem

$$E(Y) = e^{\mu + \frac{1}{2}\sigma^2}$$

The time, in minutes, to repair an automatic component insertion machine used in circuit card assembly is know to follow a log normal distribution with parameters $\boldsymbol{\mu} = 5$ and $\boldsymbol{\sigma} = 1.5$.

What is the expected time for a repair?

A. 1408 B 314 C. 108 D. 457

Lognormal Problem

The time, in minutes, to repair an automatic component insertion machine used in circuit card assembly is know to follow a log normal distribution with parameters $\boldsymbol{\mu} = 5$ and $\boldsymbol{\sigma} = 1.5$.

What is the probability that a repair can be performed within 45 minutes?

A. 0.7881 B 0.2119 C. 0.4022 D. 0.3310

Related Assignments

Practice Problems: Ch 7 Problems 3, 4

Assignment: Minitab Exercise 3