IIT ROORKEE

NPTEL ONLINE CERTIFICATION COURSE

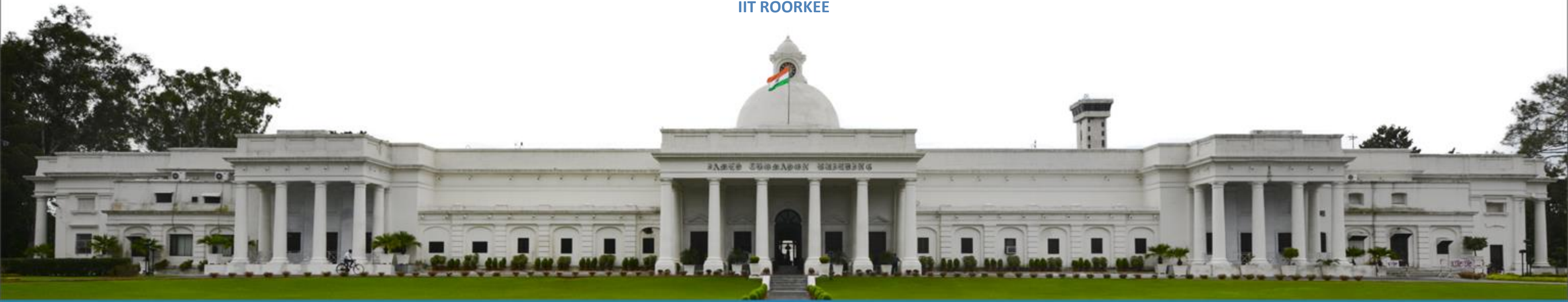
Data Analytics with Python

Lecture 10: Probability Distributions-III

Dr. A. Ramesh

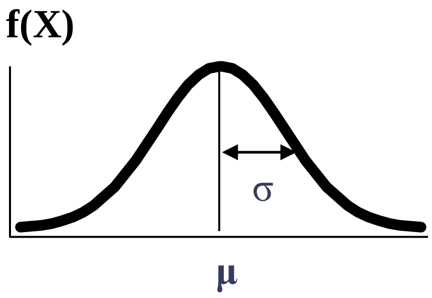
DEPARTMENT OF MANAGEMENT

IIT ROORKEE



The Normal Distribution: Properties

- ‘Bell Shaped’
- Symmetrical
- Mean, Median and Mode are equal
- Location is characterized by the mean, μ
- Spread is characterized by the standard deviation, σ
- The random variable has an infinite theoretical range: -∞ to +∞



Mean = Median = Mode

The Normal Distribution: Density Function

The formula for the normal probability density function is

$$f(X) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{(X-\mu)}{\sigma}\right)^2}$$

Where e = the mathematical constant approximated by 2.71828

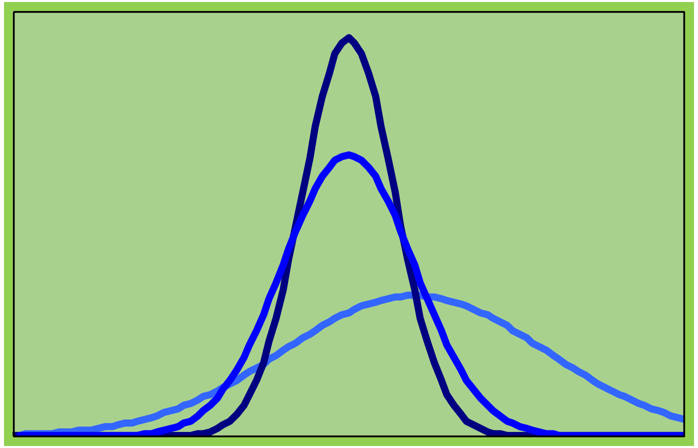
π = the mathematical constant approximated by 3.14159

μ = the population mean

σ = the population standard deviation

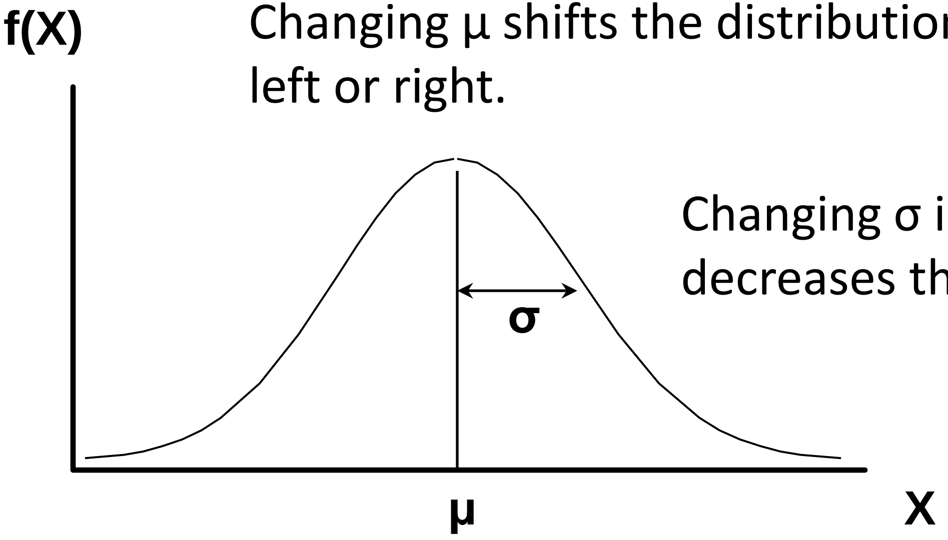
X = any value of the continuous variable

The Normal Distribution: Shape



By varying the parameters μ and σ, we obtain different normal distributions

The Normal Distribution: Shape



Changing μ shifts the distribution left or right.

Changing σ increases or decreases the spread.

The Standardized Normal Distribution

- Any normal distribution (with any mean and standard deviation combination) can be transformed into the standardized normal distribution (Z).
- Need to transform X units into Z units.
- The standardized normal distribution has a mean of 0 and a standard deviation of 1.

The Standardized Normal Distribution

- Translate from X to the standardized normal (the “Z” distribution) by subtracting the mean of X and dividing by its standard deviation:

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

The Standardized Normal Distribution: Density Function

- The formula for the standardized normal probability density function is

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

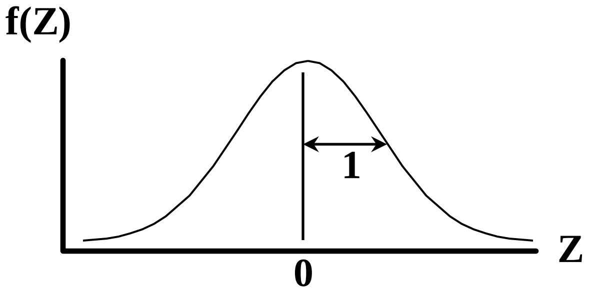
Where e = the mathematical constant approximated by 2.71828

π = the mathematical constant approximated by 3.14159

Z = any value of the standardized normal distribution

The Standardized Normal Distribution: Shape

- Also known as the “Z” distribution
- Mean is 0
- Standard Deviation is 1



Values above the mean have positive Z-values, values below the mean have negative Z-values

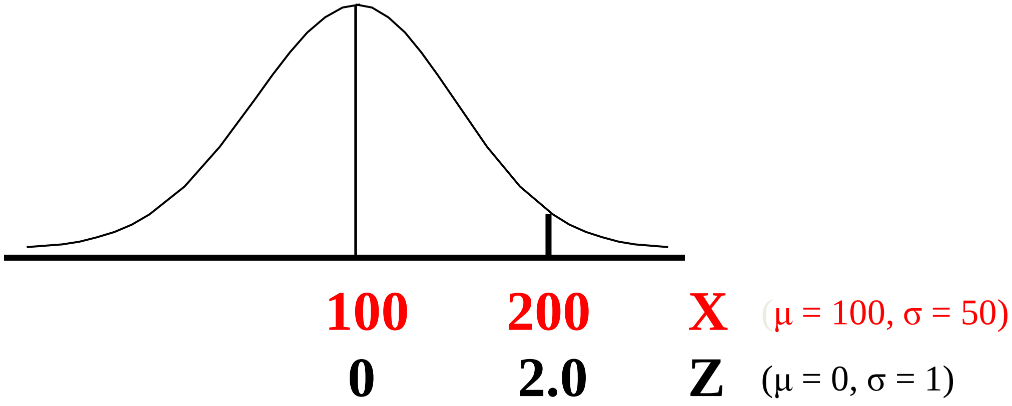
The Standardized Normal Distribution: Example

- If X is distributed normally with mean of 100 and standard deviation of 50, the Z value for $X = 200$ is

$$Z = \frac{X - \mu}{\sigma} = \frac{200 - 100}{50} = 2.0$$

- This says that $X = 200$ is two standard deviations (2 increments of 50 units) above the mean of 100.

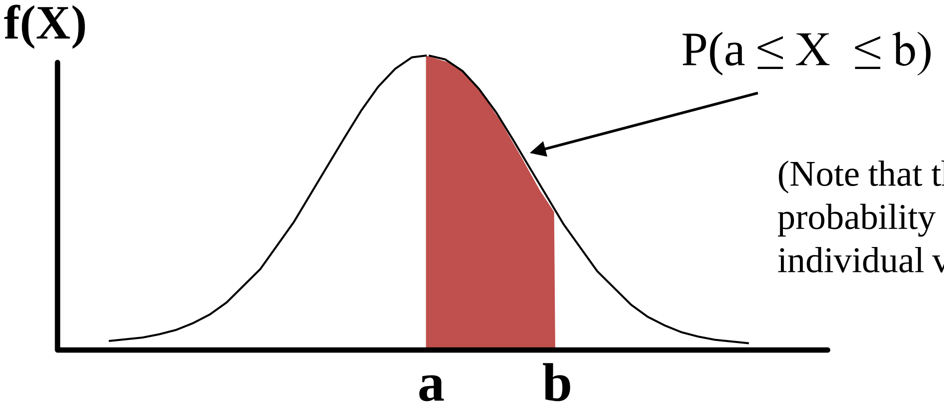
The Standardized Normal Distribution: Example



Note that the distribution is the same, only the scale has changed. We can express the problem in original units (X) or in standardized units (Z)

Normal Probabilities

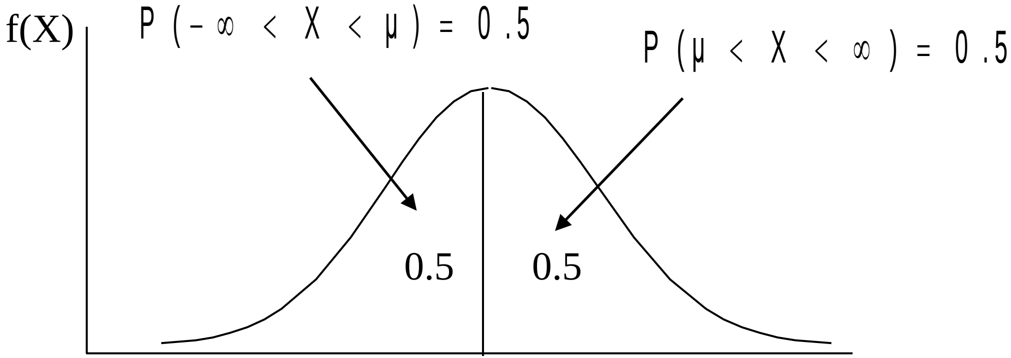
Probability is measured by the area under the curve



(Note that the probability of any individual value is zero)

Normal Probabilities

The total area under the curve is 1.0, and the curve is symmetric, so half is above the mean, half is below.

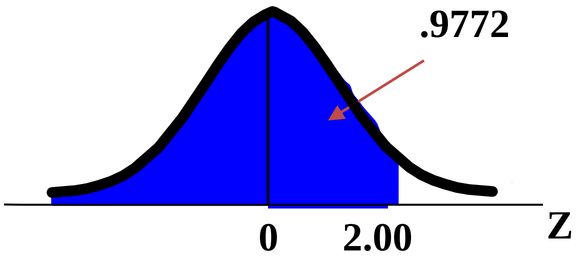


$$P(-\infty < X < \infty) = 1.0$$

Normal Probability Tables

Example:

$P(Z < 2.00) = .9772$



Normal Probability Tables

The column gives the value of Z to the second decimal point

Z	0.00	0.01	0.02 …
0.0			
0.1			
⋮			
2.0	**.9772**		

The row shows the value of Z to the first decimal point

The value within the table gives the probability from Z = $-\infty$ up to the desired Z value.

P(Z < 2.00) = .9772

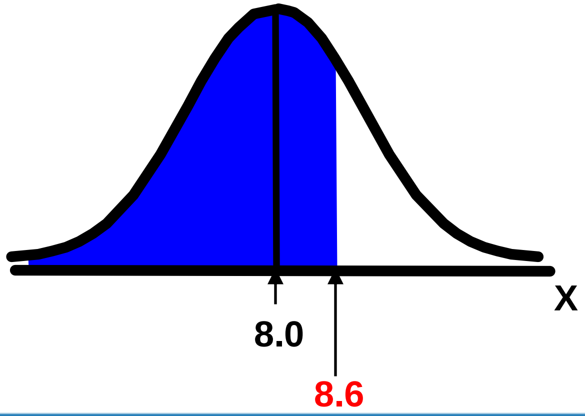
Finding Normal Probability Procedure

To find $P(a < X < b)$ when X is distributed normally:

- Draw the normal curve for the problem in terms of X.
- Translate X-values to Z-values.
- Use the Standardized Normal Table.

Finding Normal Probability: Example

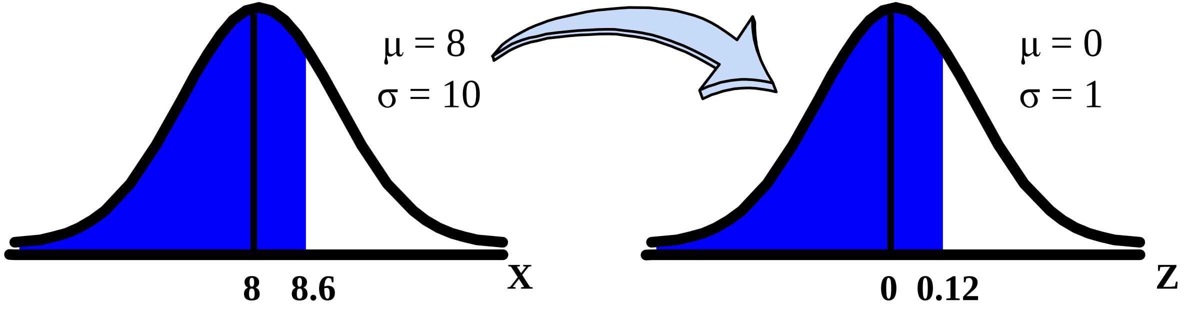
- Let X represent the time it takes (in seconds) to download an image file from the internet.
- Suppose X is normal with mean 8.0 and standard deviation 5.0
- Find $P(X < 8.6)$



Finding Normal Probability: Example

- Suppose X is normal with mean 8.0 and standard deviation 5.0. Find P(X < 8.6).

$$Z = \frac{X - \mu}{\sigma} = \frac{8.6 - 8.0}{5.0} = 0.12$$



P(X < 8.6) P(Z < 0.12)

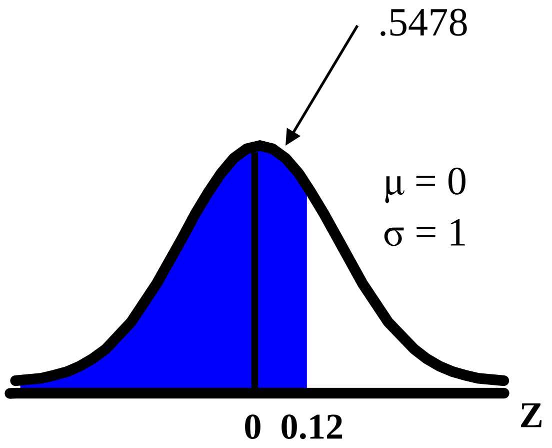
Finding Normal Probability: Example

Standardized Normal Probability Table (Portion)

Z	.00	.01	.02
0.0	.5000	.5040	.5080
0.1	.5398	.5438	.5478
0.2	.5793	.5832	.5871
0.3	.6179	.6217	.6255

$P(X < 8.6)$

$= P(Z < 0.12)$

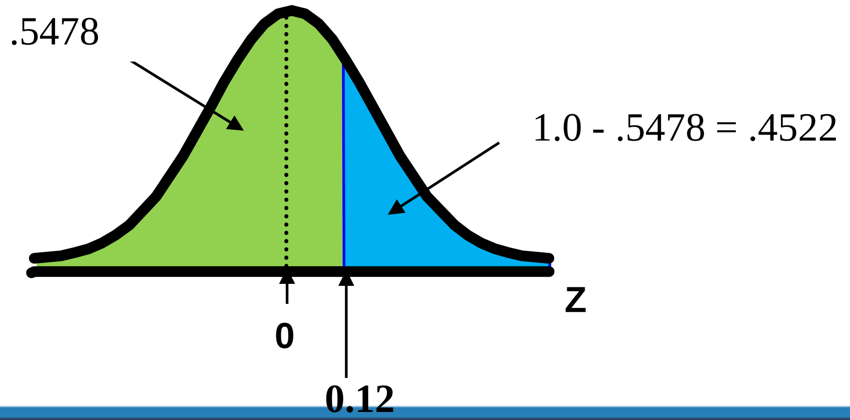


Finding Normal Probability: Example

- Find P(X > 8.6)…

$$P(X > 8.6) = P(Z > 0.12) = 1.0 - P(Z \le 0.12)$$

$$= 1.0 - .5478 = .4522$$



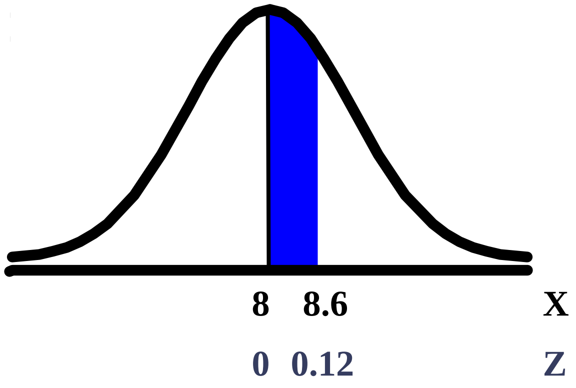
Finding Normal Probability: Between Two Values

- Suppose X is normal with mean 8.0 and standard deviation 5.0. Find P(8 < X < 8.6)

Calculate Z-values:

$$Z = \frac{X - \mu}{\sigma} = \frac{8 - 8}{5} = 0$$

$$Z = \frac{X - \mu}{\sigma} = \frac{8.6 - 8}{5} = 0.12$$



$$P(8 < X < 8.6)$$

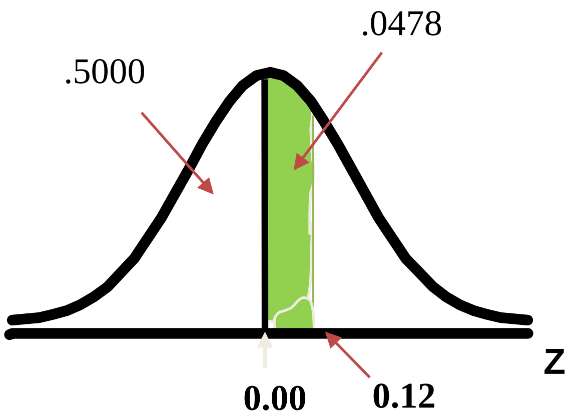
$$= P(0 < Z < 0.12)$$

Finding Normal Probability Between Two Values

- Standardized Normal Probability
- Table (Portion)

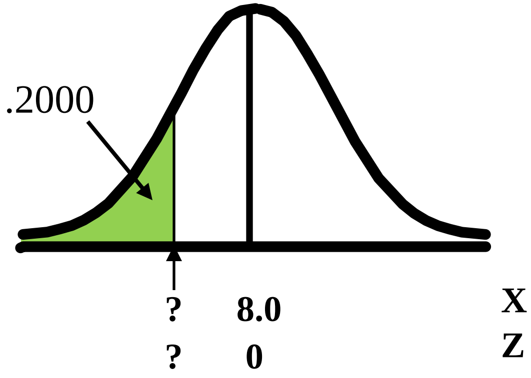
Z	.00	.01	.02
0.0	.5000	.5040	.5080
0.1	.5398	.5438	.5478
0.2	.5793	.5832	.5871
0.3	.6179	.6217	.6255

$$P(8 < X < 8.6)$$
$$= P(0 < Z < 0.12)$$
$$= P(Z < 0.12) - P(Z \le 0)$$
$$= .5478 - .5000 = .0478$$



Given Normal Probability: Find the X Value

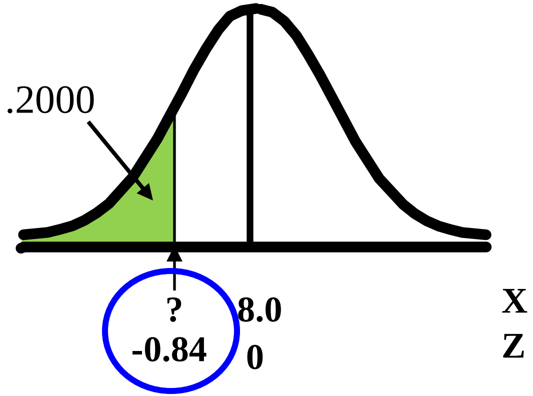
- Let X represent the time it takes (in seconds) to download an image file from the internet.
- Suppose X is normal with mean 8.0 and standard deviation 5.0
- Find X such that 20% of download times are less than X.



Given Normal Probability, Find the X Value

- First, find the Z value corresponds to the known probability using the table.

Z	….	.03	.04	.05
-0.9	….	.1762	.1736	.1711
-0.8	….	.2033	.2005	.1977
-0.7	….	.2327	.2296	.2266



Given Normal Probability, Find the X Value

- Second, convert the Z value to X units using the following formula.

$$
\begin{aligned}
X &= \mu + Z\sigma \\
&= 8.0 + (-0.84)5.0 \\
&= 3.80
\end{aligned}
$$

So 20% of the download times from the distribution with mean 8.0 and standard deviation 5.0 are less than 3.80 seconds.

Assessing Normality

- It is important to evaluate how well the data set is approximated by a normal distribution.
- Normally distributed data should approximate the theoretical normal distribution:
  - The normal distribution is bell shaped (symmetrical) where the mean is equal to the median.
  - The empirical rule applies to the normal distribution.
  - The interquartile range of a normal distribution is 1.33 standard deviations.

Assessing Normality

- Construct charts or graphs
  - For small- or moderate-sized data sets, do stem-and-leaf display and box-and-whisker plot look symmetric?
  - For large data sets, does the histogram or polygon appear bell-shaped?
- Compute descriptive summary measures
  - Do the mean, median and mode have similar values?
  - Is the interquartile range approximately 1.33 σ?
  - Is the range approximately 6 σ?

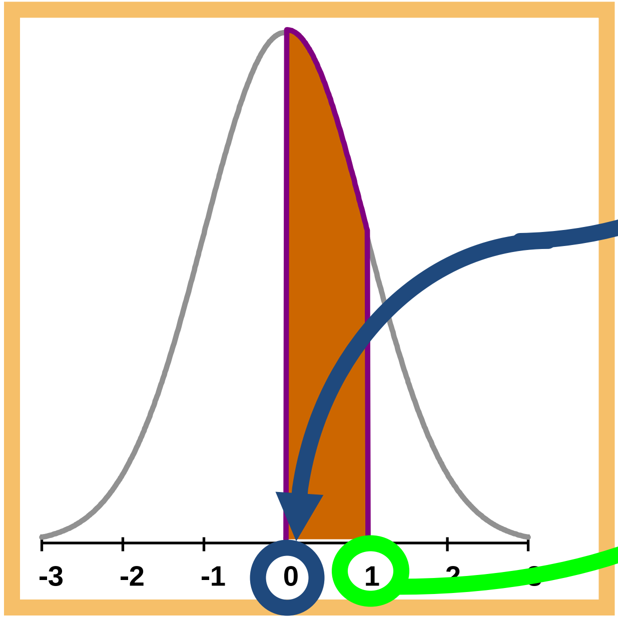
Assessing Normality

- Observe the distribution of the data set
  - Do approximately 2/3 of the observations lie within mean ± 1 standard deviation?
  - Do approximately 80% of the observations lie within mean ± 1.28 standard deviations?
  - Do approximately 95% of the observations lie within mean ± 2 standard deviations?

Z Table

	Second Decimal Place in Z									
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.10	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.20	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.30	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.90	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.00	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.10	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.20	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
2.00	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
3.00	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.40	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.50	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998

Table Lookup of a Standard Normal Probability



$$P(0 \le Z \le 1) = 0.3413$$

Z	0.00	0.01	0.02
0.00	0.0000	0.0040	0.0080
0.10	0.0398	0.0438	0.0478
0.20	0.0793	0.0832	0.0871
1.00	0.3413	0.3438	0.3461
1.10	0.3643	0.3665	0.3686
1.20	0.3849	0.3869	0.3888

Applying the Z Formula

$$X \text{ is normally distributed with } \mu = 485, \text{ and } \sigma = 105$$

$$P(485 \le X \le 600) = P(0 \le Z \le 1.10) = .3643$$

For X = 485,

$$Z = \frac{X - \mu}{\sigma} = \frac{485 - 485}{105} = 0$$

For X = 600,

$$Z = \frac{X - \mu}{\sigma} = \frac{600 - 485}{105} = 1.10$$

Z	0.00	0.01	0.02
0.00	0.0000	0.0040	0.0080
0.10	0.0398	0.0438	0.0478
1.00	0.3413	0.3438	0.3461
1.10	0.3643	0.3665	0.3686
1.20	0.3849	0.3869	0.3888

Thank You