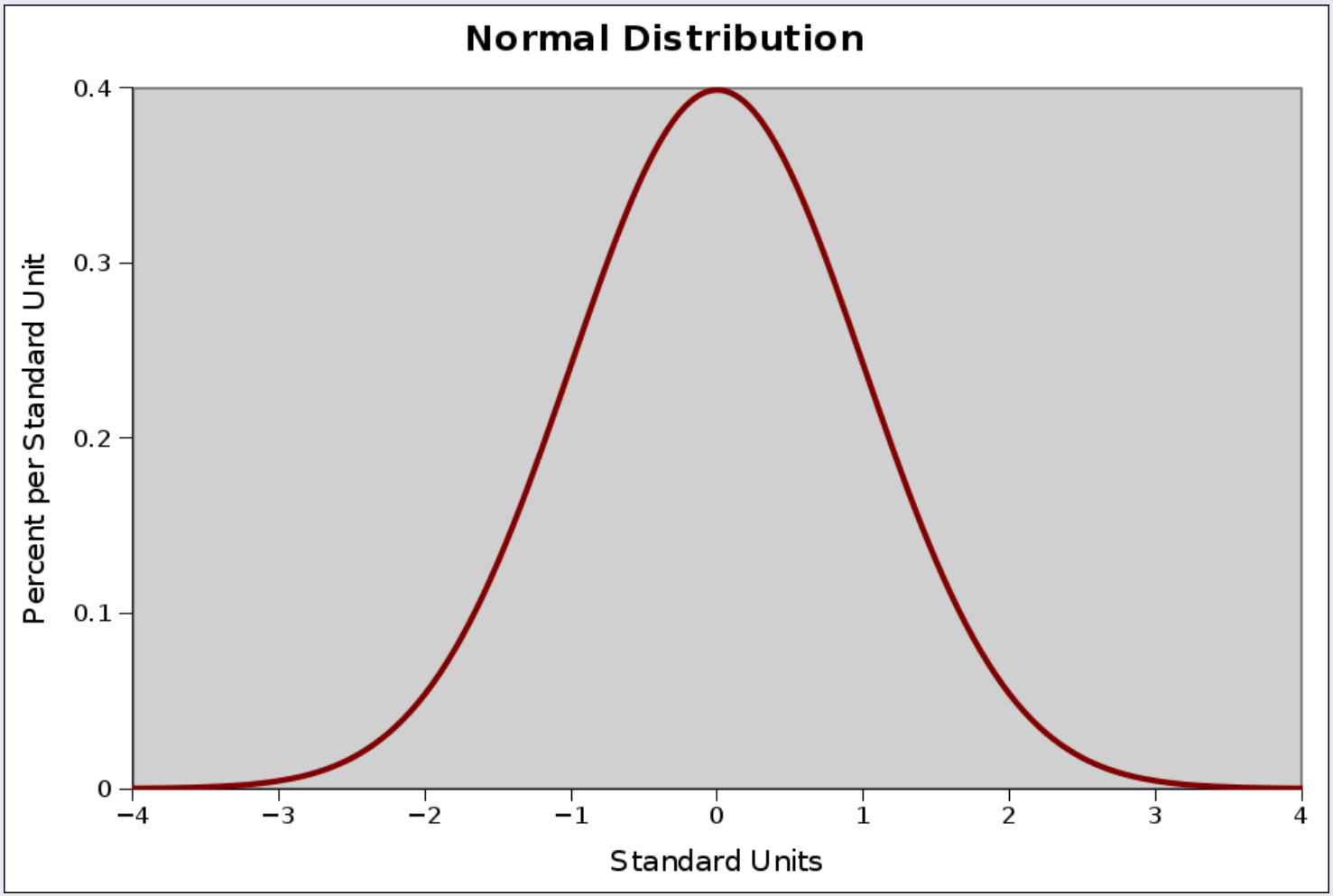
Unit 3: The Normal Approximation for Data

Marius Ionescu

09/13/2011

Fact

The normal distribution is a very common approximate distribution



Properties of the Normal Distribution

Fact

Properties of the Normal Distribution

Fact

1. *It has a single bump*

Properties of the Normal Distribution

Fact

1. *It has a single bump*
2. *It is symmetric about the average*

Properties of the Normal Distribution

Fact

1. *It has a single bump*
2. *It is symmetric about the average*
3. *Its shape depends only on average and SD*

Properties of the Normal Distribution

Fact

1. *It has a single bump*
2. *It is symmetric about the average*
3. *Its shape depends only on average and SD*
4. *68% of the area lies within 1 SD of the average*

Properties of the Normal Distribution

Fact

1. *It has a single bump*
2. *It is symmetric about the average*
3. *Its shape depends only on average and SD*
4. *68% of the area lies within 1 SD of the average*
5. *95% lies within 2 SD*

Properties of the Normal Distribution

Fact

1. *It has a single bump*
2. *It is symmetric about the average*
3. *Its shape depends only on average and SD*
4. *68% of the area lies within 1 SD of the average*
5. *95% lies within 2 SD*
6. *The height is given by*

$$\frac{1}{\sqrt{2\pi SD}} e^{-\frac{1}{2}\left(\frac{x-Avg}{SD}\right)^2}.$$

Standard Units

Definition

If X is approximated by a normal distribution with average Avg and SD, the the standard units are obtain by setting average to zero and scaling SD to one:

$$Z = \frac{X - Avg}{SD}.$$

Standard Units: Example

Example

Suppose that the average of one midterm is 65 and SD is 10. How many SDs away from the average is your score if it is

Standard Units: Example

Example

Suppose that the average of one midterm is 65 and SD is 10. How many SDs away from the average is your score if it is

- 68

Standard Units: Example

Example

Suppose that the average of one midterm is 65 and SD is 10. How many SDs away from the average is your score if it is

- 68
- 78

Standard Units: Example

Example

Suppose that the average of one midterm is 65 and SD is 10. How many SDs away from the average is your score if it is

- 68
- 78
- 50

Standard Units: Example

Example

Suppose that the average of one midterm is 65 and SD is 10. How many SDs away from the average is your score if it is

- 68
- 78
- 50
- 60

Standard Units: Example

Example

Suppose that the average of one midterm is 65 and SD is 10. How many SDs away from the average is your score if it is

- 68
- 78
- 50
- 60

Standard Units: Example

Example

Suppose that the average of one midterm is 65 and SD is 10. How many SDs away from the average is your score if it is

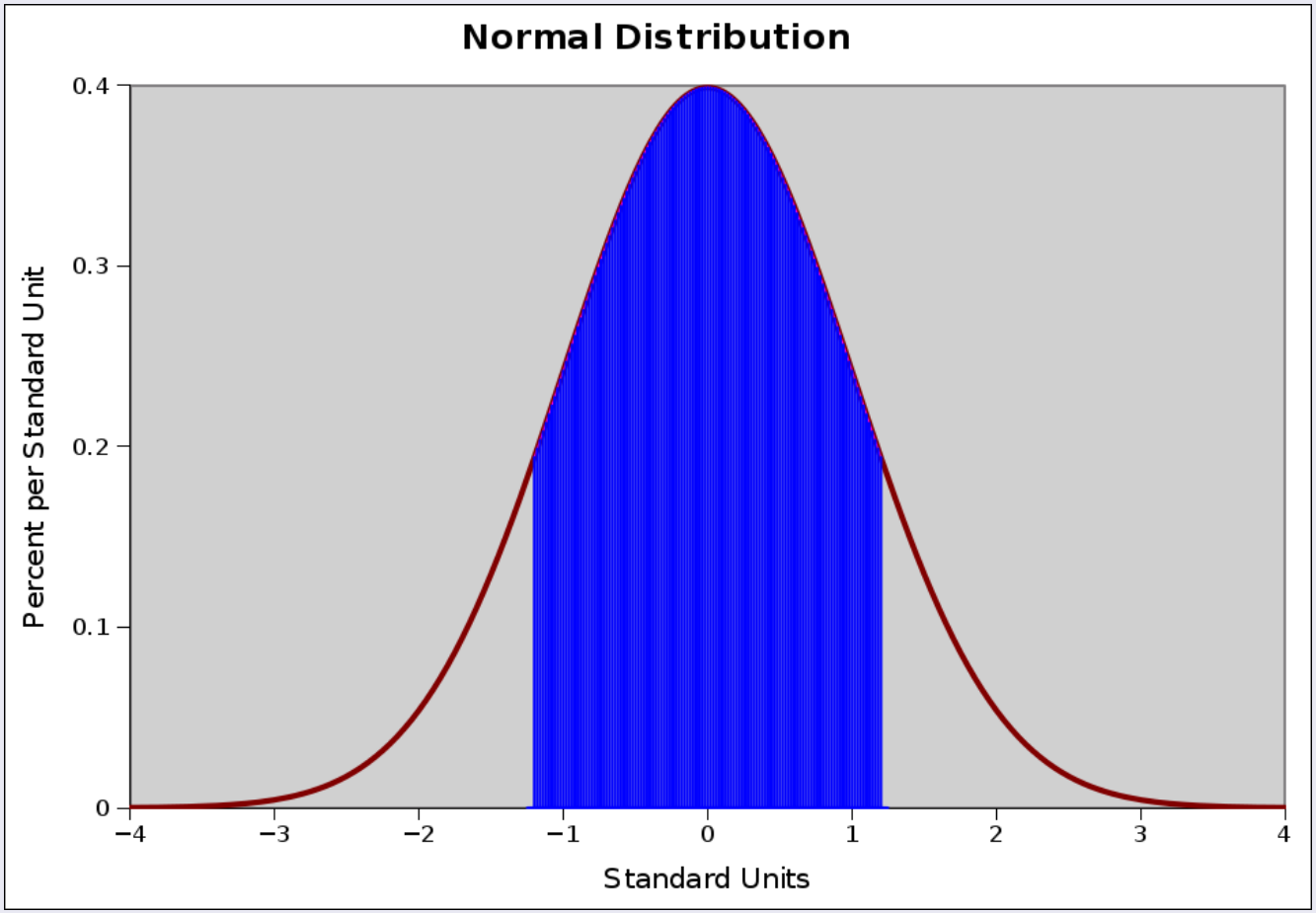
- 68
- 78
- 50
- 60

What was your score if you were 1.75 SD above the average?

Finding areas under the curve

Fact

We use a table to find the area under the normal curve: $A(z)$=Area between $-z$ and $+z$



Examples

Example

Examples

Example

- $A(1) = 68.27\%$

Examples

Example

- $A(1) = 68.27\%$
- $A(2) =$

Examples

Example

- $A(1) = 68.27\%$
- $A(2) =$

Examples

Example

- $A(1) = 68.27\%$
- $A(2) = 95.45\%$

Examples

Example

- $A(1) = 68.27\%$
- $A(2) = 95.45\%$

Example

Find the area:

Examples

Example

- $A(1) = 68.27\%$
- $A(2) = 95.45\%$

Example

Find the area:

- between $-z$ and 0

Examples

Example

- $A(1) = 68.27\%$
- $A(2) = 95.45\%$

Example

Find the area:

- between $-z$ and 0
- between $-z$ and w

Examples

Example

- $A(1) = 68.27\%$
- $A(2) = 95.45\%$

Example

Find the area:

- between $-z$ and 0
- between $-z$ and w
- below 0

Examples

Example

- $A(1) = 68.27\%$
- $A(2) = 95.45\%$

Example

Find the area:

- between $-z$ and 0
- between $-z$ and w
- below 0
- above w

Examples

Example

- $A(1) = 68.27\%$
- $A(2) = 95.45\%$

Example

Find the area:

- between $-z$ and 0
- between $-z$ and w
- below 0
- above w
- below $-w$

Example

Example

Suppose that the Math SAT has a normal distribution with an average score of 500 and SD=100.

Example

Example

Suppose that the Math SAT has a normal distribution with an average score of 500 and SD=100.

- What percent of scores are above 650?

Example

Example

Suppose that the Math SAT has a normal distribution with an average score of 500 and SD=100.

- What percent of scores are above 650?
- What percent of scores are between 600 and 650 or between 350 and 400?

Example

Example

Suppose that the Math SAT has a normal distribution with an average score of 500 and SD=100.

- What percent of scores are above 650?
- What percent of scores are between 600 and 650 or between 350 and 400?
- What percent of scores are between 600 and 650?

Example

Example

Suppose that the Math SAT has a normal distribution with an average score of 500 and SD=100.

- What percent of scores are above 650?
- What percent of scores are between 600 and 650 or between 350 and 400?
- What percent of scores are between 600 and 650?
- What percentile would 650 be?

Example

Example

Suppose that the Math SAT has a normal distribution with an average score of 500 and SD=100.

- What percent of scores are above 650?
- What percent of scores are between 600 and 650 or between 350 and 400?
- What percent of scores are between 600 and 650?
- What percentile would 650 be?
- What percentile would 400 be?

Example

Example

Suppose that the Math SAT has a normal distribution with an average score of 500 and SD=100.

- What percent of scores are above 650?
- What percent of scores are between 600 and 650 or between 350 and 400?
- What percent of scores are between 600 and 650?
- What percentile would 650 be?
- What percentile would 400 be?
- If you score in the 70th%ile, what is your score?

Example

Example

For a certain group of people the 25^{th}%ile is 62.2 inches, the 75^{th}%ile is 65.8 inches. The histogram follows a normal distribution. Find the 90^{th}%ile.

Measurement Error

Fact

Any measurement has error associated with it.

Measurement Error

Fact

Any measurement has error associated with it.

- *Measurement=Exact Value + Variation*

Measurement Error

Fact

Any measurement has error associated with it.

- *Measurement=Exact Value + Variation*
- *Error can be of two types:*

Measurement Error

Fact

Any measurement has error associated with it.

- *Measurement=Exact Value + Variation*
- *Error can be of two types:*
  1. *Bias*

Measurement Error

Fact

Any measurement has error associated with it.

- *Measurement=Exact Value + Variation*
- *Error can be of two types:*
  1. *Bias*
  2. *Chance Error*

Measurement Error

Fact

Any measurement has error associated with it.

- *Measurement=Exact Value + Variation*
- *Error can be of two types:*
  1. *Bias*
  2. *Chance Error*
- *Measurement = Exact Value + Bias + Chance Error*

Chance Error

Fact

Chance Error

Fact

- *Chance Error is due to randomness.*

Chance Error

Fact

- *Chance Error is due to randomness.*
- *It has zero average*

Chance Error

Fact

- *Chance Error is due to randomness.*
- *It has zero average*
- *If bias is zero then we can find the exact value by measuring lots of times and taking the average.*

How can we spot bias?

Fact

There is no way to tell if bias exists without comparison with another kind of experiment.

How can we spot bias?

Fact

There is no way to tell if bias exists without comparison with another kind of experiment.

Example

To determine the age of a tree one can do the following experiments:

- carbon dating
- tree ring counting
- gene variation analysis

Outliers

Definition

Outliers

Definition

- If one or two data points are wild results, statisticians often throw them out.

Outliers

Definition

- If one or two data points are wild results, statisticians often throw them out.
- They assume that something fundametally different occured for that point.

Outliers

Definition

- If one or two data points are wild results, statisticians often throw them out.
- They assume that something fundametally different occured for that point.
- For example, a bias for that measurement that is not in other measurements.