STA 1013 : Statistics through Examples

Lecture 11: Normal Distribution (Introduction, Empirical Rule)

Hwiyoung Lee

September 25, 2019

Department of Statistics, Florida State University

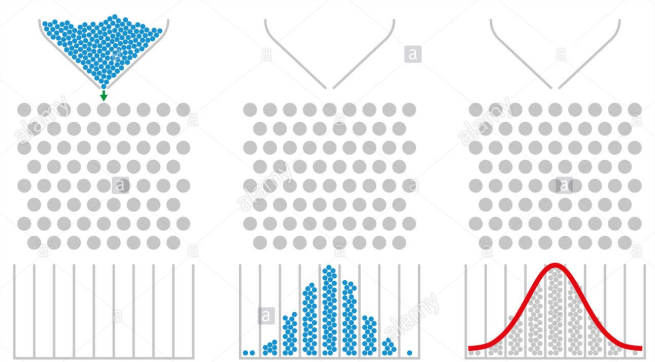
What is Normal?

Galton Board

- Beads are dropped from the top, and bounce either left or right as they hit the pegs.
- Eventually they are collected into bins at the bottom

Galton Board

The height of bead columns accumulated in the bins approximate a **bell curve**.



https://www.mathsisfun.com/data/quincunx.html

The Normal Distribution

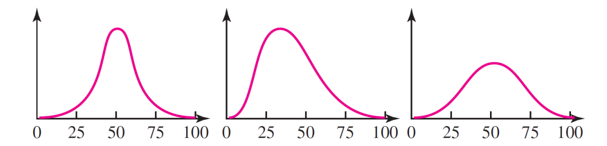
If we overlay the histogram with a smooth curve, the shape of this smooth distribution has **three** important characteristics:

1. Single peaked (Unimodal)
2. Symmetric around its single peak
3. "Bell-shaped" distribution

The smooth distribution, with these three characteristics, is called a **Normal distribution**.

Example : Normal distribution

1. Identify the distribution that is not normal



(a)

(b)

(c)

2. Decide whether the statement makes sense, or does not make sense

> **IQ scores.** The mean of a normally distributed set of IQ scores is 100, and 60% of the scores are over 105.

Example : Normal distribution

3. Which of the following variables would you expect to have a normal or nearly normal distribution?

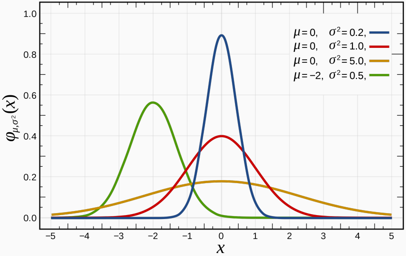
   a. Scores on a very easy test
   b. Shoe sizes of a random sample of adult women

4. Many states have lotteries that involve the random selection of digits 0, 1, 2, $\cdots$, 9. Is the distribution of those digits a normal distribution? Why or why not?

The Normal Distribution (More rigorous)

Mathematical definition :

$$\varphi(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$



The normal distribution can be fully described with just two numbers: its **mean** and its **standard deviation**.

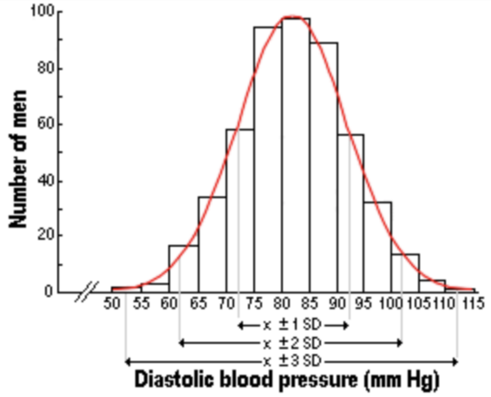
Normal Distribution

- In probability theory, the normal (or **Gaussian**) distribution is a very common continuous probability distribution

- The distributions of many natural phenomena are at least approximately normally distributed.
- It is very important in many fields of science

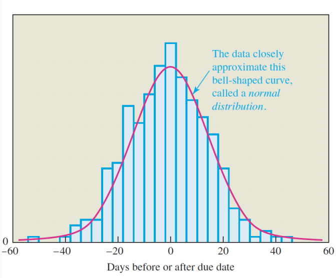
Example of Normal distribution

The distribution of diastolic blood pressure measurements among 500 men with mean=82 mm Hg and standard deviation=10 mm Hg.



Example of Normal distribution (Birth data)

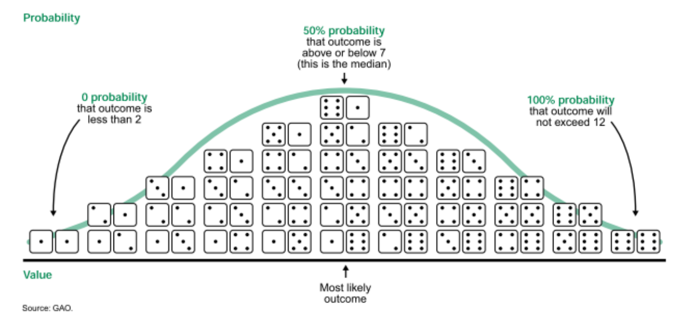
- The figure below is a histogram for a distribution of 300 natural births at Providence Memorial Hospital; the data is based on how births would be distributed without medical intervention.



- Negative numbers represent births prior to the due date, zero represents a birth on the due date, and positive numbers represent births after the due date.

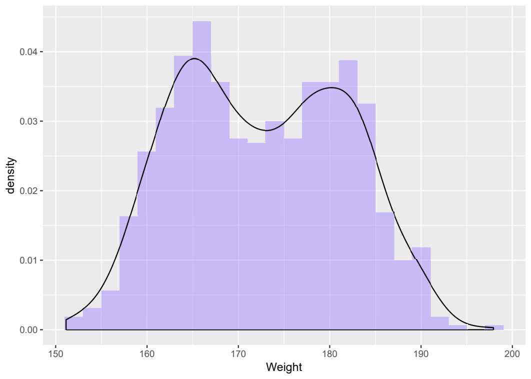
Example of Normal distribution

The distribution below is from a dice-rolling simulation in which 2 dice (sum of 2 dice) were rolled together 10,000 times.



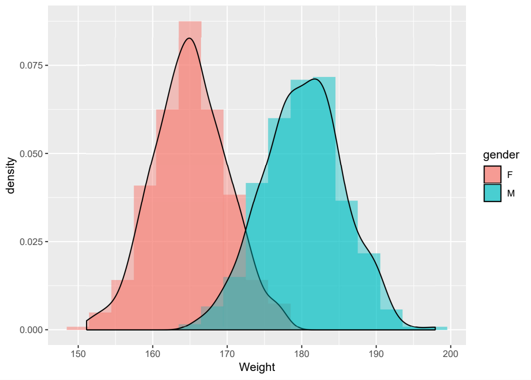
https://www.geogebra.org/m/UsoH4eNl

Example of Normal distribution



- Weight data : mean : 172.6, sd : 9.1468
- Normal distribution ? **No**

Example of Normal distribution



- Male : mean : 180.2, sd : 5.34261
- Female : mean : 165.0, sd : 4.9002

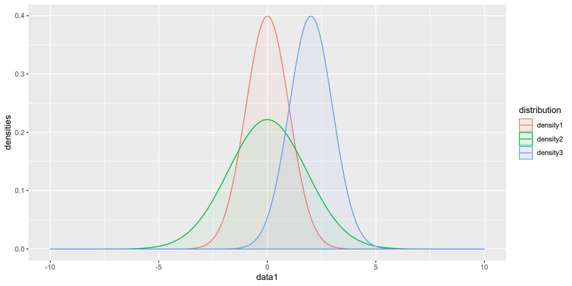
When Can We Expect a Normal Distribution?

Consider a human characteristic such as weight.

- Most men or women have weights clustered near the mean weight (for their sex), so a data set of weight has a peak at the mean weight.
- But as we consider weights increasingly far from the mean on either side, we find fewer and fewer people.
- This "tailing off" of weights far from the mean produces the two tails of the normal distribution.
- More generally, we can expect a data set to have a nearly normal distribution.

Various Normal distributions

Normal distributions are identified by two numbers (μ, σ)

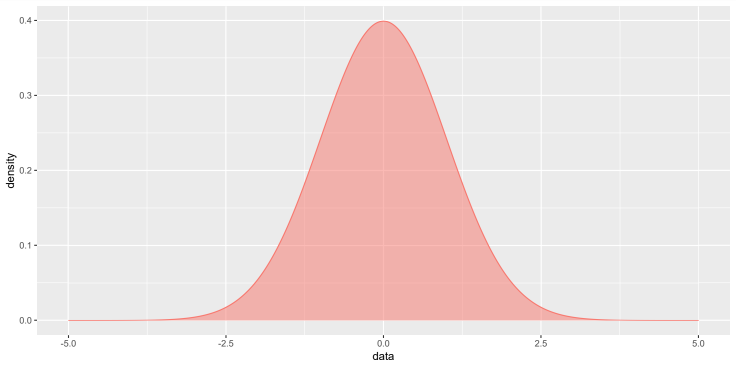


- The red and the green have the same mean, but differnt variances (or standard deviations)
- The red and the blue have the same variance, but different means

https://academo.org/demos/gaussian-distribution/

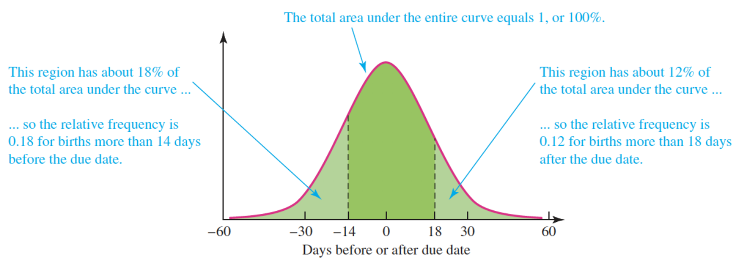
Probabilities of normal distribution

The Normal Distribution



- **The total area under the normal distribution curve must equal 1, or 100%**

The Normal Distribution (Birth data)



- Total area under this normal curve is 1
- The shaded region to the left of -14 days represents about 18% of the total area under the curve
  $\rightarrow$ 18% of births are more than 14 days early
- The shaded region to the right of 18 days represents about 12% of the total area under the curve
  $\rightarrow$ 12% of births are more than 18 days late

Example (Birth data)

1. Estimate the percentage of births occurring between 0 and 60 days after the due date.

2. Estimate the percentage of births occurring between 14 days before and 14 days after the due data

3. Estimate the percentage of births occurring between 0 and 18 days after the due data

4. Estimate the percentage of births occurring between 14 days before and 18 days after the due data

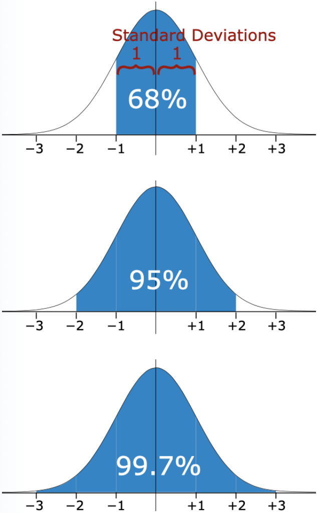
Empirical Rule

Empirical Rule

The 68-95-99.7 rule for a Normal Distribution

- About 68% of the data values fall within 1 standard deviation of the mean.
- About 95% of the data values fall within 2 standard deviations of the mean.
- About 99.7% of the data values fall within 3 standard deviations of the mean.

Properties of the Normal distribution



68% of values are within
1 standard deviation of the mean

95% of values are within
2 standard deviations of the mean

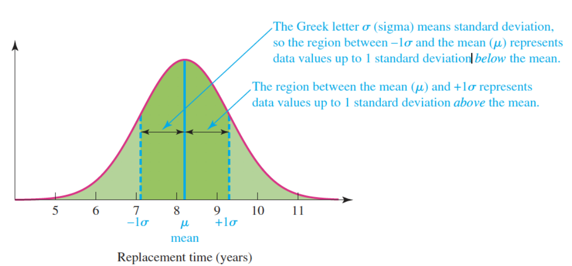
99.7% of values are within
3 standard deviations of the mean

Example

Consumer Reports survey

Consider a Consumer Reports survey in which participants were asked how long they owned their last TV set before they replaced it. Based on the survey, the distribution of replacement times has a mean (μ) of about 8.2 years, and the standard deviation (σ) of the distribution is about 1.1 years.

Example : Consumer Reports survey



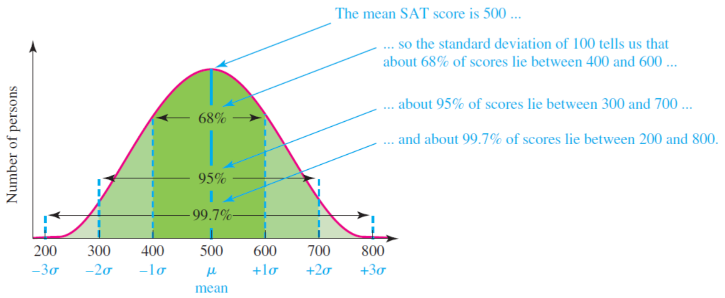
- About two thirds(68%) of the values lie within 1 standard deviation of the mean, which in this case is between ( ) years and ( ) years.
- Similarly, about 95% of the values lie within 2 standard deviations of the mean, which in this case is between ( ) years and ( ) years.

Example

SAT Score

The tests that make up the verbal (critical reading) and mathematics parts of the SAT (and the GRE, LSAT, and GMAT) are designed so that their scores are normally distributed with a mean of $\mu = 500$ and a standard deviation of $\sigma = 100$. Interpret this statement.

Example : SAT Scores



Example : Detecting Counterfeits

Vending Machine

Vending machines can be adjusted to reject coins above and below certain weights. The weights of legal U.S. quarters have a normal distribution with a mean of 5.67 grams and a standard deviation of 0.07 gram. If a vending machine is adjusted to reject quarters that weigh more than 5.81 grams and less than 5.53 grams, what percentage of legal quarters will be rejected by the machine?

Example

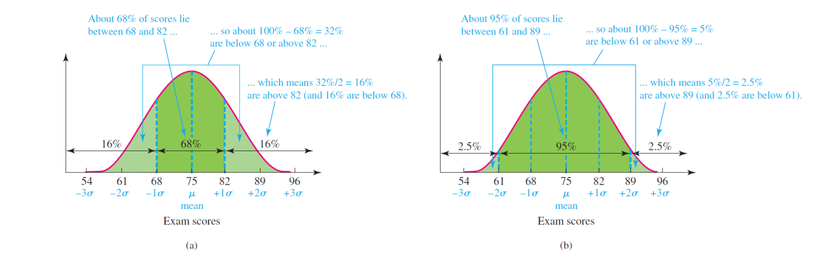
Consider an exam taken by 1,000 students for which the scores are normally distributed with a mean of $\mu = 75$ and a standard deviation of $\sigma = 7$.

1. How many students scored above 82 ?

2. How many students scored below 61 ?

3. How many students scored between 61 and 82 ?

Example : Exam



Example

Heart rate

You measure your resting heart rate at noon every day for a year and record the data. You discover that the data have a normal distribution with a mean of 66 and a standard deviation of 4. On how many days was your heart rate below 58 beats per minute?

Exercise

Height

95%of students at school are between 1.1m and 1.7m tall. Assuming this data is normally distributed can you calculate the mean and standard deviation?