Lesson 3
Probability distributions

Manuella Lech Cantuaria
Victoria Blanes-Vidal
The Maersk Mc-Kinney Moller Institute
SDU Applied AI and Data Science

1

LECTURE PLANNING

Lesson	Week	Date	TOPICS	Teacher
1	35	1/Sep	Introduction to the course Descriptive statistics – Part I	MLC
2	36	8/sep	Descriptive statistics – Part II	MLC
3	37	15/Sep	Probability distributions	MLC
4	38	22/Sep	Hypothesis testing (one sample)	VBV
5	39	29/Sep	Hypothesis testing (two samples)	VBV
6	40	6/Oct	ANOVA one-way	VBV
7	41	13/Oct	R class (Introduction to R and descriptive statistics) **Point-giving activity (in class)**	MLC+VBV
-	42	20/Oct	NO CLASS (Autum holidays)	
8	43	27/Oct	R class (hypothesis testing + ANOVA)	MLC
9	44	3/Nov	ANOVA two-way	VBV
-	45	10/Nov	NO CLASS	
10	46	17/Nov	Regression analysis	VBV
11	47	24/Nov	Notions of experimental design and questions **Point-giving activity (in class)**	VBV+MLC
12	48	1/Dec	Multiple regression	MLC

Descriptive statistics

Inferential statistics

2

2

Lesson 3: Probability distributions

3.1. Probability

3.2. Random variables

3.3. Probability distributions and cumulative probability distributions

3.4. Types of probability distributions: Discrete and Continuous

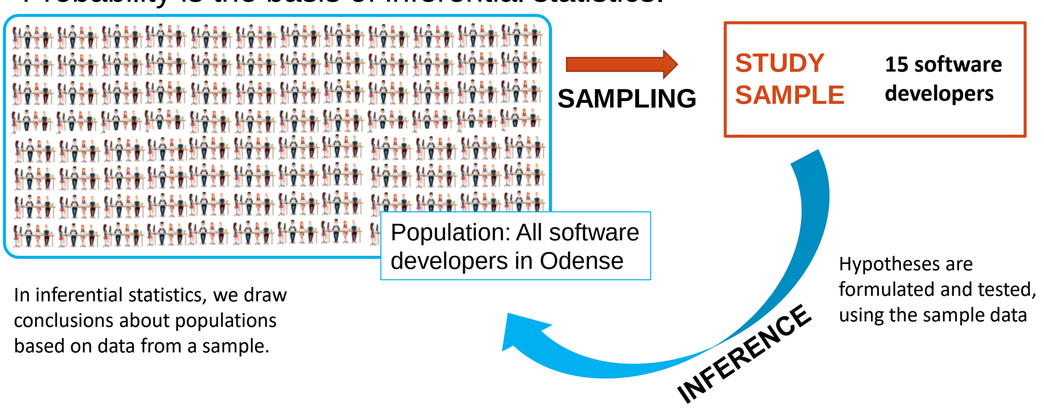
3.5. Some distributions used in inferential statistics

3

3

3.1. Probability

- Probability is the basis of inferential statistics.



- Hypotheses are tested by using probability.

4

3.1. Probability

A **probability experiment** is a chance process that leads to well-defined results called outcomes.

An **outcome** is the result of a single trial of a probability experiment.

Experiment: tossing a coin

Outcome: heads

5

Sample space - let's try it out!

A **sample space** is the set of all possible outcomes of a probability experiment.

Experiment	Sample space
Toss one coin	
Roll a die	
Answer a true/false question	
Toss two coins	

6

There are two basic interpretations of probability:

1. Classical probability
2. Empirical or experimental probability

7

Classical probability

Classical probability uses sample spaces to determine the numerical probability that an event will happen.
No need to perform the experiment to determine that probability.

The probability of any event E is

$$\frac{\text{Number of outcomes in } E}{\text{Total number of outcomes in the sample space}}$$

This probability is denoted by

$$P(E) = \frac{n(E)}{n(S)}$$

Probability values range from 0 to 1.

Example: When a coin is tossed, there is a 50/50 chance of obtaining a head or tail.

$$\text{P(head)} = \frac{n(E)}{n(S)} = \frac{1}{2}$$

8

Empirical probability

It relies on an actual experiment to determine the likelihood of outcomes. It can also be called "experimental probability"

$$P(E) = \frac{\text{frequency for the class}}{\text{total frequencies in the distribution}} = \frac{f}{n}$$

This probability is called *empirical probability* and is based on observation.

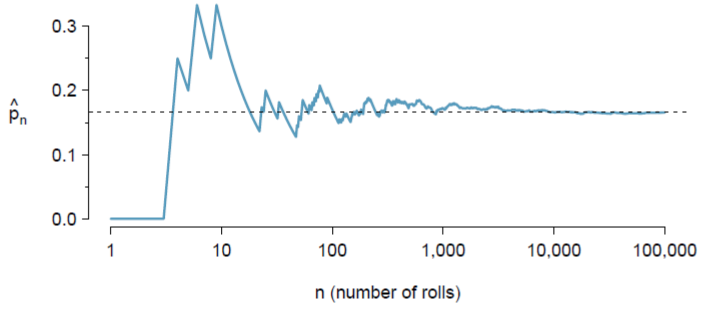
Example: Tossing a coin 100 times to see how many heads you get

9

Law of large numbers

As the number of experiments increases, the proportion $\hat{p}_n$ of occurrences (i.e. empirical probability) with a particular outcome converges to the probability P (i.e. classical probability) of that outcome.

Sample space for rolling a die: 1, 2, 3, 4, 5, or 6
P(X = 1) = 1/6



The proportion tends to get closer to the probability 1/6 (i.e. 0.167) as the number of rolls increases.

10

3.2. Random Variables

A **variable** is any characteristic, observed or measured.

When the value of a variable is the outcome of a statistical experiment, that variable is a **random variable.**

3.2. Random Variables

An example: Suppose you flip a coin two times.
This statistical experiment can have four possible outcomes: HH, HT, TH, and TT.

1st time	2nd time	1st time	2nd time	1st time	2nd time	1st time	2nd time
H	H	H	T	T	H	T	T

Now, let the variable X represent the "number of Heads that result from this action".

X = number of heads

The **random variable** X can take on the values 0, 1, or 2.

3.3. Probability Distribution

For a defined population, every **random variable** has an associated **distribution** that defines the **probability** of occurrence of each possible value of that variable.

A **probability distribution** (function) is a list of the probabilities of the values (simple outcomes) of a random variable.

13

3.3. Probability Distribution

An example: Suppose you flip a coin two times. This action can have four possible outcomes

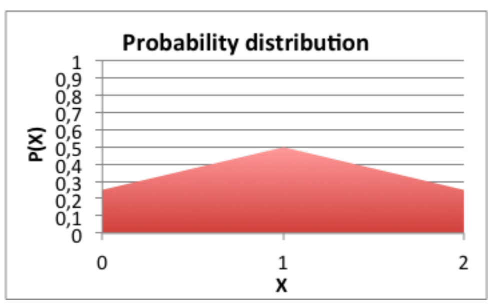



1st time	2nd time	1st time	2nd time	1st time	2nd time	1st time	2nd time
H	H	H	T	T	H	T	T

The variable X (random variable) represent the "number of Heads that result from this action".

The **probability distribution** of X is:

X	P(X=x)
0	1/4 = 0.25
1	2/4= 0.50
2	1/4 = 0.25



14

3.3 Cumulative probability Distribution

A **cumulative probability** refers to the probability that the value of a random variable falls within a specified range.

Example: If we flip a coin two times, we might ask: What is the probability that the coin flips would result in one or fewer heads (P(X $\leq$ 1))?
The answer would be a cumulative probability.

The **cumulative probability distribution** of X is:
P(X=0) + P(X=1)

X	P(X=x)	P(X≤x)
0	1/4 = 0.25	0.25
1	2/4= 0.50	0.25+0.5=0.75

15

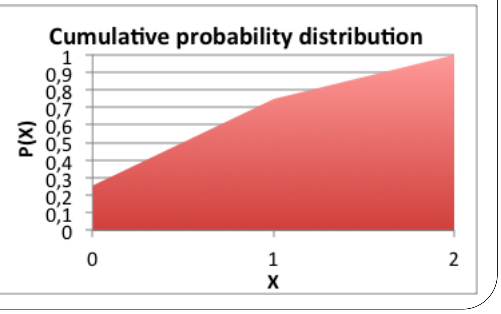
3.3. Cumulative probability Distribution

A **cumulative probability** refers to the probability that the value of a random variable falls within a specified range.

Example: If we flip a coin two times, we might ask: What is the probability that the coin flips would result in two or fewer heads (P(X $\leq$ 2))?
The answer would be a cumulative probability.

The **cumulative probability distribution** of X is:
P(X=0) + P(X=1) + P(X=2)

X	P(X=x)	P(X≤x)
0	1/4 = 0.25	0.25
1	2/4= 0.50	0.25+0.5=0.75
2	1/4 = 0.25	0.75+0.25=1

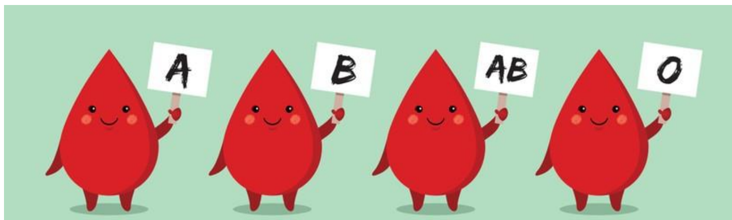


16

Let’s practice

In a sample of 50 people, 21 had type O blood, 22 had type A blood, 5 had type B blood, and 2 had type AB blood. Set up a frequency distribution and find the following probabilities.

a. A person has type O blood.
b. A person has type A or type B blood.
c. A person has neither type A nor type O blood.
d. A person does not have type AB blood.



17

3.4. Types of probability distributions

Two Types of Random Variables

- A **discrete random variable** can assume a countable number of values.
  - E.g. number of students present

1, 2, 3, 4, 5, 6, 7, 8…

- Discrete probability distributions

- A **continuous random variable** can assume any value along a given interval.
  - E.g. height of the students

1.54, 1.63, 1.71m …..

- Continuous probability distributions

20

20

Discrete Distributions

There are some specific discrete distributions that are used over and over in practice, thus they have been given special names.

There is a random experiment behind each of these distributions.

- Bernoulli Distribution
- **Binomial Distribution**
- Negative Binomial
- Poisson Distribution
- Geometric Distribution
- Multinomial Distribution

21

Discrete Distributions

- Bernoulli Distribution
- **Binomial Distribution**
- Negative Binomial
- Poisson Distribution
- Geometric Distribution
- Multinomial Distribution

Yes-No responses (e.g. Flip a coin once).

Sums of Bernoulli responses
(e.g. Flip a coin five times)

Number of trials to k^{th} event
(e.g. Flip a coin until we get 10 heads)

Points in given space (e.g. The mean number of defective products produced in a factory in one day is 21. What is the probability that in a given day there are exactly 12 defective products?)

Number of Bernoulli trials needed to get one success.

Multiple possible outcomes for each trial (e.g. tossing a dice)

22

Discrete Distributions

- Bernoulli Distribution
- **Binomial Distribution**
- Negative Binomial
- Poisson Distribution
- Geometric Distribution
- Multinomial Distribution

In this course we will only look closely at binomial distributions

23

Binomial Distribution

- The experiment consists of **n identical trials** (simple experiments).
- Each trial results in one of **two outcomes** (success or failure)
- The probability of success on a single trial is equal to π and π remains the same from trial to trial.
- The trials are independent, that is, the outcome of one trial does not influence the outcome of any other trial.
- The random variable y is the number of successes observed during n trials.

$$P(y) = \frac{n!}{y!(n-y)!}\pi^{y}(1-\pi)^{n-y}$$

Factorial

n!=1x2x3x…x n

24

Example: We flip a coin 5 times (n=5). We are interested on the number of times we get heads (the random variable y is the number of heads) and π = the probability of success (heads) each time we flip a coin (π=0.5))

$$P(y) = \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y}$$

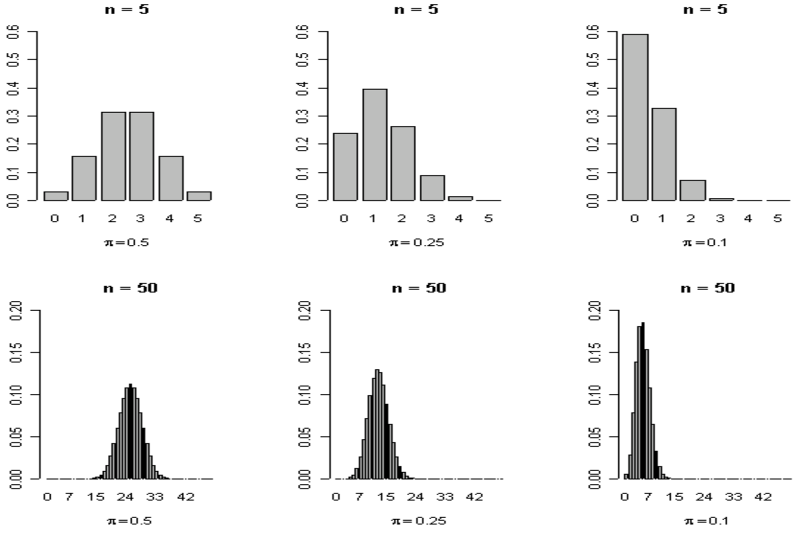
n = 5 **π = 0.5**

y	y!	n!/((y!)(n-y!))	π^y	(1-π)^(n-y)	P(y)
0	1	1	1	0.03125	0.03125
1	1	5	0.5	0.0625	0.15625
2	2	10	0.25	0.125	0.31250
3	6	10	0.125	0.25	0.31250
4	24	5	0.0625	0.5	0.15625
5	120	1	0.03125	1	0.03125

If we toss the coin five times, the probability of e.g. getting heads 2 times is 0.31, and the probability of getting heads all five times is 0.03

25

Binomial probability density function forms



As **n** goes up, the distribution looks more bell shaped.
As **π** goes closer to 0.5, the more symmetric is the distribution (less skewed)

27

3.4. Types of probability distributions

Two Types of Random Variables

- A **discrete random variable** can assume a countable number of values.
  - E.g. number of students present

1, 2, 3, 4, 5, 6, 7, 8…

- Discrete probability distributions

- A **continuous random variable** can assume any value along a given interval.
  - E.g. height of the students

1.54, 1.63, 1.71m …..

- Continuous probability distributions

28

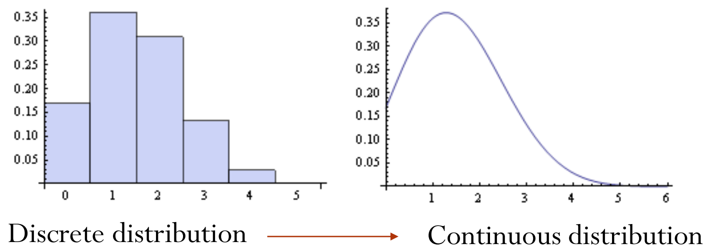
Continuous Distributions

- A continuous random variable is one which takes an infinite number of possible values.
- Continuous random variables are usually measurements. E.g. height, weight, the amount of sugar in an orange.
- Many continuous variables, such as the examples just mentioned, have distributions that are bell-shaped, and these are sometimes called "*approximately normally distributed variables*".
- If a random variable is a continuous variable, its probability distribution is called a **continuous probability distribution**.

29

Continuous Distributions

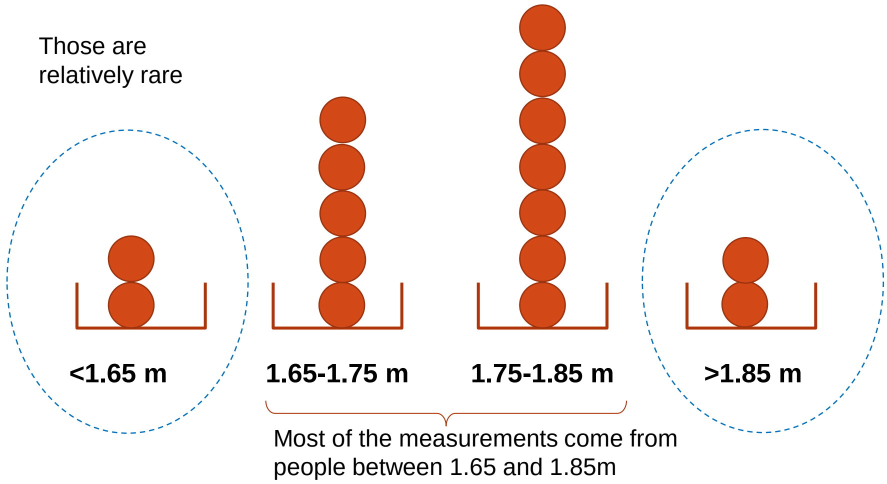
- When moving from discrete to continuous distributions, the random variable will no longer be restricted to integer values (0, 1, 2, 3…), but will now be able to take on any value in some interval of real numbers.
- Graphically, we will be moving from the discrete bars of a histogram to the curve of a continuous function.



30

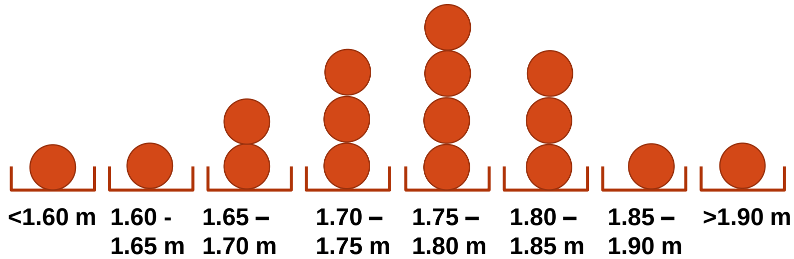
Continuous Distributions - example

Height of students in our class



31

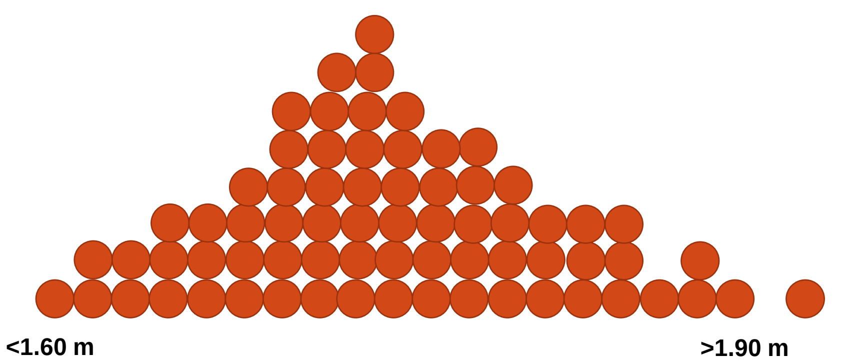
Continuous Distributions - example



32

Continuous Distributions - example

The more people we have, and the higher the number of categories, the more precise can we be in relation to the ”height distribution” of the students in our class



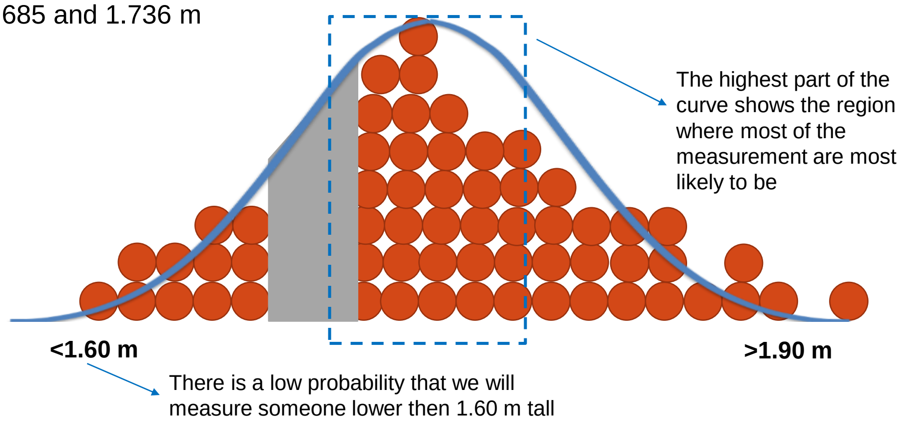
33

Continuous Distributions - example

We can use a curve to approximate the histogram and calculate the probabilities

Both the curve and the histogram show us how the probabilities of the measurements are distributed.

One of the advantages of the curve is that we are not limited to the boudaries of pre-established categories. We can for example calculate the probability of someone's height being between 1.685 and 1.736 m



34

Continuous Distributions

The probability that a continuous random variable will assume a particular value is zero, since the number of values which may be assumed by the random variable is infinite.

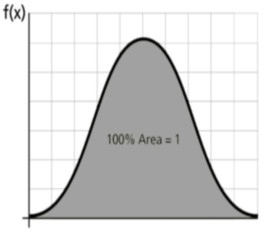
(e.g. The probability of being 1.745643234687.. m high is 0, however the probability of being between 1.74 and 1.75 is not 0)

We can only say the height of this person is in between two numbers, because there is infinite possibilities in between

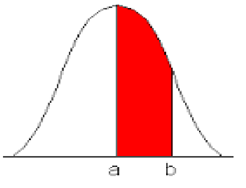
35

Continuous Distributions

- Most often, the equation used to describe a **continuous probability distribution** is called a **probability density function** (pdf).
- For a continuous probability distribution, the density function has the following properties:

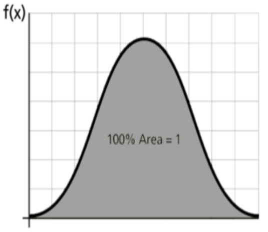


  - Since the continuous random variable is defined over a continuous range of values, the graph of the density function will also be continuous over that range.
  - The area limited by the curve of the density function and the x-axis is equal to 1.
  - The probability that a random variable assumes a value between *a* and *b* is equal to the area under the density function bounded by *a* and *b*.

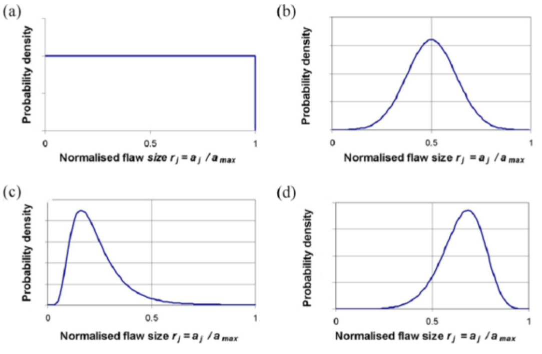


36

Continuous Distributions



In this example, we focused on a bell-shaped distribution, however, probability distributions can assume many different shapes



Those can be evaluated by:
- Skewness
- Kurtosis

37

3.5. Some distributions used in inferential statistics

Some continuous distributions used in inferential statistics:

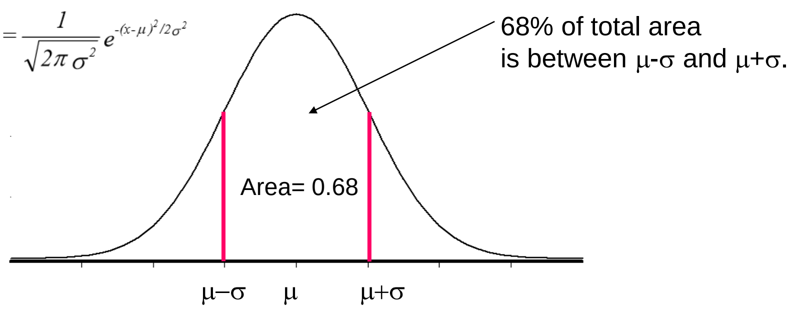
- Normal distribution
- Chi square distribution
- F distribution
- t distribution

38

Normal Distribution

A symmetric distribution defined on the range -∞ to + ∞ whose shape is defined by two parameters, the **mean**, denoted μ, that centers the distribution, and the **standard deviation**, σ, that determines the spread of the distribution.

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



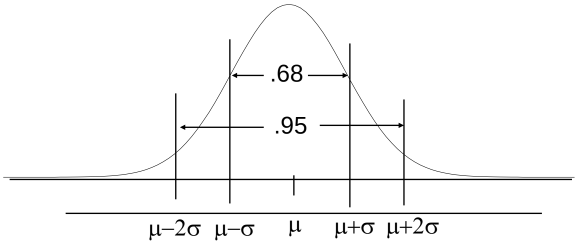
$$P(\mu-\sigma < X < \mu+\sigma) = .68$$

39

Properties of the Normal Distribution

- Symmetric, bell-shaped density function.
- 68% of area under the curve between $\mu \pm \sigma$.
- 95% of area under the curve between $\mu \pm 2\sigma$.
- 99.7% of area under the curve between $\mu \pm 3\sigma$.

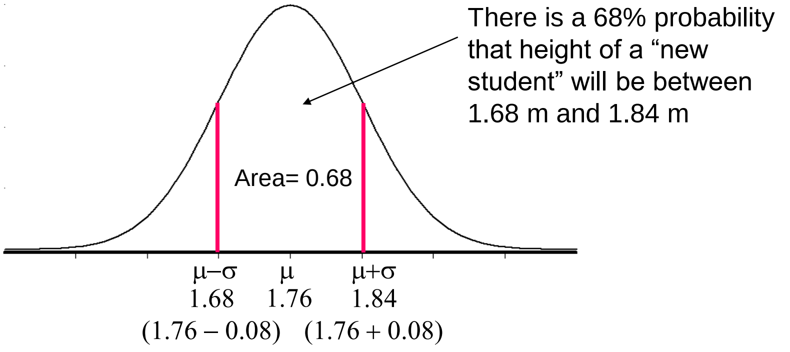
$Y \sim N(\mu, \sigma)$



40

Normal Distribution - example

In our example where we measured the height of the students in our class.
Let's considered we measured the height of 90 students, founding a mean of 1.76 m with standard deviation of 8 cm (0.08 m).

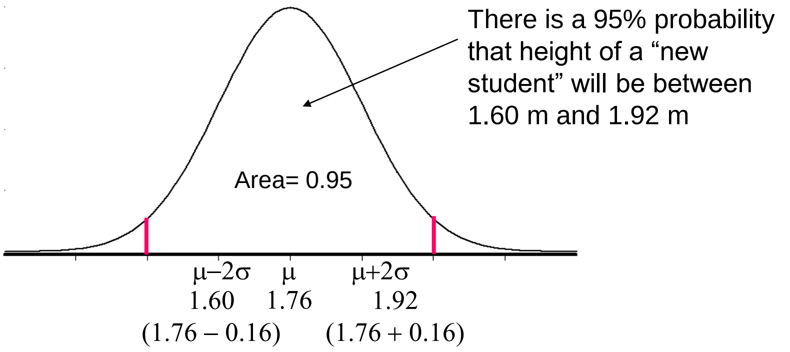


$$P(1.68 < X < 1.84) = .68$$

41

Normal Distribution - example

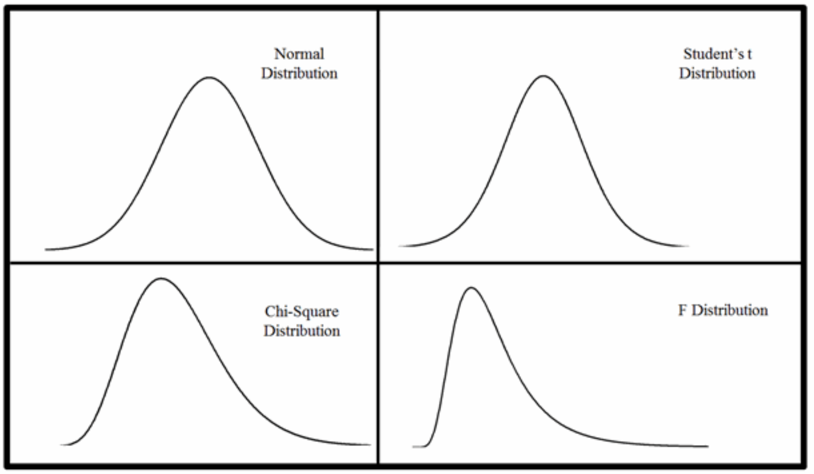
In our example where we measured the height of the students in our class.
Let's considered we measured the height of 90 students, founding a mean of 1.76 m with standard deviation of 8 cm (0.08 m).



$$P(1.60 < X < 1.92) = .95$$

42

- Chi square distribution
- F distribution
- Student t distribution



43

Last one…

The life expectancy of Timely brand watches is normally distributed with a mean of four years and a standard deviation of eight months. That means that the life expectancy of 95% of the watches will be between how many months?

44

Questions?

Now it is time for you to practice at the exercise’s class!
Rooms: U165, U166, U167, U171, U172.

46