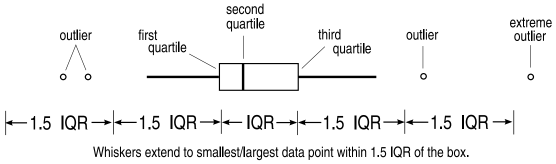
Box plots provide a compact, high level, representation of a data set. Box plots are good for comparing data sets. To compute a box plot you need:

(1) the smallest and largest data point.
(2) the median.
(3) the lower and upper quartiles.



5 Lesson (Box Plots)
Ten measurements are given below. Construct a box plot for these measurements.

$$57,\ 65,\ 66,\ 68,\ 70,\ 70,\ 72,\ 73,\ 75,\ 89$$

4 Probability Theory

Probability theory is used to study random experiments.

An experiment conducted in Jimma, Arba Minch or Mekele should have the same outcome, so long as the controls on an experiment are kept exactly the same.

random experiment: experiment whose outcome is random.

No matter how hard we try to keep the controls of a random experiment the same, the experiment produces varying outcomes. Each outcome of a random experiment has a given probability (likelihood) of occurring.

21 Example (Random Experiment)
What is the random experiment performed at the beginning of many team competitions?

Solution:

22 Definition (Assigning Probabilities)
Probabilities can be assigned in one of three ways:

(i) symmetry

(ii) observed relative frequencies of past outcomes

(iii) expert opinion (degree of belief)

23 Example (Assigning Probabilities Using Symmetry)
Assign probabilities using symmetry for a two, fair coin toss.

24 Example (Assigning Probabilities Using Past Observations)
Assign probabilities using past observation for the number of enjera's sold each day by a store.

25 Example (Assigning Probabilites Using Expert Opinion)
Assign probabilities using expert opinion for the probability a new drama will be popular.

26 Example (Random Experiment)
Pick a number at random from 1 to 10. What is the probability the number is (a) 7 (b) odd (c) prime (d) odd or prime (e) odd and prime.

Solution:

27 Theorem (Probability of the Complement of an Event)
Let A^c represent the complement of an event. Then

$$P(A^c) = 1 - P(A).$$

28 Example (Using Probability Theorems)
Compute the probability of getting anything but two heads in a fair two coin toss.

Solution:

29 Example (Probability of Union of Events)
In order to test the effectiveness of a flu vaccine, 100 people were selected at random and questioned to determine if they (i) received the flu vaccination and (ii) got the flu. The responses are summarized in the table below:

	$\mathbf{F}$ (flu)	$\mathbf{F}'$(did not get flu)
$\mathbf{V}$ (vaccinated)	10	30
$\mathbf{V}'$(not vaccinated)	15	45

Determine the probability that a person selected at random:
(a) was vaccinated.
(b) got the flu.
(c) got the vaccination and got the flu.
(d) got the vaccination or got the flu.

Solution:

30 Example (Independent Events)
Fair two coin toss.

A_1 - first coin is H.
A_2 - second coin is H.

A_1 and A_2 are independent events. Why?
Ans. $P(A_2|A_1) = P(A_2)$ and $P(A_1|A_2) = P(A_1)$

31 Example (Independent Events)
Randomly select M&M's from a bag and eat them.

A_1 - first M&M is green
A_2 - second M&M is green

A_1 and A_2 are not independent events.

product of probabilities: for independent events, $P(A_1 \cap A_2) = P(A_1)P(A_2)$.

32 Example (Product of Probabilities)
Two white and two black balls are in a box. (Sketch) Two balls are selected at random *with* replacement. Compute the probability of selecting two black balls.

Solution:

5 Random Variables

random variable: a function that assigns a numerical value to each outcome of a random experiment.

33 Example (Die Roll Random Variable)
Each outcome of a die roll is a pattern of dots. A natural random variable to use is the function that assigns each dot pattern a number equal to the number of dots.

34 Example (Coin Toss Random Variable)
$X(H) = 0, X(T) = 1.$
(Could define $X(H) = 1, X(T) = 0$ also.)

The above random variables are discrete random variables.

35 Example (Continuous Random Variable)
X - number of minutes a flight is delayed.

Random variables are characterized by probability distribution functions (PDF's).

Probability Density/Distribution Functions (PDFs)

- **discrete**: probability mass/distribution function
- **continuous**: probability density function

36 Example (PDF for Die Roll)
Determine the probability mass function for the random variable X where X equals the number of dots of a die roll.

Solution:

37 Example (PDF for Number of Boys)
Determine the probability mass function for the random variable X which equals the number of boys in a family with two children.

Solution:

The previous example is an example of a **discrete distribution**. Discrete distributions are sometimes call **probability mass functions**.

38 Example (Discrete PDF)
Consider the infinite population of all die rolls. We expect that a random sample taken from this infinite population will have the PDF

$$p(x) = \frac{1}{6}, \text{ for } x = 1, 2, 3, 4, 5 \text{ or } 6$$

	1	2	3	4	5	6
rel. freq	1/6	1/6	1/6	1/6	1/6	1/6

Sketch bar graph.

The previous example is an example of what is called a **uniform distribution**.

Continuous probability distributions functions (PDFs) are described by **density functions**.

39 Example (Continuous PDF)
Assume the bus arrives at the bus stop near your house every day between 6:00 and 6:30 pm. with uniform likelihood. What is the probability distribution function (PDF)?

Note: We can't directly compute a relative frequency as we did with discrete PDFs because each arrival time is essentially unique.

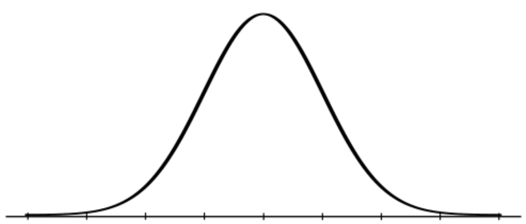
Solution:

The total area under a density function is always equal to 1. Why?

6 Normal Distribution

The most important distribution in statistics is the **normal distribution**. It is a (continuous) bell shaped PDF.

40 Definition (Normal Distribution)



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \quad -\infty \le x \le \infty$$

μ—mean
σ—standard deviation

Memorize the following rule:

41 Theorem (Empirical Rule)

- 68% of a normal population are within 1 standard deviation of the mean of the population.
- 95% of a normal population are within 2 standard deviation of the mean of the population.
- 99.7% of a normal population are within 3 standard deviation of the mean of the population.

42 Example (Normal Distribution)
Assume that the blood pressure of patients in a certain hospital are normally distributed with average 100 mm Hg and standard deviation 10 mm Hg.

(a) What proportion of the patients have blood pressure between 90 and 110 mm Hg.

(b) What proportion of the patients have blood pressure more than 120 mm Hg.

(c) What proportion of the patients have blood pressure less than 100 mm Hg.

43 Example (Confidence Interval)
Assume that the blood pressure of patients in a certain village are normally distributed. A random sample of 16 patients had an average blood pressure of 145 mm Hg and standard deviation of 40 mm Hg. Determine a 95% confidence interval for the average blood pressure in this village.

44 Example (Hypothesis Test)
Assume that the blood pressure of patients in a certain village are normally distributed. In order to test the hypothesis

$$H_0: \quad \mu = 120 \text{ mm Hg}$$
$$H_1: \quad \mu \neq 120 \text{ mm Hg}$$

a random sample of 16 patients had an average blood pressure of 145 mm Hg and standard deviation of 40 mm Hg.

45 Example (ANOVA)
To examine the effects of smoking on heart rate, nonsmokers, light smokers, moderate smokers and heavy smokers undertook sustained physical expercise. After resting for three minutes, their heart rates were measured and recorded. Does smoking affect heart rate? Test at a significance level of $\alpha = 0.05$.

Treatment	Observations						Ave
nonsmokers	69	52	71	58	59	65	62.3
light smokers	55	60	78	58	62	66	63.2
moderate smokers	66	81	70	77	57	79	71.7
heavy smokers	91	72	81	67	95	84	81.7
							69.7