# Basic Probability Concepts

Additional resources for Data Analysis using R

# Introduction

People use the term probability many times each day. For example, physician says that a patient has a 50-50 chance of surviving a certain operation. Another physician may say that she is 95% certain that a patient has a particular disease

# Definition

If an event can occur in N mutually exclusive and equally likely ways, and if m of these possess a trait, E, the probability of the occurrence of E is read as

$$P(E) = m/N$$

# Definition

**Experiment ==> any planned process of data collection. It consists of a number of trials (replications) under the same condition.**

# Definition

**Sample space:** collection of unique, non-overlapping possible outcomes of a random circumstance.

**Simple event:** one outcome in the sample space; a possible outcome of a random circumstance.

**Event:** a collection of one or more simple events in the sample space; often written as A, B, C, and so on

{ Male, Female }

# Definition

**Complement** ==> sometimes, we want to know the probability that an event will not happen; an event opposite to the event of interest is called a complementary event.
If A is an event, its complement is The probability of the complement is **AC or ┐┐A**
Example: The complement of male event is the female

$$\boldsymbol{P(A) + P(AC) = 1}$$

# Views of Probability:

## 1-Subjective:

It is an estimate that reflects a person’s opinion, or best guess about whether an outcome will occur.

Important in medicine ㅎ form the basis of a physician’s opinion (based on information gained in the history and physical examination) about whether a patient has a specific disease. Such estimate can be changed with the results of diagnostic procedures.

# 2- Objective

Classical

- **It is well known that the probability of flipping a fair coin and getting a “tail” is 0.50.**
- **If a coin is flipped 10 times, is there a guarantee, that exactly 5 tails will be observed**
- **If the coin is flipped 100 times? With 1000 flips?**
- **As the number of flips becomes larger, the proportion of coin flips that result in tails approaches 0.50**

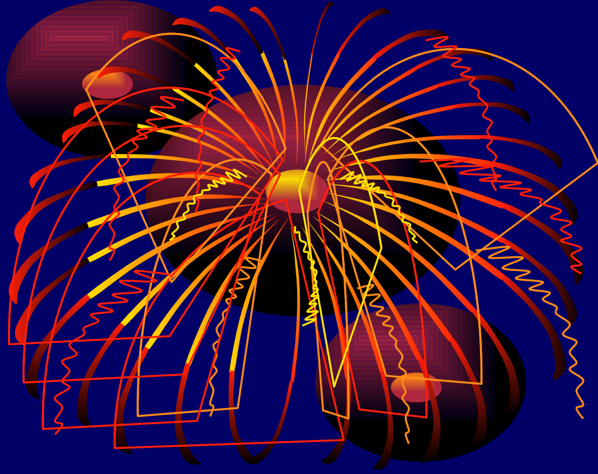
# Example: *Probability of Male versus Female Births*

Long-run relative frequency of males born in KSA is about 0.512 **(512 boys born per 1000 births)**

Table provides results of simulation: the proportion is far from .512 over the first few weeks but in the ***long run*** settles down around .512.

**Relative Frequency of Male Births over Time**

| Weeks of Watching | Total Births | Total Boys | Proportion of Boys |
|---|---|---|---|
| 1 | 30 | 19 | .633 |
| 4 | 116 | 68 | .586 |
| 13 | 317 | 172 | .543 |
| 26 | 623 | 383 | .615 |
| 39 | 919 | 483 | .526 |
| 52 | 1237 | 639 | .517 |



# 2- Objective

## Relative frequency

**Assuming that an experiment can be repeated many times and assuming that there are one or more outcomes that can result from each repetition. Then, the probability of a given outcome is the number of times that outcome occurs divided by the total number of repetitions.**

*Problem 1.*

| Blood Group | Males | Females | Total |
|---|---|---|---|
| O | 20 | 20 | 40 |
| A | 17 | 18 | 35 |
| B | 8 | 7 | 15 |
| AB | 5 | 5 | 10 |
| Total | 50 | 50 | 100 |

*Problem 2.*

An outbreak of food poisoning occurs in a group of students who attended a party

| | Ill | Not Ill | Total |
|---|---|---|---|
| Ate Barbecue | 90 | 30 | 120 |
| Did Not Eat Barbecue | 20 | 60 | 80 |
| Total | 110 | 90 | 200 |

# Marginal probabilities

Named so because they appear on the "margins" of a probability table. It is probability of single outcome

Example: In problem 1, P(Male), P(Blood group A)

$$P(\text{Male}) = \frac{\text{number of males}}{\text{total number of subjects}}$$

$$= 50/100$$

$$= 0.5$$

# Conditional probabilities

It is the probability of an event on condition that certain criteria is satisfied

Example: If a subject was selected randomly and found to be female what is the probability that she has a blood group O

Here the total possible outcomes constitute a subset (females) of the total number of subjects.

This probability is termed probability of O given F

P(O\F) = 20/50

= 0.40

# Joint probability

It is the probability of occurrence of two or more events together

Example: Probability of being male & belong to blood group AB

$P(M \text{ and } AB) = P(M \cap AB)$

$= 5/100$

$= 0.05$

$\cap$ = intersection

# Properties

- The probability ranges between 0 and 1
- If an outcome cannot occur, its probability is 0
- If an outcome is sure, it has a probability of 1
- The sum of probabilities of mutually exclusive outcomes is equal to 1

  $P(M) + P(F) = 1$

# Rules of probability

## 1- Multiplication rule

Independence and multiplication rule

**P(A and B) = P(A) P(B)**

P(A)

P(B\A)

P(B)

A and B are independent

P(B\A) = P(B)

# Example:

The joint probability of being male and having blood type O

To know that two events are independent compute the marginal and conditional probabilities of one of them if they are equal the two events are independent. If not equal the two events are dependent

P(O) = 40/100 = 0.40

P(O\M) = 20/50 = 0.40

Then the two events are independent

P(O∩M) = P(O)흥P(M) = (40/100)흥(50/100)

= 0.20

# Rules of probability

## **1- Multiplication rule**

Dependence and

**the modified multiplication rule**

$$P(A \text{ and } B) = P(A)\ P(B\backslash A)$$

P(A)

P(B\A)

P(B)

P(B\A)

A and B are not independent

$P(B\backslash A) \neq P(B)$

# Example:

The joint probability of being ill and eat barbecue

P(Ill) = 110/200 = 0.55

P(Ill\Eat B) = 90/120 = 0.75

Then the two events are dependent

P(Ill∩Eat B) = P(Eat B) 흥 P(Ill\Eat B)

= (120/200) 흥 (90/120)

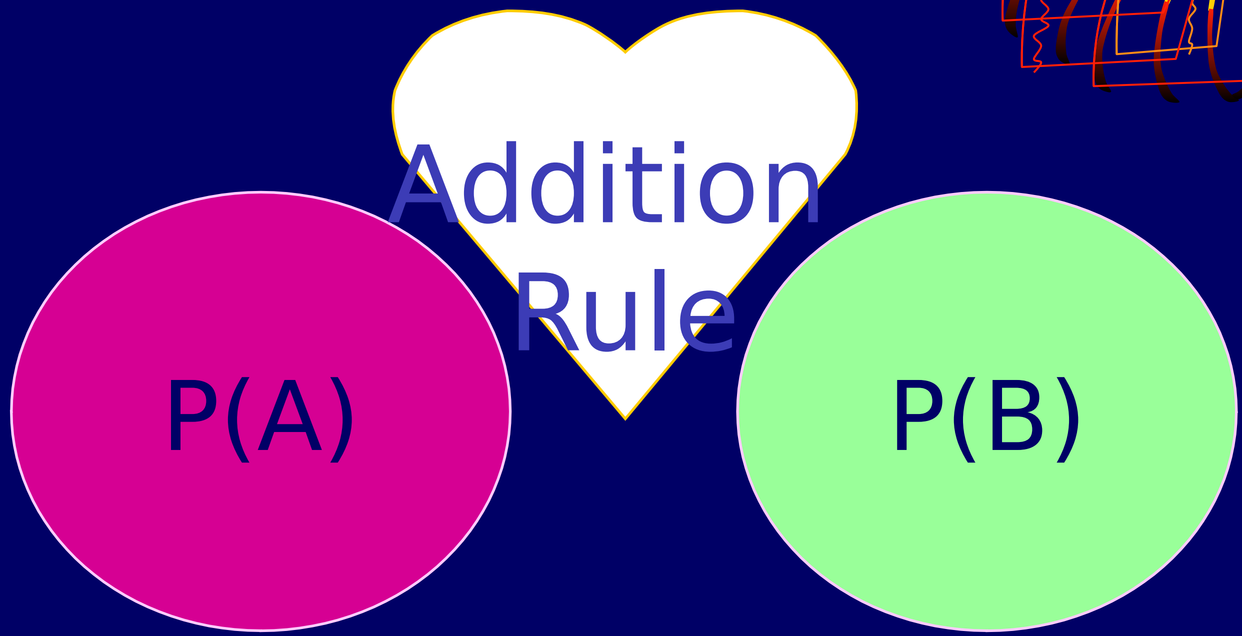
= 0.45

# Rules of probability

## 2- Addition rule

# A and B are mutually exclusive

The occurrence of one event precludes the occurrence of the other




**$P(A \text{ OR } B) = P(A \cup B) = P(A) + P(B)$**

# Example:

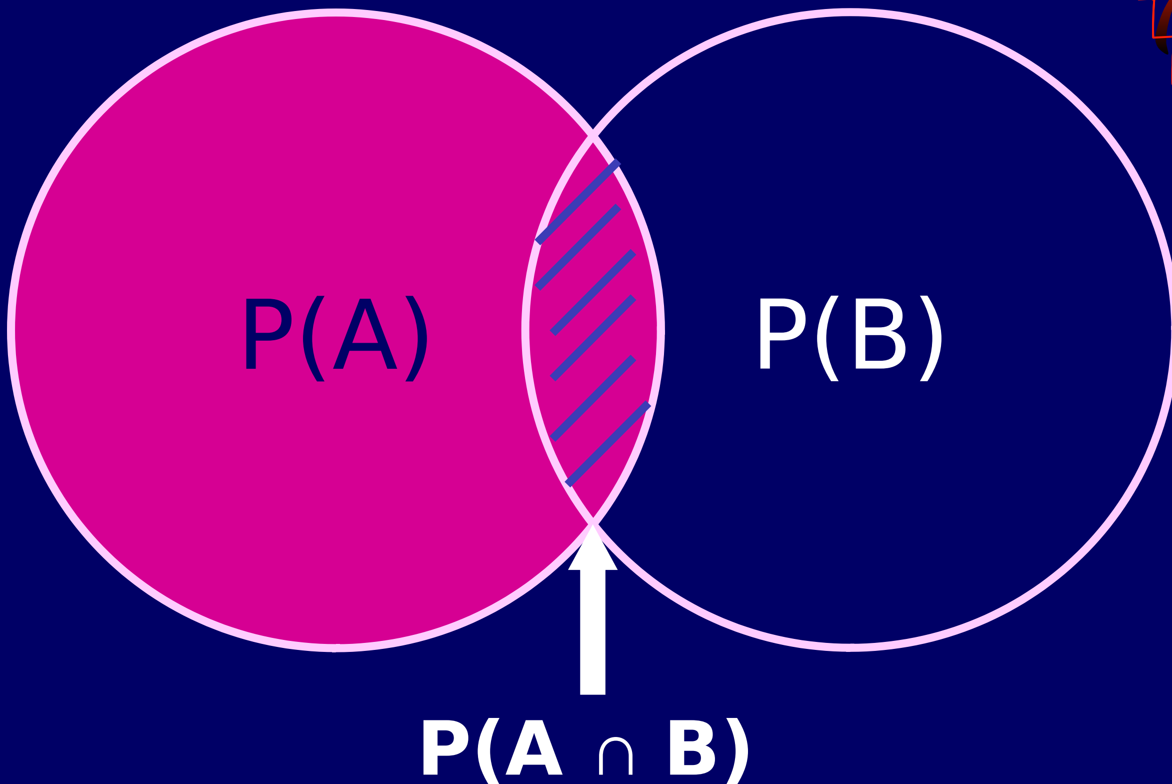
The probability of being either blood type O or blood type A

$P(O \cup A) = P(O) + P(A)$

$= (40/100)+(35/100)$

$= 0.75$

# A and B are non mutually exclusive
(Can occur together)
Example: Male and smoker


**$P(A \text{ OR } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$**

## Example:

Two events are not mutually exclusive (male gender and blood type O).

$$
\begin{aligned}
P(M \text{ OR } O) &= P(M)+P(O) - P(M \cap O) \\
&= 0.50 + 0.40 - 0.20 \\
&= 0.70
\end{aligned}
$$

# Excercises

**1. If tuberculous meningitis had a case fatality of 20%,**

**(a) Find the probability that this disease would be fatal in two randomly selected patients (the two events are independent)**

**(b) If two patients are selected randomly what is the probability that at least one of them will die?**

**(a) P(first die and second die) = 20% × 20% = 0.04**

**(b) P(first die or second die)**

**= P(first die) + P(second die) - P(both die)**

**= 20% + 20% - 4%**

**= 36%**

2. In a normally distributed population, the probability that a subject's blood cholesterol level will be lower than 1 SD below the mean is 16% and the probability of being blood cholesterol level higher than 2 SD above the mean is 2.5%. What is the probability that a randomly selected subject will have a blood cholesterol level lower than 1 SD below the mean or higher than 2 SD above the mean.

P(blood cholesterol level < 1 SD below the mean or 2 SD above the mean) = 16% + 2.5%

= 18.5%

3. **In a study of the optimum dose of lignocaine required to reduce pain on injection of an intravenous agent used for induction of anesthesia, four dosing groups were considered (group A received no lignocaine, while groups B, C, and D received 0.1, 0.2, and 0.4 mg/kg, respectively). The following table shows the patients cross-classified by dose and pain score:**

**Compute the following probabilities for a randomly selected patient:**

**1.being of group D and experiencing no pain**

**2.belonging to group B or having a**

| Pain score | Group | | | | Total |
|---|---|---|---|---|---|
| | A | B | C | D | |
| 0 | 49 | 73 | 58 | 62 | 242 |
| 1 | 16 | 7 | 7 | 8 | 38 |
| 2 | 8 | 5 | 6 | 6 | 25 |
| 3 | 4 | 1 | 0 | 0 | 5 |
| Total | 77 | 86 | 71 | 76 | 310 |

# Nightlights and Myopia

Assuming these data are representative of a larger population, what is the **approximate probability** that someone from that population who **sleeps with a nightlight** in early childhood **will develop some degree of myopia**?

| Slept with: | No Myopia | Myopia | High Myopia | Total |
|---|---|---|---|---|
| Darkness | 155 (90%) | 15 (9%) | 2 (1%) | 172 |
| Nightlight | 153 (66%) | 72 (31%) | 7 (3%) | 232 |
| Full Light | 34 (45%) | 36 (48%) | 5 (7%) | 75 |
| Total | 342 (71%) | 123 (26%) | 14 (3%) | 479 |

***Note***: 72 + 7 = 79 of the 232 nightlight users developed some degree of myopia. So the probability to be 79/232 = 0.34.

# Disclaimer:

The content may be used by the students (only for educational purpose, not for commercial purpose.)