DATA ANALYTICS

LECTURE NO: 05

Experiment, Trial, Elementary Event, Event

- **Experiment**: A process that produces outcomes
  - More than one possible outcome
  - Only one outcome per trial.
- **Trial**: One repetition of the process.
- **Elementary Event**: cannot be decomposed or broken down into other events.
- **Event**: an outcome of an experiment
  - May be an elementary event, or
  - May be an aggregate of elementary events.
  - Usually represented by an uppercase letter e.g., A, E1 etc

An example experiment

- **Experiment**: randomly select, without replacement, two families from the residents of a town.
- **Elementary event**: The sample includes families A and C
- **Event**: Each family in the sample has children in the household
- Event: The sample families own a total of four automobiles.

Tiny Town Population		
Family	Children in Household	Number of Automobiles
A	Yes	3
B	Yes	2
C	No	1
D	Yes	2

Sample Space

- The set of all elementary events for an experiment
- Methods for describing a sample space
  - Roster or listing
  - Tree diagram
  - Set builder notation
  - Venn diagram

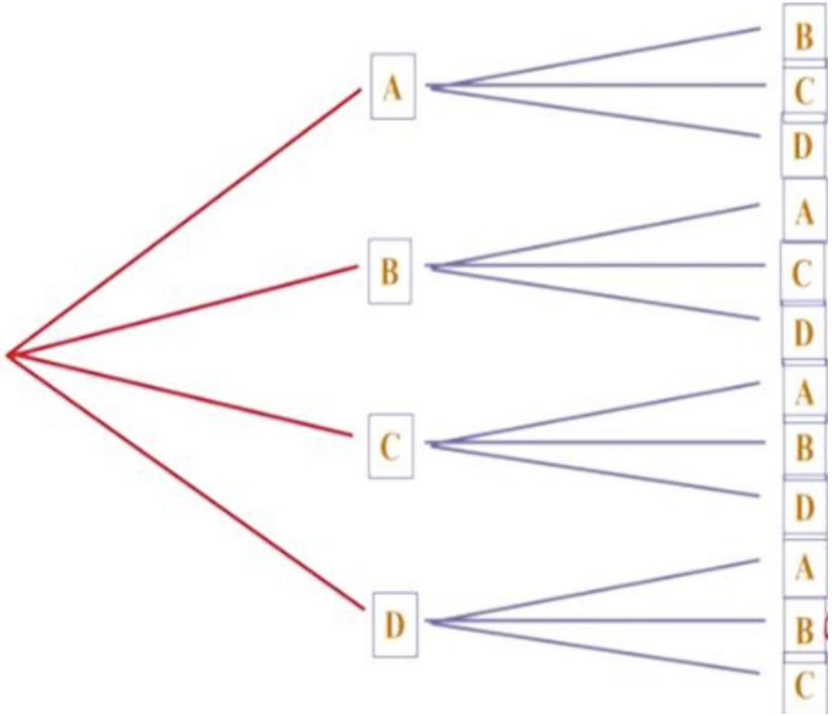
Sample Space: Roster example

- **Experiment**: randomly select, without replacement, two families from the resident of a tiny town.
- Each ordered pair in the sample space is an elementary event, for example (D,C)

Family	Children in Household	Number of Automobiles
A	Yes	3
B	Yes	2
C	No	1
D	Yes	2

Listing of Sample Space
(A,B), (A,C), (A,D), (B,A), (B,C), (B,D), (C,A), (C,B), (C,D), (D,A), (D,B), (D,C)

Sample Space: Tree diagram for random sample of two families.

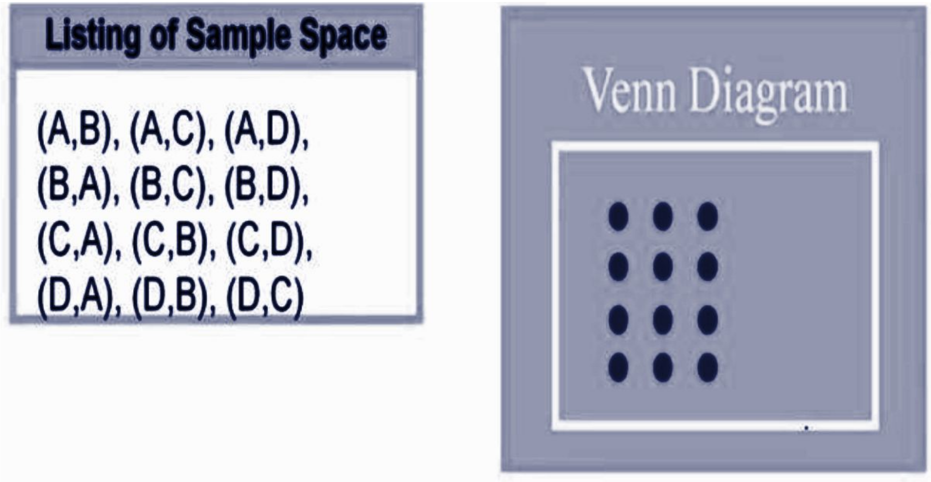


Sample Space: Set notation for randomly sample of two families

- S = {(x,y)} | x is the family selected on the first draw, and y is the family selected on the second draw.
- Concise description of large sample spaces.

Sample Space

- Useful for discussion of general principles and concepts.



Mutually Exclusive Events

- Events with no common outcomes.
- Occurrence of one event precludes the occurrence of the other event

$C = \{IBM, Apple, Dell\}$

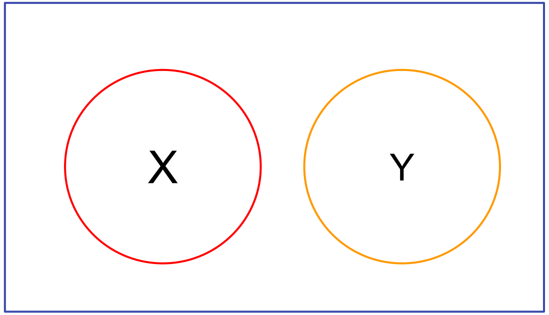
$F = \{ Banana, Grape\}$

$C \cap F = \{ \}$

$X = \{ 1, 7, 8\}$

$Y = \{ 2, 4, 5, 9, 6\}$

$X \cap Y = \{ \}$



$P(X \cap Y) = 0$

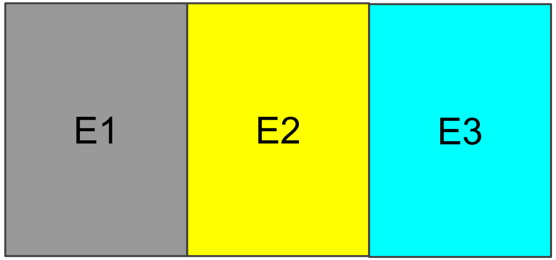
Independent Events

- Occurrence of one event does not affect the occurrence or non-occurrence of the other event.
- The condition probability of X given Y is equal to the marginal probability of X.
- The condition probability of Y given X is equal to the marginal probability of Y.

$$\boxed{P(X|Y) = P(X) \text{ and } P(Y|X) = P(Y)}$$

Collectively Exhaustive Events

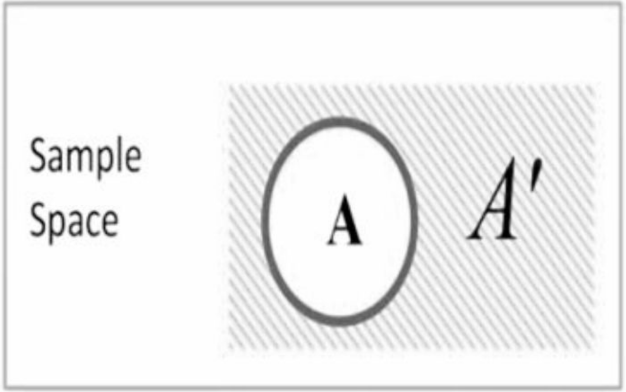
- Contains all elementary events for an experiment.



Sample Space with three
Collectively exhaustive events.

Complementary Events

- All elementary events not in the event **A** are in its complementary events.



$$P(Sample\ Space) = 1$$

$$P(A') = 1 - P(A)$$

Counting the possibilities

- mn Rule
- Sampling from a population with replacement
- **Combinations:** Sampling from a population without replacement

mn Rule

- If an operation can be done **m** ways and a second operation can be done **n** ways, then three are **mn** ways for the two operations to occur in order.
- This rule is easily extend to the k stages, with a number of ways equal to
  - $n_1.n_2.n_3\ldots n_k$
- Example : Toss two coins. The total number of sample events is **2 x 2 = 4**

Sampling from a population with replacement

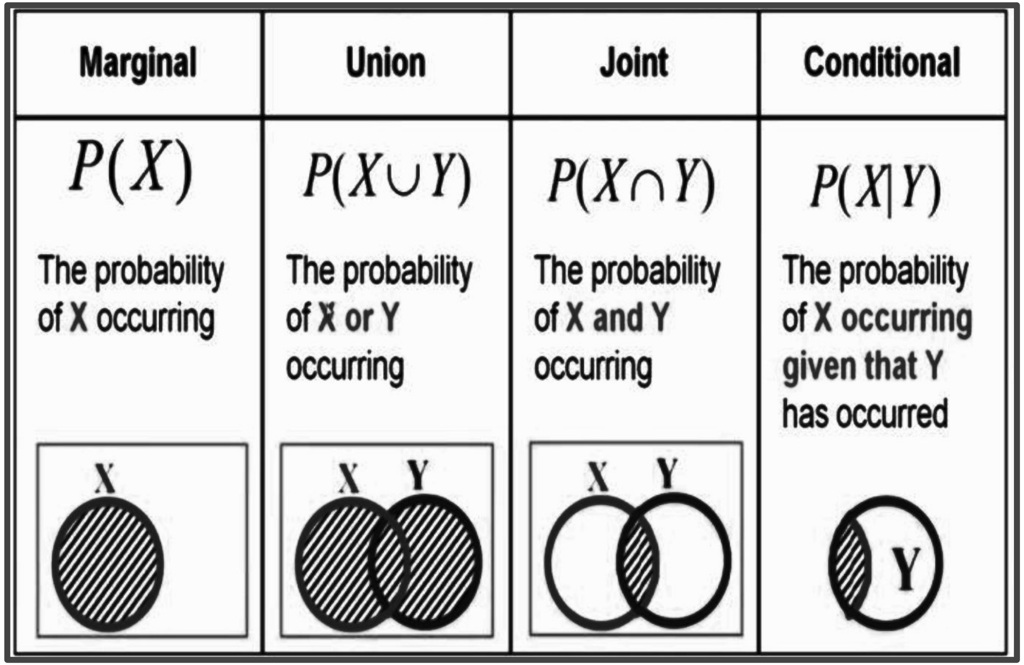
- A tray contains 1000 individual tax returns. If 3 returns are randomly selected **with replacement** from the tray, how many possible samples are there?
- $(N)^n = (1000)^3 = 1{,}000{,}000{,}000$

Combinations

- A tray contains 1000 individual tax returns. If 3 returns are randomly selected **without replacement** from the tray, how many possible samples are there?

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{1000!}{3!(1000-3)!} = 166,167,000$$

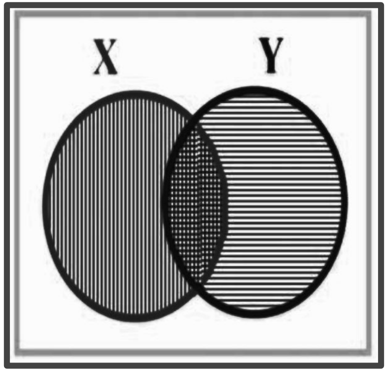
Four types of probability



Marginal	Union	Joint	Conditional
$P(X)$	$P(X \cup Y)$	$P(X \cap Y)$	$P(X \mid Y)$
The probability of X occurring	The probability of X or Y occurring	The probability of X and Y occurring	The probability of X occurring given that Y has occurred
X	X Y	X Y	Y

General law of Addition

$$P(X \cup Y) = P(X) + P(Y) - P(X \cap Y)$$



Design for improving productivity?

Problem

- A company conducted a survey for the American Society of Interior designers in which workers were asked which changes in office design would increase productivity.
- Respondents were allowed to answer more than one type of design change.

Reducing noise would increase productivity	70%
More storage space would increase productivity	67%
Both	56%

Problem

- If one of the survey respondents was randomly selected and asked what office design changes would increase worker productivity.
  - What is the probability that this person would select ***reduction noise*** or ***more storage space***?

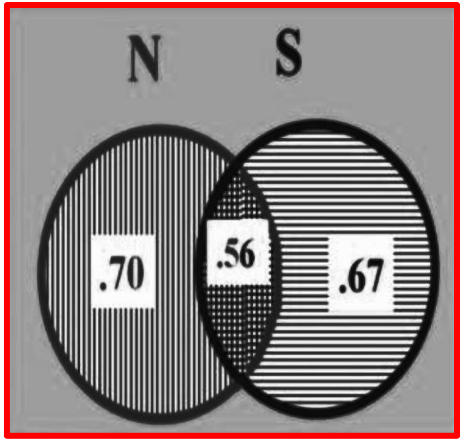
Solution

- Let **N** represent the event " reducing noise".
- Let **S** represent the event " more storage/ filling space".
- The probability of a person responding with N or S can be symbolized statistically as a union probability by using the law of addition.

P(N U S)

Solution

$$P(N \cup S) = P(N)+P(S) - P(N \cap S)$$



$$P(N) = .70$$
$$P(S) = .67$$
$$P(N \cap S) = .56$$
$$P(N \cup S) = .70 + .67 - .56$$
$$= 0.81$$

Office design problem: Probability Matrix

Increase Storage Space

Noise Reduction

	Yes	No	Total
Yes	0.56	0.14	0.70
No	0.11	0.19	0.30
Total	0.67	0.33	1.00

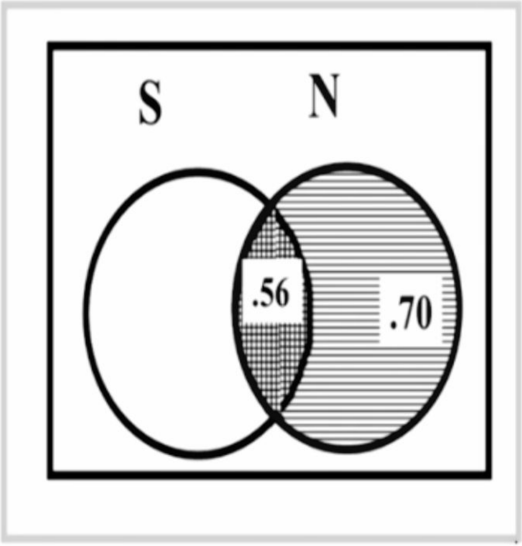
Joint Probability using a contingency Table

Event	Event		Total
	B_1	B_2	
A_1	$P(A_1 \text{ and } B_1)$	$P(A_1 \text{ and } B_2)$	$P(A_1)$
A_2	$P(A_2 \text{ and } B_1)$	$P(A_2 \text{ and } B_2)$	$P(A_2)$
Total	$P(B_1)$	$P(B_2)$	1

Joint Probabilities

Marginal (Simple) Probabilities

Law of Conditional probability



$$P(N) = .70$$

$$P(N \cap S) = .56$$

$$P(S|N) = \frac{P(N \cap S)}{P(N)}$$

$$= \frac{.56}{.70}$$

$$= .80$$

Office design problem

Increase Storage Space

Noise Reduction

	Yes	No	Total
Yes	0.56	0.14	0.70
No	0.11	0.19	0.30
Total	0.67	0.33	1.00

$$P(\bar{N}|S) = \frac{P\left(\bar{N} \cap S\right)}{P(S)} = \frac{0.11}{0.67}$$

$$= 0.164$$

Problem

- A company data reveal that 155 employees worked one of four types of positions.
- Shown here again is the raw values matrix (also called a contingency table) with the frequency counts for each category
- Subtotals and totals containing a breakdown of these employees by type of position and by gender.

Contingency Table

COMPANY HUMAN RESOURCE DATA

Gender

Type of Position

	Male	**Female**	Total
Managerial	8	3	11
Professional	31	13	44
Technical	52	17	69
Clerical	9	22	31
Total	100	55	155

Solution

- If an employee of the company is selected randomly, what is the probability that the employee is female or a professional worker?

$$P(F \cup P_w) = P(F) + P(P_w) - P(F \cap P_w)$$

$$P(F \cup P_w) = 0.355 + 0.284 - 0.084 = 0.555$$

Problem

- Shown here are the raw values matrix and corresponding probability matrix for the results of a national survey of 200 executives who were asked to identify the geographic location of their company and their company's industry type.
- The executives were only allowed to select one location and one industry type

Raw Values Matrix

<u>Geographic Location</u>

Industry Type

	Northeast D	Southeast E	Midwest F	West G	Total
Finance A	24	10	8	14	56
Manufacturing B	30	6	22	12	70
Communication C	28	18	12	16	74
Total	82	18	42	42	200

Questions

1. What is the probability that the respondent is from the Midwest (F)?
2. What is the probability that the respondent is from the communications industry (C) or from the Northeast (D)?
3. What is the probability that the respondent is from the southeast (E) or from the finance industry (A)?

Probability Matrix

Geographic Location

Industry Type

	Northeast D	Southeast E	Midwest F	West G	Total
Finance A	0.12	0.05	0.04	0.07	0.28
Manufacturing B	0.15	0.03	0.11	0.06	0.35
Communication C	0.14	0.09	0.06	0.08	0.37
Total	0.41	0.17	0.21	0.21	1.00

Mutually Exclusive Events

COMPANY HUMAN RESOURCE DATA

Gender

Type of Position

	Male	**Female**	Total
Managerial	8	3	11
Professional	31	13	44
Technical	52	17	69
Clerical	9	22	31
Total	100	55	155

$P(T \cup C) = P(T) + P(C)$

$= 69/155 + 31/155$

$= 0.645$

Law of Multiplication

- What will happen if event X and Y are independent event?

$$\boxed{P(X \cap Y) = P(X).P(Y|X) = P(Y).P(X|Y)}$$

Problem

- A company has 140 employees, of which 30 are supervisors.
- Eighty of the employees are married, and 20% of the married employees are supervisors.
- If a company employee is randomly selected, what is the probability that the employee is married and is a supervisor?

		Married		
		Yes	No	Subtotal
Supervisor	Yes	0.1143		30
	No			110
	Subtotal	80	60	140

$P(M) = 80/140 = 0.5714$

$P(S|M) = 0.20$

$P(M \cap S) = P(M).\ P(S|M)$

$= (0.5714)(0.20) = 0.1143$

Special law of multiplication for independent Events

- General Law:

$$P(X \cap Y) = P(X).P(Y|X) = P(Y).P(X|Y)$$

- Special Law:

If events X and Y are independent,

$$P(X) = P(X|Y), \text{ and } P(Y) = P(Y|X)$$

Consequently,

$$P(X \cap Y) = P(X).P(Y)$$

Law of Conditional Probability

- The conditional probability of X given Y is the **joint** probability of X and Y divided by the **marginal** probability of Y.

$$\boxed{P(X|Y) = \frac{P\left(X \cap Y\right)}{P(Y)} = \frac{P(Y|X)\,.P(X)}{P(Y)}}$$

Conditional Probability

- A conditional probability is the probability of one event, given that another event has occurred:

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

The conditional probability of A given that B has occurred

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

The conditional probability of B given that A has occurred

Where P(A and B) = joint probability of A and B
P(A) = marginal probability of A
P(B) = marginal probability of B

Problem

- Of the cars on a used car lot, 70% have air conditioning (AC) and 40% have a CD Player (CD). 20% of the cars have both.
- What is the probability that a car has a CD player, given that it has AC?
- We want to find P(CD | AC).

Probability Matrix

	CD	No CD	Total
AC	0.2	0.5	0.7
No AC	0.2	0.1	0.3
Total	0.4	0.6	1.0

$$\boxed{P(CD \mid AC) = \frac{P(CD \text{ and } AC)}{P(AC)} = \frac{2}{7} = 0.2857}$$

Solution: Decision Tree

Independent Events

- If X and Y are independent events, the occurrence of Y doesn't affect the probability of X occurring.
- If X and Y are independent events, the occurrence of X doesn't affect the probability of Y occurring.

If X and Y are independent events
$P(X|Y) = P(X)$, and
$P(Y|X) = P(Y)$

Statistical Independence

- Two events are independent if and only if:
  - $P(A|B) = P(A)$
- Events A and B are independent when the probability of one event is not affected by the other event.

Independent Events Demonstration

Geographic Location

Industry Type		Northeast D	Southeast E	Midwest F	West G	Total
	Finance A	0.12	0.05	0.04	0.07	0.28
	Manufacturing B	0.15	0.03	0.11	0.06	0.35
	Communication C	0.14	0.09	0.06	0.08	0.37
	Total	0.41	0.17	0.21	0.21	1.00

Test the matrix for the 200 executive responses to determine whether industry type is independent of geographic location.

Cont'd

$$P(A|G) = \frac{P(A \cap G)}{P(G)} = \frac{0.07}{0.21} = 0.33$$

$$P(A|G) = 0.33$$

$$Where\ as\ P(A) = 0.28$$

$$Hence\ P(A|G) \neq P(A) = 0.28$$

Revision of Probabilities: Bayes Rule

- An extension to the conditional law of probabilities.
- Enables revision of original probabilities with new information.

$$\boxed{P\left(X_1|Y\right) = \frac{P\left(Y|X_1\right)P\left(X_1\right)}{P\left(Y|X_1\right)P\left(X_1\right) + P\left(Y|X_2\right)P\left(X_2\right) + \ldots + P\left(Y|X_n\right)P\left(X_n\right)}}$$

Problem

- A particular type of printer ribbon is produced by only two companies, company A and company B.
- Suppose A produces 65% of the ribbons and B produces 35%.
- 8% of the ribbons produced by A are defective and 12% of the B ribbons are defective.
- A customer purchases a new ribbon. What is the probability that A produced the ribbon? What is the probability that B produced the ribbon?

Solution

P(A) = 0.65

P(B) = 0.35

P(def|A) = 0.08

P(def|B) = 0.12

$$P(A|def\}=\frac{P(def|A)P(A)}{P(def|A)P(A)+P(def|B)P(B)}$$

$$=\frac{(0.08(0.65))}{(0.08(0.65))+(0.12(0.35))}=0.553$$

$$P(B|def\}=\frac{P(def|B)P(B)}{P(def|A)P(A)+P(def|B)P(B)}$$

$$=\frac{(0.12(0.35))}{(0.08(0.65))+(0.12(0.35))}=0.447$$

Ribbon Problem with Bayes' Rule

Event	Prior probability $P(E_i)$	Conditional Probability $P(def \mid E_i)$	Joint Probability $P(E \cap def)$	Revised Probability $P(E_i \mid def)$
Company A	$\boxed{0.65}$	0.08	0.052	0.052/0.094=0.553
Company B	0.35	0.12	0.042	0.042/0.094=0.447

Problem

- Machines A, B, and C all produce the same two parts, X and Y. Of all the parts produced, machine A produces 60%, machine B produces 30%, and machine C produces 10%. In addition
  - 40% of the parts made by the machine A are part X.
  - 50% of the parts made by machine B are part X.
  - 70% of the parts made by machine C are part X.
- A part produced by this company is randomly sampled and is determined to be an X part.
- With the knowledge that it is an X part, revise the probabilities that the part came from machine A, B or C.

Solution

Event	Prior $P(E_i)$	Condition $P(X\|E_i)$	Joint prob; $P(X \cap E_i)$	Posterior
A	0.60	0.40	(0.60)(0.40)=0.24	0.24/0/46=0.52
B	0.30	0.50	0.15	0.15/0.46=0.33
C	0.10	0.70	0.07	0.07/0.46=0.15
			P(X) = 0.46	