Lead levels in drinking water

	Xi mg/dl	(Xi - Xbar)	(Xi - Xbar)^2
	12	2	4
	9	-1	1
	1	-9	81
	10	0	0
	18	8	64
sum	50	0	150

Sample Space (***S***): the set of all possible outcomes

Event (*E*): a collection of outcomes

Ex: Roll a die → ***S*** = {1, 2, 3, 4, 5, 6}

A: # of pips is odd A = {1, 3, 5}
B: # of pips ≥ 3 B = {3, 4, 5, 6}
C: # of pips is even C = {2, 4, 6}

Intersection $(E \cap F)$: The set of elements that belong to both *E* and *F*

Union $(E \cup F)$: The set of elements that belong to at least one of *E* or *F*

Complement $(\bar{E} \text{ or } E^c)$ All elements of ***S*** that are not in the set *E*

"Classical"* vs. Relative Frequency vs. Subjective Interpretations of Probability

Relative Frequency Definition of Probability

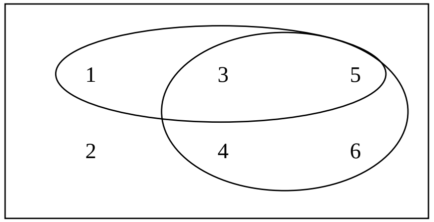
- If *n(E)*= # of times E occurs in *n* repetitions of the experiment
  Then, $P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$

Axioms of Probability

1. $0 \le P(E) \le 1$
2. $P(S) = 1$
3. If $E_1, E_2, and\ E_3$ are all mutually exclusive
   Then, $P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3)$

Ex: Roll a die → **S** = {1, 2, 3, 4, 5, 6}

A: odd # of pips A = {1, 3, 5}
B: # of pips ≥ 3 B = {3, 4, 5, 6}



$$A \cup B = \{1,3,4,5,6\} \to P(A \cup B) = \frac{5}{6}$$

$$P(A \cup B) = P(A) + P(B) = \frac{3}{6} + \frac{4}{6} = \frac{7}{6} \text{ ??}$$

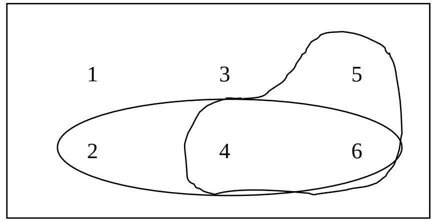
General Addition Rule for Probability

$$P(A \cup B) =$$

Conditional Probability

Ex: Roll a die → **S** = {1, 2, 3, 4, 5, 6}

C: even # of pips C = {2, 4, 6}
D: # of pips ≥ 4 D = {4, 5, 6}

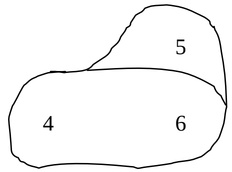


$$P(C) = \frac{3}{6},\ P(D) = \frac{3}{6}$$

What if we knew that event ***D*** occurs?

What is the probability of ***C*** if ***D*** occurs?

- If event ***D*** occurs, then it becomes our *reduced sample space*:



$$P(C \mid D) = \frac{\text{\# ways } C \cap D \text{ can occur}}{\text{\# points in } \textit{Reduced Sample Space}}$$
$$= \frac{n(C \cap D)}{n(D)}$$
$$= \frac{P(C \cap D)}{P(D)}$$

Definition of Conditional Probability

If $P(B) > 0$ then $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$

Events A and B are *independent* if $P(A \cap B) = P(A) \bullet P(B)$

Independence for A & B → conditional probabilities are the same as unconditional probabilities

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \bullet P(B)}{P(B)} = P(A)$$

Example: A card is drawn at random from an ordinary deck of cards

A: event that an ***Ace*** card is drawn
C: event that a ***Club*** card is drawn
B: event that a ***Black*** card is drawn

Are events ***A*** & ***C*** independent?

Are events ***B*** & ***C*** independent?

Sample Space (*S*) for rolling 2 die: ***n(S) = 36***

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

E. coli example

		Infected **D**	Not Infected **D'**
Test Positive	**B**	9500	100
Test Negative	**B'**	500	9900
	Total	10000	10000

In a <u>prospective study</u> individuals are followed for a period of time to see if a condition develops.

In a <u>retrospective study</u> individuals with & without a condition are asked about past exposures (or current status).

1. The effect of aspirin on heart attacks was investigated in the Physician's Health Study. In the study, over 22,000 physicians were assigned to either receive aspirin or a placebo (called the aspirin and placebo arms of the study). The number of Myocardial infarctions (MI) over the next 5 years was recorded for each arm of the study and are listed in the table below.

	MI	**No MI**	**Total**
Aspirin	139	10,898	11,037
Placebo	239	10,795	11,034
Total	378	21,693	22,071

a) According to the data from this study, what is the conditional probability (<u>risk</u>) of having an MI in the population of doctors taking the placebo? What is the value for those in the aspirin arm of the study?
b) Explain why it is appropriate to compute the risk (conditional probability) of an MI for each arm of the study?
c) Compute the ratio of the two values from part (a), the <u>relative risk</u> of an MI. State an interpretation of this relative risk value in plain language.

2. In a study to investigate variables affecting employment status, the registry of voters was used to select 5,000 employed individuals and 4,000 unemployed individuals at random from a small metropolitan area. Each individual's record was then checked to determine their education level.

	Employed	**Unemployed**	**Total**
Finished HS	4,283	3,125	7,408
Did Not Finish	717	875	1,592
Total	5,000	4,000	9,000

a) Is it appropriate to compute the risk of unemployment for each category of education? Explain why or why not?
b) Compute an appropriate conditional probability for the data based on this study.

<u>Random Variable (RV)</u>: a variable whose value is a numerical outcome determined by chance.

- X = # of heads in 2 tosses of a coin

- Y = # of pips on a tossed die

Types of Random Variables

- <u>*Continuous* Random Variables</u> have an infinite number of possible values.
- <u>*Discrete* Random Variables</u> have a finite number of possible values.

<u>Probability Distribution</u> (for a *continuous* RV) – represented as a <u>*Probability Density Curve*</u>

- Areas under a smooth curve indicate probabilities of values in a given range

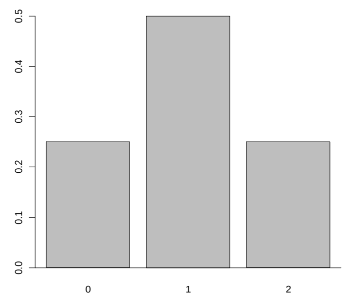
<u>Probability Distribution</u> (for a *discrete* RV) – represented graphically as a <u>*Probability Histogram*</u>

- Indicates how to assign probabilities for each of the possible values of a RV

<u>Example</u>: **X** = # of heads in 2 tosses of a coin

Outcome	**Probability**	**x**
TT	.5*.5=.25	0
TH	.5*.5=.25	1
HT	.5*.5=.25	1
HH	.5*.5=.25	2

x	**P(X=x)**
0	0.25
1	0.50
2	0.25



- Probability Distribution (for a *discrete* RV) – represented graphically as a ***Probability Histogram***
  - Indicates the possible values for a RV
  - Indicates how to assign probabilities for the possible values: $p(x) = P(X = x)$

Expected Value of a Random Variable

- a weighted average of all possible values for *X*, weighted by the probability of each value

  $E(X) = \mu$, the mean for the RV

  - $E(X) = \sum_{x=-\infty}^{\infty} x\, p(x)$ for a *discrete* RV

Example (*discrete* RV):
Y = # heads in two tosses of a fair coin

P(Y = 0) = ¼
P(Y = 1) = ½
P(Y = 2) = ¼ and zero otherwise

$$E(Y) = 0 \bullet P(Y=0) + 1 \bullet P(Y=1) + 2 \bullet P(Y=2)$$

$$= 0 \bullet \frac{1}{4} \quad + \quad 1 \bullet \frac{1}{2} \quad + \quad 2 \bullet \frac{1}{4} \quad = \quad 1$$

Properties of Expectation

*E(aX) = a*E(X)*	*E(X+Y)= E(X) + E(Y)*
E(X+a)= E(X) + a	*E(XY) = E(X) * E(Y)* if X & Y are independent

Variance of a Random Variable

- a weighted average of *squared deviations from the mean*, $[x - E(x)]^2$

$$Var(X) = E[(X-\mu)^2] = \sum_{x=-\infty}^{\infty} \left(x - E(x)\right)^2 p(x) = \sigma^2$$

Example (*continued*):
Y = # heads in two tosses of a fair coin

P(Y = 0) = ¼
P(Y = 1) = ½
P(Y = 2) = ¼ and zero otherwise

$$\mu = E(Y) = 1$$

$$Var(Y) = (0-\mu)^2 \bullet P(Y=0) + (1-\mu)^2 \bullet P(Y=1) + (2-\mu)^2 \bullet P(Y=2)$$

$$= (0-1)^2 \bullet \frac{1}{4} \quad + (1-1)^2 \bullet \frac{1}{2} \quad + (2-1)^2 \bullet \frac{1}{4} \quad = \frac{1}{2}$$

Properties of Variance

Var(X+Y) = Var(X) + Var(Y)	if X & Y are independent
*Var(X+Y) = Var(X) + Var(Y) + 2*Cov(X,Y)*	always