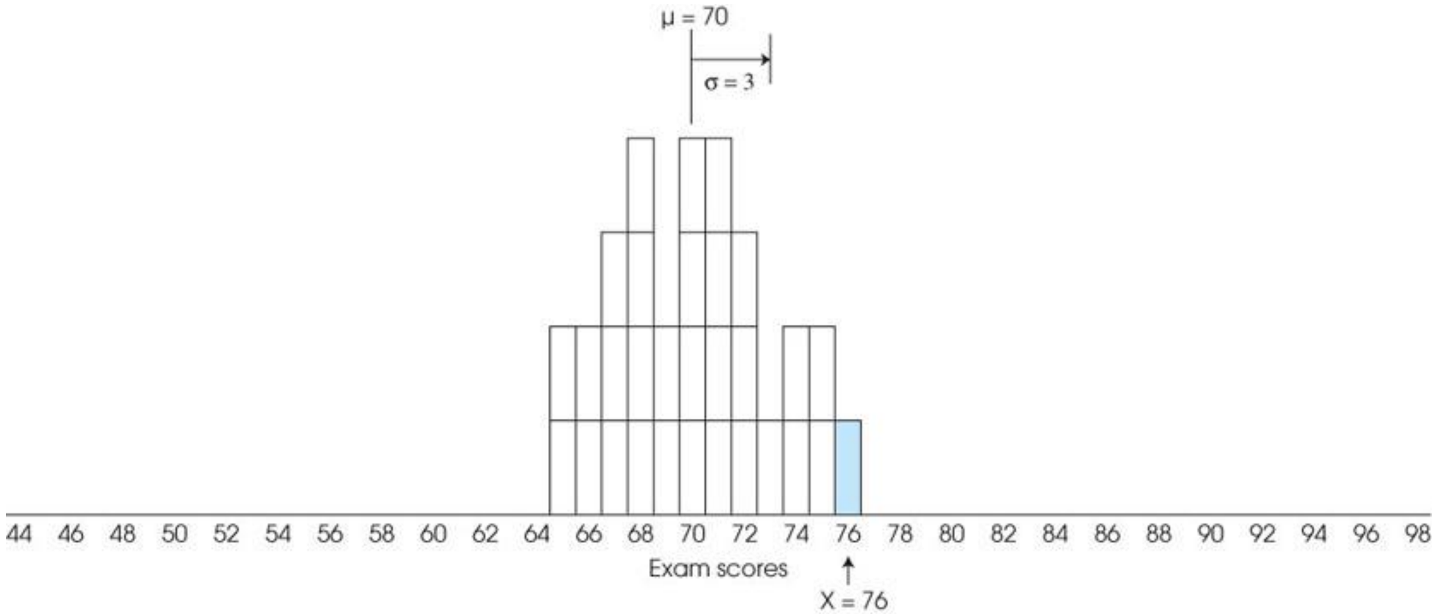
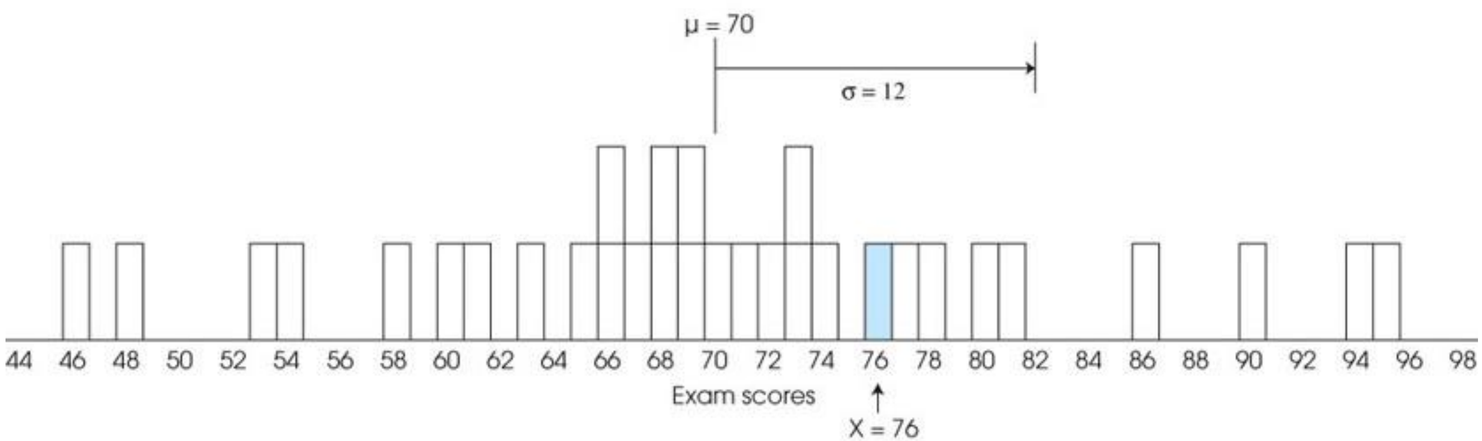
Lecture 04: z-Scores, Probability and Discrete Distributions

z-Scores: Location of Scores and Standardized Distributions

Why z-Scores? By itself, a raw score or X value provides very little information about how that particular score compares with other values in the distribution.





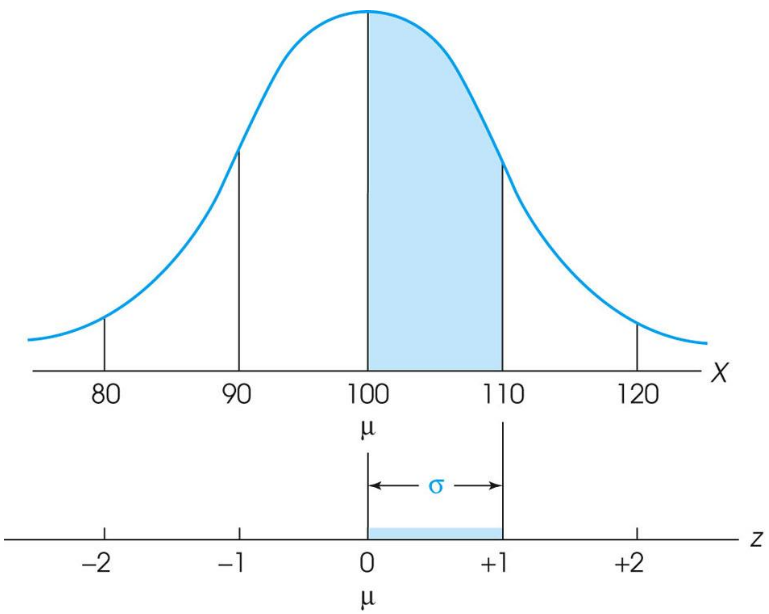
Computing *z*-Scores for a Population

- Transforming *X* to *z*:

$$z = \frac{X - \mu}{\sigma}$$

- Transforming *z* back to *X*:

$$X = \mu + z\sigma$$



z-Scores and Location

- The process of changing an X value into a z-score involves creating a signed number, called a **z-score**, such that
  - a. The sign of the z-score (+ or –) identifies whether the X value is located above the mean (positive) or below the mean (negative).
  - b. The numerical value of the z-score corresponds to the number of standard deviations between X and the mean of the distribution.
- If the raw score is transformed into a z-score, the value of the z-score tells exactly where the score is located relative to all the other scores in the distribution.

Examples

A distribution of scores has a mean of $\mu = 100$ and a standard deviation of $\sigma = 10$. What z-score corresponds to a score of $X = 130$ in this distribution?

According to the definition, the z-score will have a value of $+3$ because the score is located above the mean by exactly 3 standard deviations. Using the z-score formula, we obtain

$$z = \frac{X - \mu}{\sigma} = \frac{130 - 100}{10} = \frac{30}{10} = 3.00$$

For a distribution with a mean of $\mu = 60$ and $\sigma = 8$, what X value corresponds to a z-score of $z = -1.50$?

$$X = \mu + z\sigma = 60 - 1.50 * 8 = 48$$

Computing *z*-Scores for Samples

- Transforming *X* to *z*:

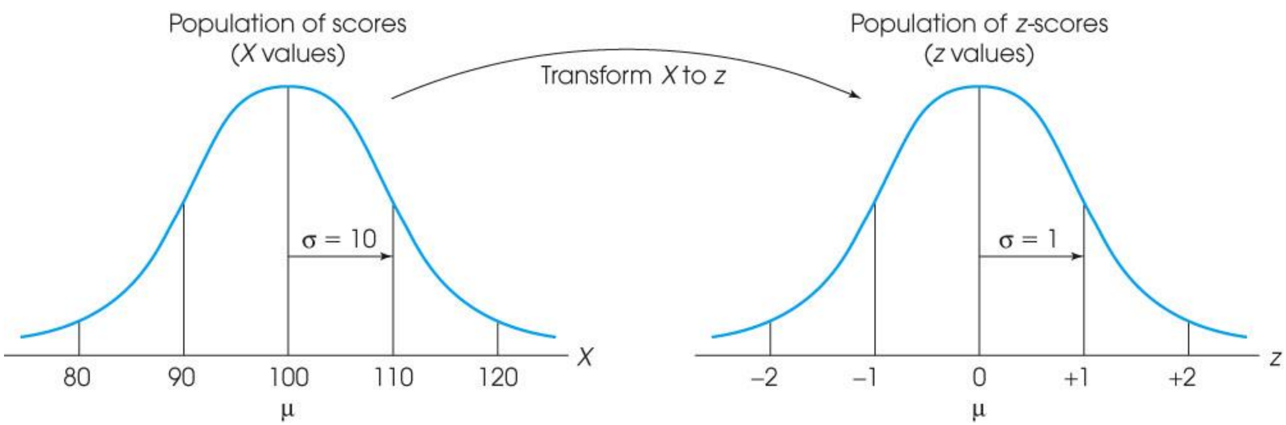
$$z = \frac{X - M}{s}$$

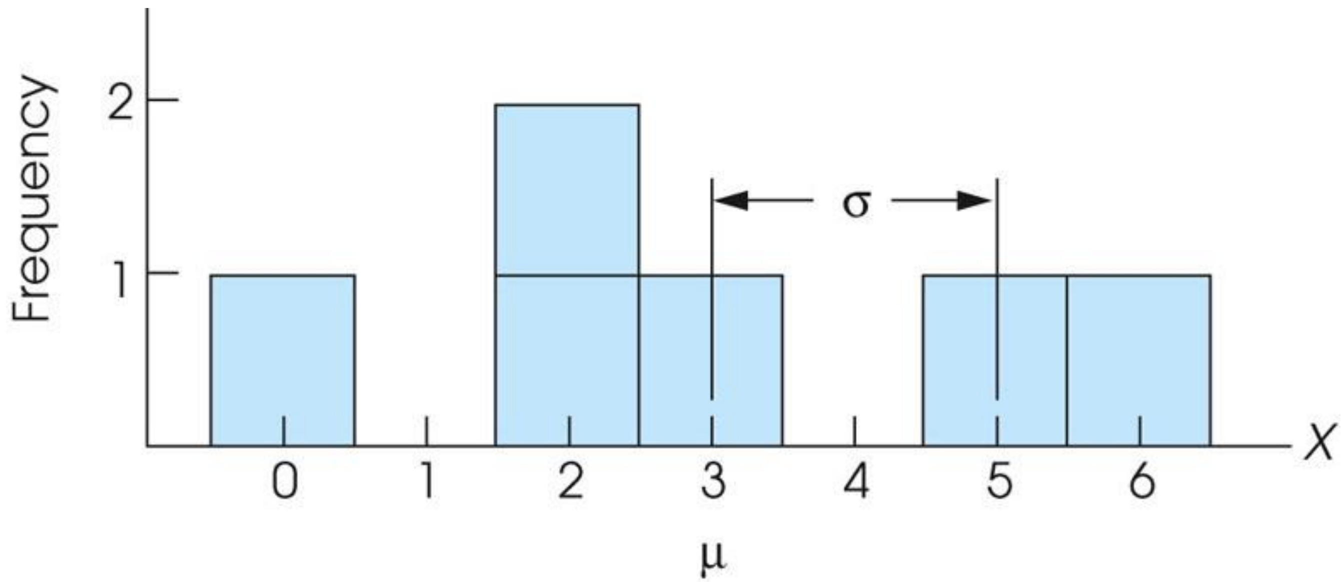
- Transforming *z* back to *X*:

$$X = M + zs$$

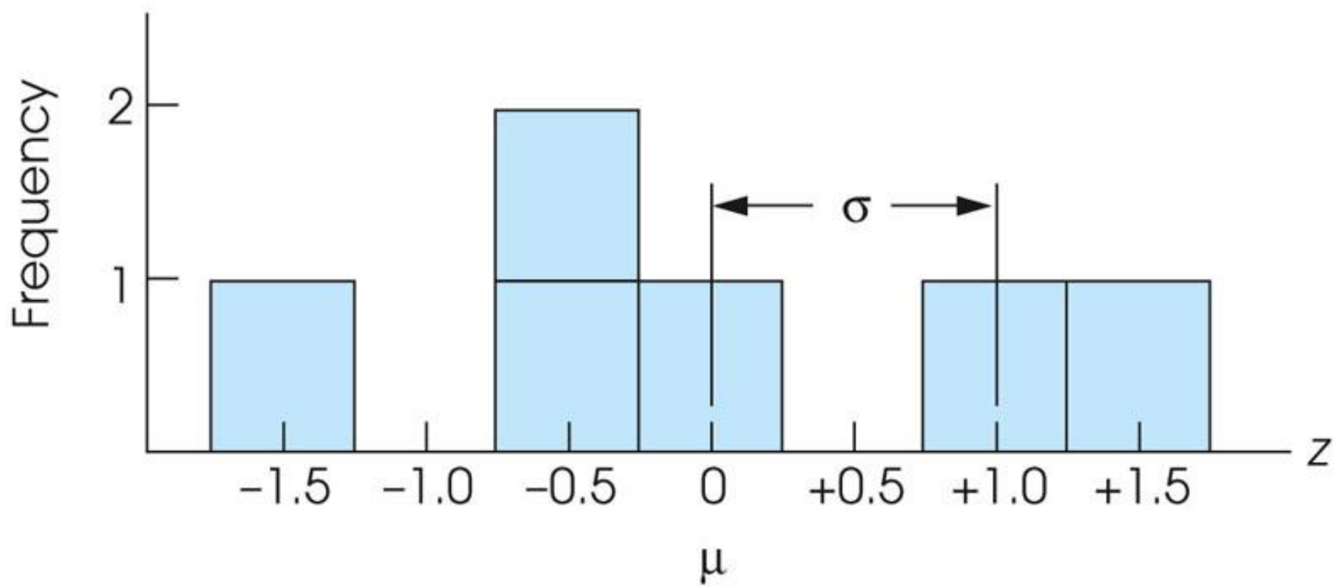
Properties of z-Score Transformation of Population Distributions

- **Shape**. The distribution of z-scores will have exactly the same shape as the original distribution of scores. The transformation does not change the location of any individual score relative to others in the distribution.
- **The Mean**. The z-score distribution will always have a mean of zero.
- **The Standard Deviation**. The distribution of z-scores will always have a standard deviation of 1.





Frequency
2
1
σ
X
0
1
2
3
4
5
6
μ



Frequency
2
1
σ
z
−1.5
−1.0
−0.5
0
+0.5
+1.0
+1.5
μ

z-Score Transformation of Sample Distributions

If all the scores in a sample are transformed into z-scores, the result is a sample distribution of z-scores. The transformed distribution of z-scores will have the same properties that exist when a population of X value is transformed into z-scores.

1. the distribution for the sample of z-scores will have the same shape as the original sample of scores.
2. the sample of z-scores will have a mean of $M_z = 0$.
3. the sample of z-scores will have a standard deviation of $s_z = 1$.

Example

X	z
0	-1.50
2	-0.50
4	$+0.50$
4	$+0.50$
5	$+1.00$

M = 3,
s = 2.

$$
\begin{aligned}
SS = \Sigma z^2 &= (-1.50)^2 + (-0.50)^2 + (+0.50)^2 + (0.50)^2 + (+1.00)^2 \\
&= 2.25 + 0.25 + 0.25 + 0.25 + 1.00 \\
&= 4.00
\end{aligned}
$$

$$s_z^2 = \frac{SS}{n-1} = \frac{4}{4} = 1.00$$

Using z-Scores for making comparisons

- A **standardized distribution** is composed of scores that have been transformed to create predetermined values for μ and σ. Standardized distributions are used to make dissimilar distributions comparable.

Example: Dave received a score of $X = 60$ on a psychology exam and a score of $X = 56$ on a biology test. For which course should Dave expect the better grade?

Dave's z-score for psychology is $$z = \frac{X - \mu}{\sigma} = \frac{60 - 50}{10} = \frac{10}{10} = +1.0$$

Dave's z-score for biology is $$z = \frac{56 - 48}{4} = \frac{8}{4} = +2.0$$

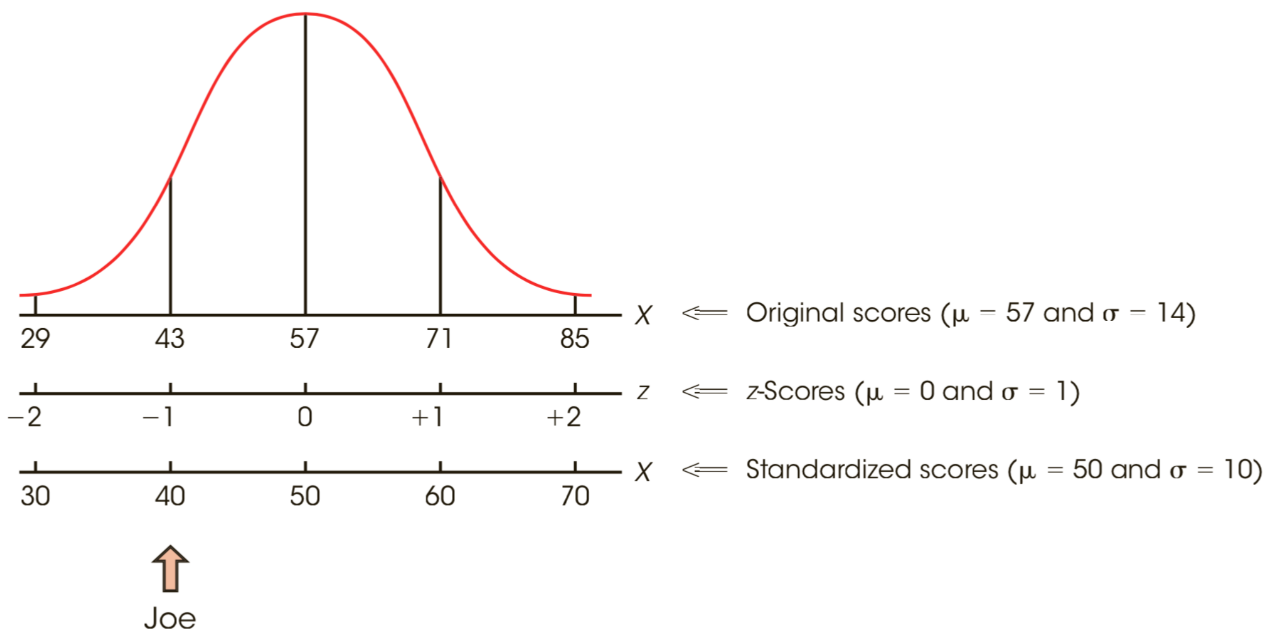
By relative class standing, Dave does much better in the biology class.

Other Standardized Distributions Based on z-Scores

- Although transforming X values into z-scores creates a standardized distribution, many people find z-scores burdensome because they consist of many decimal values and negative numbers.
- Therefore, it is often more convenient to standardize a distribution into numerical values that are simpler than z-scores.
- To create a simpler standardized distribution, you first select the mean and standard deviation that you would like for the new distribution.
- Then, z-scores are used to identify each individual's position in the original distribution and to compute the individual's position in the new distribution.

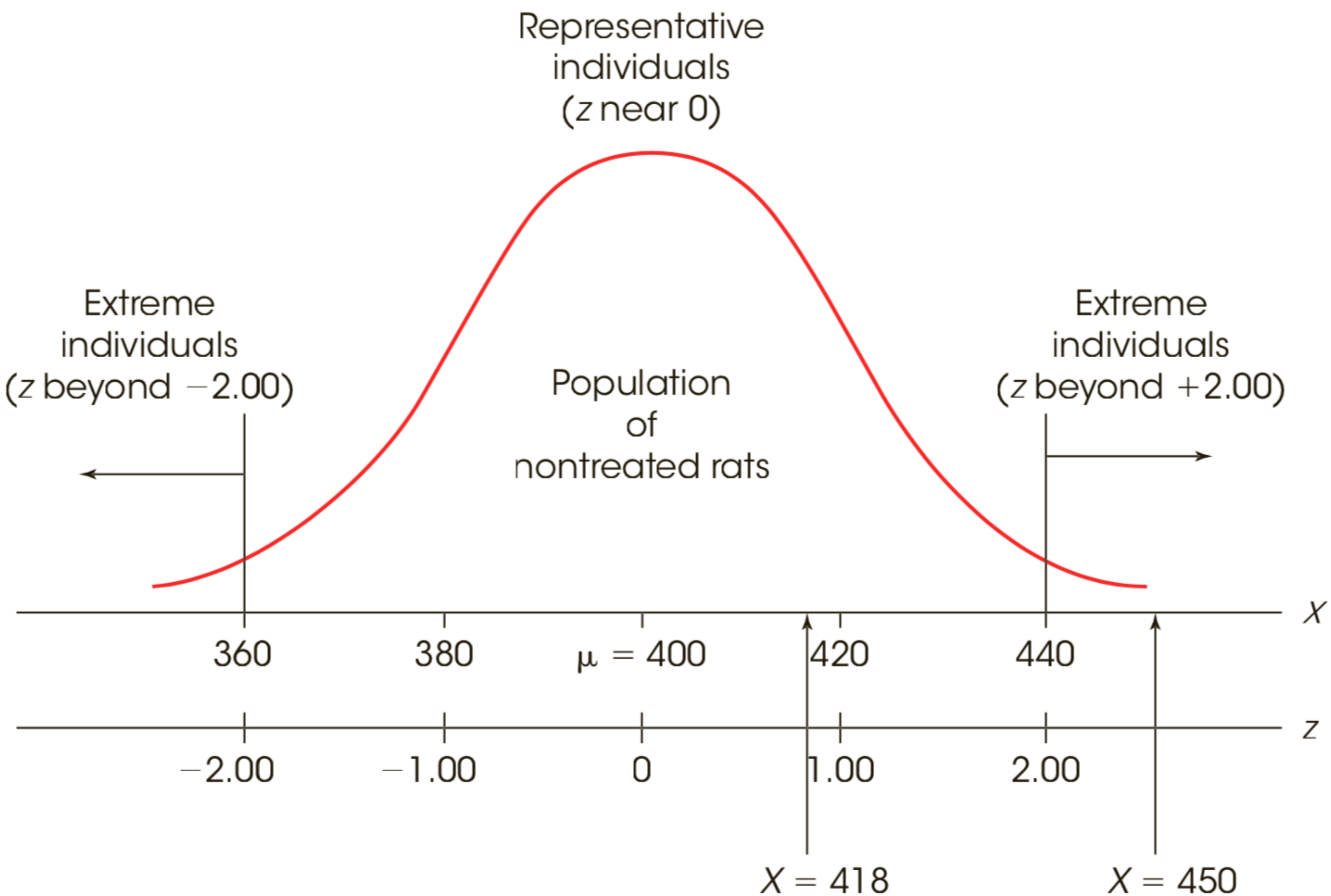
An instructor gives an exam to a psychology class. For this exam, the distribution of raw scores has a mean of $\mu = 57$ with $\sigma = 14$. The instructor would like to simplify the distribution by transforming all scores into a new, standardized distribution with $\mu = 50$ and $\sigma = 10$. To demonstrate this process, we will consider what happens to two specific students: Maria, who has a raw score of $X = 64$ in the original distribution, and Joe, whose original raw score is $X = 43$.

	Original Scores $\mu = 57$ and $\sigma = 14$		z-Score Location		Standardized Scores $\mu = 50$ and $\sigma = 10$
Maria	$X = 64$	$\rightarrow$	$z = +0.50$	$\rightarrow$	$X = 55$
Joe	$X = 43$	$\rightarrow$	$z = -1.00$	$\rightarrow$	$X = 40$



z-scores and Locations

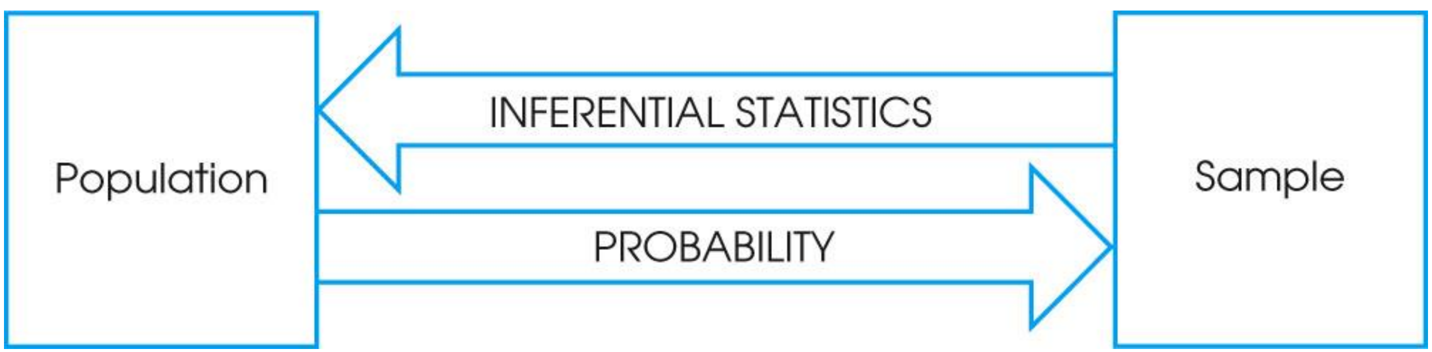
- The fact that z-scores identify exact locations within a distribution means that z-scores can be used as descriptive statistics and as inferential statistics.
  - As descriptive statistics, z-scores describe exactly where each individual is **located**.
  - As inferential statistics, z-scores determine whether a specific sample is **representative** of its population, or is extreme and unrepresentative.


Representative
individuals
(z near 0)
Extreme
individuals
(z beyond −2.00)
Population
of
nontreated rats
Extreme
individuals
(z beyond +2.00)
X
360
380
μ = 400
420
440
z
−2.00
−1.00
0
1.00
2.00
X = 418
X = 450

Introduction of probability

Probability and Inferential Statistics

- Probability is important because it establishes a link between samples and populations.
- For any known population it is possible to determine the probability of obtaining any specific sample.
- In later chapters we will use this link as the foundation for inferential statistics.



概率

Probability is the measure of the likelihood that an **event** will occur.

事件

$$P(A)$$

$$P(B)$$

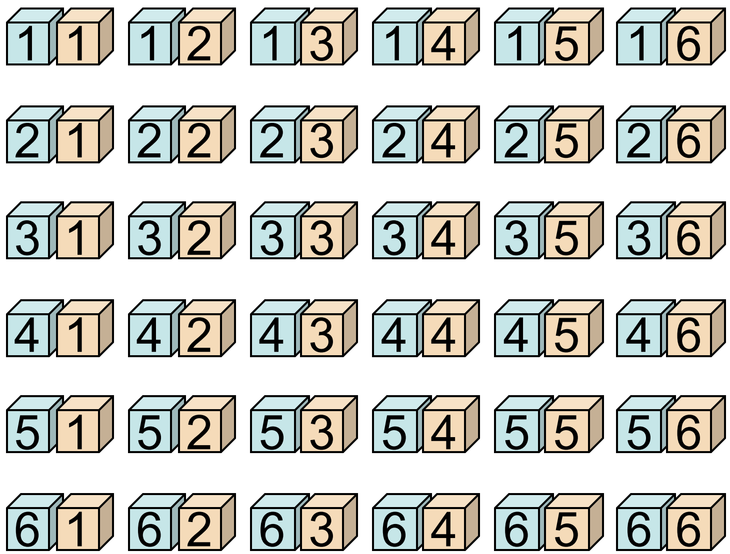
$$P(X<1)$$

$$P(X^2+Y^2=5)$$

“Pr” is also used to denote probability.

Events are sets of outcomes. $A = \{a_1, a_2, a_3\}.$

集合



Two Dice
36 outcomes

Examples

{(1,1)};

{(1,1), (2,4), (4,6)};

the total equals 9;

…

Review of set notations

S universal set
全集

$\emptyset$ null set, or
empty set
空集

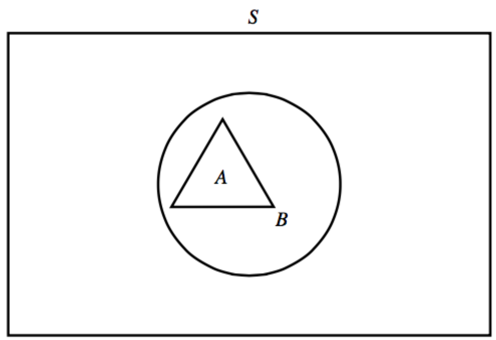
$A \subset B$ A is a subset of B
A是B的子集

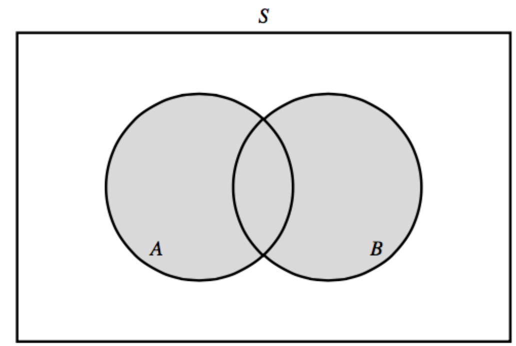
$A \cap B$ or AB
Intersection of A and B
交集

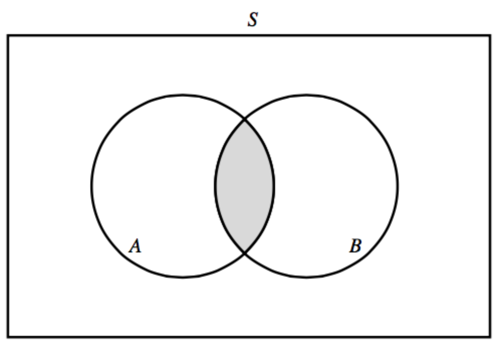
$A \cup B$ Union of A and B
并集

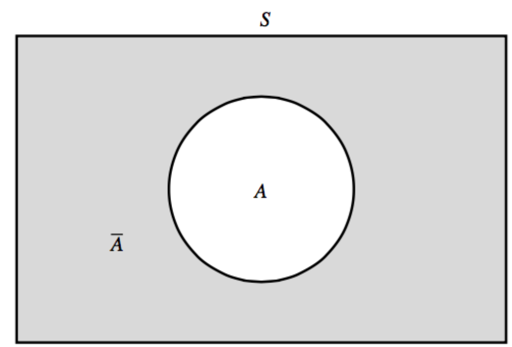
$\bar{A}$ or A^c Complement of A
补集

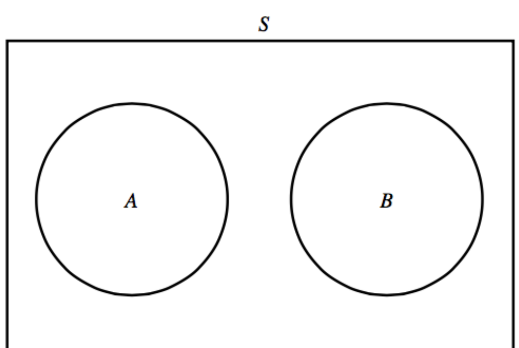
Venn diagrams









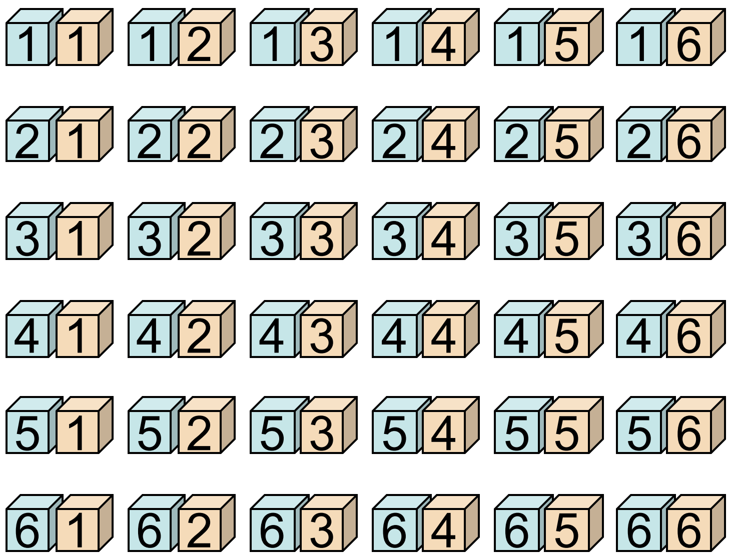


Mutually exclusive 互斥

$$A \cap B = \emptyset$$

Properties of probability

$$\boxed{P[A] = \sum_{\{i:s_i \in A\}} p_i}$$



$$p_i = 1/36$$

Two Dice
36 outcomes

Properties of probability

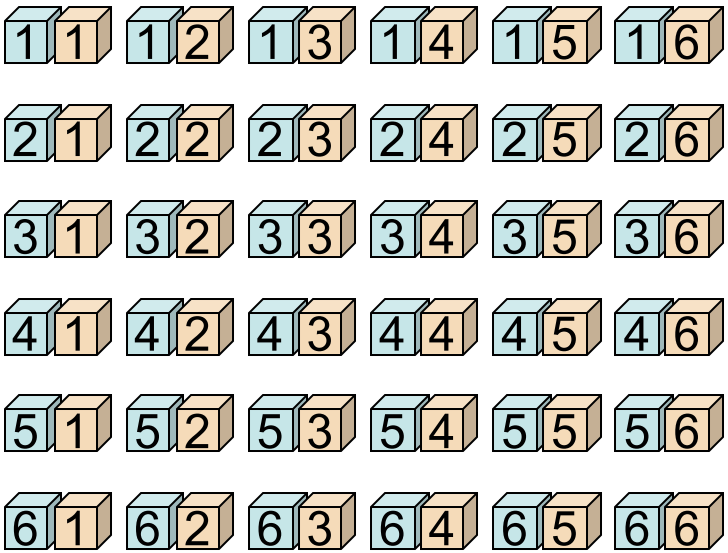
$$\boxed{P\left[A\right] = \sum_{\{i:s_i \in A\}} p_i}$$

$$P\left[S\right] = 1 \quad P\left[\varnothing\right] = 0$$

$$P\left[A\right] + P\left[A^c\right] = 1$$

$$P\left[A \cup B\right] = P\left[A\right] + P\left[B\right] - P\left[AB\right]$$

Calculate the probabilities of the events



Two Dice
36 outcomes

Examples

{(1,1)};

{(1,1), (2,4), (4,6)};

the total equals 9;

…

条件概率

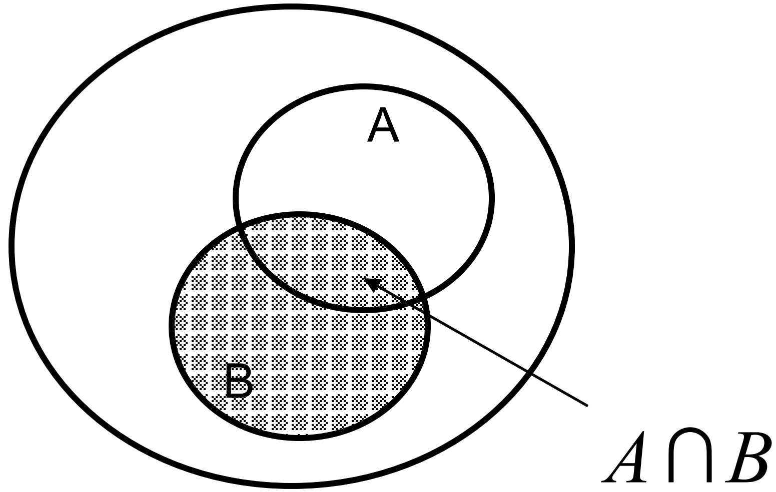
Conditional Probability

$$P[A \mid B] = P[AB] / P[B] \quad if \ P\left[B\right] \neq 0$$

Properties

$$P[AB] \quad = P\left[A \mid B\right] P\left[B\right]$$

Addition Law



$$\boxed{P\left[A \cup B\right] = P\left[A\right] + P\left[B\right] - P\left[AB\right]}$$

Independence

Two events A and B are independent precisely when

$$P[A \mid B] = P[A]$$

Question: Does this imply

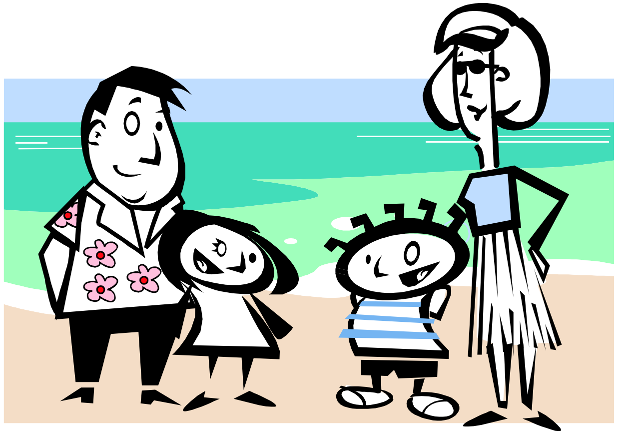
$$P[B \mid A] = P[B] \text{ ?}$$

Bayes Theorem

Thomas Bayes(?)

$$P\left[A \mid B\right] = P\left[B \mid A\right] \times \frac{P\left[A\right]}{P\left[B\right]}$$

Middleville



In Middleville, every family is happy, and every family has two children.

The children were brought by the stork and the stork delivers boys and girls with equal probability.

I pick a family at random and discover that one of the children is a boy. What is the probability that the other child is a boy?

Cancer detected

- Cancer X has a prevalence of 0.1% among the population.
- We know the doctors can correctly find it when you do have cancer with a probability of 100%.
- We know they can correctly dismiss it when you do not have cancer with a probability of 95%.

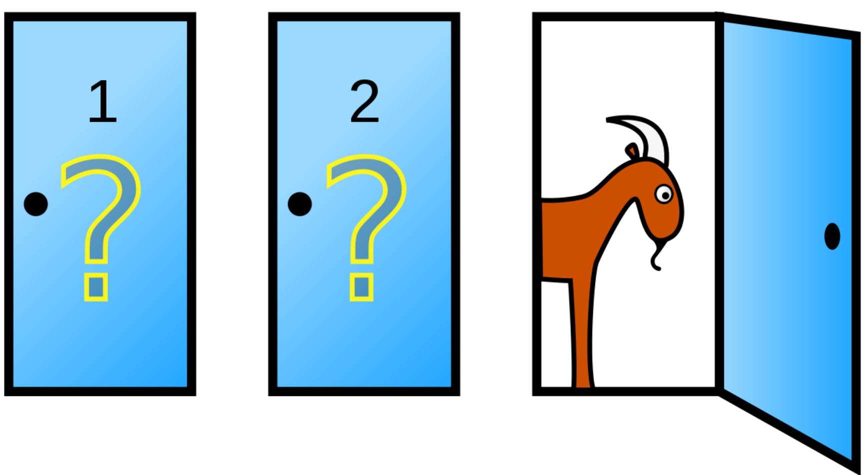
And,

- Today, the doctor say you are diagnosed with cancer

What’s the probability that you are really unlucky?

Monty Hall problem

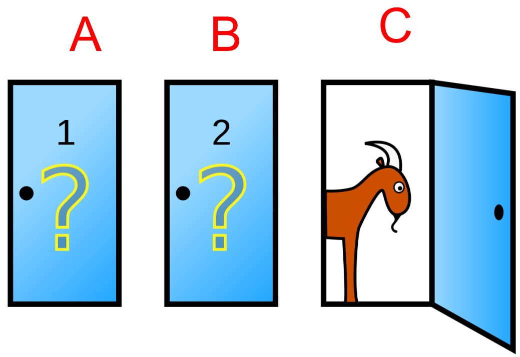
- Two goats and one car, randomly placed behind three doors.
- You are allowed to pick a door. If the car is behind, you can keep it.
- After your pick, the host Monty will open one door, revealing a goat.



Question:

- You picked 1, she revealed 3.
- Now, will you *switch*, if allowed?

With Bayesian theory



- **Prior**: p(A) = 1/3, p(B) = 1/3, p(C) = 1/3
- **Posterior**: p(A|openC) or p(B|openC) ???
- **Likelihood**: p(openC|A) = ½, p(openC|B) = 1, p(openC|C) = 0

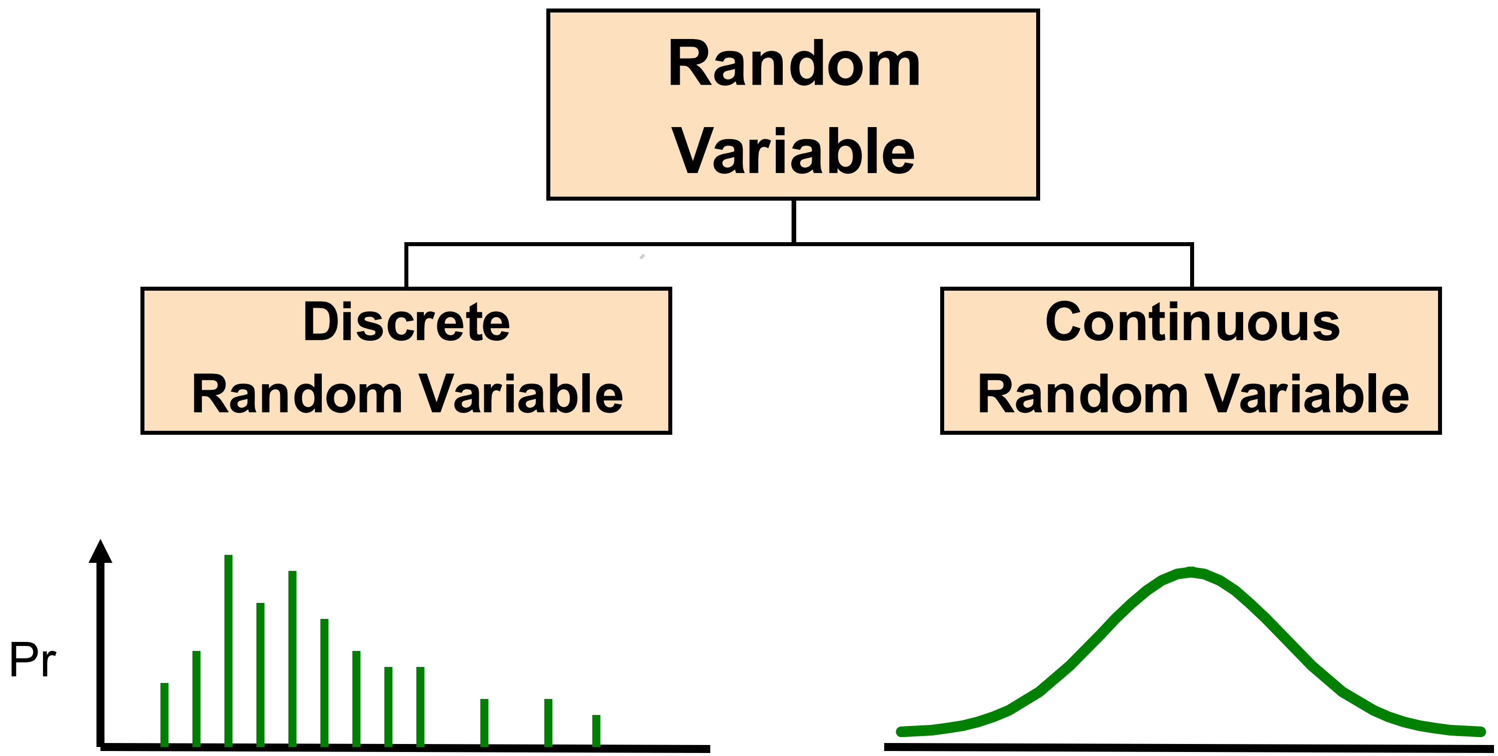
$$p(A|\text{openC}) = \frac{p(openC|A) * p(A)}{p(openC)}$$

$$= \frac{p(openC|A)*p(A)}{p(openC|A)*p(A)+p(openC|B)*p(B)+p(openC|C)*p(C)}$$

$$= \frac{1/2*1/3}{\left(\frac{1}{2}+1+0\right)*1/3} = 1/3$$

Random Variable
（RV, 随机变量）

- A **random variable** is a numeric quantity that takes different values with specified probabilities.



Discrete Random Variable（非连续性随机变量，离散随机变量）

- A random variable for which there exists a discrete set of values with specified probabilities is a **discrete random variable**.
- e.g.: Let x be the value on the roll of a die. *X* is a discrete random variable taking on values 1,2,3,4,5,6. $\Pr(X = x)$, $\Pr(X = 1)$

Capital letter: random variable
Small letter: value of random variable

Continuous Random Variable （连续性随机变量）

- A **continuous random variable** is a variable that can assume any value on a continuum (an uncountable number of values, cannot be enumerated)
- e.g. Let *Z* be the **weight** of a dice (in grams). *Z* is a continuous random variable taking on any positive real numbers.

$Pr(1<Z<1.7)=0.6$

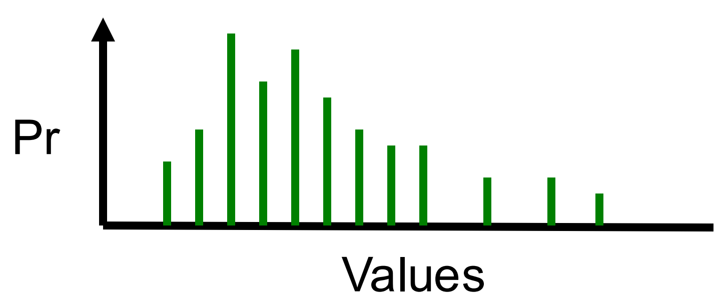
$Pr(z_1<Z<z_2)$

Transformation of Random Variables

- X and Y are random variables, then X+3, X^2, log(X), X+Y, X*Y are all random variables
- If X and Y are both discrete random variables, X+3, log(x), X^2, X+Y, X*Y are all discrete
- If X or Y is a continuous random variable, X+Y and X*Y are all continuous

Probability Mass Function（概率质量函数）for Discrete RV

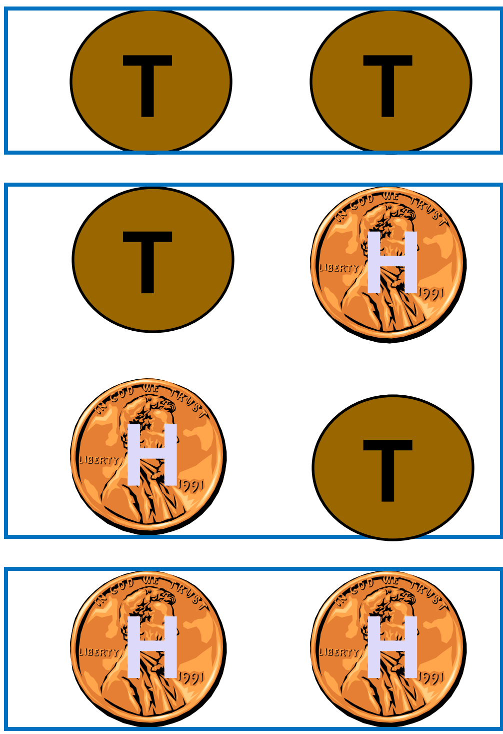
- A **Probability Mass Function** (pmf) is a mathematical relationship which assigns a probability to each possible value, *x*, of the discrete random variable *X*.
- Pr(X = x) is a function of x, also denoted p(x).
- Probability Mass Function = **Probability Distribution**（概率分布）



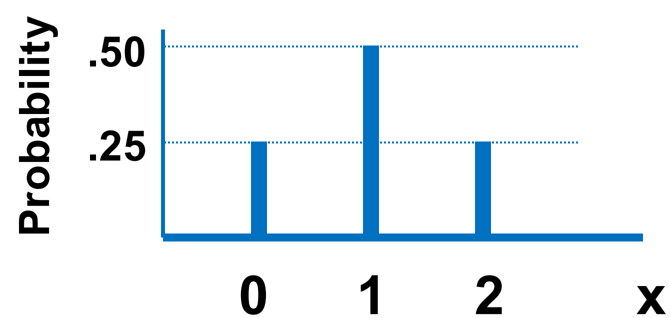
Probability Distribution

Experiment: Toss **2** Coins. Let X = # heads.

4 possible outcomes



X Value	Probability
0	1/4 = .25
1	2/4 = .50
2	1/4 = .25



Expected Value （预期值，期望值） of a Discrete RV

- Let X be a discrete random variable with the probability function p(x). The expected value of X is defined as:

$$E(X) = \sum_{i=1}^{k} x_i p(x_i)$$

- Expected value is a measure of location/centrality （位置或集中趋势） for a random variable
- The value of E(X) is usually denoted μ.

Its relation to MEAN? Analogous to center of mass?

Let the random variable *x* represent the number of boyfriends (or girlfriends) in the next 6 years of your life (starting from freshman year). Suppose the probability mass function is:

x	0	1	2	3	4	5	6
Pr(X=x)	0.129	0.264	0.271	0.185	0.095	0.039	0.017

$E(X) = 0(0.129) + 1(0.264) + 2(0.271) + 3(0.185) + 4(0.095)$
$+ 5(0.039) + 6(0.017) = 2.04$

St. Petersburg paradox

Nicholas Bernouilli, 1728:

Suppose someone offers to toss a fair coin repeatedly until it comes up heads, and to pay you $1 if this happens on the first toss, $2 if it takes two tosses to land on a head, $4 if it takes three tosses, $8 if it takes four tosses, etc. How much would you pay to play this game?

St. Petersburg paradox

Nicholas Bernouilli, 1728:

Suppose someone offers to toss a fair coin repeatedly until it comes up heads, and to pay you $1 if this happens on the first toss, $2 if it takes two tosses to land on a head, $4 if it takes three tosses, $8 if it takes four tosses, etc. How much would you pay to play this game?

What is the expected value to play this game?

St. Petersburg paradox

Nicholas Bernouilli, 1728:

Suppose someone offers to toss a fair coin repeatedly until it comes up heads, and to pay you \$1 if this happens on the first toss, \$2 if it takes two tosses to land on a head, \$4 if it takes three tosses, \$8 if it takes four tosses, etc. How much would you pay to play this game?

What is the expected value to play this game?

$\Sigma_{num\ tosses}\ \$2^{num\ tosses-1} \cdot P(num\ tosses)$

$= 1/2 * \$1 + 1/4 * \$2 + 1/8 * \$4 + 1/16 * \$8 + \ldots$

$= \$.50 + \$.50 + \$.50 + \$.50 + \ldots$

= *infinite* expected value

Variance of a Discrete RV

- The variance of a discrete random variable X, is defined as:

$$Var(X) = E(X-\mu)^2 = \sum_{i=1}^{k} (x_i - \mu)^2 \, p(x_i)$$

- Var(X) is a measure of spread, also known as population variance.
- The value of Var(X) is usually denoted σ^2.
- The standard deviation of X is the square root of Var(X), denoted SD(X) = σ.

$$\boxed{Var\left(X\right)=E\left(X-\mu\right)^2=E\left(X^2\right)-\left[E\left(X\right)\right]^2}$$

Derivation

$$Var\left(X\right)=\sum_{i=1}^{k}\left(X_i-\mu\right)^2\Pr\left(X=x_i\right)=\sum_{i=1}^{k}\left(x_i^2-2\mu x_i+\mu^2\right)\Pr\left(X=x_i\right)$$

$$=\sum_{i=1}^{n}x_i^2\Pr\left(X=x_i\right)-2\mu\sum_{i=1}^{n}x_i\Pr\left(X=x_i\right)+\sum_{i=1}^{n}\mu^2\Pr\left(X=x_i\right)$$

$$=\sum_{i=1}^{n}x_i^2\Pr\left(X=x_i\right)-2\mu^2+\mu^2=\sum_{i=1}^{n}x_i^2\Pr\left(X=x_i\right)-\mu^2$$

e.g.: Compute variance and standard deviation of random variable (# of bf or gf) whose probability mass function is given:

x	0	1	2	3	4	5	6
Pr(X=x)	0.129	0.264	0.271	0.185	0.095	0.039	0.017

μ = 2.04

$$Var(X) = \sigma^2 = \sum_{i=1}^{k} x_i^2 \Pr(X = x_i) - \mu^2$$

$$= 0^2(.129) + 1^2(.264) + 2^2(.271) + 3^2(.185) + 4^2(.095)$$

$$+ 5^2(.039) + 6^2(.017) - (2.04)^2 = 1.96$$

$$\sigma = \sqrt{1.96} = 1.40$$

Cumulative Distribution Function（cdf, 累积分布函数）of a Discrete RV

The **Cumulative Distribution Function** of a Discrete Random Variable X is defined as

$$F(x) = Pr(X \leq x)$$

Can two random variables have the same pmf but different cdfs?

e.g. Age of Participants						
x	3	4	5	6	7	9
$f(x) = Pr(X=x)$	.5	.2	.1	.1	.05	.05

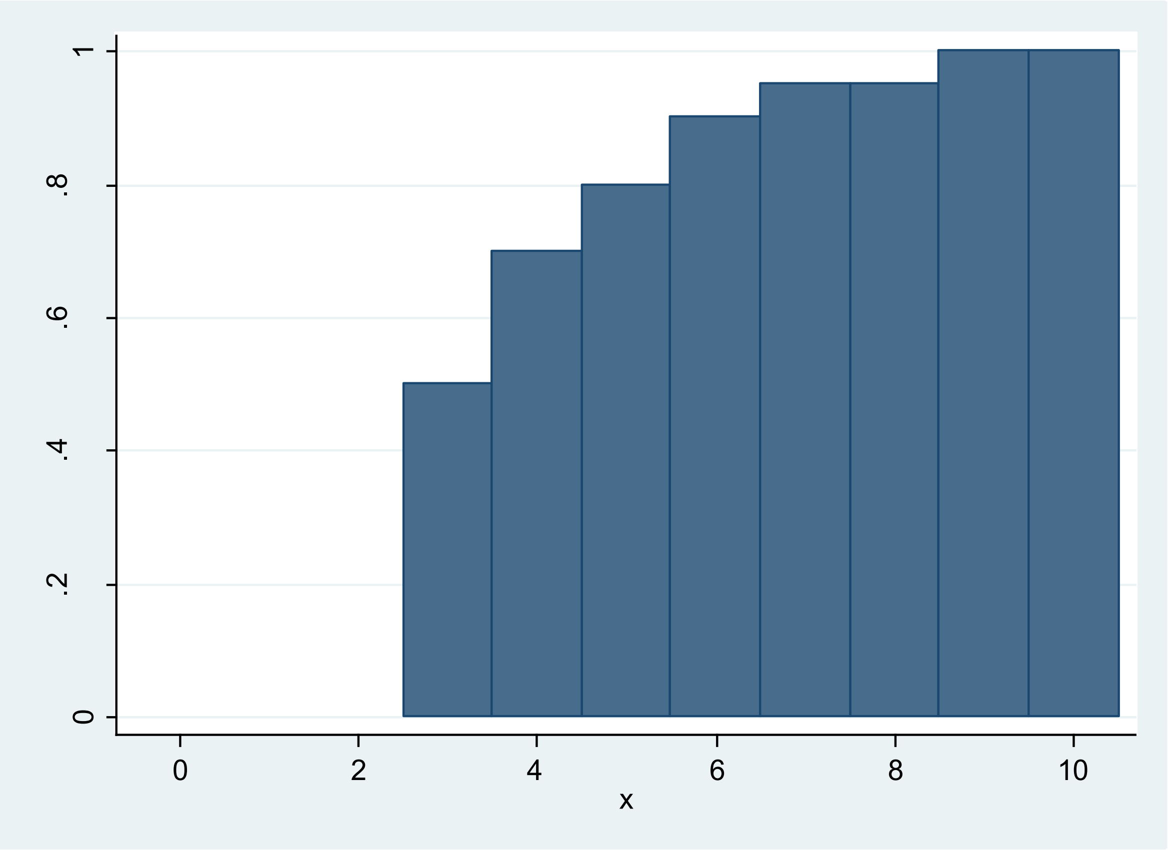
$\boldsymbol{F(x) = Pr(X \leq x)}$		
$F(x) = 0$	if $x < 3$	$\boldsymbol{x = 0,1,2}$
$F(x) = 0.5$	if $3 \leq x < 4$	$\boldsymbol{x = 3}$
$F(x) = 0.7$	if $4 \leq x < 5$	$\boldsymbol{x = 4}$
$F(x) = 0.8$	if $5 \leq x < 6$	$\boldsymbol{x = 5}$
$F(x) = 0.9$	if $6 \leq x < 7$	$\boldsymbol{x = 6}$
$F(x) = 0.95$	if $7 \leq x < 9$	$\boldsymbol{x = 7,8}$
$F(x) = 1.0$	if $9 \leq x$	$\boldsymbol{x = 9,10,\ldots}$

Evaluate $F(8)=Pr(X\leq 8)$: directly from the table above $F(8)=0.5+0.2+0.1+0.1+0.05=0.95$.

Plot of CDF

f(x) = Pr(X=x)	.5	.2	.1	.1	.05	.05
x	3	4	5	6	7	9

```
     +------------+
     |  x   F(x)  |
     |------------|
  1. |  0     0   |
  2. |  1     0   |
  3. |  2     0   |
  4. |  3    .5   |
  5. |  4    .7   |
  6. |  5    .8   |
  7. |  6    .9   |
  8. |  7   .95   |
  9. |  8   .95   |
 10. |  9     1   |
 11. | 10     1   |
     +------------+
```



Binomial Probability Distribution （二项分布）

The earliest probability distribution being studied (1713)

Suppose the probability for a coin turning heads is p.

If you toss the coin for n times, what is the probability for you to observe x heads?

Jacob Bernoulli

Binomial Probability Distribution （二项分布）

- A fixed number of observations, n
  e.g., 15 tosses of a coin; ten light bulbs taken from a warehouse
- Two mutually exclusive and collectively exhaustive categories
  - e.g., head or tail in each toss of a coin; defective or not defective light bulb
  - Generally called "success" and "failure"
  - The probability of success is p, probability of failure is $1 – p$
- Observations are independent
  The outcome of one observation does not affect the outcome of the other
- Constant probability for each observation
  e.g., Probability of getting a tail is the same each time we toss the coin

Binomial Distribution Formula

$$P(x) = \frac{n!}{x!(n-x)!} p^{x} (1-p)^{n-x}$$

P(x) = the probability of **x** successes in **n** trials, with a probability of success **p** on each trial

x = the number of 'successes' in a sample, (x = 0, 1, 2, ..., n)

n = sample size (number of trials or observations)

p = the probability of "success"

Example: Flip a coin four times, let x = # heads:

n = 4

p = 0.5

1 - p = (1 - .5) = .5

X = 0, 1, 2, 3, 4

Example: calculating a Binomial Probability

What is the probability of one success in five observations if the probability of success is .1?

x = 1, n = 5, and p = .1

$$P(x=1) = \frac{n!}{x!(n-x)!}p^{x}(1-p)^{n-x}$$

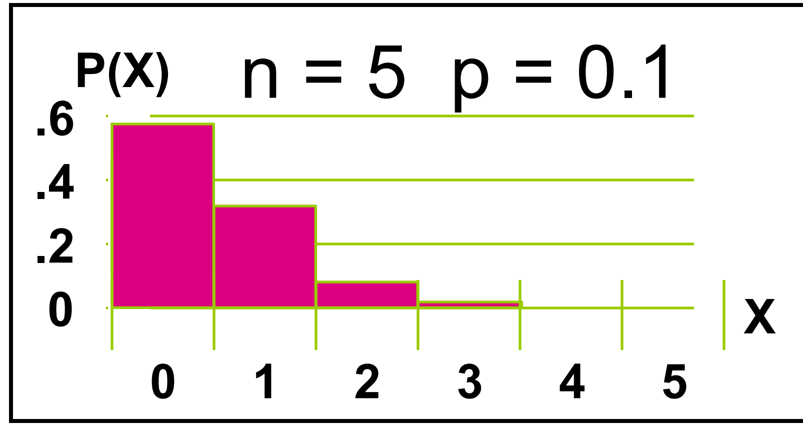
$$= \frac{5!}{1!(5-1)!}(.1)^{1}(1-.1)^{5-1}$$

$$= (5)(.1)(.9)^{4}$$

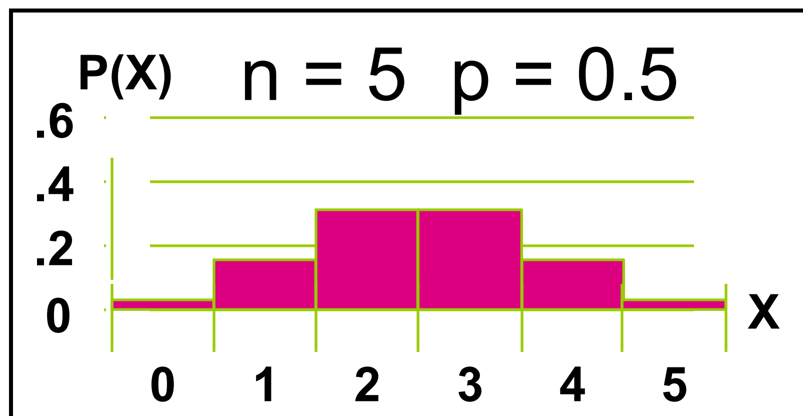
$$= .32805$$

The shapes of binomial distribution

- The shape of the binomial distribution depends on the values of p and n
  - Here, n = 5 and p = .1



  - Here, n = 5 and p = .5



Binomial Distribution Characteristics

- Mean

$$\mu = E(x) = np$$

- Variance and Standard Deviation

$$\sigma = \sqrt{np(1-p)}$$

Where n = sample size

p = probability of success

(1 – p) = probability of failure

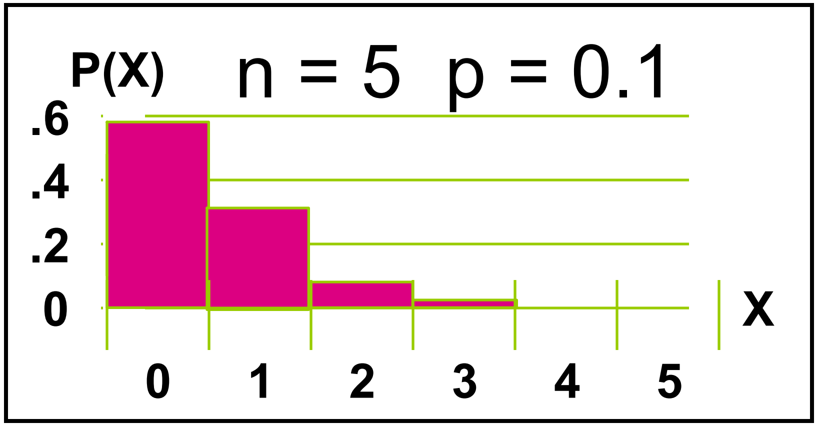
Exercise: prove this?

Binomial Characteristics

Examples:

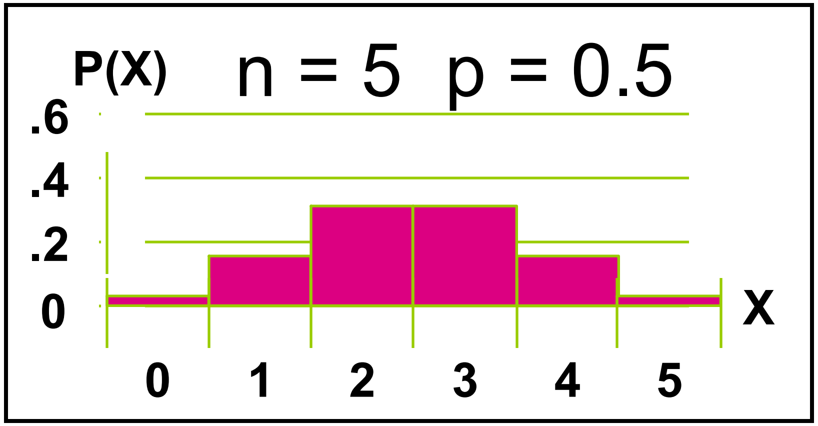
$$\mu = np = (5)(.1) = 0.5$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{(5)(.1)(1-.1)}$$
$$= 0.6708$$



$$\mu = np = (5)(.5) = 2.5$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{(5)(.5)(1-.5)}$$
$$= 1.118$$



Using Binomial Tables

The table can be found in the appendix of Wackerly et al. 2008

Or check internet:
http://mat.iitm.ac.in/home/vetri/public_html/statistics/binomial.pdf
Example: p=0.4, n=10, k=7, check the table….

A note on Binomial tables

Table goes up to p=0.5. What if p>0.5?
Reverse the roles of success and failure.
e.g. p=0.6, n=10, k=3
Look up p=0.4, n=10, k=7

Key concepts for this introduction of probability

- Z scores
- Random variable （随机变量）
- Probability distribution （概率分布）
- Probability mass function （概率质量函数）
- Cumulative distribution function （累积分布函数）
- Expected value （期望值）
- Variance & standard deviation （方差 & 标准差）
- Binomial distribution （二项分布）

What shall we know about a particular probability distribution?