

## Chapter 8 Probability Distributions and Statistics

### Section 8.1 Distributions of Random Variables

The value of the result of the probability experiment is a RANDOM VARIABLE.

If you roll a die we can let  $X$  be the number of dots showing,

$$X = \{1, 2, 3, 4, 5, 6\}$$

If we have a hand of three cards,  $X$  could be the number of clubs in the hand,

$$X = 0, 1, 2, 3$$

These examples are all finite discrete random variables.

#### Example

Roll a single die and count the number of rolls until a 6 comes up.

outcome	$Y$
6	1
2, 6	2
1, 4, 6	3
⋮	⋮

inf. disc. r.v.

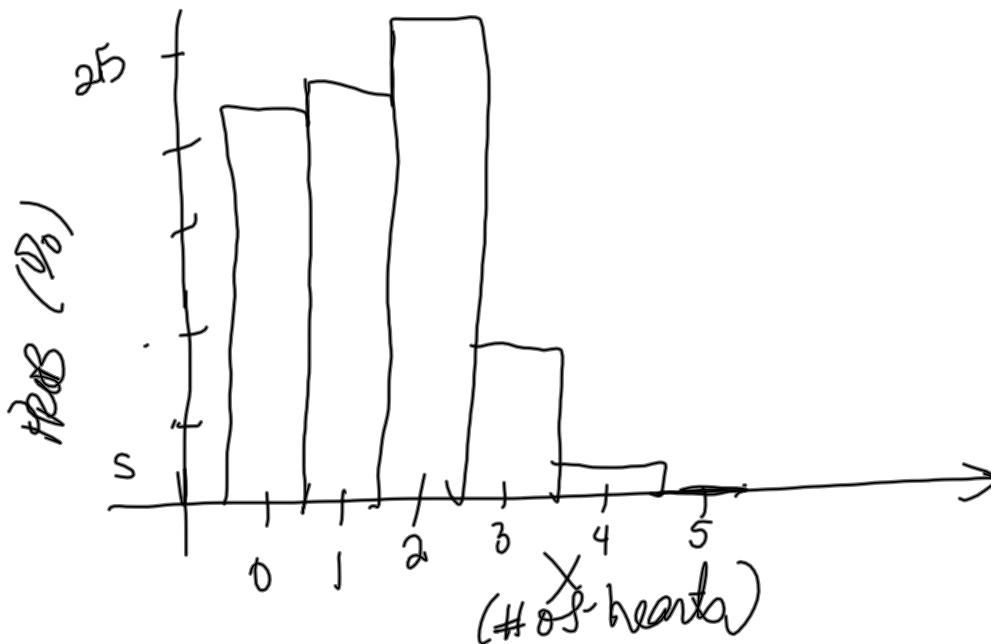
Finally, the random variable can be continuous:

time in hours studying for a test  $t \geq 0$   
 time, length, weight, temp.

Example

You are dealt a hand of 5 cards. Find the probability distribution table for the number of hearts. Graph this in a histogram

# of H X	P(X)
0	$\frac{C(13,0)C(39,5)}{C(52,5)} = \frac{575757}{2,598,960} \approx .222$
1	$\frac{C(13,1)C(39,4)}{C(52,5)} = \frac{602888}{n(5)} \approx .232$
2	$\approx .274$
3	$\approx .282$
4	$\approx .011$
5	$\approx .0005$



### 8.2 Expected Value

The expected value for the variable  $X$  in a probability distribution is

$$E(X) = X_1 \cdot P(X_1) + X_2 \cdot P(X_2) + \dots + X_n \cdot P(X_n)$$

Example

You are dealt a hand of 5 cards. What is the expected number of hearts?

$$E = 0 \left( \frac{2109}{9520} \right) + 1 \left( \frac{341}{1470} \right) + 2 \left( \quad \right) + 3 \left( \quad \right) + 4 \left( \quad \right) + 5 \left( \quad \right)$$

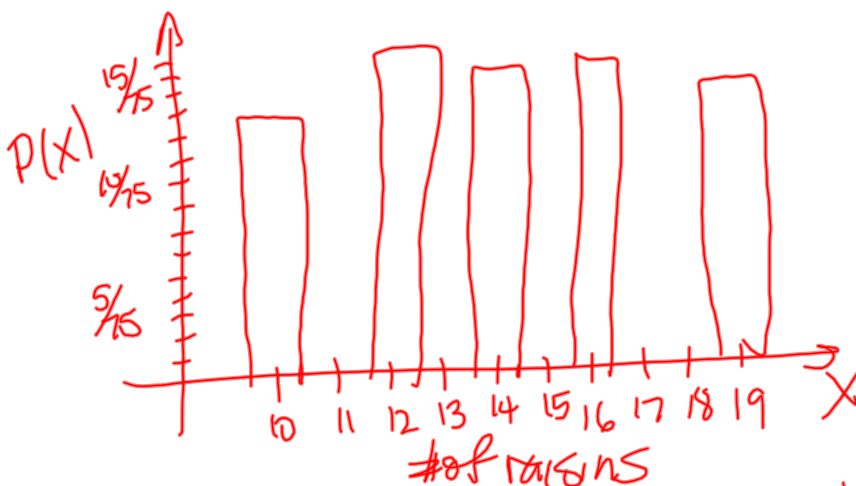
$$= \frac{615}{4681} \approx 1.305$$

Example

A sample of mini boxes of raisin bran cereal was selected and the number of raisins in each box was counted. The results are shown in the table below.

<i>freq</i>	# of boxes	13	14	15	16	17	<i>tot = 75</i>
<i>P(X)</i>	# of raisins	10	14	19	16	12	
		<i>13/75</i>	<i>14/75</i>	<i>15/75</i>	<i>16/75</i>	<i>17/75</i>	

Determine the appropriate random variable  $X$  and display the data in a probability histogram. What is the expected value of  $X$ ?



$$E = 10 \left( \frac{13}{75} \right) + 14 \left( \frac{14}{75} \right) + \dots = \frac{1071}{75} = 14.28$$

What is the MEAN of X?

$$\bar{X} = \frac{10+10+\dots+10}{13} + \frac{12+\dots+12}{17} + \dots = 14.28 = E(X)$$

$$= \frac{13 \times 10}{75} + \frac{17 \times 12}{75}$$

What was the X value that happened the most often?

10

What was the X value that was in the middle?

14 (Med on calculator)

~~What is the RANGE of X values?~~

### Histograms and Averages:

The MEAN (expected value) is where the histogram “balances”

The MODE is the tallest rectangle.

The MEDIAN is where the area is cut in half.

~~The RANGE is the number of rectangles. (remember, some may have a height of 0).~~

Example

Find the mean, median and mode of the following test scores:

77, 46, 98, 87, 84, 62, 71, 80, 66, 59,  
79, 89, 52, 94, 77, 72, 85, 90, 64, 70

L2	77	L4	3	EDIT	TESTS	1-Var Stats L3
.0625	46	-----		1:1-Var Stats		
.125	98			2:2-Var Stats		
.1875	87			3:Med-Med		
.25	84			4:LinReg(ax+b)		
.3125	62			5:QuadReg		
.375	71			6:CubicReg		
.4375				7:QuartReg		
.5						
L3=(77,46,98,87...						
1-Var Stats				1-Var Stats		
$\bar{x}=75.1$				$n=20$		
$\Sigma x=1502$				$\text{minX}=46$		
$\Sigma x^2=116532$				$Q1=65$		
$Sx=14.01465398$				$\text{Med}=77$		
$\sigma x=13.65979502$				$Q3=86$		
$n=20$				$\text{maxX}=98$		

Mean = 75.1

Median = 77

Mode = 77

Range =

Another way to measure spread? QUARTILES

$Q_1 = \frac{64+66}{2}$ 
 $M = \frac{77+77}{2}$ 
 $Q_3 = \frac{85+87}{2} = 86$   
 46 52 59 62 64 | 66 70 71 72 77 77 79 80 84 85 | 87 89 90 94 98

$Q1 = 65$      $Q3 = 86$      $IQR = Q3 - Q1 = 86 - 65 = 21$

Box and whisker plot



Example

A game consists of choosing two bills at random from a bag containing 7 one dollar bills and 3 ten dollar bills. The player gets to keep the money picked. How much should be charged to play  $P$  this game to keep it "fair" (expected value of zero)?

Event	X	P(X)
2 ones	2	$\frac{C(7,2)C(3,0)}{C(10,2)} = \frac{21}{45}$
a one and a ten	11	$\frac{C(7,1)C(3,1)}{C(10,2)} = \frac{21}{45}$
2 tens	20	$\frac{C(7,0)C(3,2)}{C(10,2)} = \frac{3}{45}$

$E = 2\left(\frac{21}{45}\right) + 11\left(\frac{21}{45}\right) + 20\left(\frac{3}{45}\right)$   
 $= 0$   
 $0 = \frac{333}{45} - P$   
 $P = \frac{333}{45}$   
 $\Rightarrow \$7.40$

Example

From a group of 2 women and 5 men a delegation of 2 is chosen. Find the expected number of women in the delegation.

Event	X	P(X)
2W 0M	2	$\frac{C(2,2)C(5,0)}{C(7,2)} = \frac{1}{21}$
1W 1M	1	$\frac{C(2,1)C(5,1)}{C(7,2)} = \frac{10}{21}$
0W 2M	0	$\frac{C(2,0)C(5,2)}{C(7,2)} = \frac{10}{21}$

$E = 2\left(\frac{1}{21}\right) + 1\left(\frac{10}{21}\right) + 0\left(\frac{10}{21}\right) = \frac{12}{21} = \frac{4}{7}$

## ODDS

If  $P(E)$  is the probability of event  $E$  occurring, then the odds in favor of  $E$  are

$$\frac{P(E)}{1-P(E)} = \frac{P(E)}{P(E^c)}, \quad P(E) \neq 1$$

We usually express the odds as a ratio of whole numbers,

$$\frac{a}{b} \rightarrow a \text{ to } b \rightarrow a:b$$

Example

If the probability that the Aggies will win a football game is 80%, what are the odds in favor of the Aggies?

$$\frac{P(E)}{P(E^c)} = \frac{.8}{1-.8} = \frac{.8}{.2} = \frac{4}{1} \Rightarrow 4:1$$

If we are given the odds we want to be able to find the probability, if the odds in favor of  $E$  are given as  $a:b$ , then

$$P(E) = \frac{a}{a+b}$$

Odds to win the Breeders' Cup Distaff  
Saturday, November 4th, 2006

Siempre 29~~5~~

Fleet Indian 2~~1~~

Round Pond 22~~1~~

Balletto 7~~5~~

Happy Ticket 21~~2~~

Spun Sugar 11~~1~~

Bushfire 18~~1~~

Summerly 30~~1~~

→  $\frac{7}{7+5} = \frac{7}{12}$

ran 12 times

8.3 Variance and Standard Deviationmean of 3 =  $\bar{x}$ 

3 3 3 3 3

2 2 3 4 4

0 0 5 5 5

 $n=5$ 

$X$	$X - \bar{x}$	$(X - \bar{x})^2$	
2	-1	1	
2	-1	1	
3	0	0	
4	1	1	
4	1	1	
TOT	0	4	

POPULATION VARIANCE,

$$\frac{\sum (x - \bar{x})^2}{n} = \frac{4}{5}$$

POPULATION STANDARD DEVIATION,

$$\sqrt{\frac{\sum (x - \bar{x})^2}{n}} = \sqrt{\frac{4}{5}} \approx .8944$$

0



What do we mean by population?

This means everyone, so if ALL the members of the population are used to find the mean, we use the symbols  $\mu$  and  $\sigma$ .

The mean from SAMPLE uses the symbol  $\bar{x}$ .

SAMPLE STANDARD DEVIATION,

$$S = \sqrt{\frac{(x - \bar{x})^2}{n-1}} = \sqrt{\frac{4}{5-1}} = \sqrt{\frac{4}{4}} = 1$$

When we have the probability, it is assumed we had the entire population to base it on, so it is appropriate to use  $\mu$  and  $\sigma$ .

Example

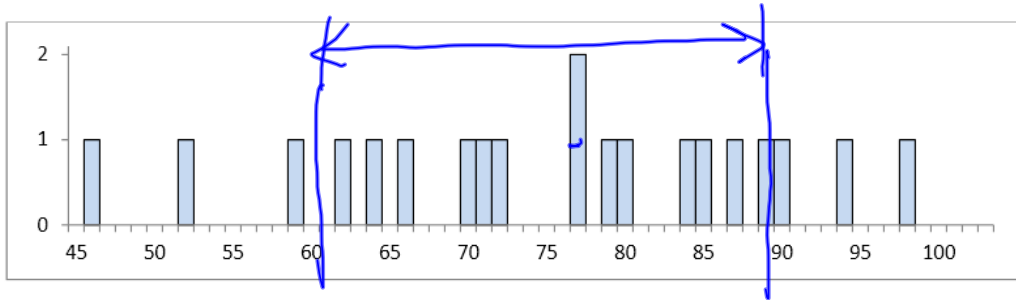
Find the mean and standard deviation for the following distribution:

1-var stats  $L_1, L_2$   
No S

$X$	$P(X)$
-2	.2
-1	.3
0	.1
1	.1
2	.3
$L_1$	$L_2$

Example

An exam has an mean of  $\mu=75$  and a population standard deviation of  $\sigma=14$ .



What is the probability that a randomly chosen data point is within 1 standard deviation of the mean?

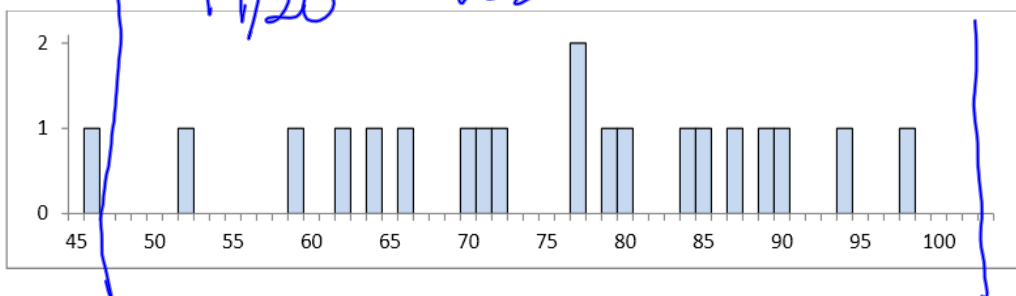
$$75 - 14 = 61, \quad 75 + 14 = 89$$

$$14/20 = 70\%$$

What is the probability that a randomly chosen data point is within 2 standard deviations of the mean?

$$75 - 2 \times 14 = 47 \quad \text{to} \quad 75 + 2 \times 14 = 103$$

$$19/20 = 95\%$$



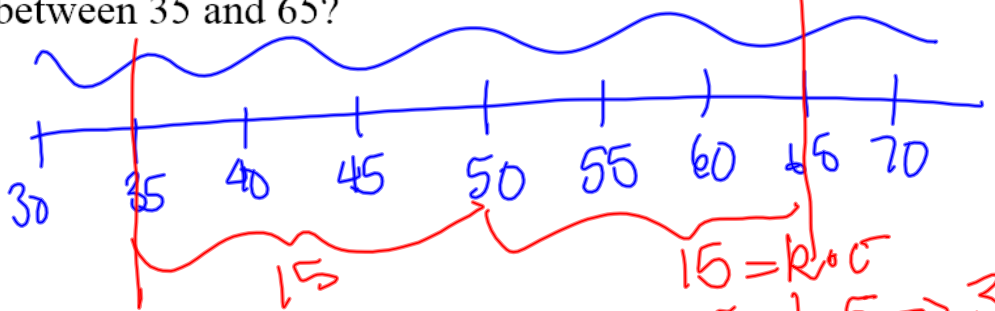
Chebychev's Theorem: For any data distribution with mean  $\mu$  and standard deviation  $\sigma$ , the probability that a randomly chosen data point is within  $\mu - k \cdot \sigma$  to  $\mu + k \cdot \sigma$  is at least  $1 - 1/k^2$ . Or,

$$P(\mu - k \cdot \sigma \leq x \leq \mu + k \cdot \sigma) \geq 1 - \frac{1}{k^2}$$

Example

A probability distribution has a mean of 50 and a standard deviation of 5.

a) What is the probability that an outcome of the experiment lies between 35 and 65?



$$P(35 \leq X \leq 65) \geq 1 - \frac{1}{3^2} = \frac{8}{9}$$

b) Find the value of  $k$  so that at least 93.75% of the data lies in the range  $50 - 5k$  to  $50 + 5k$ .

$$.9375 = \frac{15}{16} = 1 - \frac{1}{k^2} \Rightarrow k = 4$$

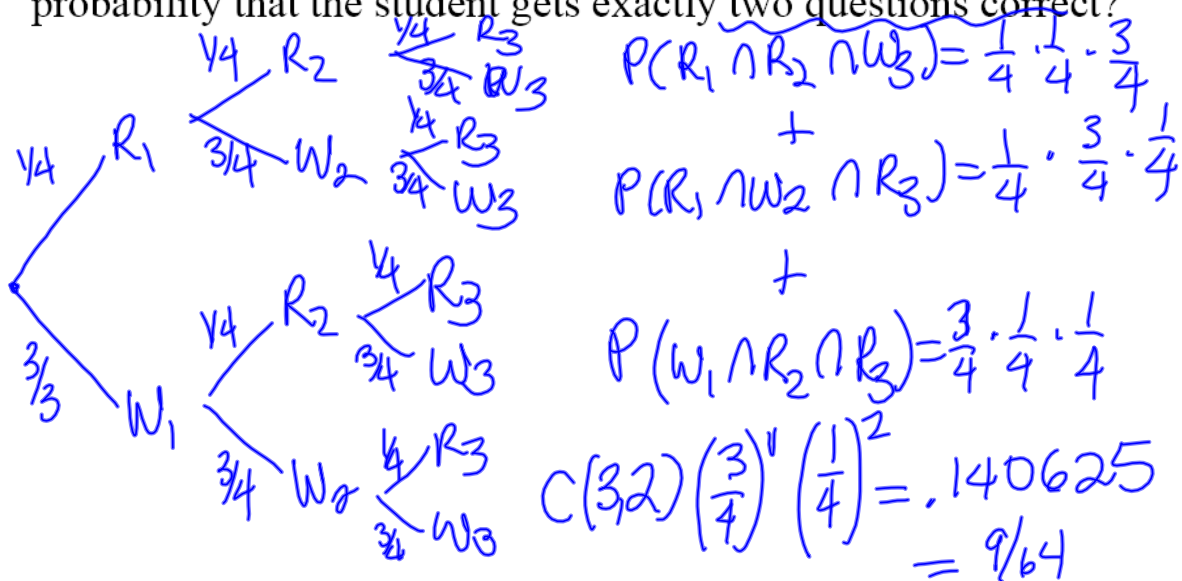
## 8.4 The Binomial Distribution

In a Bernoulli trial we have the following:

- The same experiment repeated several times.
- The only possible outcomes of these experiments are success or failure.
- The repeated trials are independent so the probability of success remains the same for each trial.

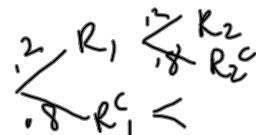
Example

A multiple choice test has 3 questions, each with 4 possible answers. A student guesses on each question. What is the probability that the student gets exactly two questions correct?

**BINOMIAL PROBABILITY:**

If  $p$  is the probability of success in a single trial of a binomial (Bernoulli) experiment, the probability of  $x$  successes and  $n-x$  failures in  $n$  independent repeated trials of the same experiment is

$$C(n, x) (p)^x (1-p)^{n-x} = \text{binompdf}(n, p, x)$$



Example

The first half of July was very dry in college station. If each day there was a 20% chance of rain, what is the probability of no rain in the first 15 days in July?

DEFINE SUCCESS: *Rain*

$n = \text{number of trials} = 15$

$x = \text{number of successes} = 0$

$p = \text{probability of success} = .2$

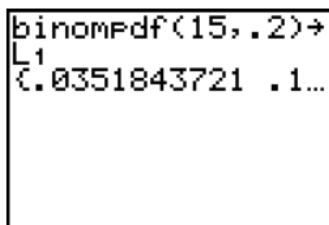
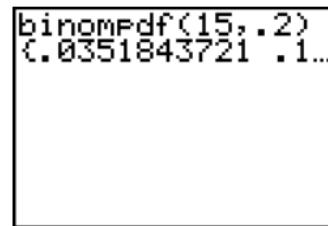
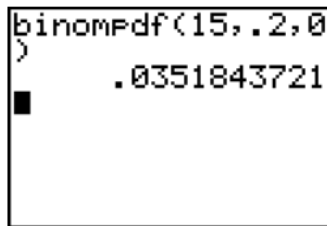
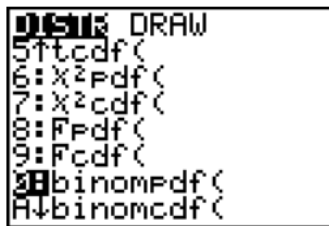
*No Rain*

$n = 15$

$X = 0$

$p = .8$

binompdf( $n, p, x$ ) on the calculator



L1	L2	L3	1
.0351843721	-----	-----	
.13194			
.2309			
.25014			
.1876			
.10318			
.04299			
L1() = .0351843720...			

What is the probability of at most 2 rain days?

$x = \text{number of successes} = 0, 1, 2$

```
L1(1)+L1(2)+L1(3)
)
.3980232093
■
```

`binomcdf(n, p, x)` will give you the sum of the probabilities from 0 to  $x$ .

```
binomcdf(15, .2, 2)
)
.3980232116
■
```

For a binomial probability question you must do the following:

- Decide that it is a Bernoulli trial.
- Define what success is.
- Find the number of times the experiment is done,  $n$ .
- Find the probability of success,  $p$ .
- Determine the number of successes you need to find,  $x$ .

Example

A new drug being tested causes a serious side effect in 5% of patients. What is the probability that in a sample of 10 patients none get the side effect from taking the drug?

define success = no side effect,

n = number of trials = 10

x = number of successes = 10

p = probability of success = .95

$$\text{binompdf}(10, .95, 10) = .5987$$

$$\text{binompdf}(10, 0, .05) = .5987$$

define success = side effect,

n = number of trials = 10

x = number of successes = 0

p = probability of success = .05

If  $X$  is a binomial random variable associated with a binomial experiment consisting of  $n$  trials with probability of success  $p$  and probability of failure  $q=1-p$ , then the mean (expected value) and standard deviation associated with the experiment are:

$$\mu=np \quad \sigma=\sqrt{npq}$$

Success=girl,  $n=6$ ,  $X$ =# of girls,  $p=.5$

Example - let the random variable  $X$  be the number of girls in a 6 child family. Find the probability distribution table, probability histogram and the mean and standard deviation for the number of girls in the family.

# of girls, $X$	$P(X)$
0	$\text{binompdf}(6, .5, 0)$
1	
2	
3	
4	
5	
6	

$$\begin{aligned} & \text{binompdf}(6, .5) \\ & \rightarrow L2 \\ & \mu = 6 * .5 = 3 \\ & \sigma = \sqrt{6 * .5 * (1-.5)} \\ & = \sqrt{1.5} \\ & (\approx 1.2247) \end{aligned}$$



## 8.5 The Normal Distribution

Continuous random variables can take on any value.

Let  $t$  = time in seconds to run a race

Let  $w$  = weight of kitten in kg

Let  $L$  = length of a week-old bean plant

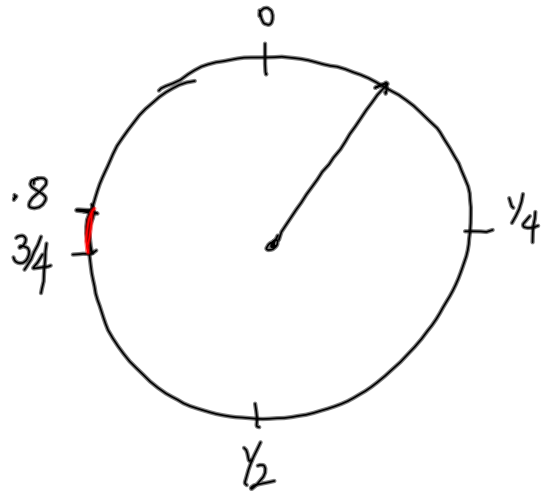
Let  $X$  = value where pointer lands.

$0 \leq X < 1$  and  $P(0 \leq X < 1) = 1$

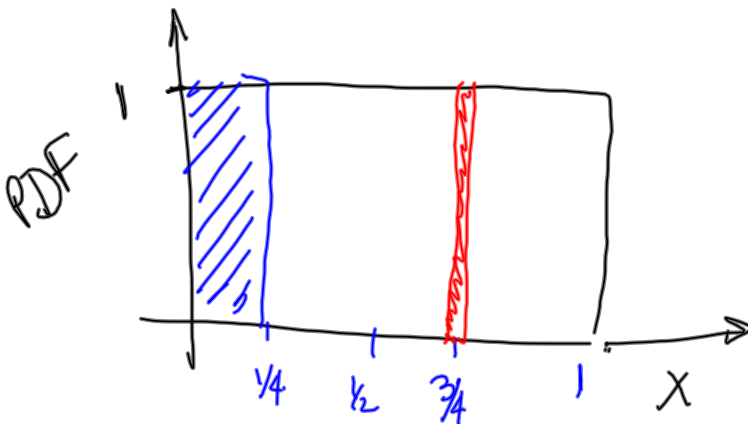
What is  $P(X = \frac{1}{2})$ ?  $= 0$

What is  $P(0 \leq X < \frac{1}{4})$ ?  $= \frac{1}{4}$

What is  $P(0.75 \leq X \leq 0.80)$ ?  $= .05$



Define a PROBABILITY DENSITY FUNCTION



Discrete finite variables - graph the probability as a histogram.

Each rectangles have a base of width 1 (centered on  $X$ ) and the height was  $P(X)$ .

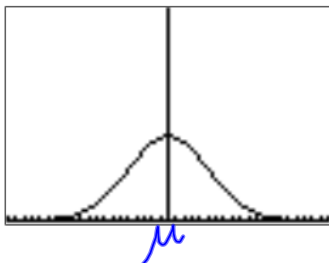
So the area, length  ~~$X$~~  height was the probability that  $X$  occurred.

AREA above our  $X$  value will be the probability that get that  $X$  value.

If we want to find the probability of a range of  $X$  values, we would add up the areas over the range of  $X$  values.

When we graph the probability distribution for a continuous variable we find a smooth curve.

Many natural and social phenomena produce a continuous distribution with a bell-shaped curve.

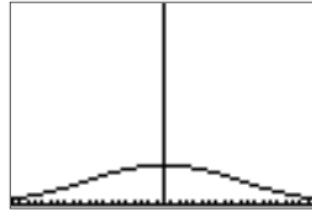
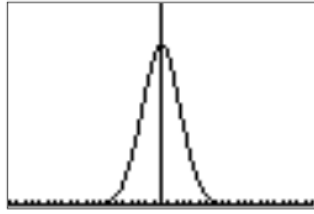


Every bell-shaped (NORMAL) curve has the following properties:

- Its peak occurs directly above the mean,  $\mu$
- The curve is symmetric about a vertical line through  $\mu$  The curve never touches the x-axis. It extends indefinitely in both directions.
- The area between the curve and the x-axis is always 1 (total probability is 1).

The shape of the curve is completely determined by  $\mu$  and  $\sigma$ ,

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



The probability that a data value will fall between  $x=a$  and  $x=b$  is given by the area under the curve between  $x=a$  and  $x=b$ .

The standard normal curve has  $\mu=0$  and  $\sigma=1$ . Use Z, NOT X.

Example *normalcdf(a,b) = area under stand norm curve from a to b*

On a standard normal curve, what is the probability that a data value is between -1 and 1,  $P(-1 < z < 1)$ ?

*normalcdf(-1,1) = 0.6827*

Example

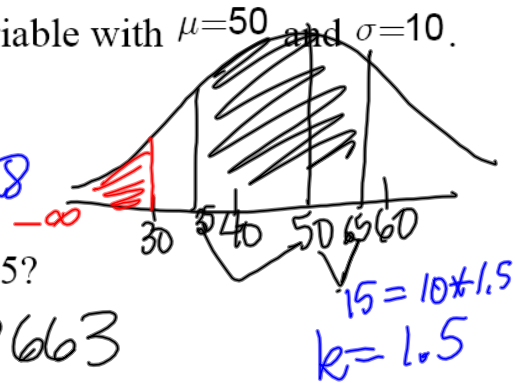
Suppose that X is a normal random variable with  $\mu=50$  and  $\sigma=10$ .

a) What is the probability that  $X < 30$ ?

*normalcdf(-1E99, 30, 50, 10) = 0.0228*

b) What is the probability that  $35 < X < 65$ ?

*normalcdf(35, 65, 50, 10) = .8663*



Remember Chebychev's theorem?

*$P(35 \leq X \leq 65) \geq 1 - \frac{1}{1.5^2} = \frac{5}{9} (\approx .56)$*

*Quiz Tuesday 8.1 - 8.3*

Example: An instructor wants to “curve” the grades in his class. The class mean at the end of the semester is 73 with a standard deviation of 12. He decides that the top 12% of the class should get an A, the next 24% should get a B, the next 36% a C, the next 18% a D and the last 10% of the class will get an F. What are the cutoffs for the grades?

A cutoff:

B cutoff:

C cutoff:

D cutoff:

## 8.6 Applications of the Normal Distribution

### Example

A machine that fills quart milk cartons is set to average 32.2 oz with a standard deviation of 1.2 oz. What is the probability that a filled carton will have more than 32 oz?

If the store receives 500 quart milk cartons, how many will have more than 32 ounces?

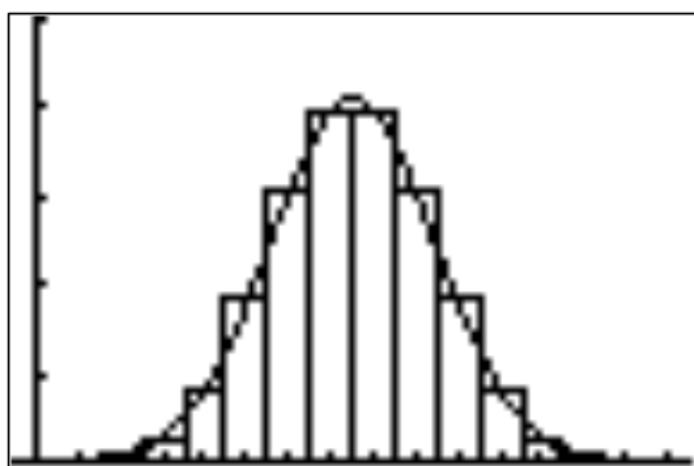
The mean height for 18 year old girls is 64.5 inches (50th percentile), with a standard deviation of 1.875 inches. These heights closely approximate the normal distribution. (this data is old)

- a) What is the probability that a woman is shorter than 5' 3"?
- b) In a group of 200 women, how many would you expect to be between 65" and 68"?
- c) What is the probability that a woman is taller than 6'?
- d) What height corresponds to the 90th percentile (that is, taller than 90% of the women)?

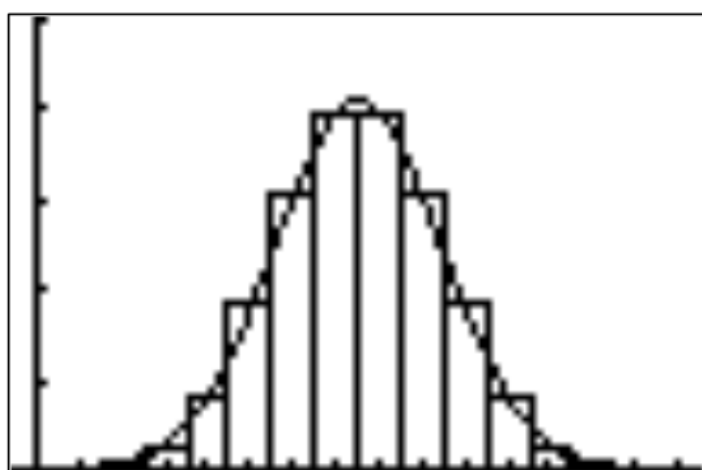
Example

Consider tossing a coin 15 times and let  $X$ =number of heads.

(a) What is the probability that you toss exactly 5 heads?



(b) What is the probability of more than 9 heads?



## THE NORMAL CURVE APPROXIMATION TO THE BINOMIAL DISTRIBUTION

At a school 1000 children are exposed to the flu. There is a 35% chance of getting the flu if you are exposed. Use the normal curve approximation to the binomial distribution to estimate the probability that

(a) more than 360 children get the flu.

(b) fewer than 320 children get the flu.

(c) between 325 and 375 children get the flu.



## The Poisson Distribution

The Poisson distribution arises when you count a number of events across time or over an area. You should think about the Poisson distribution for any situation that involves counting events such as

How many cars arrive at a toll booth in an hour?

How many items are used from an inventory in a week?

How many red blood cells are in a cc of blood?

How many fish are caught in a lake per day?

These are all events that we can try to find the probability of occurring, but not the probability that they don't occur. To study these we will use the Poisson distribution.

### When the Poisson Probability Applies

The Poisson distribution is based on four assumptions. The term "interval" refers to either a time interval or an area.

1. The probability of observing a single event over a small interval is approximately proportional to the size of that interval.
2. The probability of two events occurring in the same narrow interval is negligible.
3. The probability of an event within a certain interval does not change over different intervals.
4. The probability of an event in one interval is independent of the probability of an event in any other non-overlapping interval.

When the Poisson probability law will apply, there is a number  $\lambda$  which is the average rate that the events occur.

If the Poisson distribution applies, then the probability of  $x$  occurrences per unit measure is approximately

$$P_{\lambda}(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{with a mean of } \mu = \lambda \quad \text{and } \sigma = \sqrt{\lambda}$$

Example: Suppose on average there are 16 emergency patients on the 8A to 4P shift of a certain hospital. What is the probability that during any one hour of the shift that

- a) zero patients arrive?
- b) at most one patient?
- c) more than one patient?
- d) draw the histogram for  $x = 0$  to 6

Example: On average 90 hamburgers are sold during the lunch hour at a fast food restaurant. What is the probability that during a certain minute of the lunch hour that

- a) zero hamburgers will be sold?
- b) at most two hamburgers will be sold?
- c) more than 3 hamburgers will be sold?

The binomial distribution can be approximated by the Poisson distribution if  $p$ , the probability of success in a single trial, is small [generally less than 0.1]. In that case, using  $\lambda = np$  will be a good approximation.

Example: The probability of theft on a subway is given as 0.001. What is the probability that 10 of the next 5000 passengers will be robbed? Use both the binomial and Poisson distributions.