Chapter 3 Probability Distributions and Statistics

Section 3.1 Random Variables and Histograms

The value of the result of the probability experiment is called 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.

$X = \{1, 2, \ldots\}$

infinite discrete r.v.

outcome	Y
6	1
2, 6	2
5, 1, 6	3
⋮	

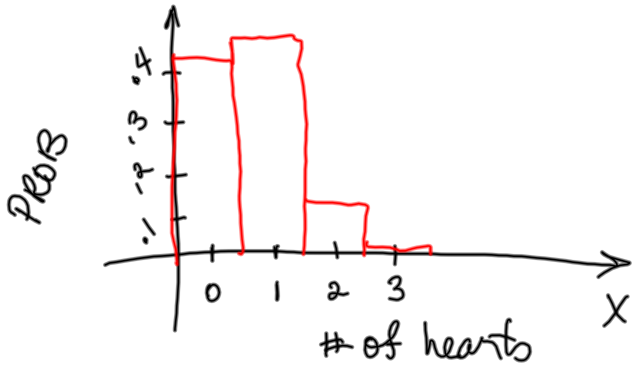
length
weight
time
Temp

Finally, the random variable can be continuous:
Let x be the number of hours per day a light is on
$0 \le x \le 24$

Example

You are dealt a hand of 3 cards. Find the probability distribution table for the number of hearts. Graph this in a histogram

EVENT	X	PROB
$0H\,3H^c$	0	$\frac{C(13,0)\,C(39,3)}{C(52,3)} = \frac{9139}{22,100} \approx (.414)$
$1H\,2H^c$	1	$\frac{C(13,1)\,C(39,2)}{C(52,3)} = \frac{9633}{22100}$ $(\approx .436)$
$2H\,1H^c$	2	$\approx .138$
$3H\,0H^c$	3	$\approx .013$



3.2 Expected Value

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

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

Example

You are dealt a hand of 3 cards. What is the expected number of hearts?

$$E = 0\left(\frac{9139}{22100}\right) + 1\left(\frac{9633}{22100}\right) + 2\left(\frac{3042}{22100}\right) + 3\left(\frac{286}{22100}\right) = \frac{3}{4}$$

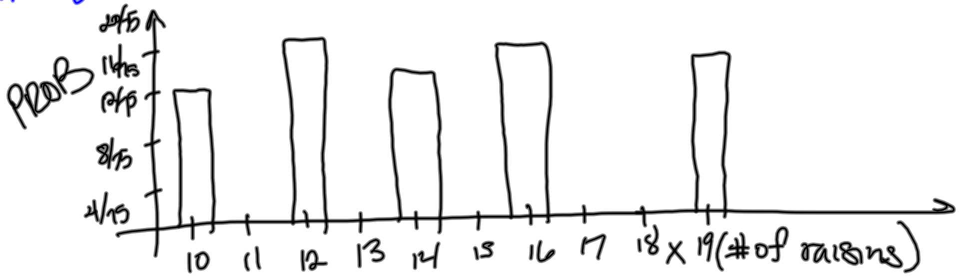
Example

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.

FREQ	# of boxes	13	14	15	16	17
X	# of raisins	10	14	19	16	12
PROB		13/75	14/75	15/75	16/75	17/75

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

# of observations = 13 + 14 + 15 + 16 + 17 = 75



$$E = 10\left(\tfrac{13}{75}\right) + 14\left(\tfrac{14}{75}\right) + 19\left(\tfrac{15}{75}\right) + 16\left(\tfrac{16}{75}\right) + 12\left(\tfrac{17}{75}\right) = 14.28$$

+

What is the MEAN of X?

$$\bar{X} = \frac{10\times 13 + 14\times 14 + 19\times 15 + 16\times 16 + 12\times 17}{75} = 14.28 = \frac{1071}{75}$$

$$= 10\left(\frac{13}{75}\right) + 14\left(\frac{14}{75}\right) + 19\left(\frac{15}{75}\right) + 16\left(\frac{16}{75}\right) + 12\left(\frac{17}{75}\right) = E(X)$$

What X value occurred the most often?

12 was observed the most ofen

MODE

A data set may have no mode, one mode or more than one mode

What X value is in the middle?

$$\underbrace{10, 10, \ldots 10, 12, 12, \ldots 12 \ldots\ldots 19}_{75 \text{ values}}$$

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.

1, 1, | 2, 2

$$\text{mean} = \frac{1+1+2+2}{4} = \frac{6}{4} = 1.5$$

$$\text{Med} = \frac{1+2}{2} = 1.5$$

No mode

1, 1, | 2, 100

$$\text{mean} = \frac{1+1+2+100}{4} = \frac{104}{4} = 26$$

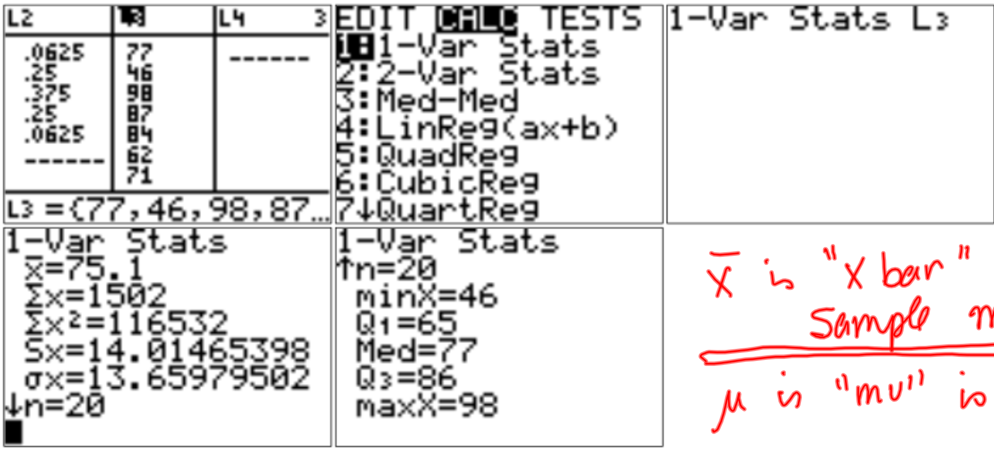
$$\text{med} = 1.5$$

$$\text{mode} = 1$$

Example

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

$\boxed{\begin{array}{l}77,\ 46,\ 98,\ 87,\ 84,\ 62,\ 71,\ 80,\ 66,\ 59,\\ 79,\ 89,\ 52,\ 94,\ 77,\ 72,\ 85,\ 90,\ 64,\ 70\end{array}}$ Put in LIST



$\bar{x}$ is "x bar" is the sample mean

μ is "mu" is the pop mean

Mean = 75.1 Median = 77

Mode = 77 Range = 98 − 46 = 52

Another way to measure spread? QUARTILES

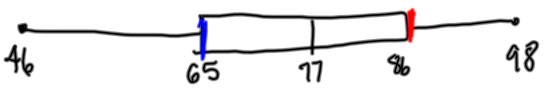
46 52 59 62 64 | 66 70 71 72 77 | 77 79 80 84 85 | 87 89 90 94 98

L M U

← 1/4 → Q1 ← 1/4 → ← 1/4 → Q3 ← 1/4 →

Q1 = 65 Q3 = 86 IQR = $Q_3 - Q_1 = 86 - 65 = 21$

Box and whisker plot



outlier? low one is a value less than $Q_1 - 1.5 \times IQR$

$65 - 1.5 \times 21 = 33.5$

high one is a value greater than $Q_3 + 1.5 \times IQR$

$86 + 1.5 \times 21 = 117.5$

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 this game to keep it "fair" (expected value of zero)?

charge p to play

EVENT	X	PROB
2 ones	2	$C(7,2)C(3,0)/C(10,2) = 21/45$
1 one 1 ten	11	$\frac{C(7,1)C(3,1)}{C(10,2)} = \frac{21}{45}$
2 tens	20	$\frac{C(7,0)C(3,2)}{C(10,2)} = \frac{3}{45}$

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

$\Rightarrow p = \frac{333}{45}$ so $\boxed{\$7.40 \text{ to play}}$

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

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

0, 1, 1, 0, 2, 0, etc 50 times

3.3 Measures of Spread

on "average" how far away from the mean is our data?

$\underbrace{3\ 3\ 3\ 3\ 3}_{\mu=3}$ $\underbrace{2\ 2\ 3\ 4\ 4}_{\mu=3}$ $\underbrace{0\ 0\ 5\ 5\ 5}_{\mu=3}$

	X	$X-\bar{X}$	$(X-\bar{X})^2$
	2	$2-3=-1$	1
	2	$2-3=-1$	1
	3	$3-3=0$	0
	4	$4-3=1$	1
	4	$4-3=1$	1
mean	3	$0/5$	$4/5=\sigma$

POPULATION VARIANCE, is σ^2

$$\sigma^2=\frac{\Sigma(x-\mu)^2}{n}=\frac{4}{5}$$

POPULATION STANDARD DEVIATION,

$$\sigma=\sqrt{\sigma^2}=\sqrt{\frac{\Sigma(x-\mu)^2}{n}}=\sqrt{\frac{4}{5}}\quad(\approx .89)$$

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 μ and σ.

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

SAMPLE STANDARD DEVIATION,

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

When we have the probability, it is assumed we had the entire population to base it on, so it is appropriate to use μ and σ.

Example
Find the mean and standard deviation for the following distribution:

L_1	L_2
X	P(x)
−4	.1
−2	.2
0	.3
2	.1
4	.3

$\mu = .6$

$\sigma = 2.6907$

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:

$\mu = np$ $\sigma = \sqrt{npq} = \sqrt{N * p * (1-p)}$

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

L1

X	P(X)
0	.015625
1	.09375
2	.234375
3	.3125
4	.234375
5	.09375
6	.015625

TRY THIS

binompdf(6, .5) → L2
1-Var Stats

$\sigma = \sqrt{6 \times .5 \times (1-.5)}$
$= \sqrt{1.5} \approx 1.2247....$

Example
A flash drive has a 0.8% probability of being defective. In a shipment of 1000 flash drives, what is the mean and standard deviation in the number of defective flash drives?

$\mu = 1000 * .008 = 8$
$\sigma = \sqrt{1000 * (.008)(1-.008)} = \sqrt{7.936}$
≈ 2.8

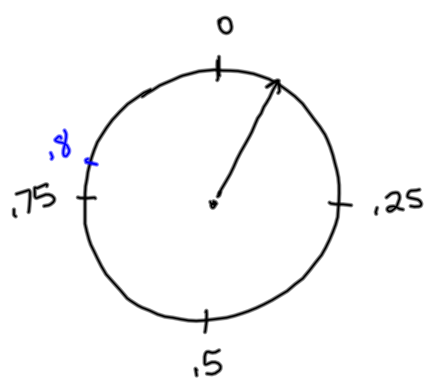
3.4 The Normal Distribution

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.

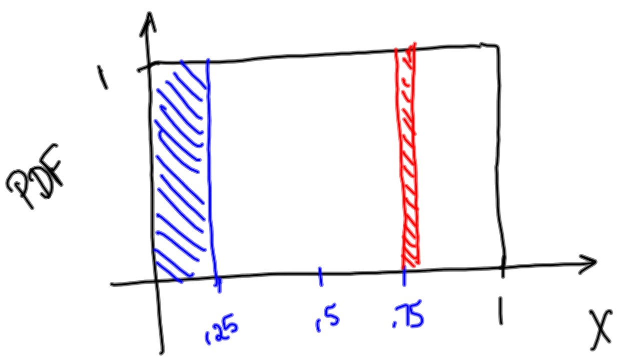
$0 \le X < 1$ and $P(0 \le X < 1) = 1$

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

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

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

Define a PROBABILITY DENSITY FUNCTION



$P(0 \le X \le .25) = .25$

$P(.75 \le X \le .80) = .05$

Discrete finite variables - graph the probability as a histogram.

Each rectangle has a base of width 1 (centered on X) and the height was P(X).

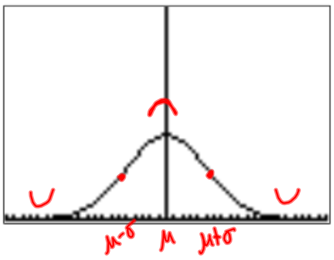
So the area, length x height was the probability that X occurred.

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.

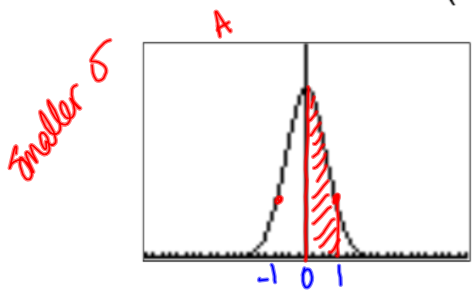


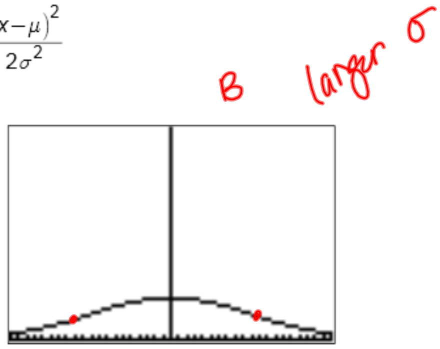
Every bell-shaped (NORMAL) curve has the following properties:

- Its peak occurs directly above the mean, μ
- The curve is symmetric about a vertical line through μThe 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 curve switches from concave down to concave up one standard deviation away from the mean.

The shape of the curve is completely determined by μ and σ,

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



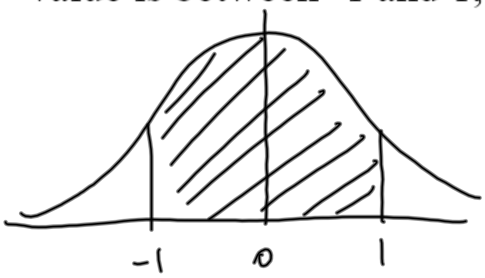


The probability that a data value will fall between x=a and x=b is given by the area under the curve between x=a and x=b.

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

Example

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

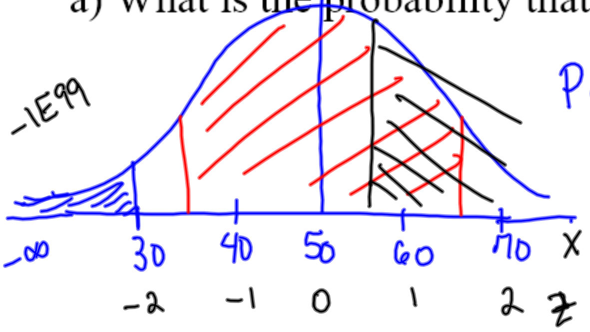


normalcdf(-1,1) = .6827

Example normalcdf(Left, right, μ, σ)

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

a) What is the probability that $X \leq 30$?



$P(X \leq 30) = \text{normalcdf}(-1E99, 30, 50, 10)$

≈ 0.0228

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

$P(35 \leq X \leq 65) = \text{normalcdf}(35, 65, 50, 10) = 0.8664$

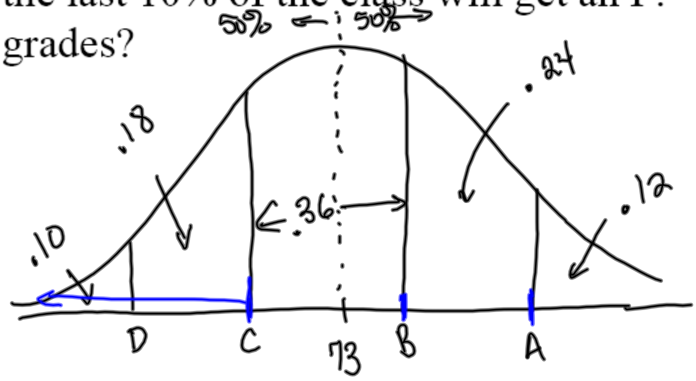
c) What is the probability that $X > 55$?

$P(X \geq 55) = \text{normalcdf}(55, 1E99, 50, 10)$

$= .3085$

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?



$\text{invNorm}(\text{area to the left}, \mu, \sigma)$

$D = \text{invNorm}(.1, 73, 12) = 57.6\ldots$

$C = \text{invNorm}(.28, 73, 12) = 66.0\ldots$

$B = \text{invNorm}(.64, 73, 12) =$

$A = \text{invNorm}(.88, 73, 12) =$

A cutoff:

B cutoff:

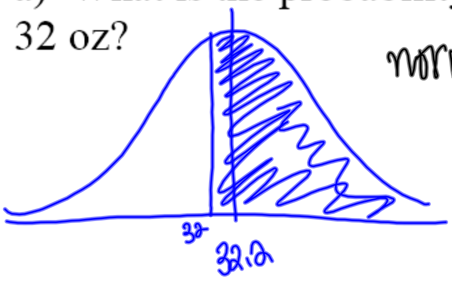
C cutoff:

D cutoff:

Example

A machine that fills quart milk cartons is set to average 32.2 oz with a standard deviation of 1.2 oz.

a) What is the probability that a filled carton will have more than 32 oz?



normalcdf(32, 1E99, 32.2, 1.2)

= .5662

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

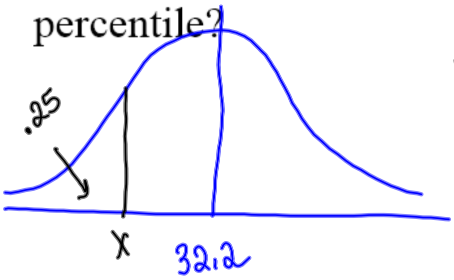
b/c it is continuous

$N = 500 * P(X>32) = 500 * P(X \geq 32)$

$= 500 * \text{normalcdf}(32, 1E99, 32.2, 1.2) = 283.09$

$\boxed{\text{283 cartons}}$

c) What is the volume of milk that corresponds to the 25$^{\text{th}}$ percentile?



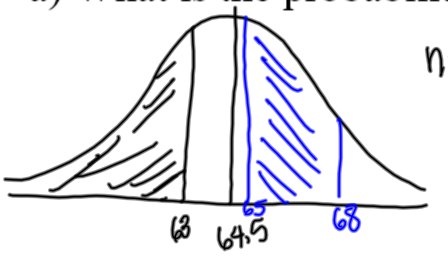
invNorm(.25, 32.2, 1.2)

= 31.3906 oz

50th percentile

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 $\overbrace{5'\,3''}^{63''}$?



normalcdf(-1E99, 63, 64.5, 1.875)
= .2119

b) In a group of 200 women, how many would you expect to be between 65" and 68"?

200 * normalcdf(65, 68, 64.5, 1.875) = 72.77
(72 or) 73 women

c) What is the probability that a woman is taller than 6'?

normalcdf(72, 1E99, 64.5, 1.875) = 3.168×10^{-5} (≈ 3 in 100,000)

$$\frac{72-64.5}{1.875} = 4$$

d) What height corresponds to the 90th percentile?

invNorm(.9, 64.5, 1.875)
= 66.9"