

Probability

- Probability is the likelihood of a particular outcome occurring.

$$\text{probability} = \frac{\# \text{ ways a given event can occur}}{\# \text{ total possible outcomes}}$$

- Example: probability of drawing a club from a deck of cards

$$p = \frac{13 \text{ clubs}}{52 \text{ total cards}} = \frac{1}{4} = .25$$

- Example: in an urn full of 10 red marbles and 12 black marbles, the probability of drawing a black marble is

$$p = \frac{12}{10+12} = \frac{12}{22}$$

Law of Large Numbers

- As we repeat an experiment a large number of times, the ratio of the number of successes in the sample to the total number of trials will approach the probability of the event.
- Example: proportion of heads should go to .5

10 coins:	7 H	3 T	$p = .7$
100 coins:	53 H	47 T	$p = .53$
1000 coins:	498 H	502 T	$p = .498$
10000 coins:	5002 H	4998 T	$p = .5002$

Types of Probabilities

- Classical/Relative Frequency

$$\text{probability} = \frac{\# \text{ ways a given event can occur}}{\# \text{ possible outcomes}}$$

- Subjective

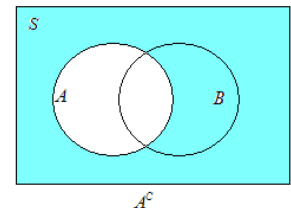
- Likelihood of an event is based on your own personal judgment

Probability Definitions

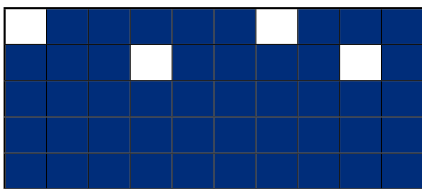
- A complement
 - All possible events that are **not** in A

- Example:

- A = it's snowing
- A^c = it's not snowing



Example of a Complement



- We have a total of 50 seats in a room, and a shaded square above represents a taken seat.
- How many people are in this room?
- It's easier to instead count the number of empty seats, of which there are 4. Therefore the number of people is $50 - 4 = 46$
- So the complement here is "all empty seats" which gives us the needed answer more easily.

Probability (HW 5.1-5.3)

- The game Scrabble contains 100 lettered tiles: 56 consonants, 42 vowels, and 2 blanks. The 42 vowels can be further classified as follows: 9 A's, 12 E's, 9 I's, 8 O's, and 4 U's. Suppose we reach into a bag containing all the tiles and select one letter.
- What's the probability of drawing a tile with a letter on it?
- If we draw a vowel, what's the probability it's in the upper half of the alphabet? (Meaning A - M)
- If the tile drawn contains a letter on it, what's the probability that letter is not a vowel?

Probability (HW 5.1-5.3)

We have an urn full of marbles of the following colors: 5 green, 4 blue, 3 yellow, 2 red, 1 white (15 total).

What's the probability of drawing a primary color? (Red, yellow, or blue)

If the marble drawn was a primary color, what's the probability it was either red or yellow?

If the marble drawn was not a primary color, what's the probability it was not green?



Probability (HW 5.1-5.3)

	Chocolate	Strawberry	Total
Male	490	871	1361
Female	513	302	815
Total	1003	1173	2176

- The ice cream HW question again!
- Find the probability that an individual selected is female and prefers strawberry ice cream.
- Find the probability that an individual selected is female or prefers strawberry ice cream.

Probability (HW 5.1-5.3)

	Chocolate	Strawberry	Total
Male	490	871	1361
Female	513	302	815
Total	1003	1173	2176

- Given that the person selected is male, find the probability they prefer strawberry.
- Given that the person prefers strawberry, find the probability they are male.

Random Variables

- Discrete
 - A countable, whole-number of possible values.
 - The number of words in a magazine article
 - The number of clubs in a drawing of 10 cards
- Continuous
 - An uncountable, infinite number of possible values, with decimals allowed.
 - The weight of an athlete
 - The time taken to complete a race

Discrete Probability Distribution

Two requirements:

1. Each individual $p(x)$ is between 0 and 1, inclusive
2. All probabilities sum to 1

The mean of a discrete distribution:

$$\text{MEAN} = \sum x \cdot p(x)$$

Also called average, or expected value

Mean of a Distribution (HW 6.1-6.2)

- Here's a table for the probability of different category hurricanes.

Category	Probability
1	0.15
2	0.32
3	
4	0.12
5	0.05

- Find the missing value and the mean/expected value of this data set.

Discrete Mean (HW 6.1-6.3)

- The Cash 4 lottery involves picking 4 numbers, each 0 to 9, and hence 10,000 combinations. If you pick the correct combination, you win \$1200.
- On a single bet, what's the probability you win \$1200?
- Make a probability distribution for this problem. X is how much you win; find the probabilities with each outcome.

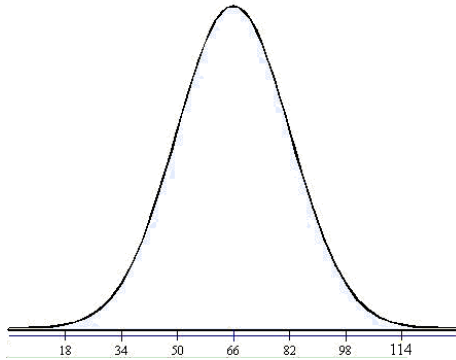
x	$p(x)$
\$0	
\$1200	
- Is this discrete or continuous?

Discrete Mean (HW 6.1-6.3)

- Find the mean (expected winnings) on this lottery.
- On one play, how much should we expect to win?

Normal Curve (Empirical Rule)

What's the mean and s.d.?

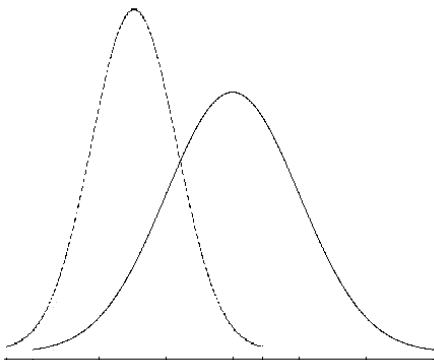


Empirical Rule (HW 6.1-6.3)

- Home rental prices are normally distributed with mean \$700 and s.d. \$150.
- Estimate the percentage of rental rates between \$400 and \$1000.
- Estimate the percentage of rental rates between \$550 and \$850.

Comparing Normal Curves

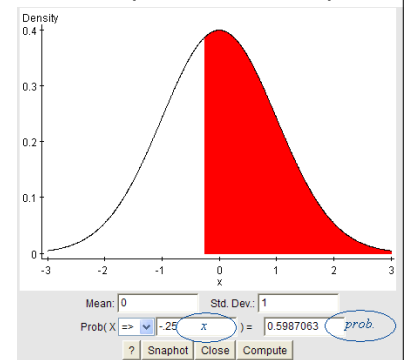
Which has a higher mean?



Normal Distribution (Continuous)

Three rules:

- Total probability/ area under the normal curve is 1
- Normal curve is symmetric
- X value / z-score goes in left box; probability goes in right box on StatCrunch



Normal Distributions

Two types of problems

1. Given an **x value**, find the **probability** above or below it
2. Given a **probability**, find the **x value** that gives that probability above or below it

– **DRAW A SKETCH!!!**

Normal Probability

- Draw a sketch of a normal curve if we were looking for the probability to the left of $z = 1.5$.

Normal Probability

- Draw a sketch to help find what z-value has probability 0.4 to the right.

Area Between Two Lines

- Find the probability within 1.2 standard deviations from the mean. Make a sketch of this.
- Try to think of a strategy for this.

Normal (HW 6.1-6.3)

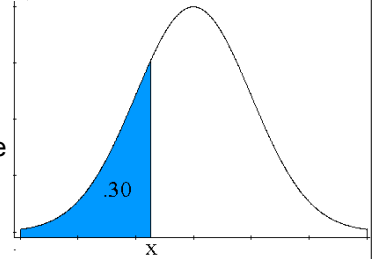
- The weight of a platypus is normally distributed with mean 4.4 and sd 0.4 pounds. Suppose the probability that a platypus' weight falls above 5 pounds is .0668. Find the probability that a randomly selected weight is between 4.4 and 5. Make a sketch as well!

Normal (HW 6.1-6.3)

- Same problem: $\mu = 4.4$, and $\sigma = 0.4$.
- What is the z-score for a platypus with a weight of 5.5 pounds?
- What is the z-score for a platypus whose weight is 2.61 standard deviations to the left of the mean?
- What weight corresponds to the above z-score?

Percentiles

- The Pth percentile is the x that gives P % below on the normal
- Always below
- Example: here the x is the 30th percentile because 30% of the data falls below x



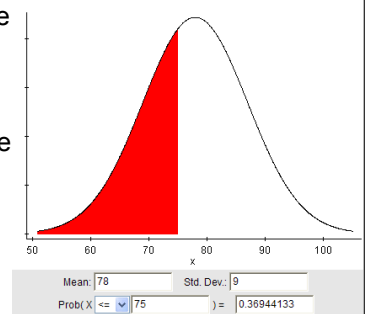
Percentiles (HW 6.1-6.3)

- The average high temperature in New York has mean 78 and s.d. 9.
- Make a sketch to help in determining the 75th percentile.

Percentiles (HW 6.1-6.3)

- What is the percentile for a day with a high temperature of 75 degrees?
- Must be a percentage and a whole number

.369 \Rightarrow 36.9%
 \Rightarrow 37%



3 Types of Distributions

1. Population
 - Distribution of all points in the population
2. Sample Data / Data
 - Distribution of one particular sample
3. Sampling
 - Distribution of the sample means of a given size n

Means Problem

	MEAN	ST. DEV.
POPULATION	μ	σ
SAMPLE DATA (DATA)	\bar{x}	s
SAMPLING DIST.	μ	$\frac{\sigma}{\sqrt{n}}$

Distribution Shapes

- Sample Data (Data Distribution) is the same shape as the population
 - If population is skewed, so is the sample data
 - If population is normal, so is the sample data
- Sampling Distribution is normal if...
 - Population is normal, or...
 - $n > 30$ (by the Central Limit Theorem)
 - Otherwise no conclusion about shape

Sampling Distribution (HW 7.1-7.2)

- 16 subjects are from a population skewed right, with mean 40 and s.d. 8.
- Shape of Sample Data
- Mean of Sampling Distribution
- St. Dev. Of Sampling Distribution
- Shape of Sampling Distribution (and why):

Sampling Distribution (HW 7.1-7.2)

- 100 subjects are from a population skewed left, with mean 40 and s.d. 8.
- Shape of Sample Data
- Mean of Sampling Distribution
- St. Dev. Of Sampling Distribution
- Shape of Sampling Distribution (and why):

Mean & Standard Error Properties

As the sample size n increases...

- The mean of the sampling distribution does not change.
- The standard error (s.d. of the sampling distribution) decreases.
- Example: $\frac{1}{2} > \frac{1}{4}$ (larger denominator, smaller overall fraction)

Similarly, as the sample size decreases...

- The mean of the sampling distribution does not change.
- The standard error increases (the opposite).

Distributions (HW 7.1-7.2)

- The size of a badger colony follows a skewed left distribution with mean 20 and s.d. 4. A sample of 36 colonies is selected, and this sample has mean 16 and s.d. 5.5.
- What is the center and spread for the population?
- What is the center and spread for the sample data?
- What shape is the sample data?

Distributions (HW 7.1-7.2)

- What is the center and spread for the sampling distribution with size 36?
- What shape is the sampling distribution?
- Suppose now we adjust the sample size, and the new resulting sampling distribution has a standard error of 1. Did we use a larger or smaller sample size?
- How will the mean of the new sampling distribution change?

StatCrunch (HW 7.1-7.2)

- The average household temperature in Chattanooga is 67.6 degrees, and the s.d. is 4.2. A sample of 51 households is selected.
- What's the probability the **average of this sample** will be above 68.1? Fill in the boxes.

Mean: Std. Dev.:
 Prob(X) =

StatCrunch (HW 7.1-7.2)

- What's the probability the average of the sample will be within 1.5 degrees of the population mean?
- Hint: draw a sketch!

Proportions Problem

	MEAN	ST. ERROR.
POPULATION	p	
SAMPLE DATA (DATA)	\hat{p}	
SAMPLING DIST.	p	$\sqrt{\frac{p(1-p)}{n}}$

The sampling distribution of the sample proportion, \hat{p} , is normal when $np \geq 15$ and $n(1-p) \geq 15$

Computing Square Roots

- It is very important to type in square roots on your calculator correctly, especially with the proportions standard errors
- Example: compute $\sqrt{\frac{.20(1-.20)}{64}}$
- One safe way is to first compute $.20(1-.20)/64 = .0025$, then square root the result: $\sqrt{.0025} = .05$
- Otherwise, use parentheses wisely

Correct: $\sqrt{.20*(1-.20)/64} = .05$

Correct: $\sqrt{.20*.80/64} = .05$

Incorrect: $\sqrt{.20}(1-.20)/64 = .00559$

Incorrect: $\sqrt{.20(1-.20)}/64 = .00625$

Proportions (HW 7.1-7.3)

- 60% of students at an academy in Vancouver are female. In a random sample of 55 students, 26 of them are female.
- Let 1 = female and 0 = male.
- Identify the population distribution of gender.

X	P(X)
1	
0	

Proportions (HW 7.1-7.3)

- Identify the data distribution of gender.
- | X | P(X) |
|---|------|
| 1 | |
| 0 | |
- What is the mean & standard error of the sampling distribution of the sample proportion?
 - Is the sampling distribution approximately normal?

Proportions (HW 7.1-7.3)

- In a certain city, the percentage of kitchens with ant infestations is thought to be 20%. In a sample of 100 kitchens, 32 of them have ants.
- Find the mean and standard error of the sampling distribution for the proportion of kitchens that have ant infestations.
- Is the sampling distribution approximately normal?

Proportions (HW 7.1-7.3)

- Same question: consider a sample of 100 kitchens, 32 of which have ant infestations. Does this provide strong evidence that the population proportion actually exceeds .20? Just like you did on the homework, find the probability of observing a sample proportion this large or higher if p is actually .20.

Mean: Std. Dev.:

Prob(X) =

- The resulting probability is .0013. What is your conclusion?

Sampling Dist. (HW 7.1-7.2)

- The platypus question again! Recall that the mean weight is 4.4 pounds and sd. is 0.4 pounds.
- Fill in the normal calculator box below to find the probability that a randomly selected platypus weighs more than 4.7 pounds.

Mean: Std. Dev.:

Prob(X) =

- Now suppose we have a sample of 9 of these animals. Fill in this box to find the probability that the average of this sample is more than 4.7 pounds.

Mean: Std. Dev.:

Prob(X) =

Notation

We use different letters for population parameters versus sample parameters.

$$\begin{aligned} \mu &= \text{population mean} & \bar{x} &= \text{sample mean} \\ \sigma &= \text{population st. dev.} & s &= \text{sample st. dev.} \\ p &= \text{population proportion} & \hat{p} &= \text{sample proportion} \end{aligned}$$

Know the differences among these!

Notation (HW 7.1-7.3)

- What symbol is used to denote the population mean?
- What symbol is used to estimate the population proportion?
- What symbol is used to describe the spread in one sample?