DATA SCIENCE DISCOVERY

#17: Functions in Python and Conditional Probability
March 4th, 2020
Karle Flanagan and Wade Fagen-Ulmschneider

Control Flow II: Functions!

Up until today, we have explored just one technique to control the flow of code: a **for-loop**.

Here's the for-loop syntax in a simple form:

```
print("Three dice rolls:")
for i in range(3):
  value = random.randint(1, 6)
  print("You rolled a: ” + value)
print("Bye!")
```

The second control flow statement we'll explore is a _______________:

```
def rollDie():
  value = random.randint(1, 6)
  return value
```

Three key ideas when using **functions**:

1. [Syntax]:

2. [Indentation]:

3. [Control Flow]:

4. [Return Value]:

Puzzle #1: After running the previous code in this handout, what does this cell do?

```
rollDie()
```

Puzzle #2: After running the previous code in this handout, what does this cell do?

```
result = rollDie()
result
```

Puzzle #3: After running the previous code in this handout, what does this cell do?

```
result = rollDie() + rollDie() + rollDie()
result
```

A **function** can take in ____:

```
def rollDie(    ):




```

This means: ____! (!!!)

Puzzle #4:

```
rollDie(6)
```

Puzzle #6:

```
rollDie(25)
```

Puzzle #5:

```
rollDie(10)
```

Puzzle #7:

```
rollDie(1000000)
```

Using Functions to Simplify Complex Problems

We are building complex problems! Suppose we don't want to remember the probability of drawing a queen, we can define P_queen():

```
def P_queen():
  return (1/13)
```

What about drawing multiple queens in a row **with replacement**:

```
def P_queen(    ):


```

What about returning a simulation of events?

```
def simulate_rolling_dice( times, dieSides ):
  data = []
  for i in range( times ):
     roll = random.randint(1, dieSides)
     d = { "roll": roll }
     data.append(d)
  df = pd.DataFrame( data )
  return df
```

Conditional Probability and Tables

***Definition:* Conditional Probability:** The *conditional probability* of an event B is the probability that the event will occur given that an event A has already occurred.

Conditional Probability: $P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)} = \frac{P(A \cap B)}{P(A)} = \frac{P(A \wedge B)}{P(A)}$
* *This expression is only valid when P(A) is not equal to 0.* * In the case where events A and B are *independent* (when event A has no effect on the probability of event B), the conditional probability of event B given A is just the probability of event B: *P(B).*

★ If a situation involves 2 or more conditions, a table is helpful. ★

Example 1: Suppose an old box of pills contains 30 Ibuprofen and 70 Tylenol. Half the ibuprofen and 10% of the Tylenol are expired. Fill in the table below:

	Expired	**Not Expired**	**Totals**
Ibuprofen			
Tylenol			
Totals			

a) What's the probability of randomly drawing a pill from the box and getting one that's expired?

b) What's the probability of randomly drawing a pill from the box and getting an expired Ibuprofen?

c) In all contingency tables, the numbers in the margins are called *marginal* probabilities (when divided by the overall total) & the numbers in the inside cells are called *joint* probabilities. Your answer to ______ above is a marginal probability and your answer to ______ above is a joint probability because it gives the probability of getting both a _______________ & a _______________. (Fill in the blanks, with either a or b.)

d) Show that P(A|B) isn't the same as P(B|A). Can you think of some obvious examples that would convince others without having to show them by doing calculations?

Example 2: Suppose jurors make the right decisions about guilt and innocence 95% of the time and that 80% of all defendants are truly guilty.

	Acquitted	Convicted	Totals
Innocent			
Guilty			
Totals			**100**

a) What's the probability that a person is convicted given that they are innocent?

b) What's the probability that a person is innocent, given that they are convicted?

Example 3: For years ELISA has been the most widely used test for screening donated blood for AIDS. (See http://www.stat.ucla.edu/cases/elisa/)

If a blood sample has the AIDS virus, then there is a 99% probability that ELISA will correctly give a positive result. If the blood sample does not have the AIDS virus then there is still a 6% probability that ELISA will incorrectly give a positive result. About 1% of the blood samples ELISA tests are truly contaminated.

	Result of Test		
	Positive	**Negative**	**Totals**
Blood has AIDS virus			
Blood does NOT have AIDS virus			
Totals			**10,000**

a) What's the probability someone tested positive given that the person has the AIDS virus?

b) What's the probability someone has the AIDS virus given the person tested positive?