Midterm Review

Fall 2018 (Adapted from Coding Practice Summer 2018, Stat 140, Past Data 8 Midterms)

For all the questions below, assume we have imported numpy as `np` and `datascience`.

Coding

1. You've been buying jelly beans from a local candy store that only sells watermelon jelly beans. However, you notice that sometimes there are licorice jelly beans in your order. You suspect that the jelly bean machine is broken, and there's a 1% chance that a licorice jelly bean is produced instead of a watermelon jelly bean.

a. Suppose there is indeed a 1% chance of getting a licorice jelly bean instead of watermelon. If you pick out 50 jelly beans chosen at random from among all the jelly beans, what is the chance that you find at least one licorice jelly bean? (You may assume that jelly beans are chosen with replacement from a population in which 1% of jelly beans are licorice flavored). Use simulation to compute the probability.

Let `beans` be an array containing 99 copies of the number 0 (to represent watermelon) and 1 copy of the number 1 (to represent licorice).

```
beans = np.append(0*np.arange(99), 1)

trials = 5,000
licorice = ________________

for ______ in ____________________:
     chosen_beans = ___________________
     num_licorice = ______________________________________
     licorice = ___________________________________

chance_of_at_least_one = ____________________________________
```

b. Define the `eat` function, which should simulate taking n number of jelly beans with replacement from a population in which 1% of jelly beans are licorice flavored. The `eat` function should return the probability of getting exactly `k` licorice jelly beans.

```
beans = np.append(0*np.arange(99), 1)

def eat(k, n, trials):
```

```
""" Repeatedly pick n jelly beans and find the chance of getting
exactly k licorice."""

 k_licorice = _________________

 for ______ in ____________________:
      chosen_beans = ______________________

      num_licorice = ____________________________
      k_licorice = np.append(_____________,
                          np.count_nonzero(_____________________)

 chance_of_exactly_k_licorice = ____________________________________
```

2. The `jelly_bean` table contains the color and count of the jelly beans you want to buy.

Flavor	Count
watermelon	10
cotton candy	20
blueberry	20

The `store` table contains the jelly beans that each store sells and their price in dollars

Store	Flavor	Price
A	watermelon	0.1
A	cotton candy	0.5
A	blueberry	0.3
B	watermelon	0.2
B	cotton candy	0.3
B	blueberry	0.1
C	watermelon	0.5
C	cotton candy	0.6
C	blueberry	0.8

a. Find the total cost of buying all the jelly beans you want from store A.

```
join  = store . _____________________________________
cost = sum(_____________________________ *_____________________________)
```

b. Find the average cost of each flavor in a table, and then add 50 cents to each value

```
def add_50_cents(x):
     return _____________________________

avg_cost_table = ____________________________________________
raised_price = ____________________________________
```

Probability

1. Given a box of 6 red, 3 white, and 7 blue marbles,
    a. What is the probability that the first marble picked is a red marble?

    b. With replacement, what's the probability that the second marble is red?

    c. Without replacement, what's the probability that the second marble is red?

    d. Without replacement, what's the probability a red and a blue marble is picked?

    e. Without replacement, what's the probability that 2 red marbles are picked?

    f. Without replacement, what's the probability a red then a blue marble are picked?

    g. With replacement, what's the probability a red and a blue marble are picked?

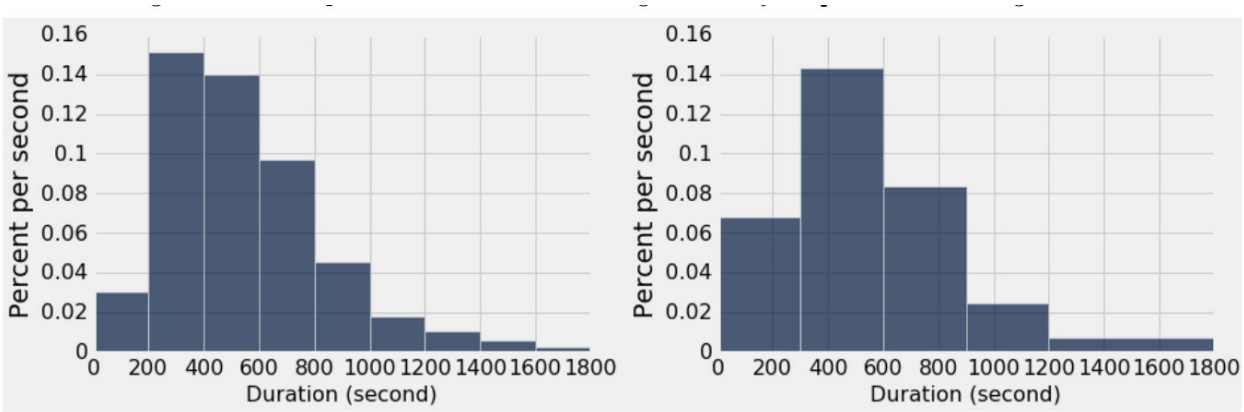
h. With replacement, what's the probability that 2 red marbles are picked?

i. With replacement, what's the probability a red then a blue marble are picked?

2. (**Spring 2017 Midterm)** A basket of 10 colored tickets contains 1 blue, 1 gold, 4 green, and 4 red tickets. If you draw 6 tickets uniformly at random with replacement, what is the chance that you draw at least one that is either blue or gold? Write your answer as a Python expression that computes the result exactly (no simulation).

3. (**Prob 140 Review Set)** A monkey hits the keys of a typewriter at random, picking each of the 26 letters of the English alphabet uniformly each time regardless of what it has picked at all other times. What is the chance that the first six letters are ORANGE, in that order?

Histograms



1. **(Spring 2017 Practice Midterm) T**he two histograms of bike trip durations below were both generated by trip.hist(...) using different bins. Write the proportion of trips that fall

into each range of durations below. Show your work. If it is not possible to tell from the histograms, instead write Not enough information.

a. Between 200 (inclusive) and 400 (exclusive) seconds

b. Between 300 (inclusive) and 900 (exclusive) seconds

c. Between 400 (inclusive) and 900 (exclusive) seconds

d. Greater than 400 seconds

e. Between 250 and 400 seconds

f. Between 200 (inclusive) and 300 (exclusive) seconds

Hypothesis Testing (based upon Practice Midterm 2017)

1. George, a Kiwibot employee, claims that the proportion of Kiwibots that deliver from Chipotle is .2. However, Stephanie doesn't believe him and says that it can't be 20% because she has never witnessed kiwibots leaving Chipotle. Their mutual friend Alex wants to settle this with an experiment
   a. What would be the null and alternative hypothesis?

      Null:

      Alternative:

b. Alex is given a random sample of 9 kiwibots. Alex says that it's not enough to make an conclusion. Do you agree or disagree? Justify your answer.

c. What would be an appropriate test-statistic to use in this hypothesis test? Is TVD an appropriate statistic?

d. Kiwibot has been kind enough to give Alex a sample of 200 orders with the information on which store orders were placed form. After 10000 iterations of sampling, Alex finds that exactly 20% of the sample come from Chipotle. What is the conclusion we can make using our test-statistic?