Data_Analyst_ND_Project0

December 10, 2016

0.1 Chopsticks!

A few researchers set out to determine the optimal length of chopsticks for children and adults. They came up with a measure of how effective a pair of chopsticks performed, called the "Food Pinching Performance." The "Food Pinching Performance" was determined by counting the number of peanuts picked and placed in a cup (PPPC).

0.1.1 An investigation for determining the optimum length of chopsticks.

Link to Abstract and Paper
the abstract below was adapted from the link

Chopsticks are one of the most simple and popular hand tools ever invented by humans, but have not previously been investigated by ergonomists. Two laboratory studies were conducted in this research, using a randomised complete block design, to evaluate the effects of the length of the chopsticks on the food-serving performance of adults and children. Thirty-one male junior college students and 21 primary school pupils served as subjects for the experiment to test chopsticks lengths of 180, 210, 240, 270, 300, and 330 mm. The results showed that the food-pinching performance was significantly affected by the length of the chopsticks, and that chopsticks of about 240 and 180 mm long were optimal for adults and pupils, respectively. Based on these findings, the researchers suggested that families with children should provide both 240 and 180 mm long chopsticks. In addition, restaurants could provide 210 mm long chopsticks, considering the trade-offs between ergonomics and cost.

0.1.2 For the rest of this project, answer all questions based only on the part of the experiment analyzing the thirty-one adult male college students.

Download the data set for the adults, then answer the following questions based on the abstract and the data set.

If you double click on this cell, you will see the text change so that all of the formatting is removed. This allows you to edit this block of text. This block of text is written using Markdown, which is a way to format text using headers, links, italics, and many other options. You will learn more about Markdown later in the Nanodegree Program. Hit shift + enter or shift + return to show the formatted text.

0.1.3 1. What is the independent variable in the experiment?

0.1.4 The length of the chopsticks.

0.1.5 2. What is the dependent variable in the experiment?

0.1.6 The food-pinching performance.

0.1.7 3. How is the dependent variable operationally defined?

0.1.8 The operational definition of the dependent variable is "the food-pinching performance (in this experiment) is the number of peanuts picked and placed in a cup (PPPC)"

0.1.9 4. Based on the description of the experiment and the data set, list at least two variables that you know were controlled.

Think about the participants who generated the data and what they have in common. You don't need to guess any variables or read the full paper to determine these variables. (For example, it seems plausible that the material of the chopsticks was held constant, but this is not stated in the abstract or data description. Because of this, chopstick material should not be cited as a controlled variable.)

0.1.10 I think variables were:

0.1.11 1) the number and composition of the experiment participants;

0.1.12 2) the age of the participants (for separation into two groups);

0.1.13 3) the time for one experiment of checking the food-pinching performance.

One great advantage of ipython notebooks is that you can document your data analysis using code, add comments to the code, or even add blocks of text using Markdown. These notebooks allow you to collaborate with others and share your work. For now, let's see some code for doing statistics.

```
In [1]: import pandas as pd

        # pandas is a software library for data manipulation and analysis
        # We commonly use shorter nicknames for certain packages.
        # Pandas is often abbreviated to pd.
        # hit shift + enter to run this cell or block of code

In [2]: path = r'~/Downloads/chopstick-effectiveness.csv'
        # Change the path to the location where the chopstick-effectiveness.csv
        # file is located on your computer.
        # If you get an error when running this block of code, be sure the
        # chopstick-effectiveness.csv is located at the path on your computer.

        dataFrame = pd.read_csv(path)
        dataFrame.head()

Out[2]:    Food.Pinching.Efficiency  Individual  Chopstick.Length
        0                     19.55           1               180
```

```
1                    27.24               2                  180
2                    28.76               3                  180
3                    31.19               4                  180
4                    21.91               5                  180
```

Let's do a basic statistical calculation on the data using code! Run the block of code below to calculate the average "Food Pinching Efficiency" for all 31 participants and all chopstick lengths.

```
In [4]: dataFrame['Food.Pinching.Efficiency'].mean()

Out[4]: 25.00559139784947
```

This number is helpful, but the number doesn't let us know which of the chopstick lengths performed best for the thirty-one male junior college students. Let's break down the data by chopstick length. The next block of code will generate the average "Food Pinching Effeciency" for each chopstick length. Run the block of code below.

```
In [5]: meansByChopstickLength =
        dataFrame.groupby('Chopstick.Length')['Food.Pinching.Efficiency']
                 .mean()
                 .reset_index()
        meansByChopstickLength

        # reset_index() changes Chopstick.Length from an index to column.
        # Instead of the index being the length of the chopsticks,
        # the index is the row numbers 0, 1, 2, 3, 4, 5.

Out[5]:    Chopstick.Length  Food.Pinching.Efficiency
        0               180                24.935161
        1               210                25.483871
        2               240                26.322903
        3               270                24.323871
        4               300                24.968065
        5               330                23.999677
```

0.1.14 5. Which chopstick length performed the best for the group of thirty-one male junior college students?

0.1.15 240 mm

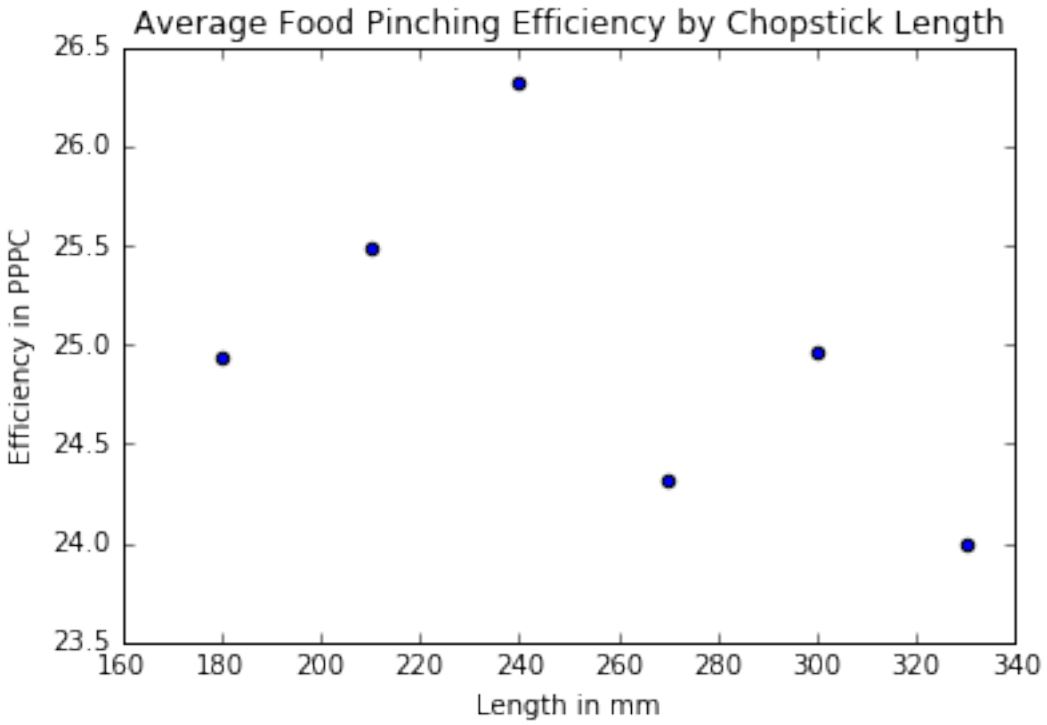
```
In [6]: # Causes plots to display within the notebook rather than in a new window
        %pylab inline

        import matplotlib.pyplot as plt

        plt.scatter(x=meansByChopstickLength['Chopstick.Length'],
                    y=meansByChopstickLength['Food.Pinching.Efficiency'])
                    # title="")
        plt.xlabel("Length in mm")
```

```
        plt.ylabel("Efficiency in PPPC")
        plt.title("Average Food Pinching Efficiency by Chopstick Length")
        plt.show()
```

```
Populating the interactive namespace from numpy and matplotlib
```



0.1.16 6. Based on the scatterplot created from the code above, interpret the relationship you see. What do you notice?

0.1.17 This dependence is not linear, and the first half of the graph has the positive direction, the second - negative. In the graph a small number of points is located, so it's hard to see the quality of correlation. It looks like a strong correlation, but with one outlier (270 mm). The maximum value of the variable "Efficiency in PPPC" reaches at 240 mm of the variable "Length in mm".

0.1.18 In the abstract the researchers stated that their results showed food-pinching performance was significantly affected by the length of the chopsticks, and that chopsticks of about 240 mm long were optimal for adults.

0.1.19 7a. Based on the data you have analyzed, do you agree with the claim?

0.1.20 Yes.

0.1.21 7b. Why?

0.1.22 The value 240 mm corresponds to the maximum value of the variable "Efficiency in PPPC", so we can expect that this length of the chopsticks is the most efficient and ergonomically optimal.

0.1.23 Let's describe the data.

```
In [7]: print meansByChopstickLength['Food.Pinching.Efficiency'].describe()
```

```
count     6.000000
mean     25.005591
std       0.830306
min      23.999677
25%      24.476694
50%      24.951613
75%      25.354919
max      26.322903
Name: Food.Pinching.Efficiency, dtype: float64
```

0.1.24 For checking the statement we can use the analysis of the null hypothesis: the length of chopsticks 240 mm does not significantly affect the value of the variable "Efficiency in PPPC".

```
In [29]: y = meansByChopstickLength['Food.Pinching.Efficiency']
         std_y = pd.Series((y - y.mean())/y.std(ddof=0))
         max_std_y = std_y.max()
         print max_std_y
```

```
1.73796608764
```

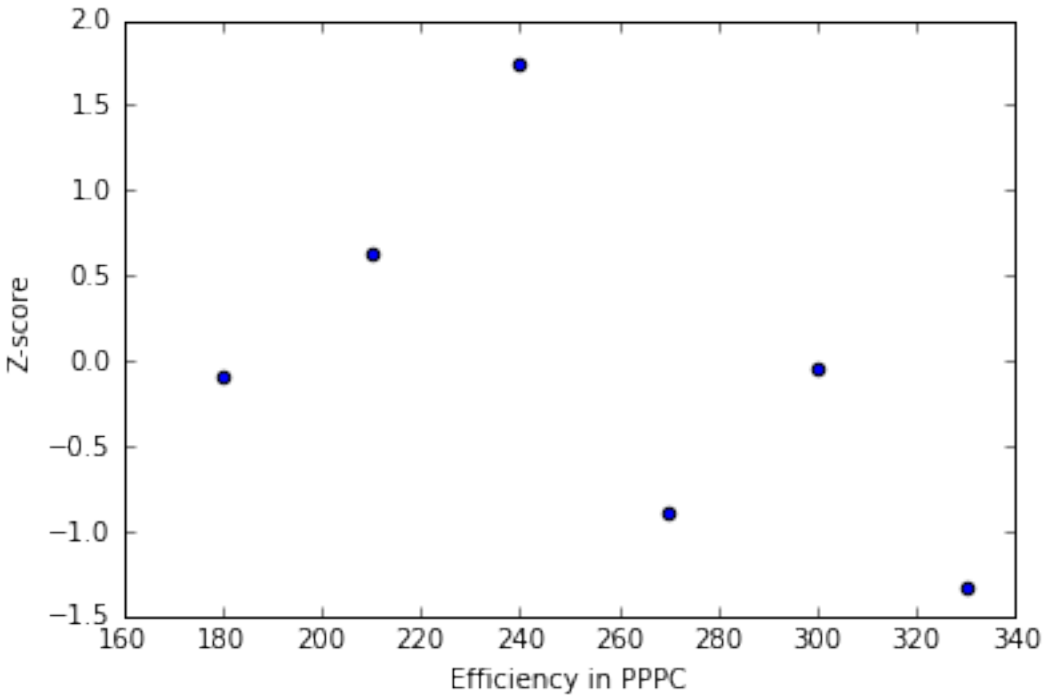
```
In [31]: df = pd.DataFrame(data={'Chopstick.Length':
                                  meansByChopstickLength['Chopstick.Length'],
```

```
                                'Food.Pinching.Efficiency':
                                meansByChopstickLength['Food.Pinching.Efficiency']
                                'Z-score': std_y})
df
```

Out[31]:

	Chopstick.Length	Food.Pinching.Efficiency	Z-score
0	180	24.935161	-0.092920
1	210	25.483871	0.631008
2	240	26.322903	1.737966
3	270	24.323871	-0.899413
4	300	24.968065	-0.049510
5	330	23.999677	-1.327130

In [75]:
```
plt.scatter(x=meansByChopstickLength['Chopstick.Length'], y=std_y)
plt.xlabel("Efficiency in PPPC")
plt.ylabel("Z-score")
plt.show()
```



0.1.25 We can see that at a length of chopsticks 240 mm variable "Effectiveness in CPR" deviates from the average value of this index on the value 1.73796608764 of the standard deviation.

0.1.26 The probability of this event is P(1.74) = 0.08186. This goes beyond the confidence interval with a level 0.1.

0.1.27 We must reject the null hypothesis at this level.