STA 3032

Python assignment.

Everyone is responsible for submitting their own jupyter notebook file in webcourses.

You may work on this assignment in groups of three or four students.

All calculations are to be completed in python. Do not calculate in excel (or other applications) and just print the answers in python.

There are 10 questions that need to be answered using python. Each chapter question is equally weighted.

Each chapter question has multiple steps to be completed to get full credit.

General notes for the assignment.

- The typical imports we have been using for class are:

```
import numpy as np
import pandas as pd
import scipy as sp
from scipy.stats import stats
import matplotlib.pyplot as plt
```

- This assignment can be completed with just those imports; however, additional imports may be used but must be included in the standard install of winpython. If you want to use an import and you are unsure if it is included with the basic install, ask me.
- All data sets are located in the folder “data” and are included in the assignment zip file
- The example print statements with question marks are to be filled in with variables containing the calculated value and not hard coded values.
    - e.g. If the print statement is => print('mean= 0.??; var=0.??');
      then the python code should look something like

```
xBar = dataFrame['ColumnName '].mean()
s = dataFrame['ColumnName '].std(ddof=1)
print('mean=',xBar, '; var=',s);
```

- If the print statement has XXX, then just type the name that is missing. For instance if the estimater is the median then the point estimate is xTilda.
    - e.g. If the print statement is => print('point estimate used was XXX ');
      then the python code should look something like

```
print('point estimate used was xTilda');
```

Chapter 1

Based on chapter 1, question 63

A sample of 77 individuals working at a particular office was selected and the noise level (dBA) experienced by each individual was determined, yielding the following data (“Acceptable Noise Levels for Construction Site Offices,” Building Serv. Engr. Research and Technology, 2009: 87–94).

Use python to organize, summarize, and describe the data.

Steps:
1) use pandas create to create a dataFrame using read_csv the csv file is 'data/Chaper1.csv'
Column header is 'noise'

2) Display the descriptive statistics for the sample 'noise'
print out at a minimum: count, mean, std, min, median, and max

3) Display Histogram of the noise levels

4) Is this data positive or negative skewed?

```
print('positive Skew');
or
print('negative Skew');
```

Chapter 2

Based on chapter 2, question 49

The accompanying table gives information on the type of coffee selected by someone purchasing a single cup at a particular airport kiosk.

	Small	Medium	Large
Regular	0.14	0.2	0.26
Decaf	0.20	0.1	0.10

Consider randomly selecting such a coffee purchaser

Steps

1) Create a dataFrame like the one above.

columns: Small, Medium, Large
indexes: Regular, Decaf

2) What is the probability that the individual purchased a small cup p(small)?

```
print('p(small)= 0.??');
```

3) What is the probability that the individual purchased a decaf coffee?

```
print('p(decaf)= 0.??');
```

4) What is p(decaf |small) ?

```
print('p(decaf|small) = 0.??');
```

5) What is p(small | decaf) ?

```
print(p(small|decaf) =0.??);
```

Chapter 3

Based on chapter 3, question 79

The article “Expectation Analysis of the Probability of Failure for Water Supply Pipes” (J. of Pipeline Systems Engr. and Practice, May 2012: 36–46) proposed using the Poisson distribution to model the number of failures in pipelines of various types. Suppose that for cast-iron pipe of a particular length, the expected number of failures is 1 (very close to one of the cases considered in the article). Then X, the number of failures, has a Poisson distribution with $\mu=1$.

Steps

1) Create a Poisson random variable with lambda = 1

2) Print the value for p(X ≤ 5);

```
print('p(X ≤ 5)=0.??');
```

3) Print the value for p(X=2);

```
print('p(X = 2)=0.??');
```

4) Print the value for p(2 ≤ x ≤ 4);

```
print('p(2 ≤ x ≤ 4)=0.??');
```

Chapter 4

Based on chapter 4, question 98
Let X= time it takes a read/write head to locate a desired record on a computer disk memory device once the head has been positioned over the correct track.
If the disks rotate once every 25 milisec, a reasonable assumption is that X is uniformly distributed on the interval [0, 25].

Steps

1) Create a uniform random variable with A=0 and B=25

2) Compute p(10 ≤ X ≤ 20)

```
print('p(10 ≤ X ≤ 20)=0.??');
```

3) Compute p(X ≥ 10)

```
print('p(X ≥ 10)=0.??');
```

4) Display E(X), V(X)

```
print('E(X)=???; V(X)=???');
```

Chapter 5

Based on Example 5.23 in the book.
The population distribution for our first simulation study is normal with mu=8.25 and sigma=0.75. See page 226 for details

Steps

1) Generate a normal random variable with mu=8.25 and sigma=0.75

Create a simulation that conducts 500 replications. Within each replication generate n=5 samples randomly selected from the normal random variable.

2) Store the sample mean from the n=5 samples for each replication in a numpy array.
    Hint: review looping and generating random samples from a normal distribution

3) Display a histogram of the sample means. Set the range to be [7,10]
    example: plt.hist(data, range=[7,10])

4) Repeat step 2 and 3 but change n=30

Chapter 6

Based on chapter 6 question 1.

The accompanying data on flexural strength (MPa) for concrete beams of a certain type:

5.9,7.2,7.3,6.3,8.1,6.8,7.0,7.6,6.8,6.5,7.0,6.3,7.9,9.0,8.2,8.7,7.8,9.7,7.4,7.7,9.7,7.8,7.7,11.6,11.3,11.8,10.7

Steps

1) Create a DafatFrame by importing data on flexural strength (MPa) for concrete beams.
The data file is located in 'data/Chaper6.csv'

2) Calculate a point estimate of the mean value of strength for the conceptual population of all beams.
    print('θHat = ???')

3) State which estimator you used.
    print('point estimate used was XXX')

4) Calculate a point estimate of the strength value that separates the weakest 50% of all such beams from the strongest 50%.
    print('θHat = ???')

5) State which estimator you used.
    print('point estimate used was XXX')

6) Calculate and interpret a point estimate of the population standard deviation.
    print('θHat = ???')

7) Which estimator did you use?
    print('point estimate used was XXX')

8) Calculate a point estimate of the proportion of all such beams whose flexural strength exceeds 10 MPa.
    print('θHat = ???')

9) State which estimator you used.
    print('point estimate used was XXX')

10) Calculate a point estimate of the population coefficient of variation
    print('θHat = ???')

11) State which estimator you used.
    print('point estimate used was XXX')

Chapter 7

Based on chapter 7, question 37

A study of the ability of individuals to walk in a straight line (“Can We Really Walk Straight?” Amer. J. of Physical Anthro., 1992: 19–27) reported the accompanying data on cadence (strides per second) for a sample of n=20 randomly selected healthy men. [Import the data from data/Chapter7.csv]

A normal probability plot gives substantial support to the assumption that the population distribution of cadence is approximately normal.

Steps

1) Use pandas create to create a dataframe using read_csv the csv file is 'data/Chaper7.csv'
Column header is 'cadence'

2) Calculate a 95% confidence interval for population mean cadence. Print the mean, lower, an upper limits

```
print('Sample Mean=???; C.I. (LowerValue???, upperValue???)');
```

Chapter 8

Based on chapter 8, question 37

The accompanying data on cube compressive strength (MPa) of concrete specimens appeared in the article “Experimental Study of Recycled Rubber-Filled High-Strength Concrete” (Magazine of Concrete Res., 2009: 549–556): 112.3, 97.0, 92.7, 86.0, 102.0, 99.2, 95.8, 103.5, 89.0, 86.7

Steps
1) use pandas create to create a dataframe using read_csv the csv file is 'data/Chaper8.csv' Column header is 'strength'

Is it plausible that the compressive strength for this type of concrete is normally distributed?
2) Print the normal probability plot.
    Hint: lookup scipy.stats.probplot, use the plot=plt option and use plt.show()

3) Are the data points along the normal line on the plot?

    print('Yes');
    or
    print('No');

Suppose the concrete will be used for a particular application unless there is strong evidence that true average strength is less than 100 MPa.

Carry out a test of appropriate hypotheses Ho: μ = 100 versus Ha: μ < 100
4) Calculate and print the test statistic for the null hypothesis
    Hint: lookup scipy.stats.ttest_1samp

    print(‘test statistic=???’)

5) Print the p value (hint: remember it's a 1 tail test)

    print(‘p value=???’)

6) Should the concrete be used? alpha = .05

    print("Fail to reject the null. This concrete should be used.");
    or
    print("Reject the null in favor of the alternative hypothesis. This concrete should NOT be used.");

Chapter 9

Based on Chapter 9 question 37
Hexavalent chromium has been identified as an inhalation carcinogen and an air toxin of concern in a number of different locales. The article “Airborne Hexavalent Chromium in Southwestern Ontario” (J. of Air and Waste Mgmnt. Assoc., 1997: 905–910) gave the accompanying data on both indoor and outdoor concentration for a sample of houses selected from a certain region.

Steps

1) Use pandas create to create a dataframe using read_csv the csv file is 'data/Chaper9.csv'

2) Create a new column 'difference' in the data frame that is the subtraction of the indoor Concentration from the outdoor Concentration.

3) Calculate a confidence interval for the population mean difference between indoor and outdoor concentrations using a confidence level of 95%
Hint: use the ConfidenceInterval method from chapter 7 question

```
print('C.I. (lower??, upper??)');
```

4) Can we be confident, at the 95% confidence level, that the true average concentration of hexavalen chromium outdoors exceeds the true average concentration indoors?

```
print('Yes we can be confident because the 95% CI does not contain zero.')
or
print(‘No we cannot because the 95% CI contains zero.')
```

Chapter 10

Based on Chapter 10 question 37
Numerous factors contribute to the smooth running of an electric motor (“Increasing Market Share Through Improved Product and Process Design: An Experimental Approach,” Quality Engineering, 1991: 361–369). In particular, it is desirable to keep motor noise and vibration to a minimum. To study the effect that the brand of bearing has on motor vibration, five different motor bearing brands were examined by installing each type of bearing on different random samples of six motors. The amount of motor vibration (measured in microns) was recorded when the motors ware running. The data for this study is in the file data/Chapter10.csv.

Let μi = true average amount of motor vibration for each of five bearing brands. Then the hypotheses are Ho:μ1 = μ2 =... = μ5 vs. Ha: at least two of the μi’s are different.

Steps

1) Use pandas create to create a dataframe using read_csv the csv file is 'data/Chaper10.csv'

2) Conduct a one-way ANOVA based on Bearing brands. Print the test statistic *f* and p value for the ANOVA test.

```
print('test statistic= ???; p value=???');
```

3) Should we reject or accept the null hypothesis at an alpha of 0.05?

```
print('Fail to reject. All bearings produce the same amount of vibrations') or
print('reject the null. At least two of the μi’s are different.')
```