STATISTICS 147 PRACTICE EXAM I

1 Data Files

The following data files may be downloaded from *iLearn* under **DataFiles**.

hallmk.dat

```
amount
23.75
26.88
21.50
32.50
28.50
29.00
30.00
27.45
29.56
26.00
```

aussie.dat

```
ACD   AS
45    52
43    55
51    60
52    58
43    62
48    49
47    55
46    54
45    60
48    59
```

agegroup.dat

```
G1  G2  G3  G4
29  20  37  28
33  21  25  29
26  30  22  34
27  28  33  36
39  20  28  21
35  23  26  20
33  23  30  25
29  23  34  24
36  21  27  33
22  25  33  32
```

2 Part 1: R

General Information:

- Complete the following using R.
- Include the following as comments at the beginning:

```
# Statistics 147 Practice Exam 1, Part 1: R
# Fall 2020
# Your name
# Your Section YY (where YY = section number: 001 or 002)
# R Question XX (where XX = question number)
```

1. *Idealab* is offering people the opportunity to receive a tablet for free. The company plans to give away 1,000 tablets. Interested persons may fill out an application on a specified Web site. A statistician believes the probability that a person who fills out an application will actually receive one of the free tablets is 0.15. The statistician requests and receives permission from *Idealab* to take a random sample of size 20 from the completed applications.

   (i) What is the probability that exactly 4 of the selected applicants will receive a free tablet? (2 pts)

   (ii) What is the probability that between 2 and 5, inclusively, of the selected applicants will receive a free tablet? (3 pts)

   (iii) What is the probability that more than 4 of the selected applicants will receive a free tablet? (3 pts)

   (iv) How many of the selected applicants would one expect to receive a free tablet? (2 pts)

2. The time a client stays logged on to a local internet service provider is normally distributed with a mean of 45 minutes and a standard deviation of 5 minutes. A client is selected at random.

   (i) What is the probability that the client stays connected for at most 43 minutes? (2 pts)

   (ii) What is the probability that the client stays connected for more than 43 minutes? (2 pts)

   (iii) What is the probability that the client stays connected for a time of between 42 and 47 minutes? (4 pts)

   (iv) What connection time represents the 97.5th percentile? (i.e., What connection time exceeds 97.5% of all other connection times?) (2 pts)

3. Let $y = 2x^3 + 4$, for $x = 0, 1, 2, 3, 4, 5, 6$.

   (i) Generate the sequence 0 through 6 and store in x.

   (ii) Generate the values of y.

4. A local *Hallmark* store manager claims that the average its customers spend on Halloween cards and knick-knacks is $30. To check the claim, the manager takes a random sample of 10 customers yielding the following data:

```
23.75   26.88   21.50   32.50   28.50
29.00   30.00   27.45   29.56   26.00
```

   The data is located in the file **hallmk.dat**. The actual data starts on Line 2.

   (i) Read in and print out the data.

   (ii) The average amount spent by the customers is _____.

   (iii) The median amount spent by the customers is _____.

   (iv) The standard deviation for the amount spent by the customers is _____.

   (v) The variance for the amount spent by the customers is _____.

   (vi) The minimum amount spent by the customers is _____.

   (vii) The maximum amount spent by the customers is _____.

   (viii) Create a histogram for the amount spent by the customers. Use the following colors

```
colors1 <- c("blue","green","yellow","red","brown","gray","purple")
```

   and breaks

```
breaks1 <- seq(20,34,2)
```

5. A dog enthusiast was interested in comparing the mean weights of the *Australian Cattle Dog (ACD)* and the *Australian Shepherd (AS)*. The dog enthusiast obtained random samples from the national registry of dogs yielding the following data:

Australian Cattle Dog	45	43	51	52	43	48	47	46	45	48
Australian Shepherd	52	55	60	58	62	49	55	54	60	59

The data is located in a file named **aussie.dat**. The actual data begins on Line 2.

(i) Read in and print out the data.

(ii) Plot the weights of the *Australian Cattle Dog* versus the weights of the *Australian Shepherd*. Sketch the plot. Does there appear to be any pattern present?

6. An experiment was conducted to examine the effect of age on heart rate when a person is subjected to a specific amount of exercise. Ten male subjects were randomly selected from four age groups (10-19, 20-39, 40-59, and 60-69). Each subject walked on a treadmill at a fixed grade for 12 minutes. The increase in heart rate (difference between after and before exercise, was recorded, in beats per minute. The data are shown in the table below: (The data is located in a file named **agegroup.dat**.) Read in and print out the data, making sure the columns are accessible individually.

3 Part 2: SAS

General Information:

- ♣ Complete the following using SAS.
- ♣ Save your SAS program in a file named *pr1_su20.sas*
- ♣ Include the following titles:

```
title1 'Statistics 147 Practice Exam 1, SAS';
title2 'Summer 2020';
title3 'Your name';
title4 'Section Y';  (Y = 001 or 002)
title4 'Question X';     (X = question number)
```

- ♣ You should place comments in your SAS program file.

1. Let $y = 2x^3 + 4$, for $x = 11, 12, 13, 14, 15, 16, 20$. Enter the values of x and use SAS to calculate the values of y.
2. Let $y = \sqrt{m^2 + n^2}$ for m = 1,2 and n = 1,2,3. Use nested DO loops to generate the values of m and n and compute the values of y.
3. Consider the following data:

x	1	3	5	7	9	11	13
y	1	8	24	40	80	20	150

(i) Plot x versus y. Sketch the plot/include the plot in a PDF.

(ii) Do any of the data points appear not to fit with the other data points? If so, answer yes and circle the point on your plot. If not, answer no.

4. Refer to R Question 4. Read in and print out the data using SAS.

   Recall: The data is located in the file, *hallmk.dat*. Use an `infile` statement to enter the data!

   Write the appropriate code to complete the following:

   (i) The average amount spent by the customers is _____.

   (ii) The median amount spent by the customers is _____.

   (iii) The standard deviation for the amount spent by the customers is _____.

   (iv) The variance for the amount spent by the customers is _____.

   (v) The minimum amount spent by the customers is _____.

   (vi) The maximum amount spent by the customers is _____.

   (vii) Create a histogram (3D vertical bar chart) for the amount spent by the customers with a starting midpoint of 20, an ending midpoint of 36 and an increment of 4. Sketch the histogram. List the frequency next to each of the bars.

   (viii) Create a histogram (3D vertical bar chart) for the amount spent by the customers with a starting midpoint of 20, an ending midpoint of 34 and an increment of 1. Sketch the histogram. List the frequency next to each of the bars.

   (ix) Which of the two histograms seems more appropriate for this data? Justify your answer.

5. Refer to R Question 5.

   Recall: Since the data is located in the file *aussie.dat*, use an infile statement to enter the data!

   (i) Read in the data using nested DO loops. Print your data as a check!

   (ii) Find the mean and variance for each dog breed.

   (iii) Using the **set** and **if** commands, create a new temporary SAS dats set, called just_sheph, that just consists of the *Australian Shepherd* data.

6. Refer to R Question 6.

   Recall: Since the data is located in the file *agegroup.dat*, use an `infile` statement to enter the data!

   (i) Read in the data using nested do loops. Print the data.

   (ii) Use an **if-then-else** structure to name the columns as follows:

Group	**GrName**
G1	10 to 19
G2	20 to 39
G3	40 to 59
G4	60 to 69

   (iii) Using the **set** and **if** commands, create a new temporary SAS dats set, called **just2**, that just consists of the **G2** data. (Be sure to print your data!)

   (iv) Using the **set** and **if** commands, create a new temporary SAS dats set, called **just2and3**, that just consists of the **G2** and **G3** data. (Be sure to print your data!)

7. Refer to R Question 1. Use SAS to complete the following.

   (i) What is the probability that exactly 4 of the selected applicants will receive a free tablet?

   (ii) What is the probability that between 2 and 5, *inclusively*, of the selected applicants will receive a free tablet?

   (iii) What is the probability that more than 4 of the selected applicants will receive a free tablet?

   (iv) How many of the selected applicants would one expect to receive a free tablet?

8. Refer to R Question 2. Use SAS to complete the following.

(i) What is the probability that the client stays connected for at most 43 minutes?

(ii) What is the probability that the client stays connected for more than 43 minutes?

(iii) What is the probability that the client stays connected for a time of between 42 and 47 minutes?

(iv) What connection time represents the 97.5th percentile? (i.e., What connection time exceeds 97.5% of all other connection times?)