Handout 2

1. In a Pew Research Center poll of 745 randomly selected adults, 589 (or 79.1%) said that it is morally wrong to not report all income on tax returns.

   a) In words, what is the parameter of interest in this survey?

   b) The estimate from the survey is likely biased. Give the main potential source of bias (selection, non-response, response) and state whether the resulting estimate will be tend to be too large or too small.

   c) Referring to the Pew Research Poll from the previous page: Suppose you want to conduct a survey to see if people of the same socioeconomic status (e.g. low income, middle class, wealthy) might have similar opinions on this issue, but responses might vary from one socioeconomic class to another. Describe a probability sampling design for carrying out this survey.

2. Three instructors are comparing scores on their final; each had 99 students. The distribution of scores in each class is shown below. Note that, in class C, one student got a score of 2, one student got a score of 3, and so forth, all the way through 99.

Class A

Score	Frequency
1	1
50	97
99	1

Class B

Score	Frequency
1	49
50	1
99	49

Class C

Score	Frequency
1	1
2	1
.	.
.	.
.	.
98	1
99	1

a) Which class had the largest mean? or are they the same? Explain. You do not have to calculate the mean, but reason your answer instead.

b) Which class had the largest SD? Which one had the smallest? or are they the same? Explain. You do not have to calculate the SD to answer this question but give your reasoning clearly.

c) Which class had the largest inter-quartile range (Q3-Q1)? Which one had the smallest? Or are they the same? Explain. In order to be able to answer this question, you will need to calculate Q1 and Q3 for each class.

3. Hearing anecdotal reports that some patients undergoing treatment for the eating disorder anorexia seemed to be responding positively to the antidepressant Prozac, medical researchers conducted an experiment to investigate this. They found 93 women being treated for anorexia who volunteered to participate. For one year, 49 randomly selected patients were treated with Prozac and the other 44 were given an inert substance called a placebo. At the end of the year, patients were diagnosed as healthy or relapsed, as summarized in the table:

	Prozac	Placebo	**Total**
Healthy	35	32	67
Relapse	14	12	26
Total	49	44	**93**

Does this result provide evidence that Prozac might be helpful in treating anorexia? Explain.

4. Here are costs of 10 electric smoothtop ranges rated very good or excellent by a consumer magazine.

$ 650, 750, 650, 1250, 850, 700, 1100, 900, 900 and 700

Find the following:

a) Mean

b) Median

c) Interquartile range (IQR)

d) Standard deviation

5. How many points do football teams core in championships games? The total number of points scored by both teams in each 11 championship games are given below:

31, 41, 59, 36, 69, 61, 23, 44, 55, 47 and 36

Find the

a) Five Number Summary

b) Interquartile range (IQR)

c) Construct a boxplot of the data and comment on the shape of the distribution. Are there any outliers in the data?

6. The table shows the 5-Number summary for the run times in minutes of the 250 highest grossing movies of last year.

Minimum Value	72
First Quartile (Q_1)	104
Median (Q_2)	113
Third Quartile (Q_3)	119
Maximum Value	131

Are there are any outliers in the data? Explain.

7. Use the **class data** posted on Canvas for this question. The **Code Book** also posted on Canvas provides a list of the variables in the dataset and their descriptions. Use the data to answer the following questions.

   a) How many observations are in the dataset?

   b) Write an informative summary describing features of the study participants with respect to the variables age (of students), gender, and class standing. Reference appropriate numerical and graphical summaries as needed. Limit your answer to no more than five sentences.

c) Is there an association between the variables gender and Intro-Extra? In order to answer this question correctly, provide answers to the following questions below:

- i) Create a contingency table.
- ii) What proportion of males are introvert?
- iii) What proportion of females are introvert?
- iv) Based on the above results, do you think there is an association between the two variables?
- v) Can the results obtained be generalized? Argue carefully.

d) Explore the relationship between age (of students) and class standing. Comment on the center, spread, shape and outliers. In addition, provide any useful suggestions.

8. For the following variables, would you expect the distribution to be uniform, unimodal or bimodal? Symmetric or skewed? Explain.

   a) Ages of people at a little league game.

   b) Weights of female babies born in a particular hospital over the course of a year.

   c) The length of the average hair on the heads of students in a large class.

   d) The life span of the general population.