2028 - Basic Statistical Methods
Homework 1 - Descriptive Statistics 100 points total.

This homework is due Tuesday Sept 1st in class.

- Please remember to staple if you turn in more than one page.
- Please make sure to **SHOW ALL WORK** in order to receive full credit.

1. (10 points.) Suppose you are trying to estimate the proportion of all US adults who support the idea of same–sex marriages. You randomly select 1200 adults and ask them their opinion (support/object/no opinion). For each of the following (i-iv), choose the label (a, b, c, or d) that best describes it:

   (i) The 1200 selected adults that were asked
   (a) population **(b)** sample **(c)** parameter **(d)** statistic

   (ii) The proportion of all US adults who support the idea of same–sex marriages
   (a) population **(b)** sample **(c)** parameter **(d)** statistic

   (iii) The proportion among the 1200 adults selected, that support the idea of same–sex marriages
   (a) population **(b)** sample **(c)** parameter **(d)** statistic

   (iv) All of US adults
   (a) population **(b)** sample **(c)** parameter **(d)** statistic

2. (20 points.) **Numerical summaries.** (6-53 in textbook.) The ?cold start ignition time? of an automobile engine is being investigated by a gasoline manufacturer. The follow- ing times (in seconds) were obtained for a test vehicle: 1.75, 1.92, 2.62, 2.35, 3.09, 3.15, 2.53, 1.91.

   (a) Calculate the sample mean, sample variance, and sample standard deviation.

   (b) Construct a box plot of the data by hand.

3. (20 points.) **More numerical summaries.** (modified 6-55 in textbook.) The nine measurements that follow are furnace temperatures recorded on successive batches in a semiconductor manufacturing process (units are ??F): 953, 950, 948, 955, 951, 949, 957, 954, 955.

   (a) Calculate the sample mean, sample variance, and standard deviation.

   (b) Find the range and median median. How much could the largest temperature measurement increase without changing the median value?

   (c) Find Q_1 and Q_3 and the IQR.

   (d) Now we have a new data point added which is 1500. Recalculate the sample mean, sample variance, standard deviation, median, range, Q_1 and Q_3, and IQR. How would you comment on this data point? Which numerical summaries are robust to the outlier and why?

4. (20 points.) **Comparative study of student heights.**

   (a) The female students in an undergraduate engineer- ing core course at ASU self-reported their heights to the nearest inch. The data are below. Construct a stem-and-leaf diagram for the height data and comment on any important features that you notice. Calculate the sample mean, the sample standard deviation, and the sample median of height.

   Female height: 62 64 66 67 65 68 61 65 67 65 64 63 67 68 64 66 68 69 65 67 62 66 68 67 66 65 69 65 70 65 67 68 65 63 64 67 67

   (b) In the same class, the male students self- reported their heights as follows. Construct a comparative stem-and-leaf diagram by listing the stems in the center of the display and then placing the female leaves on the left and the male leaves on the right. Comment on any important features that you notice in this display.

   Male height: 69 67 69 70 65 68 69 70 71 69 66 67 69 75 68 67 68 69 70 71 72 68 69 69 70 71 68 72 69 69 68 69 73 70 73 68 69 71 67 68 65 68 68 69 70 74 71 69 70 69

   (c) Construct comparative box plots. Write an interpretation of the information that you see in these plots.

5. (30 points.) **Smiling?** Dale Carnegie stated that smiling helps win friends and influence people. Research on the effects of smiling has backed this up and shown that a smiling person is judged to be more pleasant, attractive, sincere, sociable, and competent than a non-smiling person.

   There is evidence that smiling can attenuate judgments of possible wrongdoing. This phenomenon termed the "smile-leniency effect" was the focus of a study by Marianne LaFrance & Marvin Hecht in 1995. These researchers were interested in two questions: (a) Does smiling really increase leniency? (b) Are different types of smiles differentially effective?

   Data. Subjects in the experiment were asked to assume the role of a student member of a college disciplinary panel and judge a student accused of cheating. Each subject received a file that contained (a) a letter ostensibly from the chairperson of the Committee on Discipline, (b) a summary of the evidence, (c) background information on the suspect including prior academic performance and a color picture portraying one of the four facial expressions, and (d) rating scales to indicate the judgments. Subjects answered five questions about the likelihood of the suspect's guilt and how severe the punishment should be. These questions were combined into one "leniency score." This score is such that the higher is the score, the higher is the lenience.

   Four groups of subjects were tested. Each group saw one of the three types of smiles or a neutral-expression control. Subjects from a sample of 136 college students were randomly assigned to the four conditions with the constraint that there was an equal number of subjects (34) in each group.

The data are attached together with the homework document. The name of the data file is 'smiles.txt'. Once you have saved the data file in the working directory, read the data in R using the commands

```
smiles=read.table("smiles.txt")
names(smiles)=c("groups","scores")
attach(smiles)
```

The first command reads the data in R, the second assigns names to each column and the third makes the components in the data readable by their specified names.

(Part 1) Construct histograms and stem-and-leaf plots for each of the four categories. Comment on the shape of the distribution of the observations in each of the four categories. Interpret and compare.

Instructions on how to use R:

i. To obtain the stem-and-leaf plots for all four groups jointly you can run the following R command:

```
tapply(scores,groups,stem)
```

which applies the function *stem* in R to each of the four groups.

ii. To obtain the histograms for all four groups jointly you can run the following R command:

```
splitgroup=split(scores,groups)
attach(splitgroup)
par(mfrow=c(2,2))
hist(false,main="")
hist(felt,main="")
hist(miserable,main="")
hist(neutral,main="")
```

which first splits the data by group, then divides the figure into four panels and within each panel, plots the histogram of one group. Note that you apply the *hist* function to each of the four groups.

(Part 2) Obtain the 5-numerical summaries and the corresponding boxplots for all four categories. Interpret and compare.

Instructions on how to use R:

i. To obtain the 5-numerical summary for each group you will use the R functions *summary*, which summarizes the data without providing the variance, and $\widehat{\text{var}}$ which will give you the variance. The R code below is only for one group. Repeat for the other three groups.

```
summary(false)
var(false)
sqrt(var(false))
```

ii. The R function that you could use to construct a boxplot is *boxplot*. There are two ways to construct the boxplots for all categories jointly for comparison. One way is to use a similar code as you used for histograms.

```
par(mfrow=c(2,2))
boxplot(false)
title("false")
boxplot(felt)
title("felt")
boxplot(miserable)
title("miserable")
boxplot(neutral)
title("neutral")
```

A different approach is to simply use the boxplot function as follows:

```
par(mfrow=c(1,1))
boxplot(scores~groups)
```

Try to understand the input to the boxplot function. One way to read the help menu for a function is to use:

```
par(mfrow=c(1,1))
help(boxplot)
```

(Part 3) Based on the descriptive statistics and the different graphical displays, summarize your findings for the data analysis in the context of the problem.