## ECO220Y: Homework 2

Required Exercises: <u>Chapter 5</u>: 1, 3, 5, 7, 9, 11, 37, 47, 79, 87 (**Note:** The first sets of exercises at the end of each chapter correspond to specific sections. These are best done *as you complete each section* (i.e. before you finish the reading).)

Required Problems:

(1) If you add up the bar <u>heights</u> what is total for a frequency histogram? If you add up the bar <u>heights</u> what is total for a relative frequency histogram? If you add up the <u>area</u> of the bars what is the total for a density histogram?

(2) Data record hours of work in a day (accurate to the nearest second) for a sample of 78 employees. For each of these histograms, approximately what *percentage* of the sample worked 12 or more hours? Show your work for each.

(a)

(b)

(c)
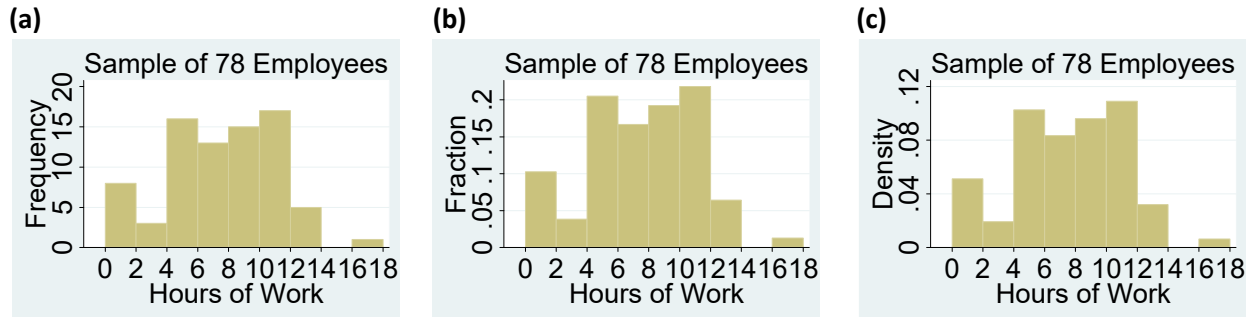
    (d) Explain any differences in your answers to Parts (a) – (c).

(3) A complete census of all 20,000 employees reveals that only 5 percent work 12 or more hours. How does this differ from answers for Question (2)? Give a plausible explanation using concepts from class and the readings.

(4) Consider the following tabulation. (These are the same data on hours of work from Question 2 above.) As summaries of these data, compare and contrast this tabulation with the histograms.

```
      hours |      Freq.      Percent         Cum.
------------+---------------------------------------
          0 |          7         8.97          8.97
   1.984956 |          1         1.28         10.26
   3.025122 |          1         1.28         11.54
   3.874622 |          1         1.28         12.82
   3.891577 |          1         1.28         14.10
   4.153069 |          1         1.28         15.38
   4.301862 |          1         1.28         16.67
        ...
   13.11456 |          1         1.28         97.44
    13.2879 |          1         1.28         98.72
   16.62028 |          1         1.28        100.00
------------+---------------------------------------
      Total |         78       100.00
```

(5) What are the key differences between a bar chart and a histogram?

(6) Consider a random sample of 337 loan applications: 202 never defaulted on a previous loan, 97 defaulted one time on previous loan, and 38 defaulted twice on a previous loan. What is the sample standard deviation for the number of times the borrower has previously defaulted for this sample of 337 loan applications?

(7) Which histogram shows the sample with the largest sample s.d.? Smallest? Explain the intuition.
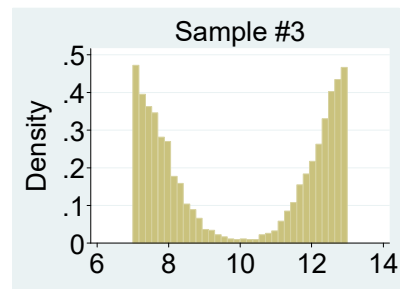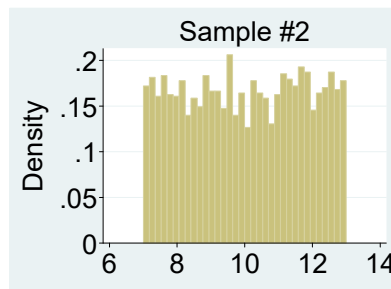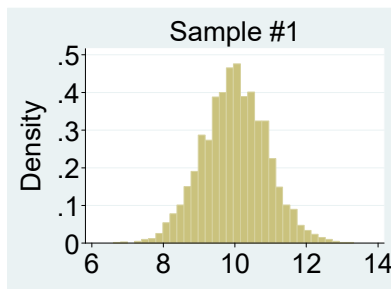

(8) Considering this histogram, could the variance of these data be 5? Explain.
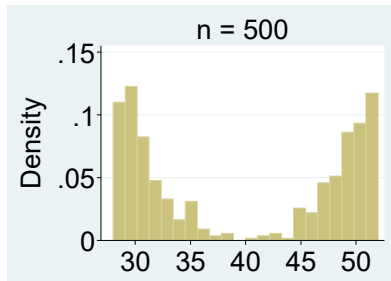
(9) The SAT is a standardized test, which most people your age in the U.S. would have taken prior to entering university. Well over 1 million students write it each year. SAT scores are one factor (along with essays, high school grades, family connections/alumni status, etc.) that affects university admissions. Top universities typically admit students with very high SAT scores whereas less selective institutions admit students even if their SAT scores are not great (and/or even if their high school grades are not great). Students get multiple SAT scores: one is the Math SAT score, where the best possible score is 800 points. Consider the Math SAT score for a random sample of 492 high school students who wrote the SAT.

(a) Given this histogram, what is the approximate mean and standard deviation? Make sure to give your reasoning and explain your answers. Include the units.
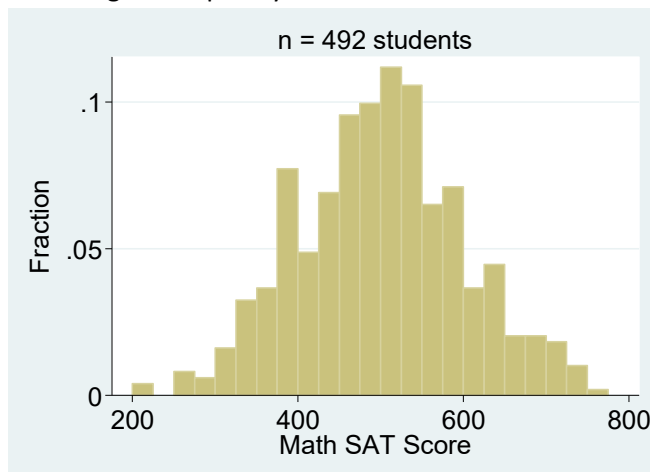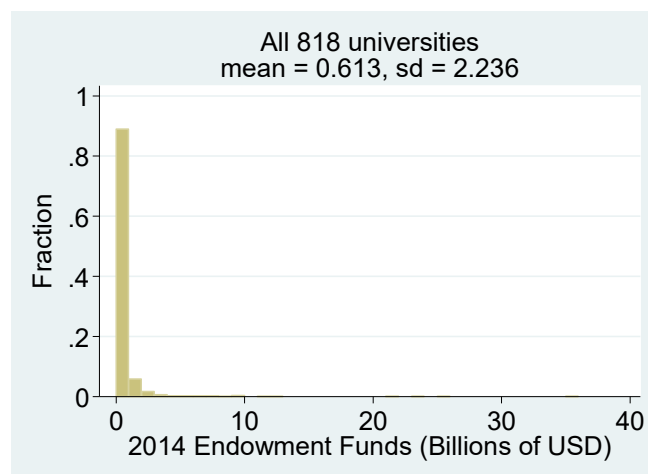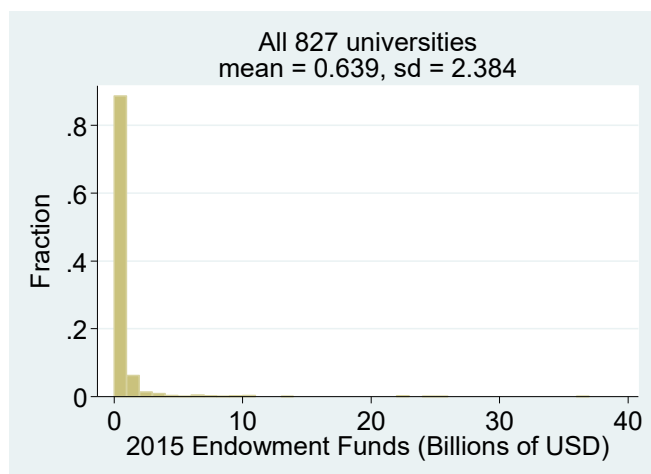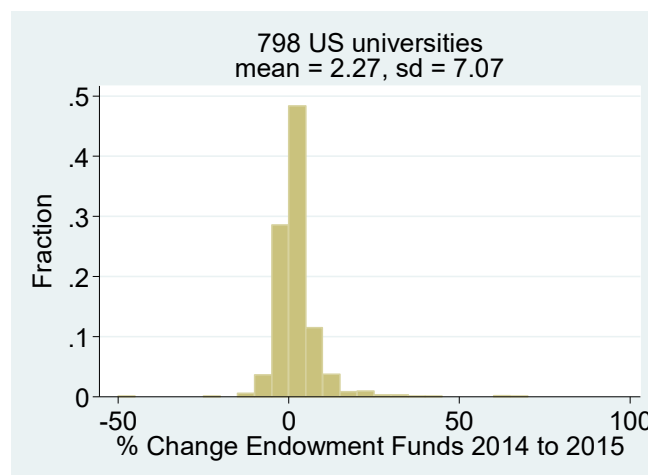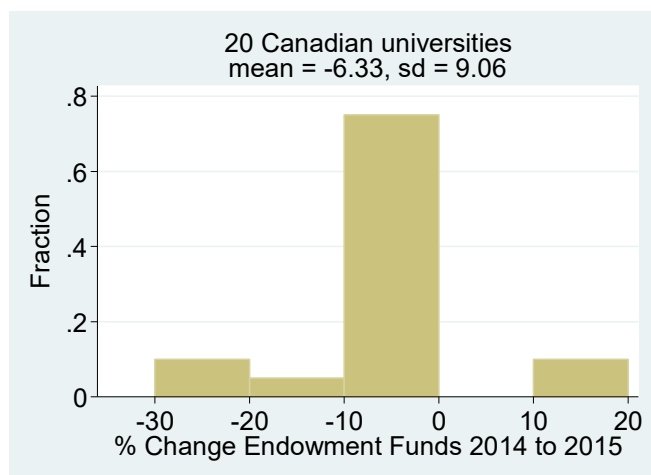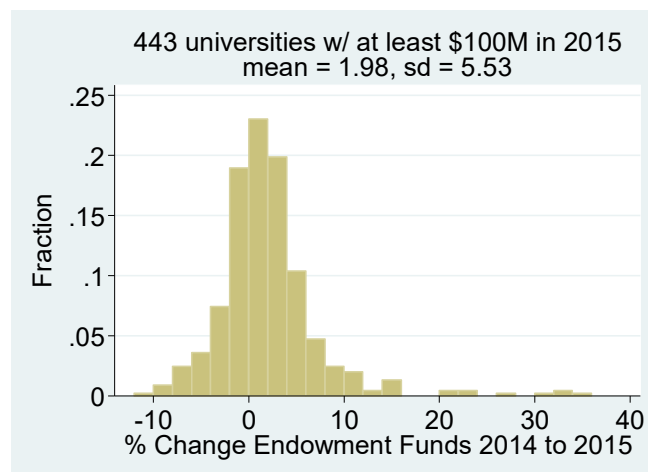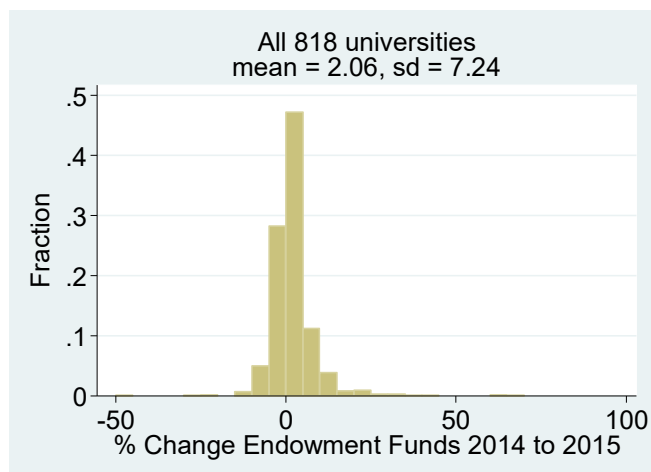

(b) If the histogram above were a *density* histogram, what would the height be of the bar with the bin that ends at 400? (That bar stands out a bit in the graph above.) Make sure to show your work.
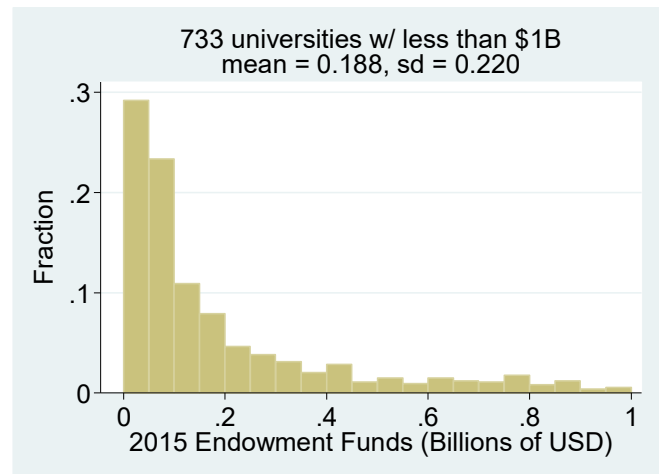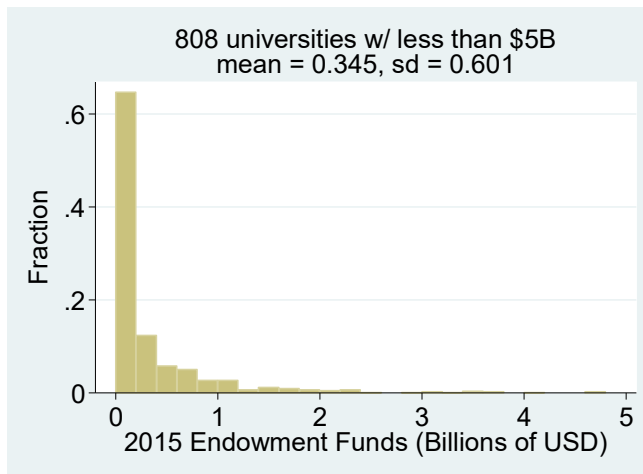
(c) Suppose that instead of a random sample of 492 high school students that wrote the SAT, you took a random sample of 50 students that were admitted into a top university (and wrote the SAT). In which ways should you expect a histogram of this second sample look different from above? Why?

(10) Consider recent data on university endowments produced by the National Association of College and University Business Officers and Commonfund Institute in a 2016 report entitled "U.S. and Canadian Institutions Listed by Fiscal Year (FY) 2015 Endowment Market Value and Change* in Endowment Market Value from FY2014 to FY2015."

(a) Browse the report making sure to see the note about the change in endowment values from 2014 to 2015: http://www.nacubo.org/Documents/EndowmentFiles/2015_NCSE_Endowment_Market_Values.pdf.

(b) Consider the eight histograms given next. Which are positively skewed? In those cases, what does that imply about the median (not reported) compared to the mean? Are any of the histograms close to Normal? Explain.

All 818 universities
mean = 2.06, sd = 7.24

443 universities w/ at least $100M in 2015
mean = 1.98, sd = 5.53

20 Canadian universities
mean = -6.33, sd = 9.06

798 US universities
mean = 2.27, sd = 7.07

All 827 universities
mean = 0.639, sd = 2.384

All 818 universities
mean = 0.613, sd = 2.236

808 universities w/ less than $5B
mean = 0.345, sd = 0.601

733 universities w/ less than $1B
mean = 0.188, sd = 0.220

Extra Problems (for those wanting more practice):

(11) Using the descriptive terms for histograms given in lecture, how would you describe this graph? Roughly how many observations are negative?

n = 205
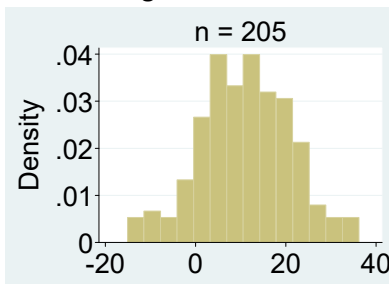
(12) Using the descriptive terms for histograms given in lecture, how would you describe this graph? (The bin width is 1.25.) Roughly how many observations are negative? Roughly how many zeros are there in the data?

n = 62
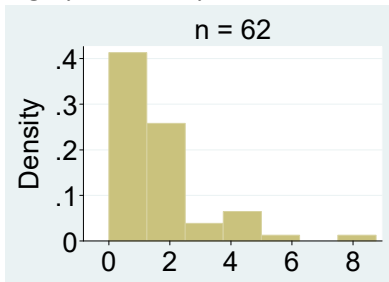
(13) Following those discussed in Lecture 2, consider the following additional example of a population and a random sample from that population.

Population
mu=1.000, med=1.000

Sample, n = 500
mean=1.017, med=1.013
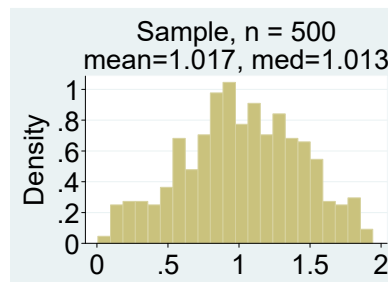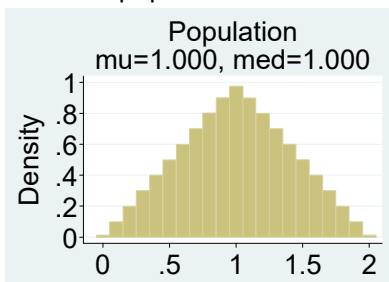
(a) Why is the population mean exactly equal to the population median? Is this true for all populations?

(b) Why isn't the sample mean exactly equal to the sample median?

(c) How would you describe the shape of the histogram on the right? If the sample size were only 50 instead of 500, how might your answer change?

(14) For this histogram, what is the *approximate* sample mean and sample standard deviation?

n = 578