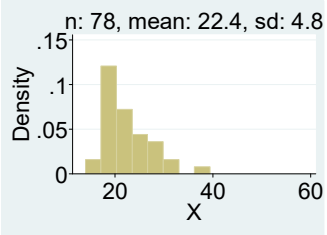
Homework 3: ECO220Y

Required Exercises: Chapter 5: 13, 15, 17, 27, 29, 50, 57, 69, 71, 73, 79

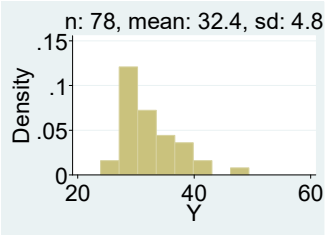
Required Problems:

(1) An Assistant Manager has evaluated employees each month for the last twelve months. Each month an employee's performance is scored from 1 to 5 where 5 is "exceptional" and 1 is "needs improvement." At year end these monthly scores are summed up and reported to the Manager. The best possible score would be 60 (=12 months * 5 points). The Manager decides that the Assistant Manager has been overly harsh in scoring the 78 employees and wishes to adjust the scores. Call X the unadjusted annual employee performance score and Y the adjusted scores.

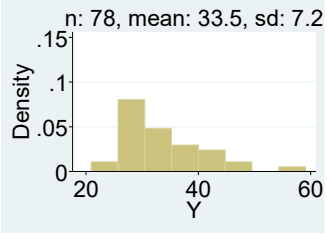


(a) Describe the shape of the unadjusted scores distribution (X).

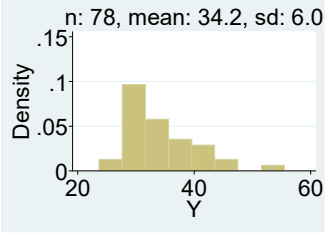
(b) Suppose the Manager adjusts these by adding 10 points to each employee's score. What is the equation that describes the adjustment? (Use X and Y notation.) Comparing the histogram of the adjusted scores (to the right) with the histogram of the unadjusted scores, how does this adjustment affect the shape of the distribution? The mean? The s.d.? In your opinion, is this fair for all of the employees?



(c) Suppose the Manager increases each employee's score by 50 percent. What is the equation that describes the adjustment? Comparing the histogram of the adjusted scores (to the right) with the histogram of the unadjusted scores, how does this adjustment affect the shape of the histogram? The mean? The s.d.? In your opinion, is this fair for all of the employees?



(d) Suppose the Manager adjusts these by adding 5 points to each employee's score and then also increasing the score by 25 percent. What is the equation that describes the adjustment? Comparing the histogram of these adjusted scores (to the right) with the histogram of the unadjusted scores, how does this adjustment affect the shape of the histogram? The mean? The s.d.? In your opinion, is this fair for all of the employees?



(e) A linear transformation is one that takes the form Y = a + bX where a and b are constants. A linear transformation does not change the shape of the histogram: if it is skewed it will remain skewed; if it is symmetric it will remain symmetric; if it is Bell shaped it will remain Bell shaped; if it is bi-modal it will remain bi-modal. Are the adjustments in (b) – (d) examples of linear transformations? Would changes in units of measurement (e.g. dollars versus thousands of dollars) also be examples of linear transformations?

(2) Consider the multiple-choice question below. Suppose 19 students answered (A), 6 answered (B), 19 answered (C), 45 answered (D), and 56 answered (E). What is the approximate mean and s.d. of the fraction of readings completed? What is the median and interquartile range (IQR)?

Multiple-choice question:
What fraction of today's required readings did you complete? (Choose the answer that best applies).
(A) none **(B)** less than half **(C)** about half **(D)** more than half **(E)** all of it

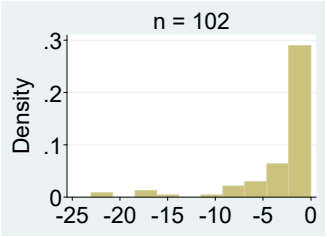
(3) Given the tabulation, find the mean, median, mode, 25th percentile, 75th percentile, IQR, and range.

x	Freq.	Percent	Cum.
0	6	3.77	3.77
.1	7	4.40	8.18
.2	16	10.06	18.24
.3	8	5.03	23.27
.4	10	6.29	29.56
.5	12	7.55	37.11
.6	12	7.55	44.65
.7	14	8.81	53.46
.8	9	5.66	59.12
.9	10	6.29	65.41
1	55	34.59	100.00
Total	159	100.00	

(4) This STATA summary is of the variable age from gasoline data posted on Professor Adonis Yatchew's website. What is unusual about these data? About how many people in the sample are 30 years old or less? How many people are 19 years old? 20 years old? Has this sample been taken from a Normal (bell shaped) population?

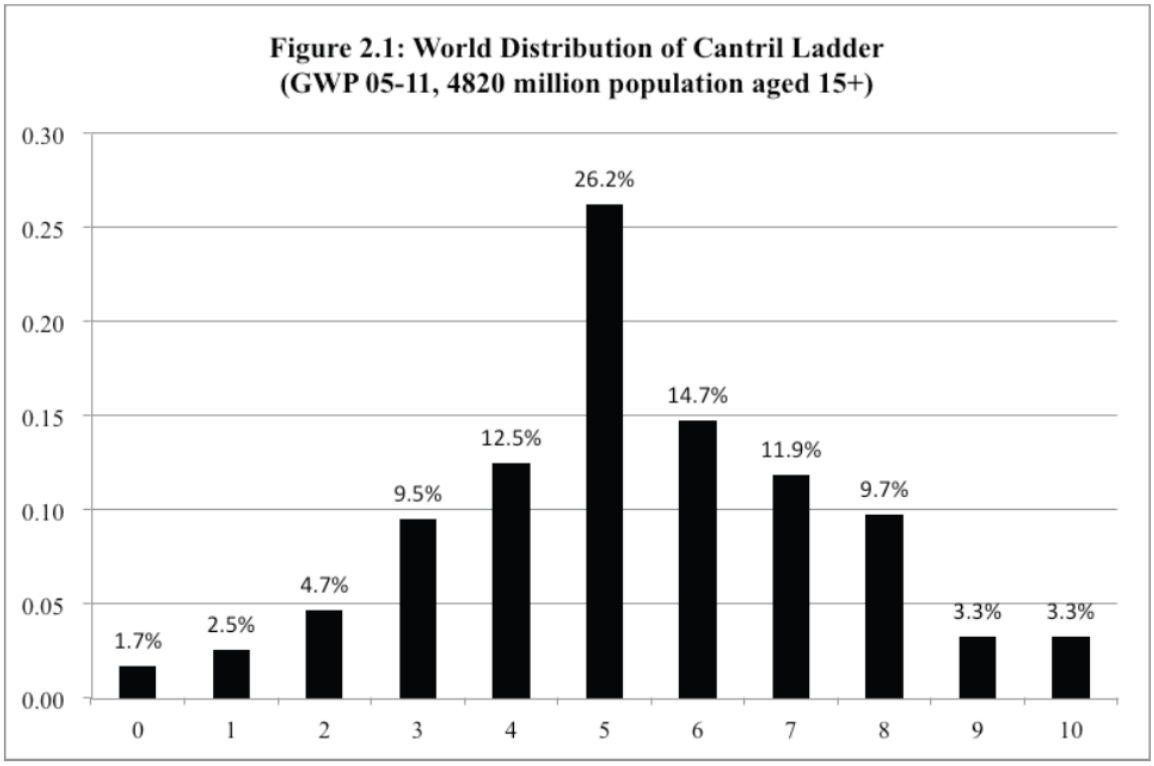
```
. summarize age, detail
                            age
-------------------------------------------------------------
      Percentiles      Smallest
 1%           24              19
 5%           29              20
10%           31              20       Obs                6491
25%           37              20       Sum of Wgt.        6491

50%           46                       Mean           47.46033
                      Largest          Std. Dev.       12.4329
75%           59              66
90%           66              66       Variance        154.577
95%           66              66       Skewness       .0843854
99%           66              66       Kurtosis       1.859655
```

(5) Give a complete qualitative description of the histogram and sketch the box plot.



(6) Consider the information below from the first "World Happiness Report." It's online (http://www.earth.columbia.edu/sitefiles/file/Sachs%20Writing/2012/World%20Happiness%20Report.pdf) but you do not need to read this report to complete this question. On page 11 it states "In the Gallup World Poll respondents are asked (using fresh annual samples of 1,000 respondents aged 15 or over in each of more than 150 countries) to evaluate the quality of their lives on an 11-point ladder scale running from 0 to 10, with the bottom rung of the ladder (0) being the worst possible life for them and 10 being the best possible." This is called the Cantril Ladder. Here is a figure showing

the distribution of world results (weighted by each country's population level). For context, the average happiness in Canada is about 7.5, U.S. 7.3, S. Korea 5.7, India 5, and China 4.7.

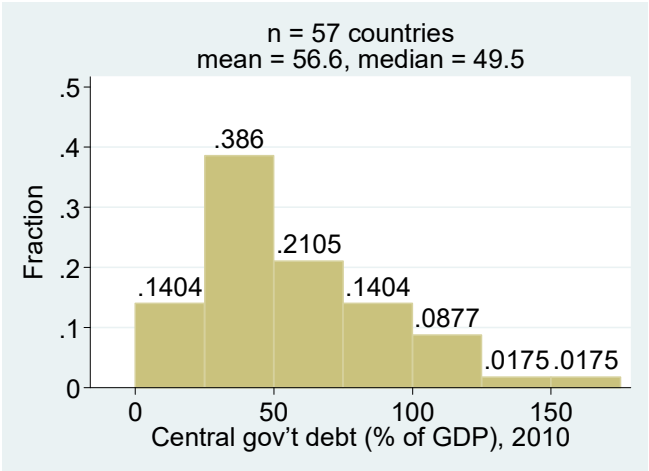


(a) Using the information in Figure 2.1, compute the mean, s.d., and interquartile range. Show your work and reasoning. [Answer with a quantitative analysis]

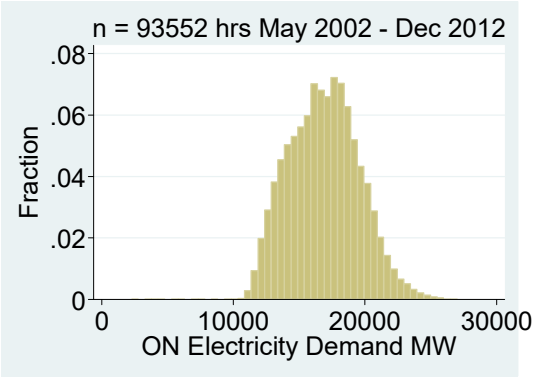
(b) Does the Empirical Rule hold for this figure? Fully explain. [Answer with a quantitative analysis and 2 – 4 sentences]

(c) Considering now the source of these data, identify *one* likely non-sampling error. Explain why you think it would be worthy of careful consideration in this particular context. [Answer with 2 – 3 sentences]

(7) Consider this summary of data download from http://data.worldbank.org/indicator/GC.DOD.TOTL.GD.ZS on May 10, 2013. Answer the following question using approximation when necessary. What is the 10th percentile? What is the 15th percentile? What is the median? What is the 75th percentile? What is the 96.5th percentile?



(8) Consider this summary of data downloaded from http://www.ieso.ca/imoweb/marketdata/marketData.asp on September 19, 2013. The variable reports Ontario's electricity usage in megawatts each hour for the period from May 2002 to December 2012. Consider both the histogram and STATA summary.




```
                          hr_elec_mw
-------------------------------------------------------------
      Percentiles      Smallest
 1%         11747           2270
 5%         12681           2365
10%         13377           2366       Obs                93552
25%         14903           2673       Sum of Wgt.        93552

50%         16889                      Mean           16901.36
                        Largest        Std. Dev.      2648.772
75%         18724          26903
90%         20346          26961       Variance        7015991
95%         21295          27000       Skewness        .171537
99%         23290          27005       Kurtosis       2.684314
```

(a) What kind of data are these: time series, cross sectional, or panel?

(b) Can we tell from the above summaries whether these data are stationary?

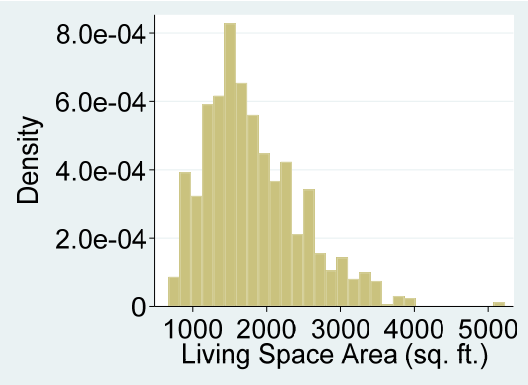
(c) Using the summaries of the electricity data above, which of these statements is FALSE?

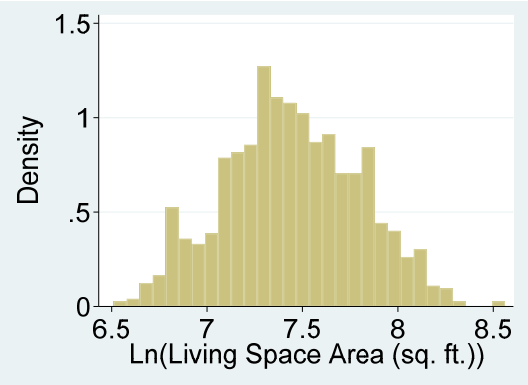
(A) there are no obvious outliers
(B) the 75th percentile is 18,724 MW
(D) the range is greater than 20,000 MW
(C) more than 50% of observations lie within 1 s.d. of the mean
(E) the mean, median, sd, range, and IQR are all measured in MW

(9) Suppose you are enrolled in a course where the instructor reports your "z-score" instead of your marks on assignments. She tells you that your z-score is your standardized test score. Suppose you have a z-score of 0, your friend Wei has a z-score of -2.15, and your friend Tina has a z-score of 1.86. What do these numbers mean? The instructor has reported that the class average is 69 percent and the class standard deviation is 12 percent. You and your friends are still getting used to the concept of a z-score and you'd like to know what your actual percentage scores are on the assignment (e.g. 80%, the cut-off for an A-). What is your percentage mark on the assignment, Wei's, and Tina's?

(10) Look at "Table 111-0002, Tax filers with charitable donations by sex and age" from Statistics Canada: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1110000201. (An older version was the source of the Lecture 3 example.) Make sure that you understand all of the numbers reported in that table.

(11) Recall Exercise 50 in Chapter 5 of our textbook. Below are the STATA summary and histograms of the original variable: living space for a sample of 1,057 homes. See also the histogram of the natural log of that variable. Additionally, see the tabulation of the living area variable for those houses larger than 3,500 square feet.





Living Area

Percentiles		Smallest		
1%	800	672		
5%	924	690		
10%	1056	720	Obs	1,057
25%	1342	728	Sum of Wgt.	1,057
50%	1675		Mean	1819.498
		Largest	Std. Dev.	662.9414
75%	2223	3981		
90%	2733	3982	Variance	439491.3
95%	3106	5114	Skewness	.9535273
99%	3720	5228	Kurtosis	4.170664

Living Area	Freq.	Percent	Cum.
3504	1	5.88	5.88
3515	1	5.88	11.76
3530	1	5.88	17.65
3535	1	5.88	23.53
3542	1	5.88	29.41
3597	1	5.88	35.29
3720	1	5.88	41.18
3721	1	5.88	47.06
3726	1	5.88	52.94
3820	1	5.88	58.82
3853	1	5.88	64.71
3944	1	5.88	70.59
3956	1	5.88	76.47
3981	1	5.88	82.35
3982	1	5.88	88.24
5114	1	5.88	94.12
5228	1	5.88	100.00
Total	17	100.00	

(a) Check that the STATA output and histogram match up with what the textbook provided in Exercise 50.

(b) Draw an *exact box plot* of living area in square feet. (Hint: Review Section 5.8 in your textbook.)

(c) Which values of the STATA summary for the natural log of living area can you predict exactly given the STATA

summary above for living area? In other words, can you say what the percentiles will be? What the 4 smallest observations will be? What the 4 largest observations will be? What the mean will be? What the standard deviation and variance will be? (Note: We ignore the entries for "Sum of Wgt.," "Skewness," and "Kurtosis.")

(d) Using the STATA summary below, draw an *exact box plot* of the natural log of living area.

ln_livingarea

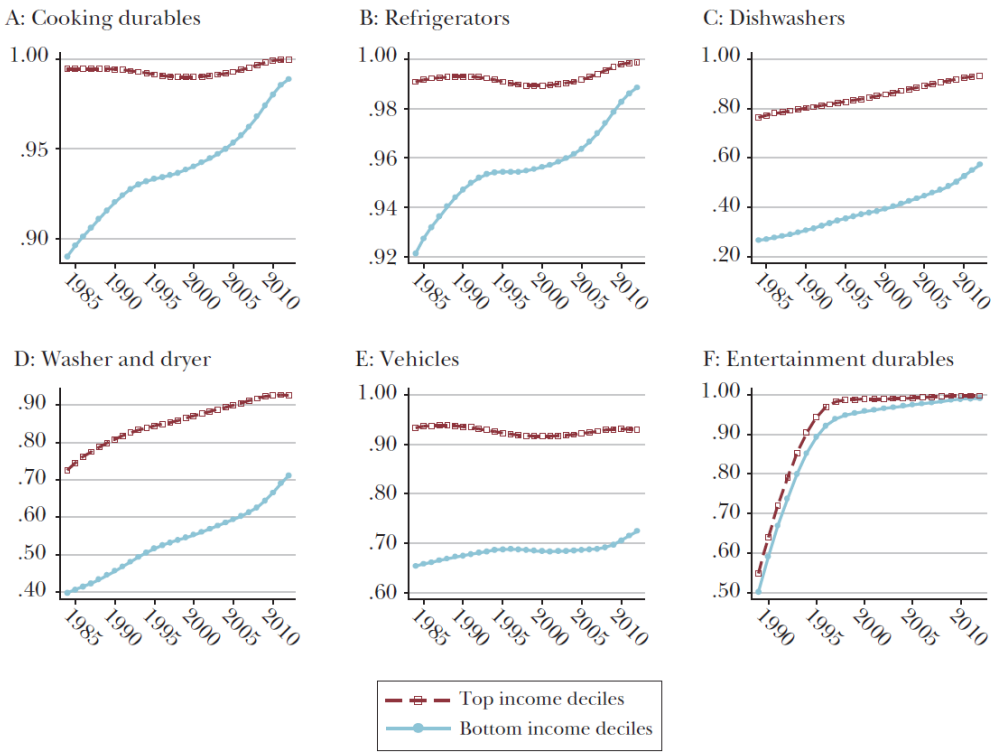
	Percentiles	Smallest		
1%	6.684612	6.510258		
5%	6.828712	6.536692		
10%	6.962244	6.579251	Obs	1,057
25%	7.201916	6.590301	Sum of Wgt.	1,057
50%	7.423568		Mean	7.443179
		Largest	Std. Dev.	.3553451
75%	7.706613	8.289289		
90%	7.913155	8.289539	Variance	.1262702
95%	8.041091	8.539737	Skewness	.0432104
99%	8.221479	8.561784	Kurtosis	2.6026

(e) Discuss why the natural log transformation changes the shape of the histogram and box plot.

(12) See the figure below from Attanasio and Pistaferri (2016) "Consumption Inequality" in the *Journal of Economic Perspectives* (https://www.aeaweb.org/articles?id=10.1257/jep.30.2.3). Researchers must concisely explain which important points readers should note in any figure or table. Interpret the results in Figure 3 in 8 – 10 sentences.

Figure 3

Share Owning Durables in the Top and Bottom Income Deciles



Source: Authors using data from the Consumer Expenditure Survey
Note: For different categories of durables, the figure compares ownership rates of the bottom and the top after-tax income deciles.