HW#1

If the Analysis Toolpak in Excel is not loaded on your computer, please load it...instructions are in the THE STATISTICAL PACKAGE IN EXCEL document posted on CANVAS.

If you wish to use the Excel graphing capability to do the histogram on 25/20b, you can look at the GRAPHING WITH EXCEL document posted on CANVAS. It is fine if you wish to do the histogram by hand, or using other software.

Look at the EXCEL STATISTICAL FUNCTIONS document posted on CANVAS.

==========

Read in textbook: pp. 1-6, 15-18, 21-22, 28-31, 35-38
Do problems (the format is page number/problem number): 25/20b, 34/34a,b, 43/45 c (just use the short-cut formula to compute the sample variance)

Read the Design of Experiments handout posted under HW#1 on CANVAS.
Do the problems from that handout: 18/1,4,5,6, 21/1, 22/4,5,6

15. Do running times of American movies differ somehow from running times of French movies? The author investigated this question by randomly selecting 25 recent movies of each type, resulting in the following running times:

Am: 94 90 95 93 128 95 125 91 104 116 162 102 90 110 92 113 116 90 97 103 95 120 109 91 138
Fr: 123 116 90 158 122 119 125 90 96 94 137 102 105 106 95 125 122 103 96 111 81 113 128 93 92

Construct a *comparative* stem-and-leaf display by listing stems in the middle of your paper and then placing the Am leaves out to the left and the Fr leaves out to the right. Then comment on interesting features of the display.

16. The article cited in Example 1.2 also gave the accompanying strength observations for cylinders:

6.1 5.8 7.8 7.1 7.2 9.2 6.6 8.3 7.0 8.3
7.8 8.1 7.4 8.5 8.9 9.8 9.7 14.1 12.6 11.2

a. Construct a comparative stem-and-leaf display (see the previous exercise) of the beam and cylinder data, and then answer the questions in parts (b)–(d) of Exercise 10 for the observations on cylinders.
b. In what ways are the two sides of the display similar? Are there any obvious differences between the beam observations and the cylinder observations?
c. Construct a dotplot of the cylinder data.

17. Temperature transducers of a certain type are shipped in batches of 50. A sample of 60 batches was selected, and the number of transducers in each batch not conforming to design specifications was determined, resulting in the following data:

2 1 2 4 0 1 3 2 0 5 3 3 1 3 2 4 7 0 2 3
0 4 2 1 3 1 1 3 4 1 2 3 2 2 8 4 5 1 3 1
5 0 2 3 2 1 0 6 4 2 1 6 0 3 3 3 6 1 2 3

a. Determine frequencies and relative frequencies for the observed values of x = number of nonconforming transducers in a batch.
b. What proportion of batches in the sample have at most five nonconforming transducers? What proportion have fewer than five? What proportion have at least five nonconforming units?
c. Draw a histogram of the data using relative frequency on the vertical scale, and comment on its features.

18. In a study of author productivity ("Lotka's Test," *Collection Mgmt.*, 1982: 111–118), a large number of authors were classified according to the number of articles they had published during a certain period. The results were presented in the accompanying frequency distribution:

Number of papers	1	2	3	4	5	6	7	8
Frequency	784	204	127	50	33	28	19	19
Number of papers	9	10	11	12	13	14	15	16	17
Frequency	6	7	6	7	4	4	5	3	3

a. Construct a histogram corresponding to this frequency distribution. What is the most interesting feature of the shape of the distribution?
b. What proportion of these authors published at least five papers? At least ten papers? More than ten papers?
c. Suppose the five 15s, three 16s, and three 17s had been lumped into a single category displayed as "≥15." Would you be able to draw a histogram? Explain.
d. Suppose that instead of the values 15, 16, and 17 being listed separately, they had been combined into a 15–17 category with frequency 11. Would you be able to draw a histogram? Explain.

19. The number of contaminating particles on a silicon wafer prior to a certain rinsing process was determined for each wafer in a sample of size 100, resulting in the following frequencies:

Number of particles	0	1	2	3	4	5	6	7
Frequency	1	2	3	12	11	15	18	10
Number of particles	8	9	10	11	12	13	14	
Frequency	12	4	5	3	1	2	1	

a. What proportion of the sampled wafers had at least one particle? At least five particles?
b. What proportion of the sampled wafers had between five and ten particles, inclusive? Strictly between five and ten particles?
c. Draw a histogram using relative frequency on the vertical axis. How would you describe the shape of the histogram?

20. The article "Determination of Most Representative Subdivision" (*J. of Energy Engr.*, 1993: 43–55) gave data on various characteristics of subdivisions that could be used in deciding whether to provide electrical power using overhead lines or underground lines. Here are the values of the variable x = total length of streets within a subdivision:

1280 5320 4390 2100 1240 3060 4770
1050 360 3330 3380 340 1000 960
1320 530 3350 540 3870 1250 2400
960 1120 2120 450 2250 2320 2400
3150 5700 5220 500 1850 2460 5850
2700 2730 1670 100 5770 3150 1890
510 240 396 1419 2109

a. Construct a stem-and-leaf display using the thousands digit as the stem and the hundreds digit as the leaf, and comment on the various features of the display.
b. Construct a histogram using class boundaries 0, 1000, 2000, 3000, 4000, 5000, and 6000. What proportion of subdivisions have total length less than 2000? Between 2000 and 4000? How would you describe the shape of the histogram?

21. The article cited in Exercise 20 also gave the following values of the variables y = number of culs-de-sac and z = number of intersections:

y 1 0 1 0 0 2 0 1 1 1 2 1 0 0 1 1 0 1 1
z 1 8 6 1 1 5 3 0 0 4 4 0 0 1 2 1 4 0 4

EXERCISES Section 1.3 (33–43)

33. The May 1, 2009 issue of *The Montclarian* reported the following home sale amounts for a sample of homes in Alameda, CA that were sold the previous month (1000s of $):

590 815 575 608 350 1285 408 540 555 679

a. Calculate and interpret the sample mean and median.
b. Suppose the 6th observation had been 985 rather than 1285. How would the mean and median change?
c. Calculate a 20% trimmed mean by first trimming the two smallest and two largest observations.
d. Calculate a 15% trimmed mean.

34. Exposure to microbial products, especially endotoxin, may have an impact on vulnerability to allergic diseases. The article "Dust Sampling Methods for Endotoxin—An Essential, But Underestimated Issue" (*Indoor Air*, 2006: 20–27) considered various issues associated with determining endotoxin concentration. The following data on concentration (EU/mg) in settled dust for one sample of urban homes and another of farm homes was kindly supplied by the authors of the cited article.

U: 6.0 5.0 11.0 33.0 4.0 5.0 80.0 18.0 35.0 17.0 23.0
F: 4.0 14.0 11.0 9.0 9.0 8.0 4.0 20.0 5.0 8.9 21.0 9.2 3.0 2.0 0.3

a. Determine the sample mean for each sample. How do they compare?
b. Determine the sample median for each sample. How do they compare? Why is the median for the urban sample so different from the mean for that sample?
c. Calculate the trimmed mean for each sample by deleting the smallest and largest observation. What are the corresponding trimming percentages? How do the values of these trimmed means compare to the corresponding means and medians?

35. The minimum injection pressure (psi) for injection molding specimens of high amylose corn was determined for eight different specimens (higher pressure corresponds to greater processing difficulty), resulting in the following observations (from "Thermoplastic Starch Blends with a Polyethylene-Co-Vinyl Alcohol: Processability and Physical Properties," *Polymer Engr. and Science*, 1994: 17–23):

15.0 13.0 18.0 14.5 12.0 11.0 8.9 8.0

a. Determine the values of the sample mean, sample median, and 12.5% trimmed mean, and compare these values.
b. By how much could the smallest sample observation, currently 8.0, be increased without affecting the value of the sample median?
c. Suppose we want the values of the sample mean and median when the observations are expressed in kilograms per square inch (ksi) rather than psi. Is it necessary to reexpress each observation in ksi, or can the values calculated in part (a) be used directly? [*Hint:* 1 kg = 2.2 lb.]

36. A sample of 26 offshore oil workers took part in a simulated escape exercise, resulting in the accompanying data on time (sec) to complete the escape ("Oxygen Consumption and Ventilation During Escape from an Offshore Platform," *Ergonomics*, 1997: 281–292):

389 356 359 363 375 424 325 394 402
373 373 370 364 366 364 325 339 393
392 369 374 359 356 403 334 397

a. Construct a stem-and-leaf display of the data. How does it suggest that the sample mean and median will compare?
b. Calculate the values of the sample mean and median. [*Hint:* $\sum x_i = 9638$.]
c. By how much could the largest time, currently 424, be increased without affecting the value of the sample median? By how much could this value be decreased without affecting the value of the sample median?
d. What are the values of $\bar{x}$ and $\tilde{x}$ when the observations are reexpressed in minutes?

37. The article "Snow Cover and Temperature Relationships in North America and Eurasia" (*J. Climate and Applied Meteorology*, 1983: 460–469) used statistical techniques to relate the amount of snow cover on each continent to average continental temperature. Data presented there included the following ten observations on October snow cover for Eurasia during the years 1970–1979 (in million km^2):

6.5 12.0 14.9 10.0 10.7 7.9 21.9 12.5 14.5 9.2

What would you report as a representative, or typical, value of October snow cover for this period, and what prompted your choice?

38. Blood pressure values are often reported to the nearest 5 mmHg (100, 105, 110, etc.). Suppose the actual blood pressure values for nine randomly selected individuals are

118.6 127.4 138.4 130.0 113.7 122.0 108.3 131.5 133.2

a. What is the median of the *reported* blood pressure values?
b. Suppose the blood pressure of the second individual is 127.6 rather than 127.4 (a small change in a single value). How does this affect the median of the reported values? What does this say about the sensitivity of the median to rounding or grouping in the data?

39. The propagation of fatigue cracks in various aircraft parts has been the subject of extensive study in recent years. The accompanying data consists of propagation lives (flight hours/10^4) to reach a given crack size in fastener holes intended for use in military aircraft ("Statistical Crack

is stretched out compared with the cancer box ($f_s = 18$ vs. $f_s = 11$), and the positions of the median lines in the two boxes show much more skewness in the middle half of the no-cancer sample than the cancer sample. Outliers are represented by horizontal line segments, and there is no distinction between mild and extreme outliers. ■

EXERCISES Section 1.4 (44–61)

44. The article "Oxygen Consumption During Fire Suppression: Error of Heart Rate Estimation" (*Ergonomics*, 1991: 1469–1474) reported the following data on oxygen consumption (mL/kg/min) for a sample of ten firefighters performing a fire-suppression simulation:

29.5 49.3 30.6 28.2 28.0 26.3 33.9 29.4 23.5 31.6

Compute the following:

a. The sample range
b. The sample variance s^2 from the definition (i.e., by first computing deviations, then squaring them, etc.)
c. The sample standard deviation
d. s^2 using the shortcut method

45. The value of Young's modulus (GPa) was determined for cast plates consisting of certain intermetallic substrates, resulting in the following sample observations ("Strength and Modulus of a Molybdenum-Coated Ti-25Al-10Nb-3U-1Mo Intermetallic," *J. of Materials Engr. and Performance*, 1997: 46–50):

116.4 115.9 114.6 115.2 115.8

a. Calculate $\bar{x}$ and the deviations from the mean.
b. Use the deviations calculated in part (a) to obtain the sample variance and the sample standard deviation.
c. Calculate s^2 by using the computational formula for the numerator S_{xx}.
d. Subtract 100 from each observation to obtain a sample of transformed values. Now calculate the sample variance of these transformed values, and compare it to s^2 for the original data.

46. The accompanying observations on stabilized viscosity (cP) for specimens of a certain grade of asphalt with 18% rubber added are from the article "Viscosity Characteristics of Rubber-Modified Asphalts" (*J. of Materials in Civil Engr.*, 1996: 153–156):

2781 2900 3013 2856 2888

a. What are the values of the sample mean and sample median?
b. Calculate the sample variance using the computational formula. [*Hint*: First subtract a convenient number from each observation.]

47. Calculate and interpret the values of the sample median, sample mean, and sample standard deviation for the following observations on fracture strength (MPa, read from a graph in "Heat-Resistant Active Brazing of Silicon Nitride: Mechanical Evaluation of Braze Joints," *Welding J.*, August, 1997):

87 93 96 98 105 114 128 131 142 168

48. Exercise 34 presented the following data on endotoxin concentration in settled dust both for a sample of urban homes and for a sample of farm homes:

U: 6.0 5.0 11.0 33.0 4.0 5.0 80.0 18.0 35.0 17.0 23.0
F: 4.0 14.0 11.0 9.0 9.0 8.0 4.0 20.0 5.0 8.9 21.0 9.2 3.0 2.0 0.3

a. Determine the value of the sample standard deviation for each sample, interpret these values, and then contrast variability in the two samples. [*Hint*: $\sum x_i = 237.0$ for the urban sample and $= 128.4$ for the farm sample, and $\sum x_i^2 = 10{,}079$ for the urban sample and 1617.94 for the farm sample.]
b. Compute the fourth spread for each sample and compare. Do the fourth spreads convey the same message about variability that the standard deviations do? Explain.
c. The authors of the cited article also provided endotoxin concentrations in dust bag dust:

U: 34.0 49.0 13.0 33.0 24.0 24.0 35.0 104.0 34.0 40.0 38.0 1.0
F: 2.0 64.0 6.0 17.0 35.0 11.0 17.0 13.0 5.0 27.0 23.0 28.0 10.0 13.0 0.2

Construct a comparative boxplot (as did the cited paper) and compare and contrast the four samples.

49. A study of the relationship between age and various visual functions (such as acuity and depth perception) reported the following observations on the area of scleral lamina (mm^2) from human optic nerve heads ("Morphometry of Nerve Fiber Bundle Pores in the Optic Nerve Head of the Human," *Experimental Eye Research*, 1988: 559–568):

2.75 2.62 2.74 3.85 2.34 2.74 3.93 4.21 3.88 4.33 3.46 4.52 2.43 3.65 2.78 3.56 3.01

a. Calculate $\sum x_i$ and $\sum x_i^2$.
b. Use the values calculated in part (a) to compute the sample variance s^2 and then the sample standard deviation s.

50. In 1997 a woman sued a computer keyboard manufacturer, charging that her repetitive stress injuries were caused by the keyboard (*Genesky v. Digital Equipment Corp.*). The injury awarded about $3.5 million for pain and suffering, but the court then set aside that award as being unreasonable