MATH1180
Homework 2
Due Friday, February 3

Please read Chapter 3 (*Tables, Graphs & Numerical Summaries*) of the Baglivo text.

Please submit solutions to the following problems. You *must* use graph paper this week, so that the graphs you draw are clear and neat. You may download a graph paper PDF file from the course website.

When submitting homework, please remember the following:
- Show all work leading to each solution.
- *Staple* all sheets together. A paper clip is not acceptable.
- Do not submit crossed-out or sloppy work.
- Do not submit ripped or torn pages.
- Be sure to submit your own work.

PROBLEM 1 (AF). Identify each of the following variables as categorical or quantitative.
(*a*) Number of children in a family.
(*b*) Amount of time in a football game before the first points are scored.
(*c*) A person's college major (English, History, Chemistry, etc.).
(*d*) A person's favorite flavor of yogurt.

PROBLEM 2 (AF). Which of the following variables are continuous (when measurements are as fine as possible) and which are discrete?
(*a*) Age of a person's mother.
(*b*) Number of pets in a family.
(*c*) The cooking time for preparing dinner.
(*d*) The latitude and longitude of a city.
(*e*) The population size of a city.

PROBLEM 3 (MM). In 1993, there were 90,523 deaths from accidents in the United States. Among these were 41,893 deaths from motor vehicle accidents; 13,141 from falls; 3,807 from drowning; 3,900 from fires; and 7,382 from poisoning.
(*a*) Find the percentage of accidental deaths from each of these causes, using 1 decimal-place accuracy. Find the percentage of accidental deaths that were due to other causes, using 1 decimal-place accuracy.
(*b*) Construct a well-labeled bar chart of the relative frequency distribution of accidental deaths in the United States in 1993. Be sure to include all categories in your chart. Describe the distribution.

PROBLEM 4 (RG). A simple random sample of Girl Scouts were asked to name their favorite Girl Scout cookie. Their responses are in Table 1 on the following page.
(*a*) Replace each "???" in the table with a frequency or relative frequency. Compute relative frequencies to 2 decimal places. Copy the table to your answers, and submit the completed table.
(*b*) Which type of cookie is the most popular?

Flavor	Frequency	Relative Frequency
S'mores	???	16.75%
Thin Mints	22	???
Samoas	61	???
Tagalongs	58	???
Shortbread	???	23.00%
Do-si-dos	83	???
Lemonades	17	4.25%
Total	???	100.00%

Table 1. Cookie Preferences

Problem 5 (RG). Here are the weights in grams of twelve randomly chosen textbooks:

107.6 654.3 371.4 640.0 468.7 686.6 349.0 425.4 710.5 591.0 345.5 844.1

Use this data to find sample 20^{th}, 40^{th}, 60^{th}, and 80^{th} percentiles of the distribution.

Problem 6 (JSE, 4(1), 1996). As part of a study on the relationship between simple body measurements and body fat, researchers computed the body mass index (BMI) of 80 male subjects. BMI is the ratio of an individual's weight in kilograms to his height in meters squared:

$$\text{BMI} = (\text{Weight in kilograms})/(\text{Height in meters})^2.$$

The following table gives the values of BMI for the 80 male subjects, arranged in increasing order:

20.1	20.3	20.4	20.6	21.0	21.3	21.3	21.5	21.5	21.6	21.6	21.6	21.7	21.8
21.9	22.1	22.2	22.3	22.8	22.9	23.0	23.1	23.3	23.4	23.6	23.6	23.6	23.7
23.8	23.9	24.2	24.3	24.4	24.4	24.9	25.0	25.1	25.1	25.4	25.5	25.6	25.6
25.6	26.1	26.1	26.2	26.3	26.3	26.4	26.6	26.7	27.3	27.4	27.4	27.5	27.5
27.6	27.6	27.9	28.1	28.3	28.5	29.0	29.2	29.2	29.3	29.7	29.7	29.7	29.7
29.8	29.8	29.9	30.1	30.3	31.1	31.8	31.8	32.2	33.9				

(a) Construct a well-labeled (density) histogram of these data using the following class intervals:

$$[20, 22), [22, 24), [24, 26), [26, 28), [28, 30), [30, 32), [32, 34].$$

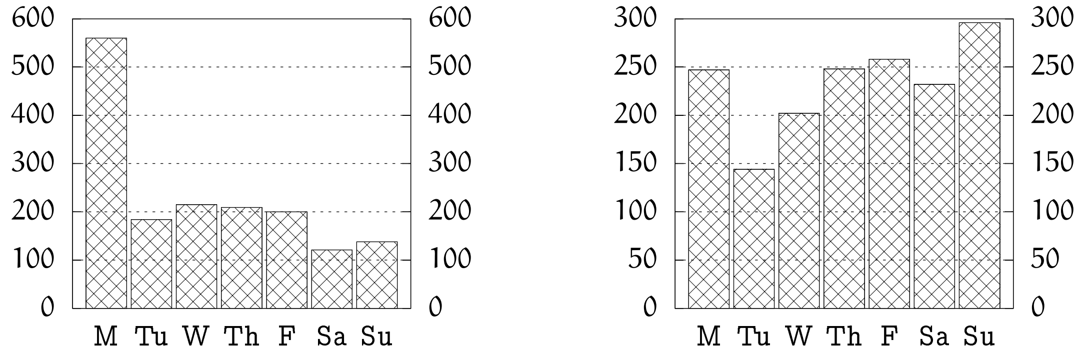
Describe the distribution.

(b) Persons with a BMI of 30 or more are considered to be obese. What proportion of the men in this study would be classified as obese?

Problem 7 (Inquiry 3:3–39). As part of a study to understand patterns in hospital use in the greater Pittsburgh area, data was collected over a six-month period. The bar charts in Figure 1 on the next page contain information about admissions and discharges by day of week during this period for all patients who were admitted for non-emergency reasons.

(a) Which day has the most admissions? What might explain this?

(b) Describe how the distribution of the day on which patients are discharged from the hospital differs from that of the day on which they are admitted. What do you think explains the difference?



Number of patients admitted by day of week

Number of patients discharged by day of week

FIGURE 1. Hospital usage in Pittsburgh, non-emergency admissions and discharges

PROBLEM 8 (MM). After examining the results of a large observational study—which seemed to suggest that calcium supplements could reduce blood pressure in adult men—researchers designed and carried out a double-blind, placebo-controlled study of calcium supplements using a sample of 21 male subjects. Ten men were randomly assigned to use a calcium supplement each day for twelve weeks. The remaining 11 men used a placebo (which looked identical to the calcium pill) each day for twelve weeks. Each man's seated systolic blood pressure in millimeters of mercury (mmHg) was measured before the treatments began and again after twelve weeks. The blood pressure distributions in the two groups should have been similar at the beginning of the experiment. Here are the initial blood pressure readings for the two groups:

Calcium Group:

107	110	123	129	112	110	107	112	136	102

Placebo Group:

123	109	112	102	98	114	119	112	110	117	130

(*a*) Construct a back-to-back stem plot of the data. Does your plot show any major differences in the groups before treatments began?

(*b*) Compute the median blood pressure level in each group. Are the medians close in value?

PROBLEM 9 (PG). The declared concentrations of nicotine in milligrams (mg) for 35 brands of Canadian cigarettes are given in the following table, arranged in increasing order:

0.09	0.10	0.20	0.30	0.40	0.40	0.50	0.89	0.90	0.90	1.00	1.00
1.00	1.00	1.10	1.10	1.10	1.10	1.20	1.20	1.20	1.20	1.20	1.20
1.20	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.30	1.40	1.40	

(*a*) Find the median and interquartile range of the nicotine data.

(*b*) Construct a box-and-whisker plot of the data and identify any outliers.

PROBLEM 10 (PG). In an investigation of the risk factors for cardiovascular disease, levels of serum cotinine—a metabolic product of nicotine—were recorded for a group of smokers and a group of nonsmokers. The relevant frequency distributions are displayed below:

Cotinine Level (ng/ml)	*Smokers*	*Nonsmokers*
0–13	78	3300
14–49	133	72
50–99	142	23
100–149	206	15
150–199	197	7
200–249	220	8
250–299	151	9
300+	412	11
Total	1539	3445

(a) Compute relative frequency distributions of serum cotinine levels for smokers and nonsmokers, using percentages with 1 decimal-place accuracy.

(b) Plot the distributions from part (a) as polygon plots on the same set of axes. Use solid line segments for the smokers and dashed line segments for the nonsmokers. Describe the shape of each distribution. What can you say about the distribution of recorded cotinine levels in each group?

(c) For all individuals in the study, smoking status was self-reported. Do you think any of the subjects might be misclassified? Why or why not?

PROBLEM 11 (PG). In Massachusetts, eight individuals experienced an unexplained episode of vitamin D intoxication that required hospitalization; it was thought that these unusual occurrences might be the result of excessive supplementation of dairy milk. Blood levels of calcium and albumin—a type of protein—for each subject at the time of hospital admission are provided below:

Calcium (mmol/l)	2.92	3.84	2.37	2.99	2.67	3.17	3.74	3.44
Albumin (g/l)	43	42	42	40	42	38	34	42

(a) Find the mean, median and standard deviation of the recorded calcium levels.

(b) Find the mean, median and standard deviation of the recorded albumin levels.

(c) For healthy individuals, the normal range of calcium values is 2.12 to 2.74 mmol/l, while the range of albumin levels is 32 to 55 g/l. Do you believe that patients suffering from vitamin D intoxication have normal blood levels of calcium and albumin? Why?