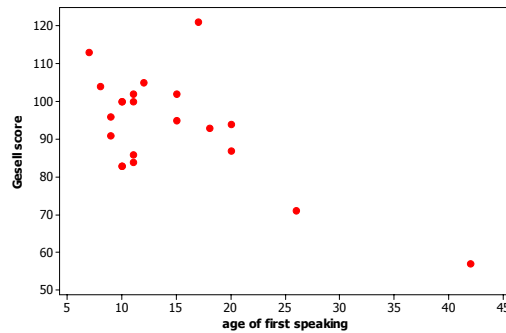


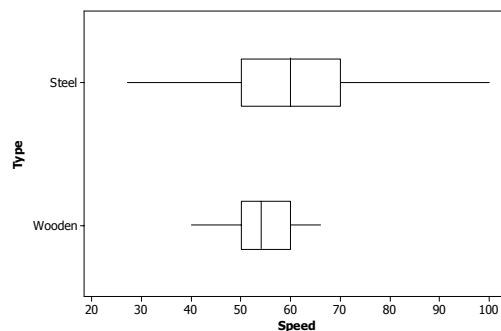
1. (8 pts) Recall the following scatterplot that we analyzed in class, with regard to the question of whether there is a relationship between the age at which a child first speaks and the child's intelligence as measured by the Gesell aptitude score:



Summarize the conclusion to be drawn from analyzing these data.

There appears to be a moderate negative association between aptitude score and age of first speaking, but this is due entirely to two unusual cases: children who took very long to speak (42 months in one case) and had very low aptitude scores. If we ignore those children and only look at those who spoke in a more typical range of ages, then there is virtually no association between aptitude score and age of first speaking

2. (20 pts, 4 pts each) The Roller Coaster Database is an on-line resource that contains lots of data on hundreds of roller coasters around the world. The following boxplots display the distributions of top speeds (in miles per hour) for coasters in the United States, classified by whether the coaster is wooden or steel:



a) What proportion of the steel coasters have a top speed of 60 miles per hour or higher? (No explanation is needed.)

50%, because the median speed is 60 mph for steel coasters.

b) What proportion of the wooden coasters have a top speed of 60 miles per hour or higher? (No explanation is needed.)

25%, because the upper quartile of the speeds is 60 mph for wooden coasters.

c) Which type of coaster (steel or wooden) has a higher proportion of coasters with a top speed above 50 miles per hour, or are they the same, or is it impossible to tell from the boxplot? (No explanation is needed.)

They are the same (25%), because the lower quartile of the speeds is 50 mph for both types of coasters.

d) Which type of coaster (steel or wooden) has more variability in its speeds, or are they the same, or is it impossible to tell from the boxplot? (No explanation is needed.)

Steel has more variability in speeds, as seen by the wider box and longer whiskers.

e) Which type of coaster (steel or wooden) has more roller coasters, or are they the same, or is it impossible to tell from the boxplot? (No explanation is needed.)

It is impossible to tell.

3. (12 pts) Suppose that the durations of human pregnancies follow a normal distribution with mean 266 days and standard deviation 16 days.

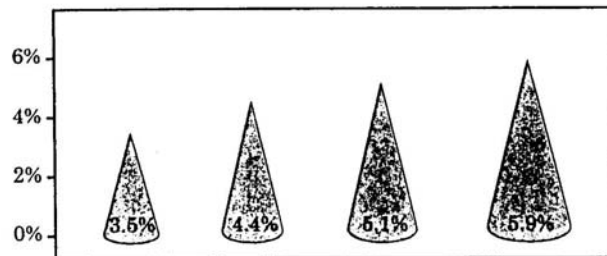
a) (8 pts) Suppose that a woman gives birth 290 days after the last visit of her husband, who has been away in the military. Assume for now that her husband was the father of the baby. What percentage of human pregnancies last for 290 or more days?

The z-score is $(290-266)/16 = 1.5$. Table B reveals that a z-score of 1.50 corresponds to the 93.32 percentile, so 6.68% of human pregnancies last longer than 290 days.

b) (4 pts) Suppose that durations of horse pregnancies follow a normal distribution with mean 336 days and standard deviation 3 days. Who is more likely to have a pregnancy that lasts for within +/- 6 days of the mean: a horse or a human? Explain briefly.

Falling within +/- 6 days of the mean is within +/- two standard deviations for the horse, so 95% of horse pregnancies will fall within that range, much larger than for a human.

4. (6 pts) The following graph compares interest rates for four different types of investments. Is this graph a fair and correct representation of these interest rates? Explain briefly. (The four percentages given are hard to read; they are 3.5%, 4.4%, 5.1%, and 5.9%.)



This is not a fair representation, because the widths as well as the heights of the cone change. Therefore, the areas of the cones are not proportional to the values being represented, and our eyes respond to the areas.

5. (12 pts) The 2001-2002 National Health and Nutrition Examination Survey asked people over the age of 60: how many close friends do you have? The answers for the 1840 respondents appear in the following table. (For example, 82 people answered that they have 0 close friends, and 13 people answered that they have 50 close friends, and nobody answered that they have 18 close friends.)

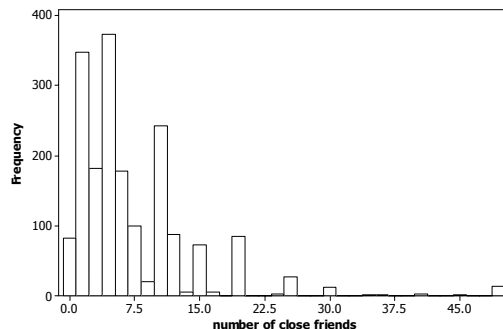
Answer	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Count	82	152	195	181	148	225	177	41	59	20	235	7	88	4

Answer	14	15	16	17	19	20	24	25	30	35	36	40	45	50
Count	2	73	3	2	1	84	3	27	12	1	1	3	1	13

a) (8 pts) Determine the five-number summary of these data.

The minimum (fewest number of close friends) is 0, and the maximum (largest number of close friends) is 50. To find the median, remember that the sample size is 1840 people, so the median is in position $(1840+1)/2$, which is the average of the 920th and 921st values. Counting the counts reveals that these values are at the value 5. Then the lower quartile is the median of the bottom 920 values, so it is the average of the 460th and 461st values, which is 3. Similarly, the upper quartile is the average of the 460th and 461st values from the top, which is the value 10. So, the five-number summary is: (0, 3, 5, 10, 50).

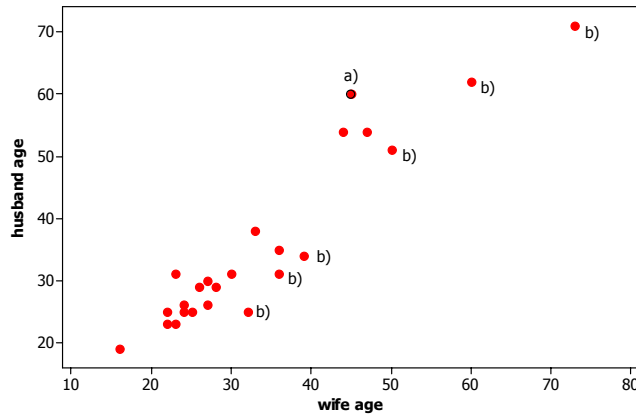
Consider the following histogram of these data:



b) (4 pts) Would you expect the mean to be larger than the median, smaller than the median, or about the same as the median? Explain briefly.

The distribution is sharply skewed to the right, so the mean will be larger than the median.

6. (12 pts) A student of mine went to the county courthouse and recorded the husband's age and wife's age from a random sample of marriage licenses. The data are displayed in the following scatterplot:



a) (4 pts) Circle the point corresponding to the couple whose husband was older than his wife by the most years.

b) (4 pts) There are six couples for which the wife is older than the husband. Put a box around the points corresponding to these couples.

c) (4 pts) Make an educated guess for the value of the correlation coefficient between husband's age and wife's age.

The association between husband's age and wife's age is positive and quite strong. The value of the correlation coefficient is .944.

7. (12 pts) The federal minimum wage is currently \$5.15 per hour.

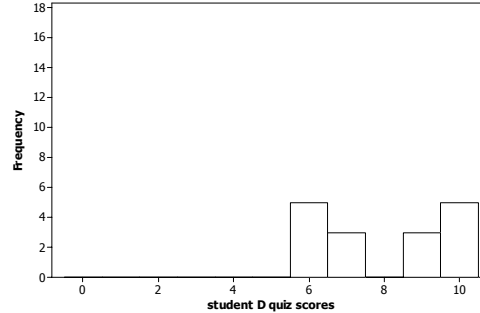
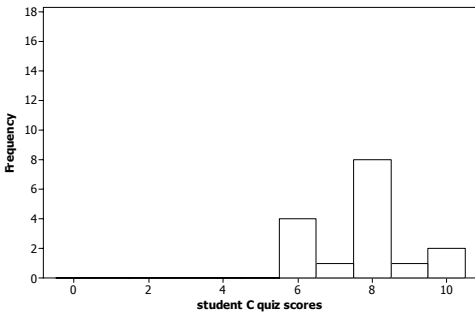
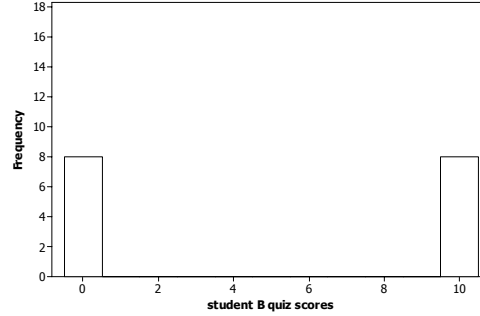
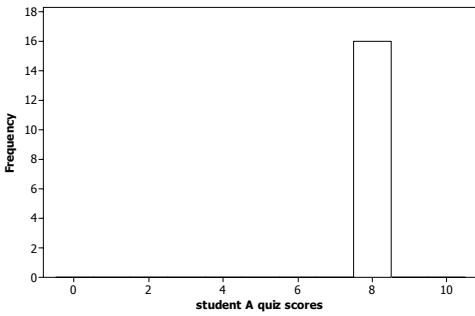
a) (6 pts) Determine what this would be in constant 1960 dollars. (Use the March 2006 CPI, which is 199.8, and the 1960 annual average CPI, which is 29.6.)

The current minimum wage of \$5.15 in 1960 dollars would be $5.15(29.6)/(199.8)$, which is about 0.763 dollars, or about 76.3 cents.

b) (6 pts) The minimum wage in 1960 (in current dollars) was \$1.00 per hour. Comment on how the real buying power of the federal minimum wage compares between then and now.

The real buying power of the minimum wage has decreased, because today's wage is worth 76 cents in 1960 dollars, but the minimum wage was actually one dollar in 1960.

8. (16 pts) You have taken 16 quizzes in this class so far. The following histograms show the distributions of these 16 quiz scores for four hypothetical students (named A, B, C, and D):



a) (8 pts) Without doing any calculations, arrange these students in order from smallest standard deviation to largest standard deviation. (No explanations are needed; just put each letter in the correct space.)

Smallest: Second smallest: Second largest: Largest:

The smallest standard deviation is with student A, because there's no variability there (that student got an 8 on every quiz). The second smallest is with student C, who usually got an 8 but sometimes differed a bit. The second largest is with student D, who often got 6 or 10. The largest standard deviation is with student B, who scores were very spread out (half zeroes and half tens).

b) (4 pts) Determine the value of the *smallest* standard deviation.

Student A has a standard deviation of zero, because every quiz score is the same.

c) (4 pts) Determine the value of the *largest* standard deviation.

The largest standard deviation is with student B. The mean score is 5, and the deviation from the mean is 5 for every quiz, and so the squared deviation is 25 for every quiz. Adding sixteen of these gives a total of $16(25) = 400$. Dividing by 15 and then taking the square root gives a standard deviation of about 5.164.

9. (2 pts) Does the United States have a single, centralized statistical agency? (No explanation is needed. Just answer "yes" or "no.")

No. (See page 323 of your text.)