

1. a) Fill in the following table (the age of the head of household in the CPS of March 1995):

Age (years)	Percent of People	Height of Histogram
0-30	20.1	
30-40	23.2	
40-50	19.3	
50-60	13.1	
60-70	11.1	
70-80	9.1	
80-90	4.1	

b) Draw and label the histogram from (a).

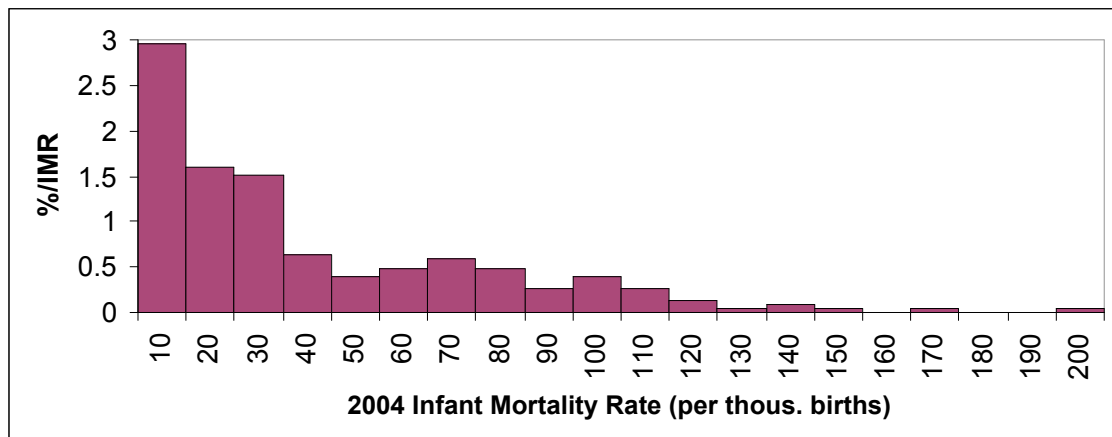
c) Estimate the percent of head's of household whose age is 40-45.

2. The poverty rate of the US states, according to the US Census Bureau, for 2001 are:

6.5, 6.7, 7.1, 7.2, 7.3, 7.4, 7.4, 7.9, 8.0, 8.1, 8.4, 8.5, 8.5, 8.7, 8.7, 8.9, 9.4, 9.4, 9.6, 9.6, 9.7, 9.7, 10.1, 10.1, 10.3, 10.5, 10.5, 10.7, 11.4, 11.5, 11.8, 12.5, 12.6, 12.6, 12.7, 12.9, 13.3, 13.8, 14.1, 14.2, 14.6, 14.9, 15.1, 15.1, 15.9, 16.2, 16.4, 17.8, 18.0, 19.3

Write down the five number summary (minimum, Q1, median, Q3, maximum).

3. For the following histogram:



a) Describe the histogram.

b) Which is bigger, the mean or median?

4. Underscoring the value of good prenatal care, new research suggests that early infection may create a cognitive vulnerability that appears later during stress on the immune system. Researchers at the University of Colorado at Boulder have reported that rats who experienced a one-time infection as newborns didn't learn as well as adult rats who were not infected as pups, after their immunity was challenged. The research is in February's Behavioral Neuroscience, published by the American Psychological Association (APA).

In the study, a team led by Staci Bilbo, PhD, injected a group of 49 rat pups on postnatal Day 4 either with the common intestinal bacteria *E. coli*, with salt water, or with nothing. In rats, Day 4 is like the third trimester in human pregnancies, a time when the brain grows significantly.

Once the pups grew up, Bilbo and her colleagues tested the adults' memory about 60 days after birth. To test memory, each rat was allowed to explore a novel experimental chamber for several minutes on the first day. On the second day, each rat was placed back into the chamber and was shocked for two seconds and then put back in its home cage. After 24 hours, the researchers again put each rat in the experimental chamber and recorded whether the rats froze -- was immobilized by fear -- or stayed active.

a) What is the possible cause and effect that the study is looking at?

b) Was this an observational study or a controlled experiment? Explain.

c) What type of variable are the possible cause and effect?

5. If the mean of a normal population is 210 with a standard deviation of 25, find:

a) $P(X > 190)$

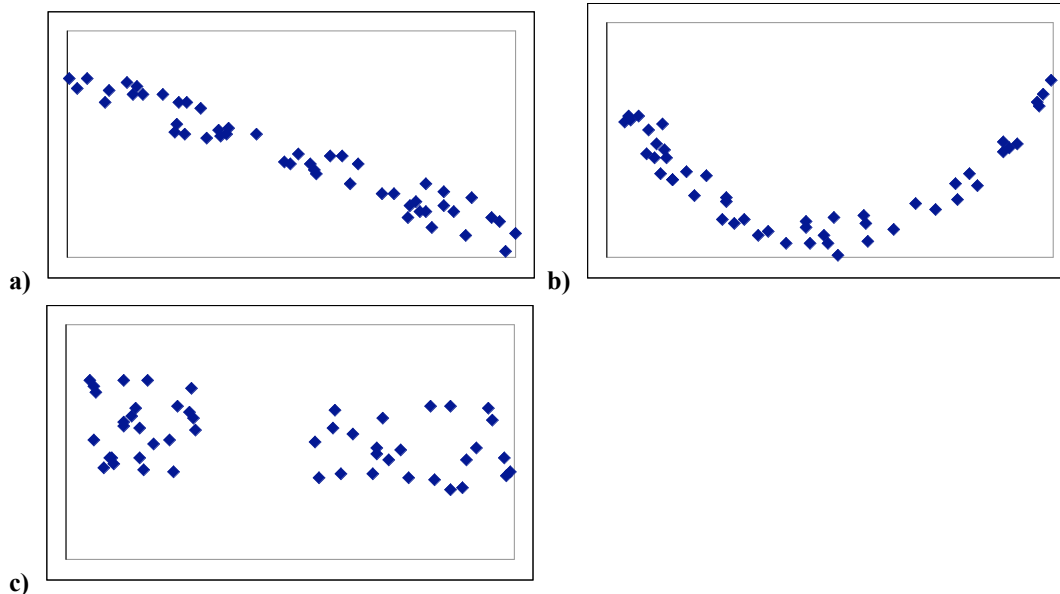
b) $P(X < 175)$

c) $P(200 < X < 260)$

d) $P(175 < X < 200)$

e) the 10th and 60th percentile.

6. Find the mean, median, and standard deviation of: 21, 15, 23, 24, and 22.
7. Suppose the correlation between the amount of education and income for men aged 25-54 is 0.44. If you calculated the average educational level and income for each state and found the correlation between the two for the states, would you expect the correlation to be more or less than 0.44? How about if we only look at men between 30 and 35?
8. An observational study finds that the correlation between two things is 0.95. Since this is a strong correlation, does it mean one variable causes the other? Explain.
9. People who are regular drinkers of alcohol are more likely to die in accidents. Whether the people are risk takers is a possible confounding factor. Explain what this means.
10. In Massachusetts, the correlation between the price of a house and its distance to Boston is negative. This means that as you get closer to Boston (distance goes down), what tends to happen to the price of houses?
11. Does the correlation coefficient summarize each of the following graphs well?



12. The current population survey from March 1995 has the following summary for age of the head of household and the number of hours they work:

Age: mean=46.76, sd=17.81
 Hours: mean=25.93, sd=22.93 $r=-.4$

- a) Find the regression line for predicting the hours of work given age.
 b) Estimate the hours worked if the head of the household is age 2.
 c) Do you think the estimate in (b) is good? Explain.
13. The Canadian Census Bureau reports that the average family income in 2000 was \$66,160 (in Canadian dollars). If the standard deviation was \$48,000 and you choose 400 people at random,
 a) find the expected value and standard error of the AVERAGE.
 b) estimate the probability that the AVERAGE of the sample is more than \$70,000.
14. Given the numbers 2, 5, 7, and 10
 a) Find the expected value and standard deviation.
 b) If you choose 500 numbers at random from this distribution, find the expected value and standard error of the SUM.

15. The 1990 US Census found that about 2% of people in the US are of Swedish descent. If you randomly chose 2000 adults then:

- a) find the expected value and standard error for the count of the sample of 2000.
- b) use the normal approximation to estimate (where X is the number of adults of Swedish descent) $P(35 < X \leq 50)$

16. A game has 10,000 tickets. One ticket wins \$9,500, while the other 9999 lose \$1.

- a) Find the expected value and standard deviation of the 'winnings' for the game.
- b) If 1,000,000 people play (with replacement), give the expected value and standard error of the total winnings.
- c) If 1,000,000 people play (with replacement), give the expected value and standard error of the average winnings.
- d) If 1,000,000 people play (with replacement), estimate the probability that the total is more than \$0.

17. Which will be more accurate: a random sample of 2% from Wyoming (population about 480,000); a sample of 1% from Maine (population about 1,200,000)? Give the sample size of both states and explain.

18. The Washington Post did a poll to find out how US adults felt about health care. The October 9-13, 2003 poll randomly selected 1000 adults from the US and found, among other things, that 83% had some form of health insurance.

- a) What is the population here?
- b) Give a statistic.
- c) Do you think this poll will be accurate? Explain.

19. A box has 100 marbles: 50 clear (numbered 1-40); 30 brown (numbered 1-30); 20 green (numbered 1-20). If you choose at random without replacement, find:

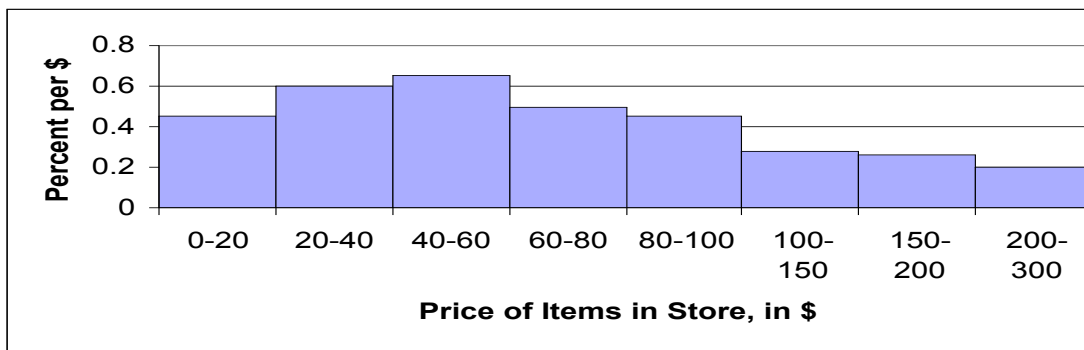
- a) if you choose one, $P(\text{brown})$.
- b) if you choose one, $P(\text{the marble is numbered 25} \mid \text{brown})$.
- c) if you choose four, $P(\text{all four are clear})$.
- d) if you choose four, $P(\text{at least one is clear})$.

20. The following is the age and gender breakdown of a town. If you choose one person at random, find:

	Age: 0-20	20-60	60 and over	
Male	110	500	240	
Female	100	510	340	

- a) $P(\text{female})$
- b) $P(20-60)$
- c) $P(\text{female} \mid 60 \text{ and over})$.
- d) $P(0-20 \mid \text{male})$.

21. Below is a poorly drawn histogram (the heights are: 0.45, 0.6, 0.65, 0.5, 0.45, 0.28, 0.26, 0.2).



The average of the price of the 1000 items is \$116.35 (with a standard deviation of \$82.42).

- a) Find the percent of items whose price is from \$100 to \$300.
- b) Is the median more or less than the mean?
- c) If the store added 100 items that cost \$120 each, what would be the average of all 1100 items?