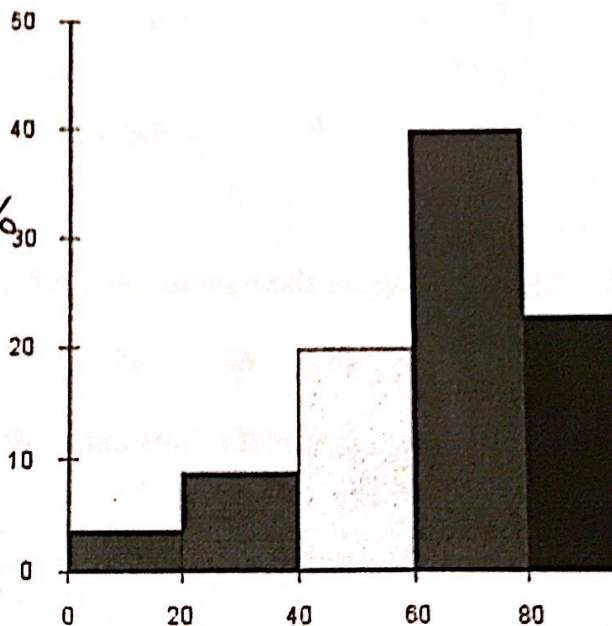


Chapter 1-2: Distributions of Data

1. The following histogram is of 95 students final exam scores.

Number of students



a) What percentage of students scored an 80 or higher?

about 24% $\frac{21}{95} \approx 22\%$

b) What percentage of students scored below a 60?

about 33% $\frac{20+8+3}{95} = \frac{31}{95} = 33\%$

c) Describe the shape of the data (including the peak).

skewed left

d) What is the median score on this final exam?

median = middle = 50%
about 60-80

e) What is the range of the data?

range scores = 100 - 0 = 100

2. Which of the following datasets is most likely to be skewed left?

a) The salaries of all National Football League players.

B b) The scores of students (out of 100 points) on a very easy exam in which most get nearly perfect scores but a few do very poorly.

c) The prices of homes in a large city.

d) The scores of students (out of 100 points) on a very difficult exam in which most get poor scores but a few do very well.

e) Amounts awarded by civil court juries.

3. State whether each of the following is an example of quantitative (Q) or categorical (C) data.

a) Height Q

e) Temperature Q

b) Age Q

f) Dog Breed C

c) Salary Q

g) GPA Q

d) Gender C

h) Color C

in the stats behind these birdfeeder visits, she collects data for 29 days. The following stem-and-leaf plot. Find the median number of birdfeeder visits.

Number of Birds Visiting a Feeder Each Day

0	1 2 3 5 6 7
1	5 7 9
2	1 1 3 4 5 5 6 6 6 6 7 7 7 8 9 9
3	0 0 5 7

Key 2 | 1 = 21

median = middle
of 29 day middle = day 15
median = 25 visits

5. The following are the number of words typed per minute for 10 different secretary applicants.

45 60 48 55 38 44 50 52 55 62

a) Create a stemplot of this data. (do this to the right side of the paper →)

b) Describe the shape of the stemplot.

approx. symmetric

stem	leaf
3	8
4	4, 5, 8
5	0, 2, 5, 5
6	0, 2

c) Find the 5# summary of this data.

min = 38 median = 51 max = 62
Q₁ = 45 Q₃ = 55

d) Find the range and IQR:

$$IQR = Q_3 - Q_1 = 55 - 45 = 10$$

$$Range = Max - Min = 62 - 38 = 24$$

e) Find the mean:

$$\bar{x} = 50.9$$

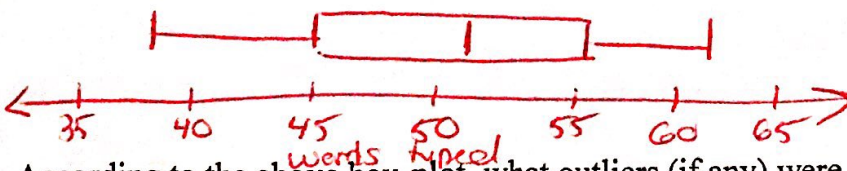
f) Find the standard deviation:

$$s = 7.45$$

$$\text{key} \Rightarrow 3 | 8 = 38$$

g) Sketch a box-plot of this data below. Label the key values.

Words Typed per minute



h) According to the above box-plot, what outliers (if any) were discovered by the calculator using the outlier rule. Show work below.

Outlier Rule:

$$(1.5)(IQR) = (1.5)(10) = 15$$

$$Q_1 - 15 = 45 - 10 = 35$$

$$Q_3 + 15 = 55 + 15 = 70$$

No outliers

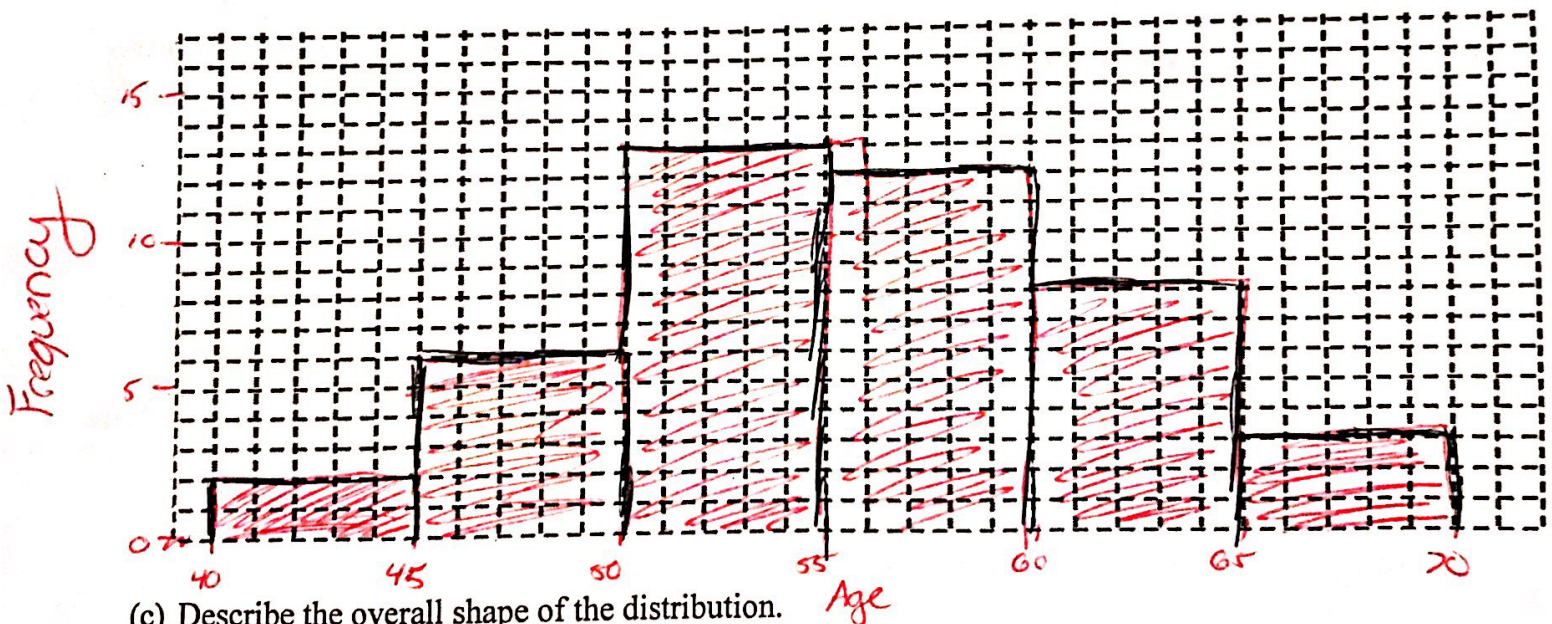
6. As of 2007, the following are the ages, in chronological order, at which U.S. presidents were inaugurated.

57, 61, 57, 57, 58, 57, 61, 54, 68, 51, 49, 64, 50, 48, 65, 52, 56, 46, 54, 49, 50, 47, 55, 55, 54, 42, 51, 56, 55, 51, 54, 51, 60, 62, 43, 55, 56, 61, 52, 69, 64, 46, 54

(a) Create a frequency distribution of the data with class widths of 5 using the table below. Start with $40 \leq \text{age} < 45$.

Class Interval	Frequency (count)
$40 \leq \text{age} < 45$	11
$45 \leq \text{age} < 50$	6
$50 \leq \text{age} < 55$	13
$55 \leq \text{age} < 60$	12
$60 \leq \text{age} < 65$	8
$65 \leq \text{age} < 70$	3
TOTAL	44

(b) Make a histogram of these data. Make sure to label your axes.



(c) Describe the overall shape of the distribution.

Symmetric

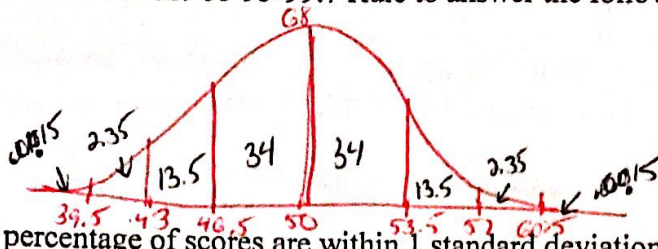
(d) Give a numerical summary of the center and spread of this distribution. Explain your choice of summary measures.

Since its symmetric, we can use the mean + standard deviation.

$\bar{x} = 54.7$ center $s_x = 6.4$ spread

Chapter 3: Normal Distributions 99.7

7. Scores from a standardized math test are normally distributed with $\mu = 50$ and $\sigma = 3.5$. Sketch the curve below and use the 68-95-99.7 Rule to answer the following questions.



a) What percentage of scores are within 1 standard deviation of the mean?

68%

b) Find the percentage of scores that are between 43 and 57.

95%

c) Find the percentage of scores that are greater than 53.5.

16%

d) Find the percentage of scores that are less than 39.5.

.15%

e) Find the percentage of scores that are greater than 46.5.

84%

Consider data on the standard normal curve. Sketch curves and show calculator instructions to answer #8 - 15.

8. $P(Z < 2.67) = .996$

9. $P(Z < 0) = .50$

normal cdf = lower = -4
upper = 2.67
 $\bar{x} = 0$
 $s = 1$



10. $P(Z < -3.49) = 0$

11. $P(Z > 1.52) = .064$

lower = -4
upper = -3.49
 $\bar{x} = 0$
 $s = 1$

lower = 1.52
upper = 4
 $\bar{x} = 0$
 $s = 1$

12. $P(Z > -0.28) = .610$

13. $P(1.09 < Z < 2.33) = .128$

lower = -.28
upper = 4
 $\bar{x} = 0$
 $s = 1$

lower = 1.09
upper = 2.33
 $\bar{x} = 0$
 $s = 1$

14. Find the z-score that is greater than approximately 70% of the distribution.

Invers Norm: area = .3 $z = -.524$
 $\mu = 0$
 $\sigma = 1$

15. Find the z-score that is the 45th percentile of the standard normal distribution.

Invers norm : area = 45
 $\mu = 0$ $\sigma = 1$
 $z = -0.126$

The lengths of a population of fish is normally distributed with $\mu = 4.8$ inches and $\sigma = 0.6$ inches. Sketch curves and show calculator instructions to answer #26 - 30.

16. The percent of fish less than 4 in is:

$P(x < 4)$ normalcdf = lower = 0 9.12%
 upper = 4
 $\mu = 4.8$
 $\sigma = .6$

17. The percent of fish less than 3.5 in is:

$P(x < 3.5) =$ lower : 0 1.51%
 upper : 4
 $\mu = 4.8$
 $\sigma = .6$

18. The percent of fish greater than 5.7 in long is:

$P(x > 5.7) \Rightarrow$ lower = 5.7 6.68%
 upper = 100
 $\mu = 4.8$
 $\sigma = .6$

19. The percent of fish greater than 4.5 in long is:

$P(x > 4.5)$ lower = 4.5 69.1%
 upper = 100
 $\mu = 4.8$
 $\sigma = .6$

20. The percent of fish between 3.9 cm and 5.2 in long is:

$P(3.9 < x < 5.2)$ lower = 3.9 68.1%
 upper = 5.2 $\mu = 4.8$
 $\sigma = .6$

21. The range of possible scores on a GRE (Graduate Record Exam) is 200 to 900. A university finds that the scores of its applicants on the GRE are approximately Normal with mean = 544 and standard deviation = 103.

(a) Use the 68-95-99.7 rule to approximate the proportion of applicants that scored higher than 441.

$544 - 103 = 441$ $50\% + 34\% = 84\%$



(b) What percent of applicants had scores between 500 and 700? Show your work.

60.0%

(c) What minimum score would a student need in order to score better than 77% of those taking the test? Show your work.

620.1

Chapter 4: Scatterplots and LSRL

22. Make a scatterplot of the following data on your calculator. Know how to make one by hand as well.

Hours in Mall	10	8	9	3	1	2	5	6	7	8	2	3
Dollars spent	108	73	87	20	10	35	50	54	68	92	15	35

Describe the scatterplot. (Overall pattern, direction, form, and strength)

Form/Pattern: Linear

Direction: Positive

Strength: Relatively strong

a) Find the least squares regression line (LSRL). $\hat{y} = 10.1x + 0.0398$

c) What does the slope and intercept of this equation represent?

Slope = The increase in \$ spent per hour at the mall

y-int = Amount of money spent w/ hours at the mall

d) What is the correlation coefficient and the coefficient of determination (r^2).

$$r = .969$$

$$r^2 = .940$$

What does it tell you about your least squares regression line?

A strong correlation between \$ and hours spent at the mall.

e) Use your LSRL to predict the amount of money a person would spend if he or she spent 4 hours at the mall.

$$\hat{y} = 10.1x + 0.0398$$

$$x = 4$$

$$\hat{y} = 10.1(4) + 0.0398 = \$40.44$$

Sampling and Experimentation: Planning and Conducting a Study

Check notes for definitions

1. Define these terms:

- a. Census
- b. Population
- c. Sample
- d. Survey
- e. Simple Random Sample (SRS)
- f. Bias in a sample
- g. Confounding
- h. Nonresponse
- i. Stratified random sample
- j. Cluster Sample
- k. Block design
- l. Match Pair Design
- m. Experiment
- n. Double-Blind Experiment
- o. Observational study

2. The Ministry of Health in the Canadian Province of Ontario wants to know whether the national health care system is achieving its goals in the province. Much information about health care comes from patient records but that source doesn't allow us to compare people who use health services with those who don't. So the Ministry of Health conducted the Ontario Health Survey, which interviewed a random sample of 61,239 people who in the Province of Ontario.

a. What is the population for this sample survey? What is the sample?

Population: Residents of Canada

Sample: 61,239 people interviewed in the Province

b. The survey found the 76% of males and 86% of females in the sample had visited a general practitioner at least once in the past year. Do you think these estimates are close to the truth about the entire population? Why or why not?

Although the statistics came from an SRS, the sample was only drawn from the Province of Ontario. This Province is populated and may have greater access to healthcare. Therefore, this may not represent the population.

c. Is this an experiment or an observation study? How can you tell?

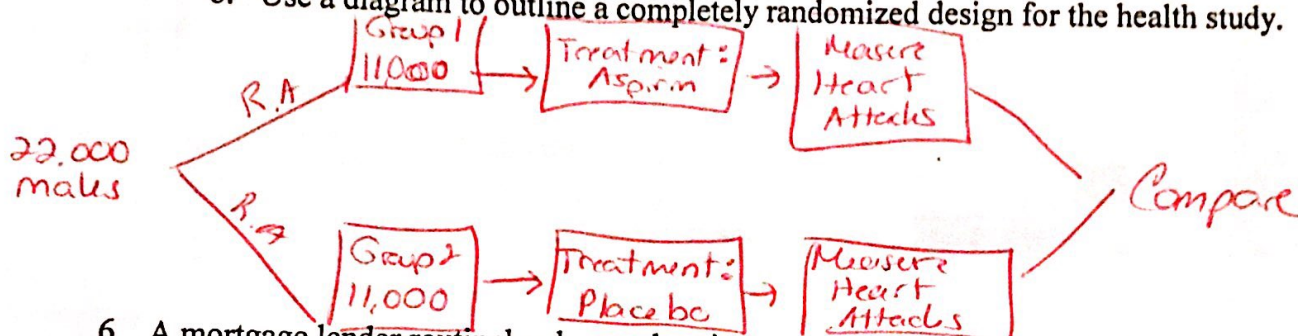
Observational study - no treatment

5. Can aspirin help prevent heart attacks? The Physicians' Health Study, a large medical experiment involving 22,000 male physicians, attempted to answer this question. One group of about 11,000 physicians took an aspirin every second day, while the rest took a placebo. After several years the study found that subjects in the aspirin group had significantly fewer heart attacks than the subjects in the placebo group.

- a. Identify the subjects, the explanatory, and the response variable in the health study.

Subjects = 22,000 male physicians
 Explanatory = aspirin daily
 Response = # heart attacks

- b. Use a diagram to outline a completely randomized design for the health study.



6. A mortgage lender routinely places advertisements in a local newspaper. The advertisements are of three different types: one focusing on low interest rates, one featuring low fees for first-time buyers, and one appealing to people who may want to refinance their homes. The lender would like to determine which advertisement format is most successful in attracting customers to call for more information. Describe an experiment that would provide the information needed to make this determination. Be sure to consider extraneous factors such as the day of the week that the advertisement appears in the paper, the section of the paper in which the advertisement appear, daily fluctuations of the interest rate and so forth. What role does randomization play in your design?

We would collect information by asking subjects to watch one of the three advertisements, we would randomly assign people into each of the advertisement groups + determine the number that would return.

7. Use of nicotine replacement therapies and the antidepressant bupropion helps people stop smoking. A double-blind, placebo-controlled experiment was conducted of sustained-release bupropion (244 subjects), a nicotine patch (244 subjects), bupropion and a nicotine patch (245 subjects), and placebo (160 subjects) for smoking cessation.

Results: The abstinence rates at 12 months were 15.6 percent in the placebo group, as compared with 16.4 percent in the nicotine patch group, 30.3 percent in the bupropion group, and 35.5 percent in the group given bupropion and the nicotine patch.

- a. How many treatments did this experiment compare?

4 treatments

- b. What is the response variable?

% that quit smoking

- c. One group received a placebo. Why not just give this group no treatment at all?

We don't want the placebo affect to cause them to think that they still have the desire to smoke.