

Exploratory Data Analysis**Chapter 1**

1. 20 randomly chosen people in San Diego were asked how many movies they saw last month. The results are in the table below:

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

- Find the mean and median.
- Based on the values you get in (a)**, what can you say about the shape of the distribution? Explain briefly.
- Now find the standard deviation and IQR.
- The persons observing 5 movies would be in what percentile of the sample
- Draw a modified boxplot of the data set.
- Show that the person going to 9 movies is an outlier.
- Another 20 people, this time in Big Bear were asked how many movies they saw last month. The results are in the table below:

4	9	2	8	6	2	6	7	8	7
5	6	8	2	5	6	7	4	9	5

Compare this study with the first study in San Diego. Make sure you compare the boxplots and all the numerical summaries.

Chapter 2

2. The average cost of an ice cream cone on the Ocean City boardwalk is \$1.97 with a standard deviation of 0.32.

- What percent of the ice cream prices are above \$2.25?
- What percentage of the ice cream prices are below \$1.50
- What percentage of the ice cream prices are between \$1 and \$2?
- What price (nearest cent) would an ice cream cone have to have to be in the bottom 2% of prices?
- What price (nearest cent) would an ice cream cone have to have to be in the 99.9th percentile?

3. Jay, James, and Joe each drive different types of cars. The price they paid for their cars are in the table below along with the mean and standard deviation of the type of car they drive. Show necessary work to determine who paid the most and least amount for their cars relative to their type. Assume that prices for each type is normally distributed.

	Jay	James	Joe
Type	Compact	SUV	Luxury
Price	\$24,857	\$37,925	\$55,290
μ	\$21,335	\$32,853	\$48,936
σ	\$2,455	\$4,065	\$5,654

Chapter 3

4. Prices of digital cameras are believed to be related to the number of megapixels the camera has. Following is a chart of the average camera price for that number of megapixels.

Megapixels: 1.5 2 3 5 6 7
Avg price: \$99 \$129 \$236 \$329 \$360 \$525

- Draw the scatterplot of the data. Describe the relationship between the number of megapixels and average price of the digital cameras.
- Find the least-squares regression line for this data and draw it on the graph.
- Explain the meaning of the slope in the LSRL for this data, in context.
- Compute the residual (nearest dollar) for a 5 megapixel camera.
- Predict the price for a 4 megapixel camera to the nearest dollar.
- Find the value of r (*correlation coefficient*) and explain its significance in the context of the problem.
- Find the value of r^2 (*coefficient of determination*) and explain its significance in the context of the problem.
- Every least-squares regression line must pass through a certain point. Find this point for this data.

- i. Is there evidence that the least-squares line is a good model for this data? Explain.
 j. Predict the average price for a 10-megapixel camera and your confidence in the truth of this value.

Study Design Chapter 5

5. A restaurant reviewer wishes to check out the cuisine in the Lansdale area. His population is 72 restaurants that are Chinese (C), Italian (I), and Mexican (M). The list is alphabetically as follows. Chain restaurants are marked with a star (*).

1	Abacus	C	19	El Sarape	M	37	Laura's	I	55	Ray's	I
2	Allegro	I	20	Fasagos	I	38	Lisa's	I	56	Rising Sun	C
3	Bacco	I	21	Frank's	I	39	Los Tacos	M	57	Sbarro's*	I
4	Baja Fresh	M	22	Fratellis	I	40	Lotus Inn	C	58	Soncini's	I
5	Boston Style*	I	23	Georgina's	M	41	Luca's	I	59	Sophie's	I
6	Brother's*	I	24	Giuseppe's*	I	42	Lu's	C	60	Spatola's*	I
7	China Garden	C	25	Golden Chopsticks	C	43	Macaroni Grille*	I	61	Star Buffet	C
8	China Moon	C	26	Golden City	C	44	Mama's	I	62	Sushi Express	C
9	China Wok	C	27	Golden Sea	C	45	Mexican Grille	M	63	Taco Bell*	M
10	Chinese Kitchen	C	28	Gourmet Wok	C	46	New Station	I	64	Tacqueria	M
11	Chong's	C	29	Great Wall	C	47	Nino's	I	65	Tamarindos	M
12	Chuck E. Cheese*	I	30	Habaneros	M	48	Palermo's	I	66	Tex Mex	M
13	Cosimo's	I	31	Hong Kong Chef	C	49	Panda garden	C	67	Tortillas	M
14	Country Garden	C	32	Hunan Best	C	50	Papa John's*	I	68	Uno*	I
15	Domino's*	I	33	Imperial palace	C	51	Pizza Hut*	I	69	Via macaroni	I
16	Domi's	I	34	Italia	I	52	Pizza Plaza	I	70	Vinny's	I
17	Eastern Dragon	C	35	Italian Delite	I	53	Pizza Time	I	71	Wok and Roll	C
18	El Cactus	M	36	Kowloon	C	54	Portofino	I	72	Yantze	C

He decides to review 12 restaurants. Using row 128 of your random digit chart choose an SRS of 12 restaurants and write their names in the chart below (complete the chart down, then across). Then determine how many Chinese, Italian, and Mexican restaurants were chosen.

The reviewer decides to stratify his choices by type of restaurants and pick 4 restaurants from each group. Use line 108 for the Chinese restaurants, line 140 for the Italian restaurants, and line 118 for the Mexican restaurants, and place your answer in the chart below.

Chinese	Italian	Mexican

- a. Explain in a sentence or two the advantage of choosing a stratified sample rather than an SRS.
 b. Explain why stratifying with 4 restaurants from each group is not an accurate representation of the restaurants in the Lansdale area.
 c. If the number of restaurants from each group were to be chosen based on the ratio of Chinese, Italian, and Mexican restaurants in the population. How many would you choose from each category?

6. A medical study of heart surgery a drug called a beta-blocker will reduce the pulse rate of a patient during surgery. The pulse rate will be measured at a specific point during the operation. The investigators will use as subjects 10 patients during heart surgery. You have a list of these patients in alphabetical order.
- Write a description for a randomized experimental design for this study.
 - Give one lurking variable that may exist in your design.

7. A suburban school has 4 classes, freshman, sophomores, juniors, and seniors. The administration wants to compare the quality and taste of the current company that provides pizza at lunch (A) with another company (B). It is interested in which pizza the students like the most, A or B. The population of the school is 1,000 students with each class having roughly the same number of students. A stratified sample of 40 students will be chosen to taste the pizza (plain pizza only) with each class having equal representation.

- Decide exactly how you will label the students in order to select your sample. Since you are going to be using a table of random numbers, think this through first.
- Use Table B, beginning at line 130, to select the first 6 students in the sample and write in the parentheses whether the student is a freshman (F), Sophomore (So), Junior (J), or Senior (Sr)

_____ (), _____ (), _____ (), _____ (), _____ (), _____ ()

- Diagram how the following experiments would be set up.
 - completely randomized
 - block design with 2 blocks (lower grades 9 and 10) and upper grades (11 and 12)
 - matched pairs (students acting as their own control)

d) A decision is made to go with the matched pairs design. Assuming that the experimenters have taken reasonable precautions, describe a lurking variable that could present itself in trying to identify what pizza (A or B) the students prefer.

Probability Chapters 6

8. 4 fair coins are tossed. Generate a sample space

Find the following probabilities

- There are exactly 2 heads
- There are at least 2 heads
- There are no heads
- There is at least one head

9. Two fair 4-sided dice (with the numbers 1-2-4-6 on them) are rolled and added. Generate a sample space.

Find the following probabilities

- The sum is even
- The sum is greater than 4
- If the sum is even, it is greater than 4
- If the sum is greater than 4, it is even

10. Suppose you spin a “wheel of fortune” with 16 equally spaced slots as shown below.

a. The wheel is spun once and the number that comes under the arrow (X) is observed. Generate a probability distribution table below. **Answers should be in fraction form.**

X	1	2	3	4	5	6	7
P(X):							

- Let A be the event: The number that comes out is divisible by 2. List the outcomes in A and find $P(A)$.
- Let B be the event: The number that comes out is divisible by 3. List the outcomes in B and find $P(B)$.
- Are the events A and B disjoint? Explain.
- Show work using the formal definition of independence to determine if events A and B are independent.

11. 300 women were randomly chosen and asked if they were married and if they were Pro-Choice or Pro-Life. The table shows the data:

	Pro Choice	Pro-Life	Total
Married	90	150	240
Single	42	18	60
Total	132	168	300

If a woman is chosen at random, find the probabilities.

- Given that she is married, she is pro-choice
- Given she is married, she is pro-life
- Given she is single, she is pro-choice
- Given she is single, she is pro-life
- Given she is pro-choice, she is married
- Given she is pro-life, she is single
- Are being married and being pro-choice independent? Explain.

12. A company has developed a drug test to detect steroid use by athletes. The test is accurate 95% of the time when an athlete has taken steroids. It is 97% accurate when an athlete has not taken steroids. Suppose that the drug test will be used in a population of athletes in which 10% have actually taken steroids. Let's choose an athlete at random and administer the drug test.

- Make a tree diagram showing the sample space of this chance process.
- What is the probability that the randomly selected athlete tests positive?
- What is the probability that an athlete who tests positive actually used steroids?

Chapter 7

13. X.R. Cise spends a lot of time in a health facility and likes to work out on an elliptical trainer. The time he daily spends on the trainer is different depending on how much a hurry he is in, but he either spends 15 minutes, 30 minutes, 40 minutes, 45 minutes, 50 minutes, or an hour. The probabilities of spending each amount of time respectively are 10%, 24%, 15%, 40%, 3%, and 8%.

- If X represents the amount of time spent on the elliptical trainer find the mean of the random variable X . Show how you got your answer.
- Find the standard deviation of X . Again, show how you got your answer.
- Assume that each minute on the elliptical burns 18 calories. Also assume that after he trains on the elliptical, he then lifts weights which burn 250 calories.

i) Find the mean of the daily calories burned

ii) Find the standard deviation of the daily calories burned

14. Here's a game. If a player rolls 2 dice and gets a sum of 2 or 12, he wins \$20. If the person gets a 7, he wins \$5. Otherwise he wins nothing. Do the following. Show how you get your answers.

- Find the expected payout for the game
- if the cost to play the game is \$3, what does a player expect to get out of this game every time he plays? Is the game fair? Explain.

15. Vista Murrieta High School has the best women's swimming team in the region. The 400-meter freestyle relay team is undefeated this year. In the 400-meter freestyle relay, each swimmer swims 100 meters. The times, in seconds, for the four swimmers this season is approximately normally distributed with means and standard deviations as shown.

- Assuming the swimmers individual times are independent, find the probability that the total team time in the 400-meter freestyle relay is less than 220 seconds (the time to beat in the championships).

Swimmer	Mean	SD
1	55.2	2.8
2	58.0	3.0
3	56.3	2.6
4	54.7	2.7

Chapter 8

16. *Ladies Home Journal* magazine reported that 66% of all dog owners greet their dog before greeting their spouse or children when they return home at the end of the workday. Assume that this claim is true. Suppose 12 dog owners are selected at random,. Let X = the number of owners who greet their dogs first.

- Explain why X is a binomial random variable.
- Only 4 of the owners in the random sample greeted their dogs first. Does this give convincing evidence against the *Ladies Home Journal* claim? Calculate an appropriate probability to support your answer.
- What would be the probability of at least 5 owners said they greeted their dog first?
- If you do another sample, what is the probability that it will take 4 dog owners before you get the first one to say they greet their dog first?