Chapter 2

2.1

An auto company is manufacturing a certain car part that should weigh precisely 3.000 grams. Over the past four days, 30 car parts have been randomly selected and weighed. Here is the ordered data in grams:

2.971	2.974	2.975	2.978	2.980	2.982
2.983	2.987	2.991	2.994	2.995	2.996
2.996	3.000	3.004	3.004	3.007	3.009
3.012	3.013	3.013	3.015	3.023	3.026
3.028	3.028	3.031	3.032	3.033	3.035

First draw a histogram for this data set of n = 30. Have the classes be [2.97,2.98), [2.98, 2.99), [2.99, 3.00), … etc. Other columns in your table should be frequency and relative frequency. Describe the histogram as right-skewed, left-skewed, or symmetric.

Two days ago, the auto company started using a new kind of plastic in the manufacturing process. Half of the above observations correspond to car parts made using the old plastic and half correspond to the new plastic.

<u>Old Plastic</u>:

2.982	2.994	2.995	3.004	3.007	3.013
3.013	3.015	3.026	3.028	3.028	3.031
3.032	3.033	3.035			

<u>New Plastic</u>:

2.971	2.974	2.975	2.978	2.980	2.983
2.987	2.991	2.996	2.996	3.000	3.004
3.009	3.012	3.023			

Use these two data sets of n = 15 to draw side-by-side stem-and-leaf plots. Have your stems be 2.97, 2.98, 2.99, etc. Describe the stem-and-leaf plots as right-skewed, left-skewed, or symmetric. Does it look like the new type of plastic has affected the weight of the car parts? Briefly explain.

2.2

The Missouri Hockey League has 20 teams with 3 players on each team who play center. These sixty centers each scored a certain number of goals last season. Here is the data in order from smallest to largest. For example, the player with the most goals had 103 and the player with the fewest number of goals had 10.

1. 10	16. 25	31. 35	46. 54
2. 11	17. 25	32. 38	47. 56
3. 12	18. 25	33. 39	48. 57
4. 14	19. 26	34. 39	49. 58
5. 14	20. 26	35. 40	50. 58
6. 15	21. 27	36. 42	51. 59
7. 16	22. 27	37. 42	52. 61
8. 16	23. 27	38. 44	53. 63
9. 19	24. 28	39. 46	54. 65
10. 21	25. 29	40. 47	55. 66
11. 23	26. 30	41. 48	56. 66
12. 24	27. 30	42. 49	57. 66
13. 24	28. 30	43. 49	58. 67
14. 24	29. 31	44. 50	59. 99
15. 24	30. 34	45. 53	60. 103

Find the values of the lower quartile, median, and upper quartile.
Draw a boxplot for the data set.

Chapter 3

3.1

A gambling game named Craps requires a player to roll two fair dice and add the two numbers that appear. Suppose one die is red and the other is blue. There are 36 equally likely outcomes in the sample space:

(red = 1, blue = 1)
(red = 2, blue = 1)
(red = 3, blue = 1)
…
(red = 5, blue = 6)
(red = 6, blue = 6)

The table below has 36 cells corresponding to these 36 outcomes. The sum corresponding to each outcome appears in the appropriate cell.

		red die					
		1	**2**	**3**	**4**	**5**	**6**
	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
blue die	**3**	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Find each of the following probabilities:

a) P(sum = 7)
b) P(at least one die lands on 6) = P(blue = 6 $\cup$ red = 6)
c) P(sum = 9 $\cap$ red = 1)
d) P(the number on the blue die is less than the number on the red die)
e) P(sum = 10 | blue = 6)
f) P(blue = 6 | sum = 10)

Identify each pair of events as “independent”, “mutually exclusive”, or “neither”. Justify your answers.

g) <u>red = 3</u> and <u>blue = 4</u>
h) <u>red = 2</u> and <u>red = 5</u>
i) <u>sum = 3</u> and <u>red = 1</u>
j) <u>red = 5</u> and <u>sum is an odd number</u>
k) <u>blue = 6 $\cup$ red = 6</u> and <u>sum is an odd number</u>

3.2

A machine on a large farm is used to sort eggs and eventually place them into cartons for shipment. After observing too many cracked eggs in cartons, the owner of the farm has hired an engineer to inspect the machine.

Eggs traveling down a conveyor belt randomly move to one of two bins named A and B. The engineer estimates that 55% of the eggs are sent to Bin A and 45% to Bin B. Based on the data he has collected, he estimates that eggs sent to Bin A have a .03 probability of cracking. He estimates that eggs sent to in Bin B have a .04 probability of cracking. Using these estimates, draw a tree diagram or probability table and then answer the following two questions:

If you select an egg at random, what is the overall probability it will be cracked?

Given that an egg is cracked, what is the probability it was sent to Bin B?

3.3

At the Creston Mine, many bulldozers are used every day. The tire pressure of every tire on every bulldozer is supposed to be 25 psi (pounds per square inch). Sadly, many tires are under-inflated when checked. If a bulldozer tire is selected at random, then Y = the tire pressure in psi rounded off to the nearest integer.

$P(Y = 22) = .10$
$P(Y = 23) = .10$
$P(Y = 24) = .30$
$P(Y = 25) = .50$

Find $\mu = E(Y)$, $\sigma^2 = Var(Y)$, and $\sigma = SD(Y)$

3.4

There is an unfair die with unusual numbers on the six faces. If Y = the outcome of one roll of the die, then the probability distribution is given below. Suppose that every time you roll this die, you win the number of dollars equal to the number that appears. There is a catch, of course, and that is you must pay 11 dollars for each roll. In the long run, do you expect to win money, lose money, or break even? Why? If you do any calculations to answer this question, show your work.

Y	P(Y=y)
0	.30
1	.20
8	.10
10	.10
20	.20
100	.10

3.5

Some balls will be drawn at random one at a time out of a basket that contains 100 balls (43 are red, 22 are green, and 35 are black). Identify each variable as "binomial", "geometric" or "neither". If your answer is "binomial", identify the values of n and p. If your answer is "geometric", identify the value of p.

A = # of draws needed to observe the first green ball (without replacement)
B = # of draws needed to observe the first red ball (with replacement)
C = # of green balls that appear out of 10 draws (with replacement)
D = # of black balls that appear out of 4 draws (without replacement)

What are the possible values of A?
What are the possible values of B?
What are the possible values of C?
What are the possible values of D?

3.6

Johnson Sawmill produces many lumber products, including boards made out of pine. Sometimes a pine board will have defects such as scratches, knots, warping, etc. Jim's Furniture Store is a particularly picky client and will only accept about 80% of the pine boards produced at Johnson Sawmill. If you select a pine board at random, the probability that Jim's Furniture will accept it is .80. Assume that consecutive boards are independent of each other.

Jim's Furniture has placed an order for 15 boards.

If Y_1 = the number of accepted boards out of the 15, what kind of discrete random variable is Y_1?

What is the probability that all 15 will be accepted?

What is the probability that exactly 12 will be accepted?

Suppose a representative from Jim's Furniture arrives at Johnson Sawmill to personally select some more pine boards. He examines them one at a time.

If Y_2 = the number of boards needed to be inspected in order to find the first one that is <u>unacceptable</u>, what kind of discrete random variable is Y_2?

What is the probability that $Y_2 = 5$? (In other words, what is the probability the representative accepts the first four boards he sees and then rejects the fifth?)

3.7

Saw #10 at Johnson Sawmill has a blade that often breaks loose. When it breaks loose, workers must stop using the saw and reattach the blade.

Y_3 = # of work stoppages per hour caused by the loose blade

Y_3 is Poisson with an average of $\lambda = 0.8$ stoppages per hour

What is the probability that there will be no stoppages for the next hour?

What is the probability that there will be two or more stoppages within the next hour?

Y_4 = hours until the next work stoppage due to the loose blade

What kind of continuous random variable is Y_4?

On average, how long do we wait until the next work stoppage?

What is the probability there will be no work stoppages for the next 2.5 hours?

3.8

On Friday afternoons, Johnson Sawmill shuts down for the weekend sometime between 5 and 6 pm.

Y_5 = the # of minutes after 5 pm the sawmill shuts down
Y_5 is Uniform on the interval (0, 60)

What is the probability Johnson Sawmill will shut down sometime between 5:23 and 5:44 this Friday afternoon?

3.9

Find the following probabilities if $Z \sim N(\mu = 0, \sigma = 1)$. You are required to draw a picture for every problem.

$P(Z < 1.67)$
$P(Z > -0.66)$
$P(-0.12 < Z < 1.03)$

Find the value of z if…

$P(Z < z) = .9099$
$P(Z > z) = .9463$
$P(-z < Z < z) = .5990$
$P(z < Z < 1.44) = .8000$

What is the 14^{th} percentile of $Z \sim N(\mu=0, \sigma=1)$?
What is the 86^{th} percentile of $Z \sim N(\mu=0, \sigma=1)$?

3.10

An engineer researching cars is interested in the population of all Ford Mustangs that were manufactured five years ago. He believes the mileage on a car randomly selected from this population is normally distributed with a mean of μ = 50,000 miles and a standard deviation of σ = 12,000 miles. Let Y = the mileage on a randomly selected 5-year-old Ford Mustang.

What is the probability the mileage on a randomly selected Mustang is…

Less than 50,480?
Greater than 22,880?
Between 50,000 and 74,000?

The probability that the mileage on a randomly selected Mustang is greater than k miles is .7995. What is k?

What is the 40^{th} percentile among the mileages?

3.11

Ned's Natural Food Store is about to purchase a large shipment of apples from an orchard. Ned is picky about the size of the apples, and wants the mean weight among the apples he is about to purchase to be very close to 150 grams.

When the shipment arrives, Ned will randomly select $n = 4$ apples and compute the sample mean weight among them. He has decided that if the sample mean weight is less than 142 grams or greater than 158 grams, then he will decline to purchase the shipment.

Ned does not know that the weight of a randomly selected apple is a normally distributed random variable with a mean of 152 grams and a standard deviation of 13 grams.

What is the probability that Ned will decline to purchase the shipment?

If Ned chooses a sample size of $n = 10$ instead, what is the probability that he will decline to purchase the shipment?

3.12

What is the degrees of freedom corresponding to each of the following probabilities? You are required to draw a picture for every problem.

$P(T > 2.920) = .05$
$P(T < -2.052) = .025$
$P(-1.796 < T < 1.796) = .90$

3.13

Each of the following data sets has a mean of 5. By simply looking at the data sets and without doing any calculations, put the data sets in a list of increasing variance/standard deviation:

A { 5, 5, 5, 5, 5, 5 }
B { –5, 0, 3, 7, 10, 15 }
C { 2, 3, 4, 6, 7, 8 }
D { 4, 6, 4, 6, 4, 6 }

3.14

Find the sample variance and standard deviation for the following data set by hand. Use both the “regular” and “convenient” formula to find the sample variance (show your work) and make sure you get the same answer each way.

{ –5, 0, 3, 7, 10, 15 }

3.15

The data set { 1, 7, 11, 26, 75 } has a sample mean of 24, a sample variance of 898, and a sample standard deviation of 29.97. Use this information and do very few additional calculations to find the sample mean, sample variance, and sample standard deviation of the following data sets:

A { –4, 2, 6, 21, 70 }
B { 101, 107, 111, 126, 175 }
C { –1, –7, –11, –26, –75 }
D { 3, 21, 33, 78, 225 }

Chapter 4

4.1

Suppose 100 people independently compute a 95% confidence interval. About how many of the intervals do you expect to capture the parameter of interest? Why?

4.2

What would be the formula for a 92% confidence interval for μ when σ is known?

4.3

The birth weights of babies born in the United States is normally distributed and has a standard deviation known to be σ = 500 grams. If the sample mean among a random sample of 30 babies is 3400 grams, what is the 99% confidence interval for μ (the population mean birth weight)?

4.4

John's confidence interval for μ was (88.8, 95.2) and σ was known to be 15 and n = 36. What was the sample mean, and what confidence level did John choose?

Suppose John wanted an error margin no larger than 2 at the same confidence level. What sample size would be necessary?

4.5

An engineer working for General Motors is interested in the population of all vehicles that have an engine of size 3.0 L or smaller, and is particularly interested in μ = the population mean weight among these vehicles. Assume the population is normally distributed. The sample mean among a random sample of 12 vehicles is 3961 pounds and the sample standard deviation is 243 pounds (σ unknown). What is the 90% CI for μ?

4.6

A new computerized telephone system has been in operation at a bank for the past year. The new system is designed to minimize the amount of time a customer will be put on hold when they call the bank. The bank president is quoted as saying "the average time a customer will be put on hold is 3 minutes." You decide to choose days and times at random and call the bank to test the president's claim. You believe the president may be underestimating the mean time. The sample mean hold time among 30 randomly selected calls is 3.2 minutes. Assume the population standard deviation σ is known to be 0.4 minutes.

Do you have significant evidence that the president's claim is false? Report H_0, H_a, the rejection region (use $\alpha = .10$), test statistic, conclusion, and p-value.

4.7

Refer to the previous problem. Suppose another person also heard the president's claim and collected their own data. Unlike you, they believe the president may be overestimating the mean time. Their sample mean hold time among 20 randomly selected calls is 2.9 minutes. They do not know the population standard deviation, and their sample standard deviation is s = 0.5 minutes.

Do they have significant evidence that the president's claim is false? Report H_0, H_a, the rejection region (use $\alpha = .05$), test statistic, conclusion, and place bounds on the p-value.

4.8

What is the rejection region for each of the following hypothesis tests on the Z-curve?

H_0: $\mu = 50$	H_0: $\mu = 50$	H_0: $\mu = 50$
H_a: $\mu < 50$	H_a: $\mu > 50$	H_a: $\mu \neq 50$
$\alpha = .05$	$\alpha = .10$	$\alpha = .07$

4.9

Place bounds on the p-value for each of these one-sample hypothesis tests on the T-curve (draw pictures):

Right-tailed, n = 6, T = 3.5
Left-tailed, n = 12, T = –1.5
Two-tailed, n = 18, T = –3.1
Two-tailed, n = 24, T = 0.1

4.10

The president of Humboldt State University is interested in p = the proportion of freshman with a car on campus. She randomly selects 90 freshman and 63 say they have a car on campus. Use this data to find a 95% confidence interval for p.

4.11

Katie is investigating a population proportion would like the error margin for her 90% confidence interval to be no larger than 1% = .01. What sample size is necessary to achieve this?

Answer the same question as above, except assume that Katie has some information indicating that the population proportion is probably somewhere around .36.

4.12

Sarah had a coin and did not know p=P(H), so to investigate she flipped the coin many times and recorded the results. Her 95% confidence interval for p was (.3804, .4196). How many times did she flip the coin, and how many of the flips were heads?

4.13

Freddy claims that he has a psychic ability to correctly predict the outcome of coin flips in advance. He admits that he can't correctly predict every single flip that he observes, but claims that he will be correct significantly more often that what would be expected by chance if he was just guessing at random.

You agree to flip a coin 50 times and Freddy is correct on 27 of the 50 flips. Freddy excitedly exclaims, "A person who is only guessing on each flip is expected to get 25 (half of 50) correct, and I got 27. This is proof that I am a psychic!" You say to Freddy, "Wait a minute - this proves nothing. Maybe you just got lucky this time. Let's do a hypothesis test with your results."

Using the data of 27 correct guesses out of 50, test Freddy's claim. Let p be the proportion of flips Freddy will correctly predict in the long run. Report H_0, H_a, the rejection region (use α = .05), test statistic, conclusion, and p-value.

4.14

Last week, a sports supply company manufactured a batch of a few thousand bowling balls that are all labeled as having a weight of 16 pounds. Engineers working for the company are confident that this population of bowling balls does have a population mean of $\mu = 16$ pounds, but are concerned about σ^2 (the population variance). Historically, an acceptable variance is $\sigma^2 = 0.02$ pounds, and the engineers are concerned that the actual value may be greater than this. A random sample of 10 bowling balls has a sample variance of $s^2 = 0.0513$. Assume the population of bowling ball weights is normally distributed. Conduct a hypothesis test for σ^2: report H_0, H_a, the rejection region (use $\alpha = .05$), test statistic, conclusion, and place bounds on the p-value.

Chapter 5

The data for these problems is posted in a Microsoft Excel file. The easiest method for doing these problems is to do the calculations and generate the charts within the Excel file (refer to the *Excel Control Chart Guide* and the Excel file that I used to create the in-class handout). You can use other software to perform the calculations and draw the charts if you wish (the data is also in a Word file that you can use to copy and paste into other software).

Your charts should look like the in-class examples: include a title, label the x axis and y axis, draw the UCL and LCL, identify any alarms, etc.

Ollie's Organic Ice Cream, Inc. has made millions of dollars creating high quality ice cream products at their factory and then shipping them all over the world. They have hired you (engineers with statistics expertise) to produce 5 control charts for various processes at the factory.

5.1

A machine at Ollie's factory adds a small amount of peanut bits to the top of each Ollie's Peanut Crunch Ice Cream Cone. Of the hundreds of cones produced per day, five are selected at random and the sample mean amount of peanut bits is measured (in grams). Workers are concerned that the machine has become erratic within the last week. You have been provided with data for the last 40 days. Use the first 20 samples as the base period and draw an X-Bar and s^2 chart *(section 5.4)* with the data. Clearly indicate the values of: Base Period Average Mean, Base Period Average Variance, UCL and LCL for the X-Bar chart, and UCL and LCL for the s^2 chart. What assumptions about the data are you required to make? Does this process appear to be in control? Comment.

5.2

Ollie's stores its ice cream in a gigantic freezer as it awaits shipment. The temperature of the freezer must be closely monitored. If the freezer runs at too warm of a temperature, the ice cream will melt. If the freezer runs at too cold of a temperature, then Ollie's electric bill will be unnecessarily high. The temperature (in °C) is recorded once an hour. Data from the last 100 hours has been provided for you. Use the first 25 observations as the base period and draw an X chart *(section 5.5)* with the data. Clearly indicate the values of: Base Period Mean, MR-Bar, UCL, and LCL. What assumptions about the data are you required to make? Does this process appear to be in control? Comment.

5.3

Ollie's purchases 2000 eggs per day for its ice cream recipes. Every morning, a truck from a local farm arrives with a shipment of 2000 eggs. Inevitably, some of the eggs arrive cracked or smashed and are unusable. Ollie's wants to monitor the number of defective eggs and if it gets out of control, they will consider purchasing eggs from a different supplier. Data from the last 80 days has been provided for you. Use the first 20 observations as the base period and draw an np chart *(section 5.6)* with the data. Clearly indicate the values of: P-Bar, UCL, and LCL. What assumptions about the data are you required to make? Does this process appear to be in control? Comment.

5.4

Ollie's most popular ice cream flavor is Cheerful Cherry Explosion because it has a bountiful amount of fresh cherries in every half-gallon pail. Ollie's randomly selects one pail per day for inspection and counts the number of cherries inside. Data from the last 90 days has been provided for you. Use the first 30 observations as the base period and draw a c chart *(section 5.7)* with the data. Clearly indicate the values of: c-Bar, UCL, and LCL. What assumptions about the data are you required to make? Does this process appear to be in control? Comment.

Chapter 6

An engineer has designed a machine that manufactures wooden pencils. The machine can be adjusted to produce the pencils at a faster or slower rate. He notices that more defective pencils tend to be produced when the machine is going faster. He designs an experiment and runs the machine for five 20-minute sessions under five different speed settings.

X = speed setting: pencils produced per minute

Y = # of defective pencils produced during the twenty minute session

X	Y
200	31
225	36
250	43
275	49
300	46

6.1 Draw the scatterplot (by hand or have a computer do it for you)

6.2 Compute SS_{XX}

6.3 Compute SS_{YY}

6.4 Compute SS_{XY}

6.5 Compute r

6.6 Compute the slope and intercept of the least-squares regression line and draw it on your scatterplot.

6.7 If the machine is set to produce pencils at 245 per minute, what is the predicted number of defective pencils (on average)?

6.8 Compute SS_{total}

6.9 Compute SS_{reg}

6.10 Compute SS_{res} and MS_{res}

6.11 Find the coefficient of determination (r^2) and report what it means regarding these two variables.

6.12 Conduct a hypothesis test for the true slope asking whether or not it equals zero. Report H_0, H_a, the rejection region (use α = .05), test statistic, and conclusion.

Chapter 7

Automobile engineers are studying the effects of snow tires and anti-lock brakes on the stopping distance for a certain vehicle on an icy road. They have two new vehicles of the same make and model that are identical in every way except one has anti-lock brakes and the other has regular brakes. They will test each vehicle five times with snow tires and regular tires. The response variable is Y = stopping distance (feet) required for the vehicle to slow down from 30 mph to 0 mph on the icy road.

Factor A: Tires
Levels: Regular ($x_1 = -1$) or Snow ($x_1 = 1$)

Factor B: Brakes
Levels: Regular ($x_2 = -1$) or Anti-lock ($x_2 = 1$)

Results (average stopping distance among $n = 5$ trials):

Regular Tires and Regular Brakes:	74 feet
Snow Tires and Regular Brakes:	56 feet
Regular Tires and Anti-lock Brakes:	66 feet
Snow Tires and Anti-lock Brakes:	52 feet

7.1 Draw the interaction plot (by hand or have a computer do it for you)

7.2 Compute the estimated effects for Factor A, Factor B, and the interaction

7.3 What conclusions do you draw?