Ch.2 Review

Chapter Summary: Modeling Distributions of Data

In this chapter, we expanded our toolbox for working with quantitative data. We learned how to describe the location of an individual within a distribution by determining its percentile or by calculating a standardized score (z-score) based on the mean and standard deviation of the distribution. We also learned that distributions with a clear overall pattern can be described using a density curve. A common density curve, the Normal curve, is a helpful model for describing many quantitative variables. Knowing how to justify that a distribution is approximately Normally distributed is an important skill. If you can show that a distribution is Normal, you can use the Normal distribution calculations to answer a number of questions about observations within the set of data. The 68-95-99.7 rule and the standard Normal table are both useful tools when performing calculations about observations in Normal distributions.

We have learned in these first two chapters that you should approach data analysis problems using four steps:

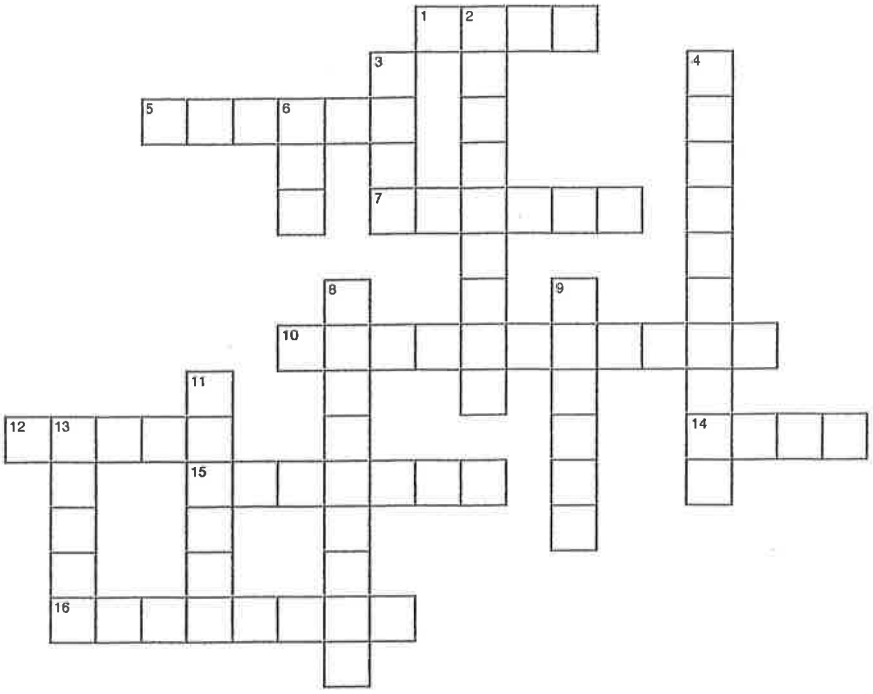
- Graph the data
- Look for an overall pattern and departures from this pattern
- Calculate and interpret numerical summaries
- If the data follow a regular overall pattern, describe the distribution with a smooth curve

The concepts introduced in this chapter will form the basis of much of our study of inference later on in the course. Standardizing data, justifying Normality, and performing Normal calculations are critical skills for statistical inference. Be sure to practice them as you will be using them a LOT!

After You Read: What Have I Learned?
Complete the vocabulary puzzle, multiple-choice questions, and FRAPPY. Check your answers and performance on each of the learning targets. Be sure to get extra help on any targets that you identify as needing more work!

Learning Target	Got It!	Almost There	Needs Work
I can find and interpret the percentile of an individual value in a distribution.			
I can estimate percentiles and individual values using a cumulative relative frequency graph.			
I can find and interpret the standardized score (z-score) of an individual value in a distribution of data.			
I can describe the effect of adding, subtracting, multiplying by, or dividing by a constant on the shape, center, and variability of a distribution of data.			
I can use a density curve to model distributions of quantitative data.			
I can use the 68-95-99.7 Rule.			
I can find the proportion of values in a specified interval in a Normal distribution.			
I can find the value that corresponds to a given percentile in a Normal distribution.			
I can determine whether a distribution of data is approximately Normal from graphical and numerical evidence.			

Chapter 2: Modeling Distributions of Data



Across

1. The balance point of a density curve, if it were made of solid material
5. The standardized value of an observation
7. These common density curves are symmetric and bell-shaped
10. A Normal _____ plot provides a good assessment of whether a data set is approximately Normally distributed
12. Another name for a cumulative relative frequency graph
14. The standard Normal table tells us the area under the standard Normal curve to the ___ of z
15. A ___ curve is a smooth curve that can be used to model a distribution
16. This Normal distribution has mean 0 and standard deviation 1

Down

2. The ____ rule is also known as the 68-95-99.7 rule for Normal distributions
3. To standardize a value, subtract the ___ and divide by the standard deviation
4. The value with p percent of the observations less than it
6. The area under any density curve is always equal to
8. We ___ data when we change each value by adding a constant and/or multiplying by a constant.
9. If a Normal probability plot shows a _____ pattern, the data are approximately Normal
11. The point that divides the area under a density curve in half
13. This mathematician first applied Normal curves to data to errors made by astronomers and surveyors

Chapter 2 Multiple Choice Practice

Directions. *Identify the choice that best completes the statement or answers the question. Check your answers and note your performance when you are finished.*

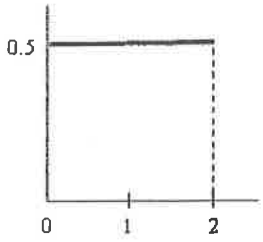
1. The 16th percentile of a Normally distributed variable has a value of 25 and the 97.5th percentile has a value of 40. Which of the following is the best estimate of the mean and standard deviation of the variable?

   (A) Mean ≈ 32.5; Standard deviation ≈ 2.5
   (B) Mean ≈ 32.5; Standard deviation ≈ 5
   (C) Mean ≈ 32.5; Standard deviation ≈ 10
   (D) Mean ≈ 30; Standard deviation ≈ 2.5
   (E) Mean ≈ 30; Standard deviation ≈ 5

2. The proportion of observations from a standard Normal distribution that take values larger than 0.75 is about

   (A) 0.2266
   (B) 0.2500
   (C) 0.7704
   (D) 0.7764
   (E) 0.8023

3. The density curve below takes the value 0.5 on the interval $0 < x < 2$ and takes the value 0 everywhere else. What percent of the observations lie between 0.4 and 1.08?



   (A) 25%
   (B) 34%
   (C) 50%
   (D) 68%
   (E) 70%

4. The distribution of the heights of students in a large class is roughly Normal. The average height is 68 inches, and approximately 99.7% of the heights are between 62 and 74 inches. Thus, the standard deviation of the height distribution is approximately equal to

   (A) 2
   (B) 3
   (C) 4
   (D) 6
   (E) 9

5. The mean age (at inauguration) of all U.S. Presidents is approximately Normally distributed with a mean of 54.6. Barack Obama was 47 when he was inaugurated, which is the 11th percentile of the distribution. George Washington was 57. What percentile was he in?

(A) 6.17
(B) 65.17
(C) 62.92
(D) 34.83
(E) 38.9

6. Which of the following statements are false?

I. The standard Normal table can be used with z-scores from any distribution
II. The mean is always equal to the median for any Normal distribution.
III. Every symmetric, bell-shaped distribution is Normal
IV. The area under a Normal curve is always 1, regardless of the mean and standard deviation.

(A) I and II
(B) I and III
(C) II and III
(D) III and IV
(E) None of the above gives the correct set of true statements.

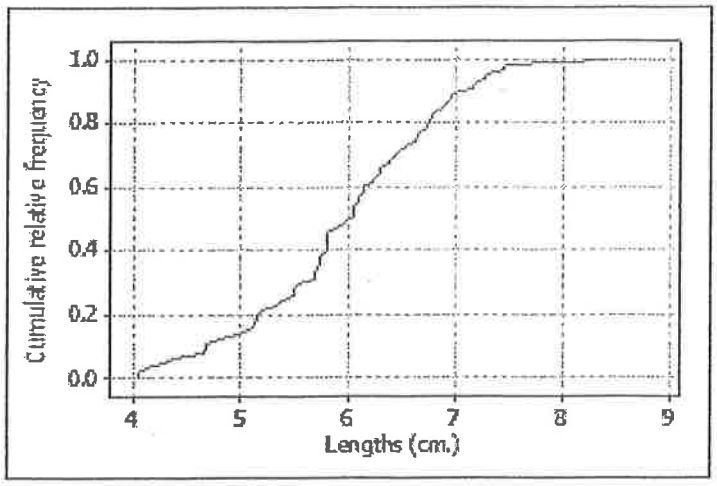
7. High school textbooks don't last forever. The lifespan of all high school statistics textbooks is approximately Normally distributed with a mean of 9 years and a standard deviation of 2.5 years. What percentage of the books last more than 10 years?

(A) 11.5%
(B) 34.5%
(C) 65.5%
(D) 69%
(E) 84.5%

8. The distribution of the time it takes for different people to solve their Strive for a Five chapter crossword puzzle is strongly skewed to the right, with a mean of 10 minutes and a standard deviation of 2 minutes. The distribution of *z*-scores for those times is

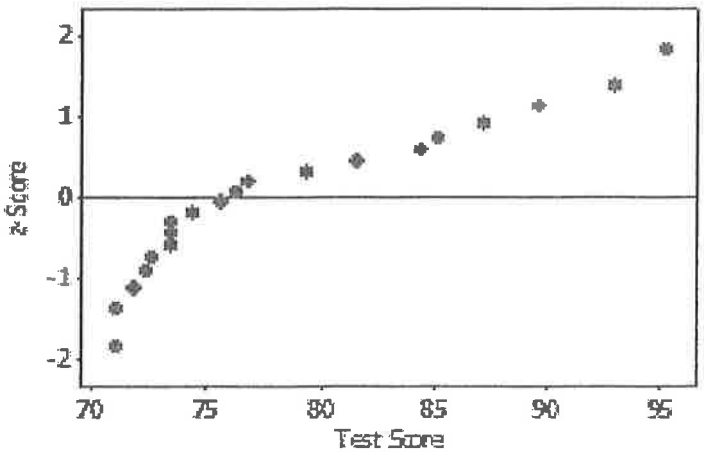
(A) Normally distributed, with mean 10 and standard deviation 2.
(B) Skewed to the right, with mean 10 and standard deviation 2.
(C) Normally distributed, with mean 0 and standard deviation 1.
(D) Skewed to the right, with mean 0 and standard deviation 1.
(E) Skewed to the right, but the mean and standard deviation cannot be determined without more information.

9. The cumulative relative frequency graph below shows the distribution of lengths (in centimeters) of fingerlings at a fish hatchery. The third quartile for this distribution is approximately:



(A) 6.7 cm
(B) 7 cm
(C) 6 cm
(D) 5.5 cm
(E) 7.5 cm

10. The plot shown below is a Normal probability plot for a set of test scores. Which statement is true for these data?



(A) The distribution is clearly Normal.
(B) The distribution is approximately Normal.
(C) The distribution appears to be skewed.
(D) The distribution appears to be uniform.
(E) There is insufficient information to determine the shape of the distribution.

Check your answers below. If you got a question wrong, check to see if you made a simple mistake or if you need to study that concept more. After you check your work, identify the concepts you feel very confident about and note what you will do to learn the concepts in need of more study.

#	Answer	Concept	Right	Wrong	Simple Mistake?	Need to Study More
1	E	68-95-99.7 Rule				
2	A	Standard Normal Table				
3	B	Area under a Density Curve				
4	A	68-95-99.7 Rule				
5	B	Standard Normal Calculations				
6	B	Properties of Normal Distributions				
7	B	Standard Normal Calculations				
8	D	Standardized Scores				
9	A	Cumulative Relative Frequency Graph				
10	C	Normal Probability Plots				

Chapter 2 Reflection

Summarize the "Big Ideas" in Chapter 2:
My strengths in this chapter:
Concepts I need to study more and what I will do to learn them:

FRAPPY! Free Response AP® Problem, Yay!

The following problem is modeled after actual Advanced Placement Statistics free response questions. Your task is to generate a complete, concise response in 15 minutes. After you generate your response, view two example solutions and determine whether you feel they are "complete", "substantial", "developing" or "minimal". If they are not "complete", what would you suggest to the student who wrote them to increase their score? Finally, you will be provided with a rubric. Score your response and note what, if anything, you would do differently to increase your own score.

Final exam grades are determined by the percent correct on the exam. A teacher's records indicate the performance on the exam is Normally distributed with mean 82 and standard deviation 5. The grades on her exam are assigned using the scale below.

Grade	Percent Correct
A	$94 \le$ percent ≤ 100
B	$85 \le$ percent < 94
C	$76 \le$ percent < 85
D	$65 \le$ percent < 76
F	$0 \le$ percent < 65

(a) Use a sketch of a Normal distribution to illustrate the proportion of students who would earn a B. Calculate this proportion.

(b) Students who earn a B, C, or D are considered to "meet standards". Based on this grading scale, what percent of students will receive a score that places them in a category other than "meets standards"?

(c) What grade would the student who scored at the 25th percentile earn on this chapter? Justify your answer.

FRAPPY! Student Responses

Student Response 1:

a) $P(B) = 0.9918 - 0.7257 = 0.2661$. 26.61 % of students earn a B.
b) P(does not meet standards) = $P(F) = P(z<3.4) = 0.0003$
c) $z = 0.25$ so the score would be $82 + 0.25(5) = 83.25$

How would you score this response? Is it substantial? Complete? Developing? Minimal? Is there anything this student could do to earn a better score?

Student Response 2:

a) $P(B) = P(0.6 \leq z < 2.4) = 0.9793 - 0.5239 = 0.4554$
b) $P(A \text{ or } F) = P(z > 2.4) + P(z < -3.4) = 0.0085$
c) A z-score of -0.6745 corresponds to the 25th percentile. So, the score would be $82 + (-0.6475)(5) = 78.63$.

How would you score this response? Is it substantial? Complete? Developing? Minimal? Is there anything this student could do to earn a better score?

FRAPPY! Scoring Rubric

Use the following rubric to score your response. Each part receives a score of "Essentially Correct," "Partially Correct," or "Incorrect." When you have scored your response, reflect on your understanding of the concepts addressed in this problem. If necessary, note what you would do differently on future questions like this to increase your score.

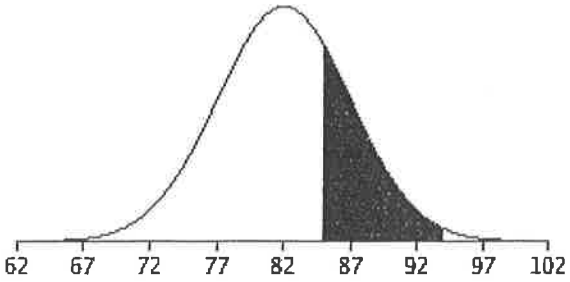
Intent of the Question

The goal of this question is to determine your ability to perform and interpret Normal calculations.

Solution

(a)

percent < 94)
= $P((85-82)/5 \le z < (94-82)/5)$
= $P(0.6 \le z < 2.4)$
= 0.9918 – 0.7257
= 0.2661



(b) P(A or F) = P(A) + P(F)
= $P(z \ge (94-82)/5) + P(z < (65-82)/5)$
= $P(z \ge 2.4) + P(z < 3.4)$
= 0.0082 + 0.0003
= 0.0085

(c) A z-score of -0.6745 corresponds to the 25th percentile.
x = mean + z(std dev)
x = 82+ (-0.6475)(5)
x = 78.63

Scoring:

Parts (a), (b), and (c) are scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is essentially correct if (1) the appropriate probability is illustrated using a labeled Normal curve and (2) the proportion is correctly computed.

Part (a) is partially correct if only one of the above elements is correct

Part (b) is essentially correct if the response (1) recognizes the need to look at grades of A and F and (2) correctly computes the tail probabilities and adds them together.

Part (b) is partially correct if the response considers only an A or an F and calculates the corresponding tail area correctly OR recognizes the need to look at A an F but only calculates one of the tail areas correctly OR approximates the probabilities using the 68-95-99.7 rule OR computes the proportion that will "meet standards" OR states the correct answer without supporting work.

Part (c) is essentially correct if (a) the correct *z*-score is identified for the 25th percentile and (2) the correct corresponding score is calculated.
Part (c) is partially correct if only one of the above elements is correct.

NOTE: If the student makes an error in part (b) and correctly uses that probability in part (c) to compute a reasonable probability, part (c) is essentially correct.

4 Complete Response
All three parts essentially correct

3 Substantial Response
Two parts essentially correct and one part partially correct

2 Developing Response
Two parts essentially correct and no parts partially correct
One part essentially correct and two parts partially correct
Three parts partially correct

1 Minimal Response
One part essentially correct and one part partially correct
One part essentially correct and no parts partially correct
No parts essentially correct and two parts partially correct

My Score:
What I did well:
What I could improve:
What I should remember if I see a problem like this on the AP Exam: