Contributor name: Ananya K. R.

Book Proposed: Introductory Statistics by Douglas S. Shafer And Zhiyi Zhang, Flat World Publishers, Edition - 1, 2012

Total Chapters: 11

Total Examples: 216

Codable Examples: 203

Chapter 1: Introduction

Example 1.1 – Codable

Example 1.2 – Codable

Chapter 2: Descriptive Statistics

Example 2.1 – Codable

Example 2.2 – Codable

Example 2.3 – Codable

Example 2.4 – Codable

Example 2.5 – Codable

Example 2.6 – Codable

Example 2.7 – Codable

Example 2.8 – Codable

Example 2.9 – Codable

Example 2.10 – Codable

Example 2.11 – Codable

Example 2.12 – Codable

Example 2.13 – Codable

Example 2.14 – Codable

Example 2.15 – Codable

Example 2.16 – Codable

Example 2.17 – Codable

Example 2.18 – Codable

Example 2.19 – Not Codable (follows the Empirical Rule to solve the problem)

EXAMPLE 19

Heights of 18-year-old males have a bell-shaped distribution with mean 69.6 inches and standard deviation 1.4 inches.

a. About what proportion of all such men are between 68.2 and 71 inches tall?
b. What interval centered on the mean should contain about 95% of all such men?

Solution:

A sketch of the distribution of heights is given in Figure 2.17 "Distribution of Heights".

a. Since the interval from 68.2 to 71.0 has endpoints $\overline{x} - s$ and $\overline{x} + s$, by the Empirical Rule about 68% of all 18-year-old males should have heights in this range.

b. By the Empirical Rule the shortest such interval has endpoints $\overline{x} - 2s$ and $\overline{x} + 2s$. Since

$$\overline{x} - 2s = 69.6 - 2(1.4) = 66.8 \quad \text{and} \quad \overline{x} + 2s = 69.6 + 2(1.4)$$

the interval in question is the interval from 66.8 inches to 72.4 inches.

Example 2.20 – Not Codable (Theoretical question)

EXAMPLE 20

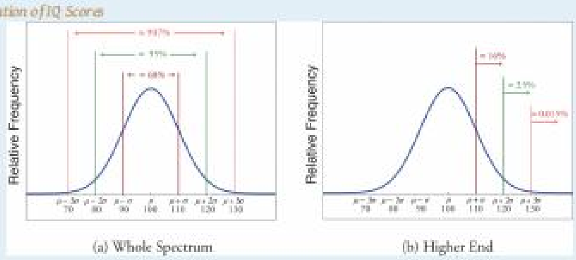
Scores on IQ tests have a bell-shaped distribution with mean $\mu = 100$ and standard deviation $\sigma = 10$. Discuss what the Empirical Rule implies concerning individuals with IQ scores of 110, 120, and 130.

Solution:

A sketch of the IQ distribution is given in Figure 2.18 "Distribution of IQ Scores". The Empirical Rule states that

1. approximately 68% of the IQ scores in the population lie between 90 and 110,
2. approximately 95% of the IQ scores in the population lie between 80 and 120, and
3. approximately 99.7% of the IQ scores in the population lie between 70 and 130.

Figure 2.18
Distribution of IQ Scores



Since 68% of the IQ scores lie *within* the interval from 90 to 110, it must be the case that 32% lie *outside* that interval. By symmetry approximately half of that 32%, or 16% of all IQ scores, will lie above 110. If 16% lie above 110, then 84% lie below. We conclude that the IQ score 110 is the 84th percentile.

The same analysis applies to the score 120. Since approximately 95% of all IQ scores lie within the interval form 80 to 120, only 5% lie outside it, and half of them, or 2.5% of all scores, are above 120. The IQ score 120 is thus higher than 97.5% of all IQ scores, and is quite a high score.

By a similar argument, only 15/100 of 1% of all adults, or about one or two in every thousand, would have an IQ score above 130. This fact makes the score 130 extremely high.

Example 2.17 – Not Codable (Theoretical Question)

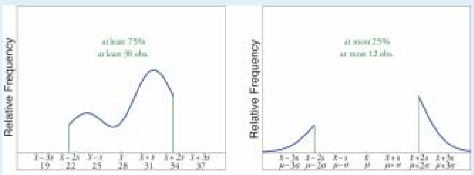
EXAMPLE 21

A sample of size $n = 50$ has mean $\overline{x} = 28$ and standard deviation $s = 3$. Without knowing anything else about the sample, what can be said about the number of observations that lie in the interval (22,34)? What can be said about the number of observations that lie outside that interval?

Solution:

The interval (22,34) is the one that is formed by adding and subtracting two standard deviations from the mean. By Chebyshev's Theorem, at least 3/4 of the data are within this interval. Since 3/4 of 50 is 37.5, this means that at least 37.5 observations are in the interval. But one cannot take a fractional observation, so we conclude that at least 38 observations must lie inside the interval (22,34).

If at least 3/4 of the observations are in the interval, then at most 1/4 of them are outside it. Since 1/4 of 50 is 12.5, at most 12.5 observations are outside the interval. Since again a fraction of an observation is impossible, x (22,34).



(a) Within $\bar{x} \pm 2s$ (b) Outside $\bar{x} \pm 2$

Example 2.18 – Not Codable (Theoretical Question)

EXAMPLE 22

The number of vehicles passing through a busy intersection between 8:00 a.m. and 10:00 a.m. was observed and recorded on every weekday morning of the last year. The data set contains $n = 251$ numbers. The sample mean is $\overline{x} = 725$ and the sample standard deviation is $s = 25$. Identify which of the following statements *must* be true.

1. On approximately 95% of the weekday mornings last year the number of vehicles passing through the intersection from 8:00 a.m. to 10:00 a.m. was between 675 and 775.
2. On at least 75% of the weekday mornings last year the number of vehicles passing through the intersection from 8:00 a.m. to 10:00 a.m. was between 675 and 775.
3. On at least 189 weekday mornings last year the number of vehicles passing through the intersection from 8:00 a.m. to 10:00 a.m. was between 675 and 775.
4. On at most 25% of the weekday mornings last year the number of vehicles passing through the intersection from 8:00 a.m. to 10:00 a.m. was either less than 675 or greater than 775.
5. On at most 12.5% of the weekday mornings last year the number of vehicles passing through the intersection from 8:00 a.m. to 10:00 a.m. was less than 675.
6. On at most 25% of the weekday mornings last year the number of vehicles passing through the intersection from 8:00 a.m. to 10:00 a.m. was less than 675.

Solution:

1. Since it is not stated that the relative frequency histogram of the data is bell-shaped, the Empirical Rule does not apply. Statement (1) is based on the Empirical Rule and therefore it might not be correct.
2. Statement (2) is a direct application of part (1) of Chebyshev's Theorem because $(\overline{x}-2s, \overline{x}+2s) = (675,775)$. It must be correct.
3. Statement (3) says the same thing as statement (2) because 75% of 251 is 188.25, so the minimum whole number of observations in this interval is 189. Thus statement (3) is definitely correct.
4. Statement (4) says the same thing as statement (2) but in different words, and therefore is definitely correct.
5. Statement (4), which is definitely correct, states that at most 25% of the time either fewer than 675 or more than 775 vehicles passed through the intersection. Statement (5) says that half of that 25% corresponds to days of light traffic. This would be correct if the relative frequency histogram of the data were known to be symmetric. But this is not stated; perhaps all of the observations outside the interval (675,775) are less than 75. Thus statement (5) might not be correct.
6. Statement (4) is definitely correct and statement (4) implies statement (6): even if every measurement that is outside the interval (675,775) is less than 675 (which is conceivable, since symmetry is not known to hold), even so at most 25% of all observations are less than 675. Thus statement (6) must definitely be correct.

Chapter 3: Basic Concepts of Probability

Example 3.1 – Codable

Example 3.2 – Codable

Example 3.3.a – Codable

Example 3.3.b – Codable

Example 3.4 – Codable

Example 3.5 – Codable

Example 3.6 – Codable

Example 3.7 – Codable

Example 3.8.a – Codable

Example 3.8.b – Codable

Example 3.8.c – Codable

Example 3.9.a – Codable

Example 3.9.b – Codable

Example 3.9.c – Codable

Example 3.10 – Codable

Example 3.11 – Codable

Example 3.12 – Codable

Example 3.13.a – Codable

Example 3.13.b – Codable

Example 3.14 – Codable

Example 3.15 – Codable

Example 3.16 – Codable

Example 3.17.a – Codable

Example 3.17.b – Codable

Example 3.18 – Codable

Example 3.19.a – Codable

Example 3.19.b – Codable

Example 3.19.c – Codable

Example 3.20.a – Codable

Example 3.20.b – Codable

Example 3.21.a – Codable

Example 3.21.b – Codable

Example 3.22.a – Codable

Example 3.22.b – Codable

Example 3.22.c– Codable

Example 3.23 – Codable

Example 3.24 – Codable

Example 3.25.a – Codable

Example 3.25.b – Codable

Example 3.26.a – Codable

Example 3.26.b – Codable

Example 3.27 – Codable

Example 3.28.a – Codable

Example 3.28.b – Codable

Example 3.28.c – Codable

Chapter 4: Discrete Random Variables

Example 4.1.a – Codable

Example 4.1.b – Codable

Example 4.2.a – Codable

Example 4.2.b – Codable

Example 4.2.c – Codable

Example 4.3 – Codable

Example 4.4.a – Codable

Example 4.4.b – Codable

Example 4.4.c – Codable

Example 4.5 – Codable

Example 4.6.a – Codable

Example 4.6.b – Codable

Example 4.6.c – Codable

Example 4.6.d – Codable

Example 4.6.e – Codable

Example 4.6.f – Codable

Example 4.6.g – Codable

Example 4.6.h – Codable

Example 4.7.a – Codable

Example 4.7.b – Codable

Example 4.7.c – Codable

Example 4.8 – Codable

Example 4.9.a – Codable

Example 4.9.b – Codable

Example 4.10.a – Not Codable (Since the solution derives values from a graph, it is not a mathematical calculation and thus, cannot be converted into a code.)

EXAMPLE 10

An appliance repairman services five washing machines on site each day. One-third of the service calls require installation of a particular part.

a. The repairman has only one such part on his truck today. Find the probability that the one part will be enough today, that is, that at most one washing machine he services will require installation of this particular part.

Let X denote the number of service calls today on which the part is required. Then X is a binomial random variable with parameters $n = 5$ and $p = 1 / 3 = 0.\overline{3}$.

a. Note that the probability in question is not $P(1)$, but rather $P(X \leq 1)$. Using the cumulative distribution table in Chapter 12 "Appendix",

$$P(X \leq 1) = 0.4609$$

Example 4.10.b – Not Codable (Since the solution derives values from a graph, it is not a mathematical calculation and thus, cannot be converted into a code.)

> b. Find the minimum number of such parts he should take with him each day in order that the probability that he have enough for the day's service calls is at least 95%.

> b. The answer is the smallest number x such that the table entry $P(X \leq x)$ is at least 0.9500. Since $P(X \leq 2) = 0.7901$ is less than 0.95, two parts are not enough. Since $P(X \leq 3) = 0.9547$ is as large as 0.95, three parts will suffice at least 95% of the time. Thus the minimum needed is three.

Chapter 5: Continuous Random Variables

Example 5.1.a – Codable

Example 5.1.b – Codable

Example 5.1.c – Codable

Example 5.2 – Codable

Example 5.3 – Codable

Example 5.4.a – Codable

Example 5.4.b – Codable

Example 5.5.a – Codable

Example 5.5.b – Codable

Example 5.6.a – Codable

Example 5.6.b – Codable

Example 5.7.a – Codable

Example 5.7.b – Codable

Example 5.8.a – Codable

Example 5.8.b– Codable

Example 5.8.c – Codable

Example 5.9.a – Codable

Example 5.9.b – Codable

Example 5.10 – Codable

Example 5.11 – Codable

Example 5.12 – Codable

Example 5.13 – Codable

Example 5.14 – Codable

Example 5.15 – Codable

Example 5.16 – Codable

Example 5.17 – Codable

Example 5.18 – Codable

Chapter 6: Sampling Distributions

Example 6.1 – Codable

Example 6.2 – Codable

Example 6.3.a – Codable

Example 6.3.b – Codable

Example 6.3.c – Codable

Example 6.4 – Codable

Example 6.5 – Codable

Example 6.6.a– Codable

Example 6.6.b – Codable

Example 6.7.a – Codable

Example 6.7.b – Codable

Example 6.8.a – Codable

Example 6.8.b – Codable

Example 6.8.c – Codable

Example 6.8.d – Not Codable (Reason: Theoretical Example with the purpose of proving a concept)

EXAMPLE 8

An online retailer claims that 90% of all orders are shipped within 12 hours of being received. A consumer group placed 121 orders of different sizes and at different times of day; 102 orders were shipped within 12 hours.

d. Based on the answer to part (c), draw a conclusion about the retailer's claim.

d. The computation shows that a random sample of size 121 has only about a 1.4% chance of producing a sample proportion as the one that was observed, $\hat{p} = 0.84$, when taken from a population in which the actual proportion is 0.90. This is so unlikely that it is reasonable to conclude that the actual value of p is less than the 90% claimed.

Chapter 7: Estimation

Example 7.1.a – Codable

Example 7.1.b – Codable

Example 7.2.a – Codable

Example 7.2.b – Codable

Example 7.3 – Codable

Example 7.4 – Codable

Example 7.5 – Codable

Example 7.6 – Codable

Example 7.7 – Codable

Example 7.8 – Codable

Example 7.9 – Codable

Example 7.10.a – Codable

Example 7.10.b – Codable

Example 7.11 – Codable

Chapter 8: Testing Hypotheses

Example 8.1 – Not Codable (Reason: Theoretical Example with the purpose of proving a concept)

EXAMPLE 1

A publisher of college textbooks claims that the average price of all hardbound college textbooks is $127.50. A student group believes that the actual mean is higher and wishes to test their belief. State the relevant null and alternative hypotheses.

Solution:

The default option is to accept the publisher's claim unless there is compelling evidence to the contrary. Thus the null hypothesis is $H_0 : \mu = 127.50$. Since the student group thinks that the average textbook price is *greater* than the publisher's figure, the alternative hypothesis in this situation is $H_a : \mu > 127.50$.

Example 8.2 –Not Codable (Reason: Theoretical Example with the purpose of proving a concept)

EXAMPLE 2

The recipe for a bakery item is designed to result in a product that contains 8 grams of fat per serving. The quality control department samples the product periodically to insure that the production process is working as designed. State the relevant null and alternative hypotheses.

Solution:

The default option is to assume that the product contains the amount of fat it was formulated to contain unless there is compelling evidence to the contrary. Thus the null hypothesis is $H_0 : \mu = 8.0$. Since to contain either more fat than desired or to contain less fat than desired are both an indication of a faulty production process, the alternative hypothesis in this situation is that the mean is *different* from 8.0, so $H_a : \mu \neq 8.0$.

Example 8.3 – Codable

Example 8.4 – Codable

Example 8.5 – Codable

Example 8.6 – Codable

Example 8.7 – Codable

Example 8.8 – Codable

Example 8.9 – Codable

Example 8.10 – Codable

Example 8.11 – Codable

Example 8.12 – Codable

Example 8.13 – Codable

Example 8.14 – Codable

Example 8.15 – Codable

Chapter 9: Two-Sample Problems

Example 9.1 – Codable

Example 9.2 – Codable

Example 9.3 – Codable

Example 9.4 – Codable

Example 9.5 – Codable

Example 9.6 – Not Codable (The solution uses an theoretical approach)

EXAMPLE 6

Perform the test of Note 9.13 "Example 5" using the *p*-value approach.

Solution:

The first three steps are identical to those in Note 9.13 "Example 5".

- Step 4. Because the test is two-tailed the observed significance or *p*-value of the test is the double of the area of the right tail of Student's *t*-distribution, with 15 degrees of freedom, that is cut off by the test statistic $T = 1.040$. We can only approximate this number. Looking in the row of Figure 12.3 "Critical Values of " headed $df = 15$, the number 1.040 is between the numbers 0.866 and 1.341, corresponding to $t_{0.200}$ and $t_{0.100}$.

  The area cut off by $t = 0.866$ is 0.200 and the area cut off by $t = 1.341$ is 0.100. Since 1.040 is between 0.866 and 1.341 the area it cuts off is between 0.200 and 0.100. Thus the *p*-value (since the area must be doubled) is between 0.400 and 0.200.

- Step 5. Since $p > 0.200 > 0.01$ $, p > \alpha$, so the decision is not to reject the null hypothesis:

  The data do not provide sufficient evidence, at the 1% level of significance, to conclude that the mean sales per month of the two designs are different.

Example 9.7 – Codable

Example 9.8 – Codable

Example 9.9 – Not Codable (The solution uses an theoretical approach)

EXAMPLE 9

Perform the test of Note 9.20 "Example 8" using the *p*-value approach.

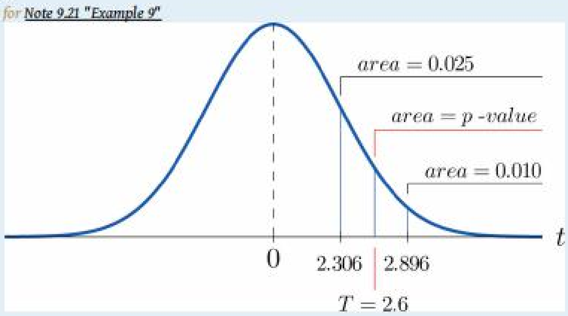
Solution:

The first three steps are identical to those in Note 9.20 "Example 8".

- Step 4. Because the test is one-tailed the observed significance or *p*-value of the test is just the area of the right tail of Student's *t*-distribution, with 8 degrees of freedom, that is cut off by the test statistic $T = 2.600$. We can only approximate this number. Looking in the row of Figure 12.3 "Critical Values of " headed $df = 8$, the number 2.600 is between the numbers 2.306 and 2.896, corresponding to $t_{0.025}$ and $t_{0.010}$.

  The area cut off by $t = 2.306$ is 0.025 and the area cut off by $t = 2.896$ is 0.010. Since 2.600 is between 2.306 and 2.896 the area it cuts off is between 0.025 and 0.010. Thus the *p*-value is between 0.025 and 0.010. In particular it is less than 0.025. See Figure 9.6.

Figure 9.6
P-Value for Note 9.21 "Example 9"



- Step 5. Since 0.025 < 0.05, $p < \alpha$ so the decision is to reject the null hypothesis:

Example 9.10 – Codable

Example 9.11 – Codable

Example 9.12 – Codable

Example 9.13 – Codable

Example 9.14 – Codable

Example 9.15.a – Codable

Example 9.15.b – Codable

Chapter 10: Correlation and Regression

Example 10.1 – Codable

Example 10.2 – Codable

Example 10.3.a – Codable

Example 10.3.b – Codable

Example 10.3.c– Codable

Example 10.3.d – Not Codable (Theoretical)

EXAMPLE 3

Table 10.3 "Data on Age and Value of Used Automobiles of a Specific Make and Model" shows the age in years and the retail value in thousands of dollars of a random sample of ten automobiles of the same make and model.

d. Interpret the meaning of the slope of the least squares regression line in the context of the problem.

The slope −2.05 means that for each unit increase in x (additional year of age) the average value of this make and model vehicle decreases by about 2.05 units (about $2,050).

Example 10.3.e – Codable

Example 10.3.f – Codable

Example 10.3.g – Not Codable (Theoretical)

EXAMPLE 3

Table 10.3 "Data on Age and Value of Used Automobiles of a Specific Make and Model" shows the age in years and the retail value in thousands of dollars of a random sample of ten automobiles of the same make and model.

g. Comment on the validity of using the regression equation to predict the price of a brand new automobile of this make and model.

The price of a brand new vehicle of this make and model is the value of the automobile at age 0. If the value $x = 0$ is inserted into the regression equation the result is always $\hat{\beta}_0$, the y-intercept, in this case 32.83, which corresponds to $32,830. But this is a case of extrapolation, just as part (f) was, hence this result is invalid, although not obviously so. In the context of the problem, since automobiles tend to lose value much more quickly immediately after they are purchased than they do after they are several years old, the number $32,830 is probably an underestimate of the price of a new automobile of this make and model.

Example 10.4.a – Codable

Example 10.4.b – Codable

Example 10.5 – Codable

Example 10.6 – Codable

Example 10.7 – Codable

Example 10.8 – Codable

Example 10.9 – Codable

Example 10.10 – Codable

Example 10.11 – Codable

Example 10.12 – Codable

Example 10.13 – Codable

Chapter 11: Chi-Square and F-Tests

Example 11.1 – Codable

Example 11.2 – Codable

Example 11.3.a – Codable

Example 11.3.b – Codable

Example 11.4.a – Codable

Example 11.4.b – Codable

Example 11.4.c – Codable

Example 11.4.d – Codable

Example 11.5.a – Codable

Example 11.5.b – Codable

Example 11.6 – Codable

Example 11.7 – Codable

Example 11.8 – Codable

Example 11.9 – Codable