Answers for Comprehension Questions for Chapter 2

Question 1: Consider the following small set of scores. Each number represents the number of siblings reported by each of the N = 6 persons in the sample:

X scores are: [0 1 1 1 2 7]

1a. Compute the mean (M) for this set of six scores.

Answer: $M = \Sigma X/N = 12/6 = 2.0$

1b. Compute the six deviations from the mean (X – M) and list these six deviations.

Answer:

[(0-2) (1-2) (1-2) (1-2) (2 -2) (7-2)]

-2 -1 -1 -1 0 +5

1c. What is the sum of the six deviations from the mean you reported in 1b? Is this outcome a surprise?

Answer: The sum of these deviations from the mean is 0. This is not a surprise; this always happens when the sum is taken for deviations from a sample mean.

d. Now calculate the sum of squared deviations (*SS*) for this set of six scores.

Answer:

$$SS = (-2)^2 + (-1)^2 + (-1)^2 + (-1)^2 + (0)^2 + (+5)^2$$

$$= 4 + 1 + 1 + 1 + 0 + 25$$

$$= 32$$

1e. Compute the sample variance, s^2, for this set of six scores.

Answer: $s^2 = SS/(N – 1) = 32/(6-1) = 32/5 = 6.4$

1f. When you compute s^2, why should you divide SS by (N-1) rather than N?

Because only (N -1), in this example, 5, of the deviations from the sample mean were free to vary. By the time we know that the first five deviations are -2, -1, -1, -1 and 0, we know that the last deviation must be +5 in order for the sum of these deviations to be 0.

Thus, when we compute SS and s^2*, we actually have only N-1 "independent" pieces of information about variability, that is, differences between individual X scores and the sample mean.*

g. Finally, compute the sample standard deviation (denoted either by *s* or *SD*).

s or SD = square root of s^2 *= square root of 6.4 = 2.53*

2. In your own words, what does an *SS* tell us about a set of data? Under what circumstances will the value of *SS* equal 0? Can *SS* ever be negative?

Answer: The value of SS tells us whether the individual X scores tend to be close to or far from the sample mean. An SS of 0 occurs when all of the X scores are equal to each other (and therefore also equal to M, the sample mean). SS cannot be negative because it is calculated by taking squared deviations, and squared deviations must be positive.

3. For each of the following lists of scores, indicate whether the value of *SS* will be negative, 0, between 0 and +15, or greater than +15. (You do not need to actually calculate *SS*.)

Sample A: *X* = [103, 156, 200, 300, 98]

Answer for Sample A: SS >15

Sample B: *X* = [103, 103, 103, 103, 103, 103]

Answer for Sample B: SS:= 0

Sample C: X = [101, 102, 103, 102, 101]

Answer for Sample C: SS < 15

4. For a variable that interests you, discuss why there is variance in scores on that variable. (In Chapter 2, e.g., there is a discussion of factors that might create variance in heart rate, HR.)

5. Assume that a population of thousands of people whose responses were used to develop the anxiety test had scores that were normally distributed with $\mu = 30$ and $\sigma = 10$. What proportion of people in this population would have anxiety scores within each of the following ranges of scores?

a. Below 20

b. Above 30

c. Between 10 and 50

d. Below 10

e. Below 50

f. Above 50

g. Either below 10 or above 50

Assuming that a score in the top 5% of the distribution would be considered extremely anxious, would a person whose anxiety score was 50 be considered extremely anxious?

Answer for 5a, what proportion of scores are expected to be below X = 20:

z = (X – 30)/ 10 = (20-30)/ 10 = -1.0

The area to the left of z = -1.0 is .1587. About .16 of the people would be expected to have scores below 20.

5b. What proportion of people would be expected to score above 30?

Answer to 5b: z = (X – 30)/10 = (30-30)/10 = 0.

The total area to the right of z = 0 is .50. About half the sample has scores above 30, that is, scores above the mean.

5c. What proportion of people would be expected to score between 10 and 50?

Answer: For X = 10, z = (10-30)/10 = -2.00

The area between z = 0 and z = -2.00 is .4772

For X = 50, z = (50 -30)/10 = +2.00

The area between z = 0 and z = +2.00 is .4772

So the total area between z = -2 and z = +2 is .4772 + .4772 = .9544.

5d. What proportion of people would be expected to score below 10?

Answer: for X = 10, z = (10 -30)/10 = -2.00

The area below z = -2.00 is .0228

Thus about .0228 of people would be expected to score below 10.

5e. What proportion of people would be expected to score below 50?

For X = 50, z = (50 – 30)/10 = +2.00

The area above z = +2.00 is .0228.

The proportion of persons expected to score below 50 is 1- .0228 = .9772

5f. What proportion of people would be expected to score above 50?

For X = 50, z = (50 – 30)/10 = +2.00

The area above z = +2.00 is .0228.

The proportion of persons expected to score above 50 is .0228

5g. What proportion of people would be expected to score either below 10 or above 50?

Answer: This is the sum of the areas in 5d and 5f, .0228 + .0228 = .0456

5h. Assuming that a score in the top 5% of the distribution would be considered extremely anxious, would a person whose anxiety score was 50 be considered extremely anxious?

Answer: Yes

6. What is a Confidence Interval, and what information is required to set up a confidence interval?

Answer: A Confidence Interval is an interval estimate for μ, the population mean; it is based on the following information:

(a) the sample mean, M

(b) Information about the standard deviation of the individual scores. If σ is known, we use σ to find a value for σ_M. If σ is not known, we use the sample standard deviation s to compute SE_M and use SE_M to estimate σ_M.

(c) N, the number of scores in the sample.

(d) A Confidence Level (usually 99%, 95%, or 90%) is arbitrarily selected.

(e) Critical values from the z (standard normal) distribution are used when σ_M is known or when N is large. If σ_M is not known and N is small, critical values from the t distribution with N – 1 df are used to identify the distances from the center of the sampling distribution of M that correspond to the middle 95% of the area (if the Confidence Level is 95%).

The lower limit of the CI is obtained as follows:

*M – (critical value from z or t distribution) ** σ_M *or* SE_M

The upper limit of the CI is obtained as follows:

*M + (critical value from z or t distribution) ** σ_M *or* SE_M

7. What is a sampling distribution? What do we know about the shape and characteristics of the sampling distribution for M, the sample mean?

Answer: A sampling distribution (for M, the sample mean) describes how values of M vary across samples of size N that are all drawn randomly from the same population. We know that the sampling distribution of M has the following characteristics:

It has a normal distribution shape (This is true even for small values of N when the original X scores are normally distributed. If the original individual X scores are not normally distributed, the sampling distribution of M is still normal in shape provided that the size of the samples, N, is reasonably large (N > 30).

It has a mean of μ *where* μ *is the mean of the population of individual X scores.*

It has a standard deviation or standard error of σ_M*, where* $\sigma_M = \dfrac{\sigma}{\sqrt{N}}$ *(When* σ*, the population standard deviation of the individual X scores, is not known, the sample standard deviation s is used to estimate it).*

8. What is SE_M? What does the value of SE_M tell you about the typical magnitude of sampling error?

a. As s increases, how does the size of SE_M change (assuming that N stays the same)?

b. As N increases, how does the size of SE_M change (assuming that s stays the same)?

Answer for 8: SE_M is calculated as follows: $SE_M = \dfrac{s}{\sqrt{N}}$

SE_M is the sample estimate of σ_M

The value of SE_M tells us about distances of values of M from the true population mean, μ. The larger SE_M is, the greater the amount of sampling error, that is, the more widely dispersed the values of M are around the population mean μ.

8a. As s increases, how does the size of $\mathrm{SE_M}$ change (assuming that N stays the same)

Answer for 8a: As s increases, other factors being equal, SE_M increases.

8b. As N increases, how does the size of $\mathrm{SE_M}$ change (assuming the s stays the same)

Answer for 8b: As N increases, if other values remain the same, the size of SE_M decreases.

9. How is a t distribution similar to a standard normal distribution score? How is it different?

Answer: Both t and the standard normal distribution are symmetrical and "bell shaped". However, a t distribution is flatter than a standard normal distribution, with thicker tails. Thus, the top 2.5% of the area in a t distribution corresponds to a distance from the mean (a t critical value) that is larger than the value of z that corresponds to the top 2.5% of the area. The shape of t depends on its df; there is a different t distribution for each possible df value. As N and df increase, t becomes closer to the normal distribution.

10. Under what circumstances should a t distribution be used rather than the normal distribution to look up areas or probabilities associated with distances from the mean?

Answer: The t distribution is used rather than the standard normal when σ is not known (and we use s to estimate σ); however, if N is large, for example, N > 100, the shape of the t distribution converges to the normal distribution shape, so for large sample sizes, the normal distribution can be used to obtain critical values or distances from the mean that correspond to specific tail areas.

11. Consider the following questions about Confidence Intervals (CI).

A researcher tests Emotional Intelligence (EI) for a random sample of children selected from a population of all students who are enrolled in a school for gifted children. The researcher wants to estimate the mean EI for the entire school.

Let's suppose that a researcher wants to set up a 95% CI for IQ scores using the following information. The population standard deviation σ for EI is not known.

The sample mean, M, is M = 130.

The sample standard deviation, s = 15.

The sample size, N, is N = 120.

The df = N – 1 = 119

Because the df is 119, the "$t_{critical}$" values that are used to indicate the "distance from the mean" that corresponds to the middle 95% of the sampling distribution are values of -1.96 (for the lower limit) and +1.96 (for the upper limit of the CI). See the table of critical t values in Appendix B. Use the df in the table that is closest to the df in the problem (in this case, df = 100 in the table is the closest to df = 119 in the problem).

The standard error of the mean, SE_M, is computed by taking s divided by the square root of N.

To set up the 95% CI, the researcher uses these equations:

Lower limit = $M - t_{critical} * SE_M$

Upper limit = $M + t_{critical} * SE_M$

For the values given above, the limits of the 95% CI are:

Lower limit = 130 – 1.96* 1.37 = 127.31

Upper limit = 130 + 1.96* 1.37 = 132.69

Now let's "experiment" to see how changing some of the values involved in computing the CI influences the width of the CI

Recalculate the Confidence Interval above to see how the lower and upper limits (and the width of the CI) change, as you vary the N in the sample (and leave all the other values the same).

11a. What are the upper and lower limits of the CI and the width of the 95% CI if all of the other values remain the same (M = 130, s = 15) , but you change the value of N to 16?

Answer for 11a: $SE_M = s/\sqrt{N} = 15/\sqrt{16} = 15/4 = 3.75$

For middle 95% of the t distribution with $df = N - 1 = 16 - 1 = 15$, $t_{critical} = 2.13$

For N = 16, Lower limit=__122.01______ Upper limit = ____137.99_____

Width (Upper limit – lower limit) = ____15.98___________________

Note that when you change N, you need to change two things: the computed value of σ_M, and the degrees of freedom used to look up the critical values for t. See the attached table for the critical values of t that are needed for different values of the df and different levels of confidence.

11b. What are the upper and lower limits of the CI and the width of the 95% CI if all of the values remain the same, but you change the value of N to 25?

Answer for 11b: $SE_M = 15/\sqrt{25} = 15/5 = 3.00$

For middle 95% of area of a t distribution with df = 25 – 1 = 24, $t_{critical} = 2.06$

For N = 25, Lower limit=___123.82___________ Upper limit = ____136.18________

Width (Upper limit – lower limit) = _______12.36________________

11c. What are the upper and lower limits of the CI and the width of the 95% CI if all of the other values remain the same (M = 130, s = 15), but you change the value of N to 49?

Answer for 11c: $SE_M = 15/\sqrt{49} = 15/\ 7 = 2.14$

For middle 95% of area of a t distribution with df = 49 – 1 = 48, $t_{critical} = 2.01$

For N = 49, Lower limit=______125.69____ Upper limit = __134.307___________

Width (Upper limit – lower limit) = ____8.61___________________

11d. Based on the numbers you reported for Ns of 16, 25, and 49: How does the width of the CI change as the N (number of cases in the sample) increases?

Answer for 11d: As N increases (other factors being equal), the width of the CI decreases.

11e. What are the upper and lower limits, and the width of this CI, if you change the "Confidence" level to 80% (and continue to use M = 130, s= 15 and N = 49)

Answer for 11e: $SE_M = 15/\sqrt{49} = 15/\ 7 = 2.14$

For middle 80% of area of a t distribution with df = 49 – 1 = 48, $t_{critical} = 1.30$

For an 80% CI, Lower limit=___127.22_____ Upper limit = ____132.782________

Width (Upper limit – lower limit) = ___5.562____________________

11f. What are the upper and lower limits, and the width of the CI, if you change the confidence level to 99%? (continue to use M = 130, s = 15 and N = 49)

Answer for 11f: For middle 99% of area of a t distribution with df = 49 – 1 = 48, $t_{critical}$ *= 2.68*

For a 99% CI, Lower limit=___124.27_______ Upper limit = __135.73__________

Width (Upper limit – lower limit) = _____11.46_________________

11g. How does increasing the level of "Confidence" from 80% to 99% affect the width of the CI?

Answer for 11g: As the level of confidence increases, the width of the CI increases.

12. In your own words: what does a Sum of Squares (SS) tell us about a set of data?

Answer: SS summarizes information about the distances of individual X scores from the sample mean.

13. Under what circumstances will the value of SS equal 0?

Answer: If all of the X scores in a sample are equal to each other and therefore equal to the sample mean, SS = 0.

Can SS ever be negative?

Answer: No, because it is a sum of squared terms; squared terms can be zero but they cannot be negative.

Why do we divide by (N -1) rather than N when we compute a variance from an SS term?

Answer: Because we only have (N – 1) independent deviations from the mean.

14. Data Analysis Project.

The N = 130 scores in the TEMPHR.SAV file are hypothetical data created by Shoemaker (1996) so that they yield results similar to those obtained in an actual study of temperature and heart rate (Mackowiak, Wasserman, & Levine, 1992).

Use the Temperature data in the TEMPHR.SAV file to do the following:

Note that temperature in degrees Fahrenheit (tempF) can be converted into temperature in degrees Centigrade (tempC) by the following: tempC = (tempF – 32) / 1.8

The following analyses can be done on tempf, tempc, or both tempf and tempc.

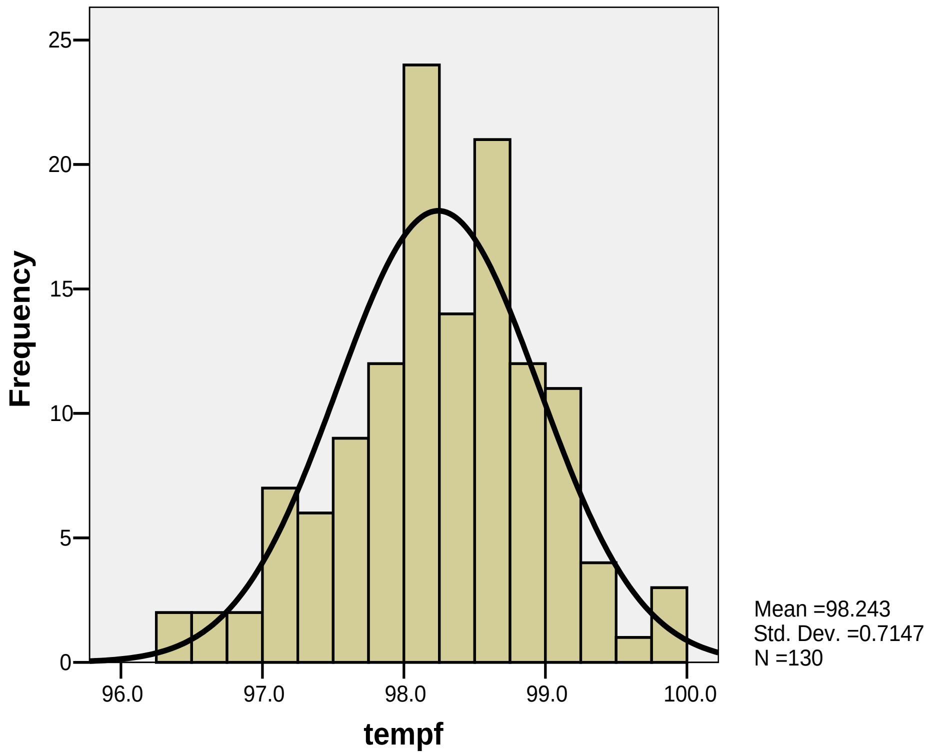
15a. Find the sample mean M, standard deviation s, and standard error of the mean SE_M for scores on temperature.

Answer for 15a from SPSS:

Statistics

		tempf	tempc
N	Valid	130	130
	Missing	0	0
Mean		98.243	36.8017
Std. Error of Mean		.0627	.03483
Std. Deviation		.7147	.39707

15b. Examine a histogram of scores on temp. Is the shape of the distribution reasonably close to normal?



Answer: Yes, this is reasonably close to a normal distribution.

15c. Set up a 95% CI for the sample mean, using your values of M, s, and N (N = 130 in this data set).

For both: use z = -1.96 and z = + 1.96 to correspond to middle 95% of the area of the sampling distribution.

Answer for Fahrenheit scale scores:

Limits of 95% CI: $M \pm 1.96 * SE_M = 98.243 \pm 1.96 * .0627$

$98.243 \pm .1229 = 98.12$ *to* 98.37

Answer for Centigrade scale scores:

Limits of 95% CI: $M \pm 1.96 * SE_M = 36.80 \pm 1.96 * .0348$

$36.80 \pm .0682 = 36.73$ *to* 36.87

15d. The temperature that is popularly believed to be “average” or “healthy” is 98.6 degrees Fahrenheit (or 37 degrees Centigrade). Does the 95% CI based on this sample include this value that is widely believed to represent an “average/ healthy” temperature? What conclusion do you draw from this?

Answer: The 95% CI from this sample does not include this “healthy” or “hypothesized population average” value (of 98.6 degrees Fahrenheit or 37 degrees Centigrade). The failure to include this “normative value” might be due to sampling error. However, if this result can be replicated across many other samples drawn from a normal healthy population, we might eventually conclude that the true population mean for body temperature is actually a little lower than the “standard” value of 98.6 degrees Fahrenheit that is often used to define normal or healthy body temperature.