LEARNING CHECK SOLUTIONS

CHAPTER 1

(p. 8)

Learning Check 1.3. Identify the independent and dependent variables in the following hypotheses:

- *Older Americans are more likely to support stricter immigration laws than younger Americans.*
- *People who attend church regularly are more likely to oppose abortion than people who do not attend church regularly.*
- *Elderly women are more likely to live alone than elderly men.*
- *Individuals with postgraduate education are likely to have fewer children than those with less education.*

What are the independent and dependent variables in your hypothesis?

Answer:

Independent	Dependent
Age	Support for stricter immigration laws
Church attendance	Opposition to abortion
Gender	Living arrangement
Educational attainment	Number of children

(p. 10)

Learning Check 1.4. Review the definitions of exhaustive and mutually exclusive. Now look at Table 1.2. What other categories could be added to each variable to be exhaustive and mutually exclusive?

Answer:

To the variable gender, we can include a transgender or gender neutral category. For the variable religion, we can include a category for those without any religion. For the marital status category, we could include a divorced category, though this would be covered under the Other category.

CHAPTER 2

(p. 38)

Learning Check 2.6. Inspect Table 2.12 and answer the following questions:

- *What is the source of this table?*
- *How many variables are presented? What are their names?*
- *What is represented by the numbers presented in the second column? In the last row of the table?*

Answer:

The source for the data is noted at the bottom of the table.

There are 10 variables listed in the first column of the table. The first variable name is "Prenatal care in first 3 months of pregnancy."

The second column corresponds to mothers who are Mexican immigrants. The numbers correspond to the percentage of these mothers who utilized each health and public assistance program.

The last row corresponds to WIC (Women, Infants, and Children Program) utilization.

CHAPTER 3

(p. 65)

Learning Check 3.1. Listed below are the political party affiliations of 15 individuals. Find the mode.

Democrat	Republican	Democrat	Republican	Republican
Independent	Democrat	Democrat	Democrat	Republican
Independent	Democrat	Independent	Republican	Democrat

Answer:

The mode is "Democrat," because this category has the highest frequency, which is 7.

(p. 69)

Learning Check 3.2. Find the median of the following distribution of an interval-ratio variable: 22, 15, 18, 33, 17, 5, 11, 28, 40, 19, 8, 20.

Answer:

First, we need to arrange the numbers: 5, 8, 11, 15, 17, 18, 19, 20, 22, 28, 33, 40.

$(N + 1)/2 = (12 + 1)/2 = 6.5$. So the median is the average of the sixth and the seventh numbers, which are 18 and 19. The median is 18.5.

$$\text{Median} = \frac{18+19}{2} = 18.5$$

(p. 74)

Learning Check 3.5. The following distribution is the same as the one you used to calculate the median in an earlier Learning Check: 22, 15, 18, 33, 17, 5, 11, 28, 40, 19, 8, 20. Can you calculate the mean? Is it the same as the median, or is it different?

Answer:

$$\text{Mean} = \frac{22+15+18+33+17+5+11+28+40+19+8+20}{12} = 19.67$$

So the mean, 19.67, is larger than the median, 18.5.

CHAPTER 4

(p. 100)

Learning Check 4.2. Why can't we use the range to describe diversity in nominal variables? The range can be used to describe diversity in ordinal variables (e.g., we can say that responses to a question ranged from "somewhat satisfied" to "very dissatisfied"), but it has no quantitative meaning. Why not?

Answer:

In nominal variables, the numbers are used only to represent the different categories of a variable without implying anything about the magnitude or quantitative difference between these categories. Therefore, the range, being a measure of variability that gives the quantitative difference between two values that a variable takes, is not an appropriate measure for nominal variables. Similarly, in ordinal variables, numbers corresponding with the categories of a variable are only used to rank-order these categories without having any meaning in terms of the quantitative difference between these categories. Therefore, the range does not convey any quantitative meaning when used to describe the diversity in ordinal variables.

(p. 103)

Learning Check 4.3. Why is the IQR better than the range as a measure of variability, especially when there are extreme scores in the distribution? To answer this question, you may want to examine Figure 4.3.

Answer:

Extreme scores directly affect the range—the difference between the highest and the lowest scores. Therefore, if a distribution has extreme (very high and/or very low) scores, the range does not provide an accurate description of the distribution. IQR, on the other hand, is not affected by extreme scores. Thus, it is a better measure of variability than the range when there are extreme scores in the distribution.

(p. 109)

Learning Check 4.4. Examine Table 4.8 again and note the disproportionate contribution of the Western region to the sum of the squared deviations from the mean (it actually accounts for about 45% of the sum of squares). Can you explain why? (Hint: It has something to do with the sensitivity of the mean to extreme values.)

Answer:

The Western region has the highest projected percentage change in the elderly population between 2008 and 2015, which is 27%. Therefore, it deviates more from the mean than the other regions. The more a category of a variable deviates from the mean, the larger the square of the deviation gets, and hence the more this category contributes to the sum of the squared deviations from the mean.

CHAPTER 5

(p. 135)

Learning Check 5.2. How many students obtained a score between 305 and 475?

Answer:

$0.4406 \times 1{,}108{,}165 = 488{,}257$

(p. 136)

Learning Check 5.3. Calculate the proportion of test takers who earned a SAT writing score of at least 400. What is the proportion of students who earned a score of 600 or higher?

Answer:

The proportion who earned a score of 400 or less is 0.2451 ($Z_{400} = -0.69$)

The proportion who earned a score of 600 or higher is 0.1251 ($Z_{600} = 1.15$)

(p. 138)

Learning Check 5.4. Which score corresponds to the top 5% of SAT writing test takers?

Answer:

We select a Z score of 1.65, corresponding to 0.45 (B) and 0.05 (C) areas. The formula is $Y = 475 + (1.65)(109) = 475 + 179.85 = 654.85$

(p. 141)

Learning Check 5.5. In a normal distribution, how many standard deviations from the mean is the 95th percentile?

Answer:

The number of standard deviations from the mean is what we call a Z score. The Z score associated with the 95th percentile is 1.65. So a score at the 95th percentile is 1.65 standard deviations above the mean.

(p. 142)

Learning Check 5.6. What is the raw SAT writing score associated with the 50th percentile?

Answer:

The raw score associated with the 50th percentile is the median.

(p. 143)

Learning Check 5.7. Review the mean math Z scores for the variable "conditions at age 7, 11, and 16" (the last column of Table 5.3). From ages 7, 11, and 16, was there an improvement in their math scores? Explain.

Answer:

Each group's mean math score is consistently lower than the overall mean, but over time, the distance from the mean is reduced (from –1.026 to –0.706). Recall that larger *Z* scores indicate a greater distance from the overall mean.

CHAPTER 6

(p. 157)

Learning Check 6.3. How does a systematic random sample differ from a simple random sample?

Answer:

In a systematic random sample, we select each case according to a predetermined number (the *K*th case). We would select each *K*th case for the study sample. For a simple random sample, there is no systematic selection. Selection could be based on a table of random numbers.

(p. 170)

Learning Check 6.6. Suppose a population distribution has a mean μ = *150 and a standard deviation* σ = *30, and you draw a simple random sample of* N = *100 cases. What is the probability that the mean is between 147 and 153? What is the probability that the sample mean exceeds 153? Would you be surprised to find a mean score of 159? Why? (Hint: To answer these questions, you need to apply what you learned in Chapter 5 about* Z *scores and areas under the normal curve [Appendix B].) Remember, to translate a raw score into a* Z *score we used this formula:*

$$Z = \frac{Y - \bar{Y}}{s}$$

However, because here we are dealing with a sampling distribution, replace Y *with the sample mean* $\bar{Y}$, $\bar{Y}$ *with the sampling distribution's mean* $\bar{Y}$, $\bar{Y}$ *and* σ *with the standard error of the mean*

$$Z = \frac{\bar{Y} - \mu_{\bar{Y}}}{s / \sqrt{N}}$$

Answer:

Z score equivalent of 147 is

$$\frac{147 - 150}{30/\sqrt{100}} = \frac{-3}{3} = -1$$

Z score equivalent of 153 is

$$\frac{153 - 150}{30/\sqrt{100}} = \frac{3}{3} = 1$$

Using the standard normal table (Appendix B), we can see that the probability of the area between the mean and a score 1 standard deviation above or below the mean is 0.3413. So the probability that the mean is between 147 and 153, both of which deviate from the mean by 1 standard deviation, is 0.6826 (0.3413 + 0.3413), or 68.26%.

The probability of the area beyond 1 standard deviation from the mean is 0.1587. So the probability that the mean exceeds 153 is 0.1587, or 15.87%.

Z score equivalent of 159 is

$$\frac{159 - 150}{30/\sqrt{100}} = \frac{9}{3} = 3$$

The probability of the area beyond 3 standard deviations from the mean, according to the standard normal table, is 0.0014. Therefore, it would be surprising to find a mean score of 159, as the probability is very low (0.14%).

CHAPTER 7

(p. 181)

Learning Check 7.1. What is the difference between a point estimate and a confidence interval?

Answer:

When the estimate of a population parameter is a single number, it is called a point estimate. When the estimate is a range of scores, it is called an interval estimate. Confidence intervals are used for interval estimates.

(p. 183)

Learning Check 7.2. To understand the relationship between the confidence level and Z, review the material in Chapter 6. What would be the corresponding Z value for a 98% confidence interval?

Answer:

The appropriate Z value for a 98% confidence interval is 2.33.

(p. 184)

Learning Check 7.3. What is the 90% confidence interval for the mean commuting time? First, find the Z value associated with a 90% confidence level.

Answer:

$$\begin{aligned} 90\%\ \text{CI} &= 7.5 \pm 1.65(0.07) \\ &= 7.5 \pm 0.12 \\ &= 7.38 \text{ to } 7.62 \end{aligned}$$

(p. 188)

Learning Check 7.4. Why do smaller sample sizes produce wider confidence intervals? (See Figure 7.5.) Compare the standard errors of the mean for the three sample sizes.

Answer:

As the sample size gets smaller, the standard error of the mean gets larger, which, in turn, results in a wider confidence interval.

(p. 194)

Learning Check 7.5. Calculate the 95% confidence interval for the CNN/ORC survey results for those who do not support anti-trans bathroom legislation (refer to page 179).

Answer:

$$.39 +\!- 1.96\ (.02) = .35 - .43$$

CHAPTER 8

(p. 216)

Learning Check 8.2. For the following research situations, state your research and null hypotheses:

- *There is a difference between the mean statistics grades of social science majors and the mean statistics grades of business majors.*
- *The average number of children in two-parent black families is lower than the average number of children in two-parent nonblack families.*
- *Grade point averages are higher among girls who participate in organized sports than among girls who do not.*

Answer:

Null Hypothesis	Research Hypothesis
Means are presumed equal for all statements.	Two-tailed test. No direction is stated.
	One-tailed test, left.
	One-tailed test, right.

(p. 219)

Learning Check 8.3. Would you change your decision in the previous example if alpha were.01? Why or why not?

Answer:

For alpha = .01, the t critical is 2.576. Also, t obtained (2.51) is less than t critical. We would fail to reject the null hypothesis.

(p. 221)

Learning Check 8.4. State the null and research hypothesis for this SPSS example. Would you change your decision in the previous example if alpha was .01? Why or why not?

Answer:

The null hypothesis is: $\mu_1 = \mu_2$

The research hypothesis is: $\mu_1 \neq \mu_2$

If alpha was set at .01, we would reject the null hypothesis. The t obtained of −3.444 (equal variances assumed) is significant at the .001 < .01.

(p. 223)

Learning Check 8.5. If alpha was changed to .01, two-tailed test, would your final decision change? Explain.

Answer:

The probability of the obtained Z is .0002 < .01. We would still reject the null hypothesis of no difference.

(p. 223)

Learning Check 8.6. If alpha was changed to .01, one-tailed test, would your final decision change? Explain.

Answer:

The probability of the obtained Z is .0062 < .01. We would still reject the null hypothesis of no difference.

(p. 225)

Learning Check 8.7. Based on Table 8.4, what would be the t critical at the .05 level for the first indicator, EC Index? Assume a two-tailed test.

Answer:

The *N*s are reported as a Note in the bottom of the table. The *df* calculation would be (78 + 113) – 2 = 189. Based on Appendix C, $df = \infty$, *t* critical is 1.960.

CHAPTER 9

(p. 235)

Learning Check 9.1. Identify the independent and dependent variables in Examples 2 and 3.

Answer:

Example 2: The independent variable is *race* and dependent variable is *receipt of public aid*.

Example 3: The independent variable is *race* and *health insurance status* is the dependent variable.

(p. 238)

Learning Check 9.2. Examine Table 9.2. Make sure you can identify all the parts just described and that you understand how the numbers were obtained. Can you identify the independent and dependent variables in the table? You will need to know this to convert the frequencies to percentages.

Answer:

The dependent variable is *homeownership* and the independent variable is *race*.

(p. 243)

Learning Check 9.4. Using the percentages reported under "Time in El Paso" in Table 9.5, calculate the cumulative percentages for each homeless group. What percentage of each group was in El Paso for a year or less? Which group has been in El Paso for a longer period of time?

Answer:

Almost 10% of Hispanic homeless were in El Paso for a year or less compared with 22.5% of Non-Hispanic homeless. The Hispanic homeless have been in El Paso longer than the non-Hispanic homeless. Seventy percent were in El Paso for more than 5 years. Only 43% of non-Hispanic homeless were in El Paso for the same period of time. Note that the total percentages do not equal 100.

	Hispanic	Cumulative % (Hispanic)	Non-Hispanic	Cumulative % (Non-Hispanic)
Less than 4 months	3.4	3.4	13.8	13.8
4–6 months	3.2	6.6	4.6	18.4

(Continued)

(Continued)

	Hispanic	Cumulative % (Hispanic)	Non-Hispanic	Cumulative % (Non-Hispanic)
7–12 months	3.2	9.8	4.1	22.5
1–2 years	6.9	16.7	11.5	34
2–5 years	11.9	28.6	17.5	51.5
Above 5 years	70.0	98.6	43.3	94.8

CHAPTER 10

(p. 272)

Learning Check 10.1. Construct a bivariate table (in percentages) showing no association between age and first-generation college status.

Answer:

Age and First-Generation College Status

	19 Years or Younger	20 Years or Older	
Firsts	41.9%	41.9%	41.9% (1,934)
Nonfirsts	58.1%	58.1%	58.1% (2,683)
	100.0%	100.0%	4,617

(p. 273)

Learning Check 10.2. Refer to the data in the previous Learning Check. Are the variables age and first-generation college status statistically independent? Write out the research and the null hypotheses for your practice data.

Answer:

Null hypothesis: There is no association between age and first-generation college status.

Research hypothesis: Age and first-generation college status are statistically dependent.

(p. 274)

Learning Check 10.3. Refer to the data in the Learning Check on page 272. Calculate the expected frequencies for age and first-generation college status and construct a bivariate table. Are your column and row marginals the same as in the original table?

Answer:

Construct a table to calculate chi-square for age and educational attainment.

	f_o	f_e	$f_o - f_e$	$(f_o - f_e)^2$	$(f_o - f_e)^2/f_e$
19/Firsts	916	1138.53	−222.53	49519.60	43.49
19/Nonfirsts	1802	1579.47	222.53	49519.60	31.35
20/Firsts	1018	795.47	222.53	49519.60	62.25
20/Nonfirsts	–881	1103.53	−222.53	49519.60	44.87

Chi-square = 181.96, with Yates correction = 181.15.

(p. 277)

Learning Check 10.5. Based on Appendix D, identify the probability for each chi-square value (df in parentheses)

Answer:

- *12.307 (15)* Between .70 and .50
- *20.337 (21)* Exactly .50
- *54.052 (24)* Less than .001

(p. 279)

Learning Check 10.6. What decision can you make about the association between age and first-generation college status? Should you reject the null hypothesis at the .05 alpha level or at the .01 level?

Answer:

We would reject the null hypothesis of no difference. Our calculated chi-square is significant at the .05 and the .01 levels. We have evidence that age is related to first-generation college status—older students are more likely to be first-generation students than younger students. Fifty-four percent of students 20 years or older are first-generation students versus 34% of students 19 years or younger.

(p. 282)

Learning Check 10.7. For the bivariate table with age and first-generation college status, the value of the obtained chi-square is 181.15 with 1 degree of freedom. Based on Appendix D, we determine that its probability is less than .001. This probability is less than our alpha level of .05. We reject the null hypothesis of no relationship between age and first-generation college status. If we reduce our sample size by half, the obtained chi-square is 90.58. Determine the p *value for 90.58. What decision can you make about the null hypothesis?*

Answer:

Even if we reduce the chi-square by half, we would still reject the null hypothesis.

CHAPTER 11

(p. 304)

Learning Check 11.1. Identify the independent and dependent variables in Table 11.1.

Answer:

Race/ethnicity is the independent variable, educational attainment is the dependent variable.

(p. 312)

Learning Check 11.3. If alpha were changed to .01, would our final decision change?

Answer:

If alpha were changed to .01, the F critical would be 5.01. Our F obtained is greater than the F critical. We would still reject the null hypothesis of no difference.

(p. 313)

Learning Check 11.4. Calculate eta-squared for the model presented in Figure 11.2.

Answer:

Based on the output $\eta^2 = 10.408/1300.902 = .008$ or .01. If we know political party identification, we can predict attitudes about ethical consumerism with 1% accuracy. This is a very weak prediction model.

(p. 316)

Learning Check 11.5. For the ANOVA model for Identity Formation, what is the F critical? What information do you need to determine the F *critical? Assume alpha = .05.*

Answer:

You would need k and N. In this case, $k = 3$ (number of categories) and $N = 486$ (total sample size). There are two degrees of freedom to calculate: df (between) $= k - 1 = 3 - 1 = 2$ and df(within) $= N - k = 486 - 3 = 483$. For an alpha of .05, the F critical is 2.99 (based on Appendix E).

CHAPTER 12

(p. 329)

Learning Check 12.1. For each of these four lines, as X *goes up by 1 unit, what does* Y *do? Be sure you can answer this question using both the equation and the line.*

Answer:

For the line $Y = 1X$, as X goes up by 1 unit, Y also goes up by 1 unit. In the second line, $Y = 2 + 0.5X$, Y increases by 0.5 units as a result of 1-unit increase in X. The line $Y = 6 - 2X$ tells that every 1-unit increase in X results in 2-unit decrease in Y. Finally, in the fourth line, Y decreases by 0.33 units as a result of 1-unit increase in X.

(p. 335)

Learning Check 12.2. Use the prediction equation to calculate the predicted values of Y *if* X *equals 9, 11, and 14. Verify that the regression line in Figure 12.4 passes through these points.*

Answer:

$$Y = -4.68 + .62(9) = .90$$

$$Y = -4.68 + .62(11) = 2.14$$

$$Y = -4.68 + .62(14) = 4.00$$

(p. 342)

Learning Check 12.3. Calculate r and r^2 for the age and Internet hours regression model. Interpret your results.

Answer: $r = -.77$ and $r^2 = .88$. The correlation coefficient indicates a strong negative relationship between age and Internet hours per week. Using information about respondent's age helps reduce the error in predicting Internet hours by 88%.

(p. 345)

Learning Check 12.4. Test the null hypothesis that there is a linear relationship between age and Internet hours. The mean squares regression is 63.66 with 1 degree of freedom. The mean squares residual is 2.355 with 8 degrees of freedom. Calculate the F *statistic and assess its significance.*

Answer: $F = 63.66/2.35 = 27.09$. This exceeds the F critical (1, 8) of 5.32 (alpha = .05). We can reject the null hypothesis and conclude that the linear relationship between age and Internet hours per week as expressed in r^2 is greater than zero in the population.

(p. 349)

Learning Check 12.5. Use the prediction equation describing the relationship between Internet hours per week and both educational attainment and age to calculate Internet hours per week for someone with 20 years of education who is 35 years old.

Answer:

$$Y = -.605 + .491(20) + -.057(35) = -.605 + 9.82 + -1.20 = 8.02 \text{ Internet hours per week}$$