[A] Circle whether each statement is true or false.

T F   1. A correlation coefficient of 10 is impossible.

T F   2. The correlation between children's heights and reading speeds is positive.

T F   3. The relationship between children's heights and reading speeds is causal.

T F   4. Small samples are less likely than large samples to be statistically significant.

T F   5. A negative residual means the predicted value is lower than the actual value.

T F   6. In a negative correlation, low values of *x* tend to be paired with low values of *y*.

T F   7. If a new hypothesis is made based on the results of a sample, a new sample must be collected in order to test the new hypothesis.

T F   8. If a data set is statistically significant, this means that the relationship found in the sample is likely to be true for the population as well.

[B] Explain each underlined error.

I analyzed the relationship between students' scores on the PreCalculus fall final and their scores on the Statistics fall final the following year. I looked at 10 students who had taken both finals. The correlation coefficient was ① r = .58, and the line of best fit was ② .56x + 92. It was a positive correlation, ③ meaning that most students did better on their Statistics final than on their PreCalculus final. The *p* value was .04, ④ so there is a strong correlation between these two variables. There are a number of possible explanations for this correlation. ⑤ One reason could be that students are more enthusiastic about Statistics than PreCalculus. Another is that Statistics is all open notes, whereas in PreCalculus students have to memorize formulas and procedures. The correlation could also be because of random variables like how much sleep they happened to get before the finals. ⑥ One possible causal reason for the correlation is that students who got low scores on their PreCalculus fall final would try harder the next year on their Statistics final.

1. Variables should be…

2. Equations must have…

3. The positive correlation actually means…

4. A *p* value below .05 does not mean a correlation is strong, it means…

5. These do not explain a correlation because…

6. This would explain a correlation that was…

[C] Fill in each blank with "a" or "e".

1. Kids who do preschool tend to do better in elementary school. However, this does not imply that preschool actually __ffects elementary school performance. It is possible that preschool has little or no __ffect on elementary school performance, and that the relationship is primarily due to confounding variables. For example, wealthier parents are more likely able to afford preschool, and more educated parents are more likely to see value in preschool. Parental wealth and education are variables that can __ffect school performance the same way they __ffect preschool attendance. This does not mean that preschool is not an __ffective way to prepare for elementary school, but it does mean that we are not able to determine the __ffect of preschool using only correlational data.

[D] The table below shows fall PreCalculus final exam scores (*x*) and fall Statistics final exam scores (*y*) for 10 students. (Their names have been changed.)

1. Find the following for the data set, and precisely explain, without using pronouns, exactly what each means about the sample and/or the population.

a) equation of the line of best fit

b) correlation coefficient

Student	x	y
Aly	195	163
Alex	177	189
Zane	105	159
Mary	189	203
Nelle	126	138
Geoff	162	211
Penny	153	210
Nathan	126	136
Tommy	171	203
Richard	141	174

c) *p* value

2. Explain how each of the following could be a reason for the correlation between PreCalculus final exam scores and Statistics final exam scores.

a) coincidence

b) causation

c) reverse causation

d) confounding variables

[E] Do the following to organize your group's reviews.

1. Make sure your name and your partners' names are at the top of your review the first day.

2. Staple the reviews in order, all facing the same way. Put the staple in the very top left corner if everyone is finished or if the review is due; otherwise put the staple in the top right corner.

[A] Circle whether each statement is true or false.

T F   1. A correlation coefficient of 10 is impossible.

T F   2. The correlation between children's heights and reading speeds is positive.

T F   3. The relationship between children's heights and reading speeds is causal.

T F   4. Small samples are less likely than large samples to be statistically significant.

T F   5. A negative residual means the predicted value is lower than the actual value.

T F   6. In a negative correlation, low values of x tend to be paired with low values of y.

T F   7. If a new hypothesis is made based on the results of a sample, a new sample must be collected in order to test the new hypothesis.

T F   8. If a data set is statistically significant, this means that the relationship found in the sample is likely to be true for the population as well.

[B] Explain each underlined error.

I analyzed the relationship between students' scores on the PreCalculus fall final and their scores on the Statistics fall final the following year. I looked at 10 students who had taken both finals. The correlation coefficient was ① r = .58, and the line of best fit was ② .56x + 92. It was a positive correlation, ③ meaning that most students did better on their Statistics final than on their PreCalculus final. The p value was .04, ④ so there is a strong correlation between these two variables. There are a number of possible explanations for this correlation. ⑤ One reason could be that students are more enthusiastic about Statistics than PreCalculus. Another is that Statistics is all open notes, whereas in PreCalculus students have to memorize formulas and procedures. The correlation could also be because of random variables like how much sleep they happened to get before the finals. ⑥ One possible causal reason for the correlation is that students who got low scores on their PreCalculus fall final would try harder the next year on their Statistics final.

1. Variables should be…

2. Equations must have…

3. The positive correlation actually means…

4. A p value below .05 does not mean a correlation is strong, it means…

5. These do not explain a correlation because…

6. This would explain a correlation that was…

[C] Fill in each blank with "a" or "e".

1. Kids who do preschool tend to do better in elementary school. However, this does not imply that preschool actually __ffects elementary school performance. It is possible that preschool has little or no __ffect on elementary school performance, and that the relationship is primarily due to confounding variables. For example, wealthier parents are more likely able to afford preschool, and more educated parents are more likely to see value in preschool. Parental wealth and education are variables that can __ffect school performance the same way they __ffect preschool attendance. This does not mean that preschool is not an __ffective way to prepare for elementary school, but it does mean that we are not able to determine the __ffect of preschool using only correlational data.

[D] The table below shows quiz points (*x*) and test points (*y*) for nine students in PreCalculus fall semester. (Their names have been changed.)

1. Find the following for the data set, and precisely explain, without using pronouns, exactly what each means about the sample and/or the population.

a) equation of the line of best fit   b) correlation coefficient

Student	*x*	*y*
Ed	255	503
Bob	247	395
Tony	230	414
Claire	281	555
Jessica	280	472
Patricia	282	548
Isabella	280	424
Marissa	245	465
Stephan	244	484

c) *p* value

2. Explain how each of the following could be a reason for the correlation between participation points and semester grades.

a) coincidence   b) causation

c) reverse causation   d) confounding variables

[E] Bonus.

1. Find and cite a print or online source that inappropriately assumes causation based on correlational data, and explain an alternative explanation for the correlation. Do not use the same one as someone else.

[A] Circle whether each statement is true or false.

T F   1. A correlation coefficient of 10 is impossible.

T F   2. The correlation between children's heights and reading speeds is positive.

T F   3. The relationship between children's heights and reading speeds is causal.

T F   4. Small samples are less likely than large samples to be statistically significant.

T F   5. A negative residual means the predicted value is lower than the actual value.

T F   6. In a negative correlation, low values of *x* tend to be paired with low values of *y*.

T F   7. If a new hypothesis is made based on the results of a sample, a new sample must be collected in order to test the new hypothesis.

T F   8. If a data set is statistically significant, this means that the relationship found in the sample is likely to be true for the population as well.

[B] Explain each underlined error.

I analyzed the relationship between students' scores on the PreCalculus fall final and their scores on the Statistics fall final the following year. I looked at 10 students who had taken both finals. The correlation coefficient was ① r = .58, and the line of best fit was ② .56x + 92. It was a positive correlation, ③ meaning that most students did better on their Statistics final than on their PreCalculus final. The *p* value was .04, ④ so there is a strong correlation between these two variables. There are a number of possible explanations for this correlation. ⑤ One reason could be that students are more enthusiastic about Statistics than PreCalculus. Another is that Statistics is all open notes, whereas in PreCalculus students have to memorize formulas and procedures. The correlation could also be because of random variables like how much sleep they happened to get before the finals. ⑥ One possible causal reason for the correlation is that students who got low scores on their PreCalculus fall final would try harder the next year on their Statistics final.

1. Variables should be…

2. Equations must have…

3. The positive correlation actually means…

4. A *p* value below .05 does not mean a correlation is strong, it means…

5. These do not explain a correlation because…

6. This would explain a correlation that was…

[C] Fill in each blank with "a" or "e".

1. Kids who do preschool tend to do better in elementary school. However, this does not imply that preschool actually __ffects elementary school performance. It is possible that preschool has little or no __ffect on elementary school performance, and that the relationship is primarily due to confounding variables. For example, wealthier parents are more likely able to afford preschool, and more educated parents are more likely to see value in preschool. Parental wealth and education are variables that can __ffect school performance the same way they __ffect preschool attendance. This does not mean that preschool is not an __ffective way to prepare for elementary school, but it does mean that we are not able to determine the __ffect of preschool using only correlational data.

[D] The table below shows literacy rates (x) and life expectancies (y) for nine nations.

1. Find the following for the data set, and precisely explain, without using pronouns, exactly what each means about the sample and/or the population.

a) equation of the line of best fit

b) correlation coefficient

Country	x	y
Argentina	97	75
Brazil	89	72
China	92	73
Colombia	90	73
Ecuador	91	75
Egypt	72	73
India	74	64
Iran	85	72
Pakistan	55	65

c) p value

2. Explain how each of the following could be a reason for the correlation between national literacy rates and life expectancies.

a) coincidence

b) causation

c) reverse causation

d) confounding variables

[E] Bonus.

1. Find and cite a print or online source that inappropriately assumes causation based on correlational data, and explain an alternative explanation for the correlation. Do not use the same one as someone else.

[A] Circle whether each statement is true or false.

T F   1. A correlation coefficient of 10 is impossible.

T F   2. The correlation between children's heights and reading speeds is positive.

T F   3. The relationship between children's heights and reading speeds is causal.

T F   4. Small samples are less likely than large samples to be statistically significant.

T F   5. A negative residual means the predicted value is lower than the actual value.

T F   6. In a negative correlation, low values of *x* tend to be paired with low values of *y*.

T F   7. If a new hypothesis is made based on the results of a sample, a new sample must be collected in order to test the new hypothesis.

T F   8. If a data set is statistically significant, this means that the relationship found in the sample is likely to be true for the population as well.

[B] Explain each underlined error.

I analyzed the relationship between students' scores on the PreCalculus fall final and their scores on the Statistics fall final the following year. I looked at 10 students who had taken both finals. The correlation coefficient was ① r = .58, and the line of best fit was ② .56x + 92. It was a positive correlation, ③ meaning that most students did better on their Statistics final than on their PreCalculus final. The *p* value was .04, ④ so there is a strong correlation between these two variables. There are a number of possible explanations for this correlation. ⑤ One reason could be that students are more enthusiastic about Statistics than PreCalculus. Another is that Statistics is all open notes, whereas in PreCalculus students have to memorize formulas and procedures. The correlation could also be because of random variables like how much sleep they happened to get before the finals. ⑥ One possible causal reason for the correlation is that students who got low scores on their PreCalculus fall final would try harder the next year on their Statistics final.

1. Variables should be…

2. Equations must have…

3. The positive correlation actually means…

4. A *p* value below .05 does not mean a correlation is strong, it means…

5. These do not explain a correlation because…

6. This would explain a correlation that was…

[C] Fill in each blank with "a" or "e".

1. Kids who do preschool tend to do better in elementary school. However, this does not imply that preschool actually __ffects elementary school performance. It is possible that preschool has little or no __ffect on elementary school performance, and that the relationship is primarily due to confounding variables. For example, wealthier parents are more likely able to afford preschool, and more educated parents are more likely to see value in preschool. Parental wealth and education are variables that can __ffect school performance the same way they __ffect preschool attendance. This does not mean that preschool is not an __ffective way to prepare for elementary school, but it does mean that we are not able to determine the __ffect of preschool using only correlational data.

[D] The table below shows millions of iPhones sold (*x*) and millions of dollars of Black Friday online sales (*y*) for eight years.

1. Find the following for the data set, and precisely explain, without using pronouns, exactly what each means about the sample and/or the population.

a) equation of the line of best fit

b) correlation coefficient

Year	*x*	*y*
2008	12	534
2009	21	595
2010	40	648
2011	72	816
2012	125	1042
2013	150	1350
2014	169	1510
2015	231	2930

c) *p* value

2. Explain how each of the following could be a reason for the correlation between iPhone sales and Black Friday online sales.

a) coincidence

b) causation

c) reverse causation

d) confounding variables

[E] Bonus.

1. Find and cite a print or online source that inappropriately assumes causation based on correlational data, and explain an alternative explanation for the correlation. Do not use the same one as someone else.