Module 3 Practice problem and Homework answers

Practice Problems
Set A
Eating sugary cereal in the morning reduces children's attention span in school.

Answer: Independent variable = consumption of sugary cereal

This hypothesis suggests a child's attention span in school during the day is influenced by his/her breakfast that morning. Therefore, attention span depends on (among other things) the child's breakfast. If you think about it, a child's attention span during the day can't influence what he or she had for breakfast this morning, so we can conclude that it is the independent variable. Anytime one variable precedes the other in time (as in this question), it is the independent variable.

Workers' morale is improved when their bosses recognize them for working overtime.

Answer: Independent variable = whether or not a worker is asked to work overtime

In this situation, the worker's morale is influenced by the boss's actions.

Support for a political candidate is higher for voters from some racial/ethnic groups than for voters from other groups.

Answer: Independent variable = race of voters

As a rule of thumb, if race is one of the variables in a hypothesis, it is the independent variable, because it can't depend on anything else.

People with college degrees are more likely to say that their favorite type of music is jazz than are people without college degrees.

Answer: Independent variable = people's education level

This hypothesis suggests that taste varies depending on people's education level. Could the reverse be true? If not, then it's fairly easy to identify the dependent and independent variables in this case.

If a _____ is drawn randomly from a _____, its characteristics _____ the characteristics of the _____.

Answer: sample, population, will very closely match, population

A truly random sample drawn from a population will have the same characteristics (other than size) of the population. This is why we can draw conclusions about populations from statistical analyses run on samples.

Practice Problems Set A, continued

Which of the following is an example of a random sample?

Incorrect: a researcher surveys the first 50 people who leave a department store about their views on a particular social issue

There is very little chance that the people who leave a particular department store at a particular time of day are representative of any population (other than perhaps the people who shop at a particular department store at a particular time of day)

Incorrect: a school principal looks through the records for all of the students in her school to determine how many of them have not submitted required paperwork

In this case, the principal collects data from the entire population of the school, not a sample.

Incorrect: a research firm obtains a list of all of the land lines in a particular county, and calls every 10^{th} number to ask about support for a particular political candidate.

This would be a good way of obtaining a random sample if every person in the population had a land line, but we know that this will very likely not be true in any population

Correct: a firm's manager puts the name of each person who works at that company in a bucket, pulls out 15 names, and then asks those 15 people to state their preference about an impending policy change.

Because every person at the company (the population) has an equal chance of being selected for the sample, this method would be a valid way to obtain a random sample.

Which of the following is an example of a null hypothesis

Answer: There is no significant relationship between the amount of money spent on family vacation and how much fun the children report having on the trip.

Because this hypothesis suggests that there is not a significant relationship between variables, it is an example of a null hypothesis.

Which of the following is an example of a two-tailed research hypothesis?

Answer: Attendance at religious services varies significantly by race.

This hypothesis doesn't predict either a direct or indirect relationship between variables, but leaves open the possibility of either. Non-directional hypotheses like this one are classified as two-tailed.

Practice Problems Set A, continued

Which of the following is an example of a one-tailed research hypothesis?

Answer: Children who spend more than 45 minutes a day playing outside are significantly more likely than those who do not to fall asleep at night within 5 minutes of going to bed.

This hypothesis predicts that one group of children is *more likely* than another group to fall asleep quickly based on the amount of time spent outside. This prediction about one group being more likely to do something than the other makes this a one-tailed hypothesis.

Set B

In a normally-distributed dataset, you find that the mean is 10, the standard deviation is 1.5, and the range is 4 – 16. What is the median?

Answer: 10

In a normally-distributed dataset, the mean, median and mode all have the same value.

In the dataset mentioned in the previous question, what value lies 2 standard deviations above the mean?

Answer: 13

When the mean is 10, and the standard deviation is 1.5, the value at one standard deviation above the mean is 11.5 (10 + 1.5), and the value at two standard deviations above the mean is 13 (11.5 + 1.5, or 10 + 2*1.5)

In the dataset mentioned above, what percentage of values falls in the range between 8.5 and 11.5?

Answer: 68%

In a normally-distributed dataset, 68% of cases fall within one standard deviation of the mean. When the mean is 10 and the standard deviation is 1.5, the value at one standard deviation above the mean is 11.5 (10 + 1.5), and the value at one standard deviation below the mean is 8.5 (10 – 1.5).

In the dataset mentioned above, what percentage of values is greater than 13?

Answer: 2.5%

In a normally-distributed dataset, 95% of cases fall within two standard deviations of the mean, 2.5% fall above the 2 standard deviation mark, and 2.5% fall below the -2 standard deviation mark. As we determined two questions ago, the value at 2 standard deviations above the mean is 13. Therefore, 2.5% of cases fall above this point in the distribution.

Practice Problems Set B, continued

As the chances of making a Type I error increase, so do the chances of making a Type II error.

Answer: false

The lower our chances of making a Type I error get (the lower the significance level that we choose is), the higher our chance of making a Type II error get. That is, the harder we make it for ourselves to reject the null hypothesis, the easier we make it to fail to reject it.

Set C

If the critical value for the research problem you are working on is greater than the value of the test statistic that you have calculated, you have found a difference between groups that is large enough to be considered statistically significant.

Answer: false

We can only conclude that there is a significant relationship between variables when the calculated value of a test statistic exceeds the critical value. When this happens, we can be confident that the result that we found is so different from 0 that it could not have occurred because of random error in the sample selection process.

If your research hypothesis predicts that men are significantly more likely than women to leave their turn signals on after completing a turn, and you find evidence that allows you to reject the null hypothesis, which of the following can you conclude:

Answer: Men are significantly more likely than women to forget to turn off their turn signals.

When you find statistical evidence that allows you the reject the null hypothesis, you can conclude that the research hypothesis is accurate (unless you have a one-tailed research hypothesis and the results were opposite of what you expected, such as women being more likely than men to forget to turn off their signals)

Which of the following pieces of information about a research project is used to decide which test statistic should be calculated?

Answer: the levels of measurement of the variables in the hypotheses

Each statistical class that we'll cover in this course was designed to be used with variables measured at particular levels of measurement. Therefore, it is imperative that you have a strong grasp on the concept of levels of measurement so that you will be able to identify which is the appropriate statistical test for each problem.

Homework Problems
Set A

As time spent exercising increases, so does time spent on productive activity during the day.

Answer: Dependent variable = how much time a person spends on productive activity at work

This hypothesis suggests that a person's daily productivity *depends* on the length of time spent exercising.

Younger people use more toothpaste when brushing their teeth than older people.

Answer: amount of toothpaste used when brushing one's teeth

This hypothesis suggests that a person's toothpaste usage is influenced by his or her age. Think about this: could age be dependent on toothpaste usage?

Caffeine consumption while studying impacts exam scores.

Answer: exam scores

If one variable is predicted to impact another, the one being impacted is the dependent variable.

Which of the problems above represents a two-tailed hypothesis?

Answer: The exercise/productive activity question predicts that as one variable increases, the other one will too.

The toothpaste/age question predicts that as one variable increases, the other will decrease. Because these predict a particular direction, they are both one-tailed hypotheses. The caffeine consumption question leaves room for either a direct or an indirect relationship, making it a two-tailed hypothesis.

Which of the following is true?

Answer: A complication of working with sample data is that a researcher can never be completely sure that the conclusions he or she draws from statistical analyses are accurate.

In order to draw a conclusion from a sample about a population, the sample must be drawn randomly so that it can represent the population. Even under the best circumstances, random sample selection might not work out perfectly, resulting in one or more segments of the population being either over or underrepresented in the sample. This potential error in sampling means that conclusions drawn from a sample might not represent what is happening in the population perfectly.

Homework Problems Set A, continued

Which of the following represents a method for obtaining a random sample?

Incorrect: Standing outside a shopping mall and talking to individuals willing to be interviewed, making sure to get an equal number of men and women, as well as an adequate number of people with children.

This method might get a random sample of people who shop in shopping malls (but probably won't), but leaves the segment of the population who does not shop in shopping malls (which could be meaningful).

Incorrect: Randomly selecting 75 telephone numbers from a list of all of the phone numbers associated with smart phones in an area code.

This method is likely to get a random sample of every person with smart phones, but does not give any people without smart phones the chance to be in the sample. A representative sample must give every person in the population an equal chance of being selected. If the sampling technique leaves someone out, then it will not produce a truly random sample.

Incorrect: Collecting data from a large introductory psychology course.

Although this method is used in many, many psychological research projects, it is not a method for obtaining a random sample. Instead, it produces a very narrow sample of young adults who are in college and interested for various reasons in psychology.

Correct: Randomly selecting 100 student ID numbers from the entire student roster of a school

This method will produce a random sample of the school's population (but not the wider population, of course), assuming every student has an ID number, because every person at the school has an equal chance of being selected.

In what circumstance would a researcher make a Type II error?

Answer: Failing to reject the null hypothesis when the research hypothesis accurately reflects the situation in the population.

When a researcher concludes that her/his research hypothesis is not supported by the data when it actually his, the result is a Type II error.

Set B

If the mean of a normally-distributed dataset is 50, and the value that falls 2 standard deviations above the mean is 70, what is the value that falls 2 standard deviations below the mean?

Answer: 30

The normal distribution is symmetrical, so if the value at two standard deviations above the mean is 20 units greater than the mean (70 – 50), then the value at two standard deviations below the mean will be 20 units less than the mean (50 – 20).

In the dataset described just above, what percentage of cases falls between 50 and 60?

Answer: 34%

If the difference between the mean and the value that falls at two standard deviations above the mean is 20 (70 – 50), then the standard deviation is 10 (20/2). Therefore, 60 falls at one standard deviation above the mean (50 + 10). We know that 68% of cases falls within 1 standard deviation of the mean (on either side). In this problem, we're focusing on only the percentage of cases that falls between the mean and one standard deviation to the right, not the left plus the right, which is the 68% area. Therefore, we can divide 68% by 2 in order to find the percentage of cases that falls between the mean and one standard deviation. (Because the normal distribution is symmetrical, the answer would have been the same if the question had asked about the percentage of cases that falls between 40 and 50.)

If the scores in the dataset described above are scores on an exam, what score would a student have to get to be in the top 2.5% of the class?

Answer: 70

First, we have to remember that 95% of cases fall between -2 and +2 standard deviations around the mean. Second, we have to remember that the normal distribution is symmetrical. Therefore, we can divide the 5% that is leftover once we've marked the 95% within 2 standard deviations of the mean by 2, giving us 2.5% (5/2 = 2.5). This tells us that 2.5% of the cases are greater than the value at 2 standard deviations above the mean, and another 2.5% are less than the value at 2 standard deviations below the mean. Therefore, the cutoff point to make it into the top 2.5% of the class, in this case, is at the value that is 2 standard deviations above the mean, which is 70.