Probability & Statistics Workbook Solutions

Sampling

krista king MATH

TYPES OF STUDIES

1. The following table shows the age and shoe size of six children. Does the data have a positive correlation, negative correlation, or no correlation?

Age	Shoe size
3	7
3	6
5	9
6	12
6	11
7	13

Solution:

The data has a positive correlation because, as the age of the child increases, so does the size of shoe. Positive correlation occurs when two variables increase or decrease together, negative correlation occurs when one variable increases while the other decreases, and no correlation would have no discernible pattern.

2. A class conducts a survey and finds that 75% of the school spends 2 or more hours on social media each day. Is the data one-way or two-way data? Is the study observational or experimental?

Solution:

The survey only shows data for one variable for a set of individuals, the amount of time spent on social media, so the data is one-way data. The survey is an observational study because it records the results without manipulation.

■ 3. The following table shows the number of classes from which students were absent and their final grade in the class. Does the data have a positive correlation, negative correlation, or no correlation?

Number of absences	0	0	1	2	3	3	3	5	5	6	7	10
Final grade	95%	97%	90%	86%	80%	74%	70%	65%	64%	58%	55%	45%

Solution:

The data has a negative correlation because, as the number of absences increases, the final grade in the class decreases. Positive correlation occurs when two variables increase or decrease together, negative correlation occurs when one variable increases while the other decreases, and no correlation would have no discernible pattern.

4. The table below shows the favorite winter activities for 50 adults. Is the data one-way data? Why or why not?

	Skiing	Snowboarding	Ice Skating
Men	9	13	6
Women	8	7	7

Solution:

This is a two-way data table because we have the two categories of individuals: men and women, and the three categories of activities: skiing, snowboarding, and ice skating. We can use this data to examine the relationship between the two categorical variables.

5. Is the following experiment an example of a double-blind experiment? If not, what could be changed to make it a double-blind experiment?

> "A soda company has developed a new flavor and wants to know how it compares in taste to competitor sodas. An employee of the soda company conducts a survey where participants are asked which soda tastes the best. The sodas are given to participants in unmarked plastic cups by the employee."

Solution:

This experiment is an example of a blind experiment since the participants don't know which soda is being targeted. However, it's not a double-blind experiment since the employee of the soda company, who is also administering the survey, knows which soda is being targeted. To make it a double-blind experiment, the employee conducting the survey should have the sodas prepared by someone else so that neither the participants nor the employee administering the experiment know which soda is being targeted.

■ 6. A new cancer drug is being used to treat cancer in children and adults. The hospital conducts a study to measure the effectiveness of the new drug. Cancer patients are placed into groups according to their age and each age range is split into two groups. One group is given traditional treatment of the cancer and the other group is given the new drug. Is the data one-way or two-way data? Is the study observational or experimental?

Solution:

The data is two-way data because there's a control group and an experimental group, grouped according to age, and the data is about the effectiveness of the drug. It's an experimental study because the experimental group is being manipulated by receiving the new drug.

SAMPLING AND BIAS

1. The zoo conducts a survey on why patrons enjoy coming to the zoo. They ask families with children about why they like to visit the zoo as they're leaving. Give a reason why the sampling method may be biased.

Solution:

The sampling method is selection biased since the zoo is only surveying families with children. An unbiased sampling method would include all zoo patrons. For example, the zoo could survey every 10th customer as they leave.

2. The owner of a restaurant gives a survey to each customer. Included in the survey is the question "Have you ever not tipped your waiter or waitress?" Give a reason why the sampling method may be biased.

Solution:

The sampling method is response biased because some people may not want to answer the question about tipping truthfully. There might be less of a response bias if the wording were changed to, "Is there ever a circumstance where it's acceptable to not tip your waiter or waitress?"

3. A health club wants to purchase a new machine and would like to know which machine members would most like to have. It creates a survey where members can rate the different machines that the health club is considering purchasing, and posts it at the reception desk for members to fill out if they choose to do so. Does the sample contain a bias? If so, what kind?

Solution:

The sampling method is biased because of voluntary response sampling. People who voluntarily participate in the survey may have different habits, opinions, or tendencies than people who choose not to participate.

4. A biologist wants to study a group of prairie dogs for parasites, but cannot examine the entire population. Which sampling method would be better in this case, a stratified random sample or a clustered random sample?

Solution:

A clustered random sample would be better. The biologist could divide the field into different sections and take a random sample from each section. This would give the biologist a representative sample of the entire population. A stratified random sample would separate the prairie dogs by gender, age, or some other variable, and the results might vary based on those values.

■ 5. A hospital is studying the health effects of obesity. They group patients into different groups according to a specific weight range and study a variety of biometrics. What type of sampling is this?

Solution:

The sampling method is a stratified random sample because people are the same weight range within each group. A simple random sample would study a group of people picked randomly with no regards to weight range. A clustered random sample might select a random sampling of people from each wing of the hospital.

■ 6. A museum wants to find out the demographics of its patrons. They set up a survey and ask every 5th customer about their age, ethnicity, and gender. What type of sampling is this?

Solution:

This sampling method is a simple random sample. Patrons are randomly selected with no regard to groups or clusters.

SAMPLING DISTRIBUTION OF THE SAMPLE MEAN

■ 1. The population of 32 year-old women in the United States have an average salary of $\$42,000$, but the distribution of their salaries is not normally distributed. A random sample of 24 women is taken. Does the sample meet the criteria to use the central limit theorem?

Solution:

Our sample space should be no more than 10% of our population, the sample should be selected randomly, and (if the population is not normal) the sample size must be greater than 30.

The sample is random, 24 is definitely less than 10% of all 32 year-old women in the United States, but 24 isn't greater than 30 and the population is not normal. So the sample does not meet the criteria to use the central limit theorem.

■ 2. There are 130 dogs at a dog show who weigh an average of 11 pounds with a standard deviation of 3 pounds. A sample of 9 dogs is taken. What is the standard deviation of the sampling distribution of the sample mean?

Solution:

Find the standard deviation of the sampling distribution of the sample mean using $\sigma = 3$ and $n = 9$, making sure to use the finite population correction factor in the standard deviation formula.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$\sigma_{\bar{x}} = \frac{3}{\sqrt{9}}\sqrt{\frac{130-9}{130-1}}$$

$$\sigma_{\bar{x}} = \frac{3}{3}\sqrt{\frac{121}{129}}$$

$$\sigma_{\bar{x}} = 0.9685$$

■ 3. A large university population has an average student age of 30 years old with a standard deviation of 5 years, and student age is normally distributed. A sample of 80 students is randomly taken. What is the probability that the mean of their ages will be less than 29?

Solution:

Our sample space should be no more than $10\,\%$ of our population, the sample should be selected randomly, and (if the population is not normal) the sample size must be greater than 30.

The sample was collected randomly. It’s safe to assume that 80 students is less than 10% of the student population at a large university. The population is normal, so the sample size doesn’t have to be greater than 30, but 80 is greater than 30 anyway. The sample space meets the conditions of normality.

Find the standard deviation of the sampling distribution of the sample mean.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{x}} = \frac{5}{\sqrt{80}}$$

$$\sigma_{\bar{x}} = 0.559$$

We want to know the probability that the sample mean $\bar{x}$ is less than 29. We need to express this in terms of standard deviations.

$$\frac{29-30}{0.559} = \frac{-1}{0.559} = -1.79$$

This means we want to know the probability of $P(z < -1.79)$. Using a z-table, a z-value of -1.79 gives 0.0367, so $P(z < -1.79) = 3.67\%$. There’s a 3.67% chance that our sample mean will be less than 29.

■ 4. A cereal company packages cereal in 12.5-ounce boxes with a standard deviation of 0.5 ounces. The amount of cereal put into each box is normally distributed. The company randomly selects 100 boxes to check their weight. What is the probability that the mean weight will be greater than 12.6 ounces?

Solution:

Our sample space should be no more than 10% of our population, the sample should be selected randomly, and (if the population is not normal) the sample size must be greater than 30.

The sample was collected randomly. It's safe to assume that 100 boxes is less than 10% of the cereal boxes in the factory. The population is normal so the sample size doesn't have to be greater than 30, but 100 is greater than 30 anyway. The sample space meets the conditions of normality.

Find the standard deviation of the sampling distribution of the sample mean.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{x}} = \frac{0.5}{\sqrt{100}}$$

$$\sigma_{\bar{x}} = 0.05$$

We want to know the probability that the sample mean $\bar{x}$ is more than 12.6 ounces. We need to express this in terms of standard deviations.

$$\frac{12.5 - 12.6}{0.05} = \frac{0.1}{0.05} = 2$$

This means we want to know the probability of $P(z > 2)$.

Using the z-table, a z-value of 2 gives 0.9772, but we need to subtract this from 1 to find the probability that the sample mean is more than 12.6 ounces.

$$P(z > 2) = 1 - 0.9772$$

$$P(z > 2) = 0.0228$$

$$P(z > 2) = 2.28\,\%$$

There's a $2.28\,\%$ chance that our sample mean will be greater than 12.6 ounces.

5. A large hospital finds that the average body temperature of their patients is 98.4°, with a standard deviation of 0.6°, and we'll assume that body temperature is normally distributed. The hospital randomly selects 30 patients to check their temperature. What is the probability that the mean temperature of these patients $\bar{x}$ is within 0.2° of the population mean?

Solution:

Our sample space should be no more than $10\,\%$ of our population, the sample should be selected randomly, and (if the population is not normal) the sample size must be greater than 30.

The sample was collected randomly. It's safe to assume that 30 patients is less than $10\,\%$ of the total patients in a large hospital. The population is normal so the sample size doesn't have to be greater than 30. The sample space meets the conditions of normality.

Find the standard deviation of the sampling distribution of the sample mean.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{x}} = \frac{0.6}{\sqrt{30}}$$

$$\sigma_{\bar{x}} = 0.1095$$

We want to know the probability that the sample mean $\bar{x}$ is within 0.2° of the population mean. We need to express 0.2° in terms of standard deviations.

$$\frac{0.2}{0.1095} = 1.83$$

This means we want to know the probability of $P(-1.83 < z < 1.83)$.

Using a z-table, a z-value of -1.83 gives 0.0336 and a value of 1.83 gives 0.9664. The probability under the normal curve between these z-scores is

$$P(-1.83 < z < 1.83) = 0.9664 - 0.0336$$

$$P(-1.83 < z < 1.83) = 0.9328$$

$$P(-1.83 < z < 1.83) = 93.28\,\%$$

There's a $93.28\,\%$ chance that our sample mean will fall within 0.2° of the population mean of 98.4°.

■ 6. A company produces volleyballs in a factory. Individual volleyballs are filled to an approximate pressure of 7.9 PSI (pounds per square inch), with a standard deviation of 0.2 PSI. Air pressure in the volleyballs is normally distributed. The company randomly selects 50 volleyballs to check their pressure. What is the probability that the mean amount of pressure in these balls $\bar{x}$ is within 0.05 PSI of the population mean?

Solution:

Our sample space should be no more than $10\,\%$ of our population, the sample should be selected randomly, and (if the population is not normal) the sample size must be greater than 30.

The sample was collected randomly. It's safe to assume that 50 volleyballs is less than $10\,\%$ of all the volleyballs produced in the factory. The population is normal so the sample size doesn't have to be greater than 30, but 50 is greater than 30 anyway. The sample space meets the conditions of normality.

Find the standard deviation of the sampling distribution of the sample mean.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_{\bar{x}} = \frac{0.2}{\sqrt{50}}$$

$$\sigma_{\bar{x}} = 0.02828$$

We want to know the probability that the sample mean $\bar{x}$ is within 0.05 PSI of the population mean. We need to express 0.05 in terms of standard deviations.

$$\frac{0.05}{0.02828} = 1.77$$

This means we want to know the probability of $P(-1.77 < z < 1.77)$.

Using a z-table, a z-value of -1.77 gives 0.0384 and a z-value of 1.77 gives 0.9616. The probability under the normal curve between these z-scores is

$$P(-1.77 < z < 1.77) = 0.9616 - 0.0384$$

$$P(-1.77 < z < 1.77) = 0.9232$$

$$P(-1.77 < z < 1.77) = 92.32\,\%$$

There's a $92.32\,\%$ chance that our sample mean will fall within 0.05 PSI of the population mean of 7.9 PSI.

SAMPLING DISTRIBUTION OF THE SAMPLE PROPORTION

1. The state representatives want to know how their constituents feel about the new tax to fund road improvements, so they send out a survey. Of the 5 million who reside in the state, $150{,}000$ people respond. 40% disapprove of the new tax and 60% are in favor of the new tax because of the improvements they've seen to the roads. Does this sample meet the conditions for inference?

Solution:

Our sample space should be no more than 10% of our population, the expected number of successes and failures should each be at least 10, and the sample should be selected randomly.

The sample space was random, but may have a bias since it was a voluntary sample. The sample space is no more than 10% of the population:

$$\frac{150{,}000}{5{,}000{,}000} = 0.03 = 3\% \leq 10\%$$

And there are more than 10 expected successes and failures.

$$150{,}000(0.6) = 90{,}000 \geq 10$$

$$150{,}000(0.4) = 60{,}000 \geq 10$$

The sample space meets the conditions for inference. However, the voluntary bias should be noted and the direction of bias taken into account.

■ 2. An ice cream shop states that only 5% of their $1{,}200$ customers order a sugar cone. You want to verify this claim, so you randomly select 120 customers to see if they order a sugar cone. Does this sample meet the conditions for inference?

Solution:

Our sample space should be no more than 10% of our population, the expected number of successes and failures should each be at least 10, and the sample should be selected randomly.

The sample space was random, and was no more than 10% of the population:

$$\frac{120}{1{,}200} = 0.1 = 10\% \leq 10\%$$

But there are not at least 10 expected successes and failures.

$$120(0.05) = 6 \ngeq 10$$

$$120(0.95) = 114 \geq 10$$

The sample space doesn't meet the conditions for inference because the success of a customer ordering a sugar cone is 6, which is less than 10.

■ 3. The zoo conducts a study about the demographics of its patrons, and wants to learn about how many groups that visit the zoo bring children under age 12. Every 10th customer or group is recorded as a "family," and classified as either "including children under 12" or "not including children under 12." The zoo collected data on 65 families, and 45 of them are classified as "not including children under 12." That day, 650 families came to the zoo. What is the standard error of the sampling distribution of the sample proportion?

Solution:

Our sample space should be no more than 10% of our population, the expected number of successes and failures should each be at least 10, and the sample should be selected randomly.

The sample space was random, and was no more than 10% of the population:

$$\frac{65}{650} = 0.1 = 10\% \leq 10\%$$

The "success" rate was $45/65 = \approx 0.69$, which means the failure rate was $1 - 0.69 \approx 0.31$. Which means there are more than 10 expected successes and failures.

$$65(0.69) = 45 \geq 10$$

$$65(0.31) = 20 \geq 10$$

We've met the conditions for inference, so we'll identify the sample size $n = 65$ and the population proportion as

$$p = \frac{45}{65} = 0.69$$

Now we can calculate standard error of the proportion.

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.69(1-0.69)}{65}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.69(0.31)}{65}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.2139}{65}}$$

$$\sigma_{\hat{p}} = 0.057365$$

■ 4. A pizza shop finds that 80% of the 75 randomly selected pizzas ordered during the week have pepperoni. What is the standard error of the proportion if the pizza shop has a total of $1{,}000$ pizzas ordered during the week?

Solution:

Our sample space should be no more than 10% of our population, the expected number of successes and failures should each be at least 10, and the sample should be selected randomly.

The sample space was random, and was no more than 10% of the population:

$$\frac{75}{1{,}000} = 0.075 = 7.5\% \leq 10\%$$

And there are more than 10 expected successes and failures.

$$75(0.8) = 60 \geq 10$$

$$75(0.2) = 15 \geq 10$$

We've met the conditions for inference, so we'll identify the sample size $n = 75$ and the population proportion as $p = 0.8$. Now we can calculate standard error of the proportion.

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.8(1-0.8)}{75}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.8(0.2)}{75}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.16}{75}}$$

$$\sigma_{\hat{p}} = 0.046188$$

5. A hospital conducts a survey and finds that 10 patients of 30 who are randomly selected on a given day have high blood pressure. There were 325 patients in the hospital that day. What is the standard error of the proportion?

Solution:

Our sample space should be no more than $10\,\%$ of our population, the expected number of successes and failures should each be at least 10, and the sample should be selected randomly.

The sample space was random, and was no more than $10\,\%$ of the population:

$$\frac{30}{325} = 0.0923 = 9.23\,\% \leq 10\,\%$$

And there are more than 10 expected successes and failures.

$$30(0.33) = 10 \geq 10$$

$$30(0.67) = 20 \geq 10$$

We've met the conditions for inference, so we'll identify the sample size $n = 30$ and the population proportion as

$$p = \frac{10}{30} = 0.33$$

Now we can calculate standard error of the proportion.

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.33(1-0.33)}{30}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.33(0.67)}{30}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.2211}{30}}$$

$$\sigma_{\hat{p}} = 0.085849$$

■ 6. A study claims that first-born children are more likely to become leaders. The study finds that 72% of $2{,}000$ first-born children are currently in or have held leadership roles in their careers. Another group of scientists wants to verify the claim, but can't survey all $2{,}000$ people, so they randomly sample 175 of the participants. What is the probability that their results are within 2% of the first study's claim?

Solution:

Our sample space should be no more than 10% of our population, the expected number of successes and failures should each be at least 10, and the sample should be selected randomly.

The sample space was random, and was no more than 10% of the population:

$$\frac{175}{2{,}000} = 0.0875 = 8.75\% \leq 10\%$$

And there are more than 10 expected successes and failures.

$$175(0.72) = 126 \geq 10$$

$$175(0.28) = 49 \geq 10$$

We've met the conditions for inference. The original study found the population proportion to be $p = 72\%$. So the standard error of the proportion will be

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.72(0.28)}{175}} \approx 0.0339$$

We need to find the probability that our results are within 2% of the population proportion $p = 72\%$. This means, how likely is it that the mean of the sampling distribution of the sample proportion falls between 70% and 74%? We need to express 2% in terms of standard deviations:

$$\frac{0.02}{0.0339} \approx 0.59$$

This means we want to know the probability of $P(-0.59 < z < 0.59)$. Using a z-table, -0.59 gives us 0.2776 and 0.59 gives us 0.7224, so the probability is

$$P(-0.59 < z < 0.59) = 0.7224 - 0.2776$$

$$P(-0.59 < z < 0.59) = 0.4448$$

There's a 44.48% chance that our sample proportion will fall within 2% of the first study's claim.

CONFIDENCE INTERVAL FOR A POPULATION MEAN

1. You want to determine the mean of calories served in a restaurant meal in America. The government has already done a study to find this mean, and they found $\sigma = 350.2$. You randomly sample 31 meals and find $\bar{x} = 1{,}500$. Construct and interpret a $95\,\%$ confidence interval for the mean number of calories in a restaurant meal.

Solution:

We have population standard deviation, so with sample mean $\bar{x} = 1{,}500$, standard error $\sigma = 350.2$, and critical values of 1.96 associated with $95\,\%$ confidence, the confidence interval is given by

$$(a, b) = \bar{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}}$$

$$(a, b) = 1{,}500 \pm 1.96 \cdot \frac{350.2}{\sqrt{31}}$$

$$(a, b) = 1{,}500 \pm 123.28$$

$$(a, b) \approx (1{,}376.72, 1{,}623.28)$$

Based on the sample, we're $95\,\%$ confident that the average number of calories in a restaurant meal was between $1{,}376.72$ and $1{,}623.28$ calories.

2. A bus travels between Kansas City and Denver. You take a sample of 50 trips and find a mean travel time of $\bar{x} = 12$ hours with standard error $s = 0.25$ hours. Construct and interpret a 95% confidence interval for the mean bus trip time in hours from Kansas City to Denver.

Solution:

We don't have population standard deviation, so we'll have to use the standard error from the sample instead. The sample size is still large (at least 30), so the confidence interval is given by

$$(a, b) = \bar{x} \pm z^* \cdot \frac{s}{\sqrt{n}}$$

$$(a, b) = 12 \pm 1.96 \cdot \frac{0.25}{\sqrt{50}}$$

$$(a, b) \approx 12 \pm 0.0693$$

$$(a, b) \approx (11.93, 12.07)$$

Based on the sample, we're 95% confident that the average bus trip from Kansas City to Denver takes between 11.93 and 12.07 hours.

3. A student wanted to know how many chocolates were in the small bags of chocolate candies her school was selling for a fundraiser. She took a simple random sample of 20 small bags of chocolate candy. From the

sample, she found an average of 17 pieces of candy per bag with a standard deviation of 2.03.

A box-plot of the data from the sample showed the distribution to be approximately normal. Compute and interpret a 95% confidence interval for the mean number of chocolate candies per bag.

Solution:

We're told in the problem that the distribution is approximately normal and that it's from a simple random sample. We have a small sample size of 20 bags of candy and an unknown population standard deviation. This means we need to use a t test-statistic in the confidence interval.

t* is the test statistic, so we'll look this up in the t-table. We need to use the confidence level and the degrees of freedom. The confidence level is 95%, and the degrees of freedom is $n-1=20-1=19$. The value we get from the t-table is 2.093.

	Upper-tail probability p									
df	**0.25**	**0.20**	**0.15**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**	**0.0005**
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	**2.093**	2.539	2.861	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
	50%	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.8%**	**99.9%**
	Confidence level C									

With a sample mean of $\bar{x}=17$, a standard error of $s=2.030$, a sample size of $n=20$, and a critical value from the t-table of 2.093, the confidence interval is

$$(a, b) = \bar{x} \pm t^* \cdot \frac{s}{\sqrt{n}}$$

$$(a, b) = 17 \pm 2.093 \cdot \frac{2.03}{\sqrt{20}}$$

$$(a, b) = 17 \pm 0.9501$$

$$(a, b) = (16.0499, 17.9501)$$

Based on the sample, we're 95% confident that the average number of chocolates per bag is between 16.0499 and 17.9501 pieces.

4. Consider the formula for a confidence interval for a population mean with an unknown sample standard deviation. How does doubling the sample size affect the confidence interval?

$$(a, b) = \bar{x} \pm t^* \cdot \frac{s}{\sqrt{n}}$$

Solution:

Doubling the sample size makes the confidence interval narrower, which means you would get a better idea of your estimate for the population mean.

The confidence interval has the formula:

$$(a, b) = \bar{x} \pm t^* \cdot \frac{s}{\sqrt{n}}$$

If we double the sample size, we multiply n by 2.

$$(a, b) = \bar{x} \pm t^* \cdot \frac{s}{\sqrt{2n}}$$

We can choose some numbers for our confidence interval just to look at what's happening. Let's randomly choose some numbers for the sample mean, sample standard deviation and sample size.

$$\bar{x} = 17$$

$$s = 2.030$$

$$n = 11$$

Let's choose a confidence interval of 95%. Then we can choose the test-statistic based on the sample size. Here we choose the test statistic for $n = 11$ as $t^* = 2.228$ and the test statistic for $2n = 2(11) = 22$ as $t^* = 2.080$.

Let's set up the confidence interval with the first sample size.

$$(a, b) = 17 \pm 2.228 \cdot \frac{2.030}{\sqrt{11}}$$

$$(a, b) = 17 \pm 1.3637$$

Now let's look at what happens when the sample size is doubled.

$$(a, b) = 17 \pm 0.9002$$

Here you can see that you're adding and subtracting a smaller amount when the sample size is doubled. This would make the confidence interval narrower, which means you would get a better idea of your estimate for the population mean.

■ 5. A magazine took a random sample of 30 people and reported the average spending on an Easter basket this year to be $\$44.78$ per basket with a sample standard deviation of $\$18.10$. Construct and interpret a 98% confidence interval for the data.

Solution:

We're told in the problem that the data is from a simple random sample. We have a large sample size of 30 people and an unknown population standard deviation. But the sample size is large enough to make the distribution approximately normal, so we can still use a z-statistic.

Let's set up the values we need for the calculation. The sample mean is $\bar{x} = \$44.78$, and the sample standard deviation is $s = \$18.10$. We also know the sample size is $n = 30$.

To find the z-value associated with a 98% confidence interval, we realize that $\alpha/2 = 2\%/2 = 1\%$. So we'll look up 0.9900 in the body of the z-table. The closest z-value is $z = 2.33$.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936

So the confidence interval will be

$$(a, b) = \bar{x} \pm z^* \cdot \frac{s}{\sqrt{n}}$$

$$(a, b) = 44.78 \pm 2.33 \cdot \frac{18.10}{\sqrt{30}}$$

$$(a, b) \approx 44.78 \pm 7.6997$$

$$(a, b) \approx (37.08, 52.48)$$

Based on the sample, we're 98% confident that the average amount spent on Easter baskets was between $\$37.08$ and $\$52.48$.

6. A confidence interval for a study is $(11.5, 18.5)$. What was the value of the sample mean?

Solution:

The sample mean is always in the middle of the confidence interval. If we find the middle of $(11.5, 18.5)$, then we know the sample mean.

$$\bar{x} = \frac{11.5 + 18.5}{2} = \frac{30}{2} = 15$$

CONFIDENCE INTERVAL FOR A POPULATION PROPORTION

1. According to a recent poll, 47% of the 648 Americans surveyed make weekend plans based on the weather. Construct and interpret a 99% confidence interval for the percentage of Americans who make weekend plans based on the weather.

Solution:

The sample proportion is $\hat{p}=0.47$ and the confidence level is 99%. The test statistic for this confidence level is $z^*=2.58$ and the sample size is $n=648$. So the confidence interval is

$$(a,b)=\hat{p}\pm z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$(a,b)=0.47\pm 2.58\sqrt{\frac{0.47(1-0.47)}{648}}$$

$$(a,b)\approx 0.47\pm 0.0506$$

$$(a,b)\approx (0.42,0.52)$$

This means that we're 99% confident that the percentage of Americans who make weekend plans based on weather is between 42% and 52%.

2. You want to determine the proportion of teenagers who own their own cell phone. You take a random sample of 100 teenagers and find that 86 of them own a cell phone. At 90% confidence, build a confidence interval for the population proportion.

Solution:

The sample proportion is $\hat{p} = 86/100 = 0.86$ and the confidence level is 90%. The test statistic for this confidence level is $z^* = 1.65$ and the sample size is $n = 100$. So the confidence interval is

$$(a, b) = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

$$(a, b) = 0.86 \pm 1.65 \sqrt{\frac{0.86(1 - 0.86)}{100}}$$

$$(a, b) \approx 0.86 \pm 0.0573$$

$$(a, b) \approx (0.80, 0.92)$$

This means that we're 90% confident that the proportion of teenagers who own their own cell phone is between 80% and 92%.

3. A biologist is trying to determine the proportion of plants in a jungle that are ferns. She takes a random sample of 82 plants and finds that 31 of them can be classified as ferns. At 95% confidence, what is the confidence interval for the population proportion?

Solution:

The sample proportion is $\hat{p} = 31/82 \approx 0.3780$ and the confidence level is 95%. The test statistic for this confidence level is $z^* = 1.96$ and the sample size is $n = 82$. So the confidence interval is

$$(a, b) = \hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$(a, b) = 0.3780 \pm 1.96 \sqrt{\frac{0.3780(1-0.3780)}{82}}$$

$$(a, b) \approx 0.3780 \pm 0.1050$$

$$(a, b) \approx (0.27, 0.48)$$

This means that the biologist can be 95% confident that the proportion of plants in the jungle that are ferns is between 27% and 48%.

4. A statistics teacher at a university conducted a study and found that 80% of university students are interested in taking a statistics class. You want to see if this proportion holds at your own university. Find the minimum sample size you can use to keep a margin of error of 0.02 at a 99% confidence level.

Solution:

The given proportion is $\hat{p} = 80\% = 0.8$. The confidence level is 99% and the test statistic for this confidence level is $z^* = 2.58$. The margin of error is $ME = 0.02$. Plug these values into the formula for margin of error from the confidence interval for a population proportion.

$$ME = z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.02 = 2.58\sqrt{\frac{0.8(1-0.8)}{n}}$$

$$0.02 = 2.58\sqrt{\frac{0.16}{n}}$$

Solve for n.

$$0.02 = 2.58\frac{\sqrt{0.16}}{\sqrt{n}}$$

$$0.02\sqrt{n} = 2.58\sqrt{0.16}$$

$$\sqrt{n} = \frac{2.58\sqrt{0.16}}{0.02}$$

$$n = \left(\frac{2.58\sqrt{0.16}}{0.02}\right)^2$$

$$n = 2{,}662.56$$

Since we need more than $2{,}662$ university students for the sample, we have to round up to $2{,}663$ students, so we can say $n = 2{,}663$.

5. Sarah is conducting a class survey to determine if the percentage of juniors in favor of having the next dance at a local bowling alley is $65\,\%$. How many juniors should she survey in order to be $90\,\%$ confident with a margin of error of 0.08?

Solution:

The pre-determined success rate is $\hat{p} = 65\,\% = 0.65$. The confidence level is $90\,\%$ and the test statistic for this confidence level is $z^* = 1.65$. The margin of error is $ME = 0.04$. Plug these values into the formula for margin of error from the confidence interval for a population proportion.

$$ME = z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.08 = 1.65\sqrt{\frac{0.65(1-0.65)}{n}}$$

$$0.08 = 1.65\sqrt{\frac{0.2275}{n}}$$

Solve for n.

$$0.08 = 1.65\frac{\sqrt{0.2275}}{\sqrt{n}}$$

$$0.08\sqrt{n} = 1.65\sqrt{0.2275}$$

$$\sqrt{n} = \frac{1.65\sqrt{0.2275}}{0.08}$$

$$n = \left(\frac{1.65\sqrt{0.2275}}{0.08}\right)^2$$

$$n \approx 96.78$$

Since we need more than 96 juniors for the sample, we have to round up to 97 juniors, so we can say $n = 97$.

■ 6. A study suggests that 10% of practicing physicians are cognitively impaired. What random sample of practicing physicians is needed to confirm this finding at a confidence level of 95% with a margin of error of 0.05?

Solution:

The sample proportion is given as 0.10. The confidence level is 95% and the test statistic for this confidence level is $z^* = 1.96$. The margin of error is $ME = 0.05$. Plug these values into the formula for margin of error from the confidence interval for a population proportion.

$$ME = z^*\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.05 = 1.96\sqrt{\frac{0.10(1-0.10)}{n}}$$

$$0.05 = 1.96\sqrt{\frac{0.09}{n}}$$

Solve for n.

$$0.05 = 1.96\frac{\sqrt{0.09}}{\sqrt{n}}$$

$$0.05\sqrt{n} = 1.96\sqrt{0.09}$$

$$\sqrt{n} = \frac{1.96\sqrt{0.09}}{0.05}$$

$$n = \left(\frac{1.96\sqrt{0.09}}{0.05}\right)^2$$

$$n \approx 138.30$$

Since we need more than 138 physicians for the sample, we have to round up to 139 physicians, so we can say $n = 139$.

W W W . K R I S T A K I N G M A T H . C O M