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. Would the data fit into a one-way or two-way table? 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 fits into a one-way table. 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 activity of 50 adults. Is it a one-way data table? 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. Will the data fit into a one-way or two-way table? Is the study observational or experimental?

Solution:

The data fits into a two-way table 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. This is also called "social desirability bias." 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 systematic sampling. 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.

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

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

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

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

■ 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.

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

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

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

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.

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

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.

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

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.

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

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

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

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.

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

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.

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

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

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

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.

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

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.

CONDITIONS FOR INFERENCE WITH THE SDSM

1. There are $1{,}000$ students at our school, and we ask 150 of them to tell us their height as they exit school at the end of the day. Have we met the conditions for inference?

Solution:

Because we're asking students as they exit the school building at the end of the day, we're not taking a truly random sample, so we're violating the random condition.

We're also sampling with replacement, and whenever we sample without replacement, we have to keep the sample size below 10% of the population. Because 150 is greater than 10% of the population, we violate the independent condition as well.

2. We randomly sample 400 boxes (with replacement) in a very large, national shipping warehouse and record their weight in kilograms. Have we met the conditions for inference?

Solution:

We've taken a random sample with replacement, which means we've met the random and independent conditions. We don't know whether the population is normally distributed, but our sample size is large, $n \geq 30$, so we've met the normal condition as well, and we can move forward with using our data to answer probability questions.

3. A cookie company makes packages of cookies, where the weight of the packages is normally distributed with $\mu = 500$ grams and $\sigma = 4$ grams. If the cookie company's production manager randomly selects 100 packages of cookies, what is the probability that the sample mean is within 7.5 grams of the population mean?

Solution:

The production manager takes a random sample, and the sample size is 100 cookie packages, so we've met the random and normal conditions for inference. The question suggests that the manager samples without replacement, but we can safely assume that 100 packages is less than 10% of the total number of packages that the company produces, which means we've met the independence condition as well.

The mean of the sampling distribution of the sample mean will be $\mu_{\bar{x}} = \mu = 500$, and the standard error will be

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

$$\sigma_{\bar{x}} = \frac{4}{\sqrt{100}}$$

$$\sigma_{\bar{x}} = \frac{4}{10}$$

$$\sigma_{\bar{x}} = 0.4$$

We want to know the probability that the sample mean $\bar{x}$ is within 7.5 grams of the population mean, $\mu = 500$. A 7.5 interval around 500 gives us the interval 492.5 to 507.5, so

$$P(492.5 < \bar{x} < 507.5) = P\left(\frac{492.5 - 500}{0.4} < z_{\bar{x}} < \frac{507.5 - 500}{0.4}\right)$$

$$P(492.5 < \bar{x} < 507.5) = P\left(-\frac{7.5}{0.4} < z_{\bar{x}} < \frac{7.5}{0.4}\right)$$

$$P(492.5 < \bar{x} < 507.5) = P(-18.75 < z_{\bar{x}} < 18.75)$$

The z-score $z = 18.75$ is far higher than the largest value in the standard z-table, which means the probability under the normal curve between $z = -18.75$ and $z = 18.75$ is virtually 100%.

Therefore, it's almost guaranteed that the sample mean will fall within 7.5 grams of the population mean $\mu = 500$.

4. A sushi chef builds a sushi roll approximately every 3 minutes, with a standard deviation of 15 seconds, every night while his restaurant is open between $5:00$ p.m. and $10:00$ p.m., Tuesday through Sunday. The time

spent to build sushi rolls is normally distributed. If the chef takes a random sample of 20 sushi rolls over the course of a week, what is the probability that the sample mean is within 5 seconds of the population mean?

Solution:

The chef takes a random sample, and the population is normally distributed, so we've met the random and normal conditions.

The restaurant is open 30 hours each week (5 hours per night for 6 nights per week), and the chef builds approximately $60 \div 3 = 20$ sushi rolls per hour, so he builds about $30 \cdot 20 = 600$ sushi rolls per week. The 20-roll sample is well below $10\,\%$ of the population, so we've met the independent condition as well.

The mean of the sampling distribution of the sample mean will be $\mu_{\bar{x}} = \mu = 3$ minutes (or 180 seconds). We'll calculate everything in seconds, and the standard error will be

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

$$\sigma_{\bar{x}} = \frac{15}{\sqrt{20}}$$

$$\sigma_{\bar{x}} \approx 3.354$$

We want to know the probability that the sample mean $\bar{x}$ is within 5 seconds of the population mean, 180. A 5-second interval around 180 gives us the interval 175 to 185, so

$$P(175 < \bar{x} < 185) = P\left(\frac{175-180}{3.354} < z_{\bar{x}} < \frac{185-180}{3.354}\right)$$

$$P(175 < \bar{x} < 185) = P\left(-\frac{5}{3.354} < z_{\bar{x}} < \frac{5}{3.354}\right)$$

$$P(175 < \bar{x} < 185) \approx P(-1.49 < z_{\bar{x}} < 1.49)$$

In the z-table, a z-value of 1.49 gives 0.9319,

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441

and a z-value of -1.49 gives 0.0681.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823

Which means the probability under the normal curve between these z-scores is

$$P(-1.49 < z_{\bar{x}} < 1.49) = 0.9319 - 0.0681$$

$$P(-1.49 < z_{\bar{x}} < 1.49) = 0.8638$$

$$P(-1.49 < z_{\bar{x}} < 1.49) = 86.4\%$$

So there's an approximately 86.4% chance that the mean $\bar{x}$ of the 20-roll sample the chef takes will fall within 5 seconds of the population mean of $\mu = 180$ seconds.

■ 5. The time spent playing video games by competitive gamers is normally distributed with $\mu = 40$ hours per week, and $\sigma = 2.5$ hours. If we take a random sample with replacement of 75 players and record the number of hours they spend playing this week, what's the probability that our sample mean is within 30 minutes of the population mean?

Solution:

We take a random sample, and the population is normally distributed, so we've met the random and normal conditions. We sample with replacement, which means we've meet the independent condition as well.

The mean of the sampling distribution of the sample mean will be $\mu_{\bar{x}} = \mu = 40$ hours, and the standard error will be

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

$$\sigma_{\bar{x}} = \frac{2.5}{\sqrt{75}}$$

$$\sigma_{\bar{x}} \approx 0.289$$

We want to know the probability that the sample mean $\bar{x}$ is within 30 minutes (0.5 hours) of the population mean, 40. A 0.5-hour interval around 40 gives us the interval 39.5 to 40.5, so

$$P(39.5 < \bar{x} < 40.5) = P\left(\frac{39.5 - 40}{0.289} < z_{\bar{x}} < \frac{40.5 - 40}{0.289}\right)$$

$$P(39.5 < \bar{x} < 40.5) = P\left(-\frac{0.5}{0.289} < z_{\bar{x}} < \frac{0.5}{0.289}\right)$$

$$P(39.5 < \bar{x} < 40.5) \approx P(-1.73 < z_{\bar{x}} < 1.73)$$

In the z-table, a z-value of 1.73 gives 0.9582,

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706

and a z-value of -1.73 gives 0.0418.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455

Which means the probability under the normal curve between these z-scores is

$$P(-1.73 < z_{\bar{x}} < 1.73) = 0.9582 - 0.0418$$

$$P(-1.73 < z_{\bar{x}} < 1.73) = 0.9164$$

$$P(-1.73 < z_{\bar{x}} < 1.73) \approx 91.6\,\%$$

So there's an approximately $91.6\,\%$ chance that the mean $\bar{x}$ of the 75-player sample we take will fall within 30 minutes of the population mean of $\mu = 40$ hours.

6. The time it takes for a roofing company to install a new roof on a single-story house normally distributed with $\mu = 6$ hours and $\sigma = 1$ hour. If the company's owner takes a random sample with replacement of 10 roofing jobs, what's the probability that his sample mean is within 45 minutes of the population mean?

Solution:

The company's owner takes a random sample, and the population is normally distributed, so he's met the random and normal conditions. Because he's sampling with replacement, he meets the independent condition as well.

The mean of the sampling distribution of the sample mean will be $\mu_{\bar{x}} = \mu = 6$ hours, and the standard error will be

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

$$\sigma_{\bar{x}} = \frac{1}{\sqrt{10}}$$

$$\sigma_{\bar{x}} \approx 0.316$$

The owner wants to know the probability that the sample mean $\bar{x}$ is within 45 minutes (0.75 hours) of the population mean, 6. A 0.75-hour interval around 6 gives us the interval 5.25 to 6.75, so

$$P(5.25 < \bar{x} < 6.75) = P\left(\frac{5.25 - 6}{0.316} < z_{\bar{x}} < \frac{6.75 - 6}{0.316}\right)$$

$$P(5.25 < \bar{x} < 6.75) = P\left(-\frac{0.75}{0.316} < z_{\bar{x}} < \frac{0.75}{0.316}\right)$$

$$P(5.25 < \bar{x} < 6.75) \approx P(-2.37 < z_{\bar{x}} < 2.37)$$

In the z-table, a z-value of 2.37 gives 0.9911,

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

and a z-value of -2.37 gives 0.0089.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110

Which means the probability under the normal curve between these z-scores is

$$P(-2.37 < z_{\bar{x}} < 2.37) = 0.9911 - 0.0089$$

$$P(-2.37 < z_{\bar{x}} < 2.37) = 0.9822$$

$$P(-2.37 < z_{\bar{x}} < 2.37) \approx 98.2\,\%$$

So there's an approximately $98.2\,\%$ chance that the mean $\bar{x}$ of the sample of 10 roofing jobs the owner takes will fall within 45 minutes of the population mean of $\mu = 6$ hours.

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:

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

And there are more than 10 expected successes and failures.

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

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

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. We want to verify this claim, so we 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

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

Now we can calculate the standard error of the proportion, remembering to add in the finite population correction factor, since we have a finite population of 650 families.

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

$$\sigma_{\hat{p}} = \sqrt{\frac{0.69(1-0.69)}{65}}\sqrt{\frac{650-65}{650-1}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.69(0.31)}{65}}\sqrt{\frac{585}{649}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.2139 \cdot 585}{65 \cdot 649}}$$

$$\sigma_{\hat{p}} \approx 0.054463$$

■ 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:

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

And there are more than 10 expected successes and failures.

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

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

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, remembering to add in the finite population correction factor, since we have a finite population of $1{,}000$ families.

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

$$\sigma_{\hat{p}} = \sqrt{\frac{0.8(1-0.8)}{75}}\sqrt{\frac{1{,}000-75}{1{,}000-1}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.8(0.2)}{75}}\sqrt{\frac{925}{999}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.16 \cdot 925}{75 \cdot 999}}$$

$$\sigma_{\hat{p}} \approx 0.044444$$

■ 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:

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

And there are more than 10 expected successes and failures.

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

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

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

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

Now we can calculate standard error of the proportion, remembering to add in the finite population correction factor, since we have a finite population of 325 patients.

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

$$\sigma_{\hat{p}} = \sqrt{\frac{0.33(1-0.33)}{30}}\sqrt{\frac{325-30}{325-1}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.33(0.67)}{30}}\sqrt{\frac{295}{324}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.2211 \cdot 295}{30 \cdot 324}}$$

$$\sigma_{\hat{p}} \approx 0.081917$$

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:

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

And there are more than 10 expected successes and failures.

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

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

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

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

$$\sigma_{\hat{p}} = \sqrt{\frac{0.72(0.28)}{175}}\sqrt{\frac{2,000-175}{2,000-1}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.2016}{175}}\sqrt{\frac{1,825}{1,999}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.2016 \cdot 1,825}{175 \cdot 1,999}}$$

$$\sigma_{\hat{p}} \approx 0.03243$$

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:

$$\frac{0.02}{0.03243} \approx 0.616713$$

This means we want to know the probability of $P(-0.62 < z < 0.62)$. Using a z-table, -0.62 gives us 0.2676 and 0.62 gives us 0.7324, so the probability is

$$P(-0.62 < z < 0.62) = 0.7324 - 0.2676$$

$$P(-0.62 < z < 0.62) = 0.4648$$

There's a 46.48% chance that our sample proportion will fall within 2% of the first study's claim.

CONDITIONS FOR INFERENCE WITH THE SDSP

■ 1. A gym owner takes a random sample of 10 local fitness instructors and asks them whether or not they train clients at multiple gyms. He finds that $\hat{p} = 70\,\%$ of them report training clients at multiple gyms. Can he proceed with a hypothesis test?

Solution:

With a sample of only 10 instructors, the gym owner finds $n\hat{p} = 10(0.7) = 7$ "successes," and $n(1-\hat{p}) = 10(0.3) = 3$ "failures" in his sample. He needs at least 5 successes and at least 5 failures in order to guarantee normality. Therefore, he fails to meet the normal condition, and can't proceed with a hypothesis test.

■ 2. A professional basketball player makes $87.5\,\%$ of his free throws. If he takes a random sample (without replacement) of 100 of his own free throws, can he move forward with a hypothesis test?

Solution:

We're told that the player takes a random sample, so we assume he's met the random condition. With the given population proportion of $p = 0.875$, we can expect $np = 100(0.875) = 87.5$ "successes" and

$n(1-p) = 100(0.125) = 12.5$ "failures." Both of these values are greater than 5, so the player has met the normal condition.

Finally, even though the player samples without replacement, we can assume that the the population of all his free throws is significantly larger than 10 times his sample size, and he therefore meets the independent condition.

Because he's met all three conditions for inference, he can proceed with a hypothesis test.

3. If the basketball player from the previous question finds a sample proportion $\hat{p} = 0.85$ in his sample of 100 free throws, calculate his test statistic. Remember that $p = 0.875$.

Solution:

When he runs a hypothesis test, the player will calculate his test statistic as

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

$$z_{\hat{p}} = \frac{0.85 - 0.875}{\sqrt{\frac{0.875(1-0.875)}{100}}}$$

$$z_{\hat{p}} = \frac{-0.025}{\sqrt{\frac{0.875(0.125)}{100}}}$$

$$z_{\hat{p}} = -0.025\sqrt{\frac{100}{0.875(0.125)}}$$

$$z_{\hat{p}} = -0.025\left(\frac{10}{\sqrt{0.109375}}\right)$$

$$z_{\hat{p}} \approx -0.76$$

■ 4. A grocery chain claims that 75% of their customers say that they are "satisfied" with their local store. We want to verify this claim, so we take a random sample of 45 of their customers and ask them whether or not they are "satisfied." How likely is it that our results are within 2% of the chain's claim?

Solution:

The sample size is $n = 45$ and $p = 0.75$, which means $1 - p = 0.25$, so

$$np = 45(0.75) = 33.75 \geq 5$$

$$n(1 - p) = 45(1 - 0.75) = 45(0.25) = 11.25 \geq 5$$

and we've verified normality. We were told in the question that our sample was random, and our sample $n = 45$ is certainly less than 10% of the total population of the grocery chain's customers, so we're not violating the 10% rule. With the conditions for inference satisfied, we can calculate the mean and standard deviation of the sampling distribution of the sample proportion. The mean is

$$\mu_{\hat{p}} = p$$

$$\mu_{\hat{p}} = 0.75$$

The standard deviation of the sampling distribution (the standard error) is

$$\sigma_{\hat{p}} = \sqrt{\frac{0.75(1-0.75)}{45}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.75(0.25)}{45}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.1875}{45}}$$

$$\sigma_{\hat{p}} \approx 0.0645$$

The question asks us for the probability that our sample proportion is within 2% of the population proportion $p = 75\%$. In other words, how likely is it that the sample proportion falls between 73% and 77%?

$$P(0.73 < \hat{p} < 0.77) \approx P\left(\frac{0.73-0.75}{0.0645} < z < \frac{0.77-0.75}{0.0645}\right)$$

$$P(0.73 < \hat{p} < 0.77) \approx P\left(-\frac{0.02}{0.0645} < z < \frac{0.02}{0.0645}\right)$$

$$P(0.73 < \hat{p} < 0.77) \approx P(-0.31 < z < 0.31)$$

In a z-table, a z-value of 0.31 gives 0.6217,

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879

and a value of -0.31 gives 0.3783.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859

Which means the probability under the normal curve of the sampling distribution of the sample proportion between these z-scores is

$$P(-0.31 < z < 0.31) \approx 0.6217 - 0.3783$$

$$P(-0.31 < z < 0.31) \approx 0.2434$$

$$P(-0.31 < z < 0.31) \approx 24.34\,\%$$

Which means there's an approximately $24\,\%$ chance that our sample proportion will fall within $2\,\%$ of the grocery chain's claim.

■ 5. A professional pickleball player claims that he wins $60\,\%$ of the points he plays in championship matches. We want to verify this claim, so we take a random sample of 25 of his points in championship matches and record whether or not he wins each point. How likely is it that our results are within 5% of the player's claim?

Solution:

The sample size is $n = 25$ and $p = 0.6$, which means $1 - p = 0.4$, so

$$np = 25(0.6) = 15 \geq 5$$

$$n(1 - p) = 25(1 - 0.6) = 25(0.4) = 10 \geq 5$$

and we've verified normality. We were told in the question that our sample was random, and our sample $n = 25$ is certainly less than 10% of the total population of points played in championship matches by the professional player, so we're not violating the 10% rule. With the conditions for inference satisfied, we can calculate the mean and standard deviation of the sampling distribution of the sample proportion. The mean is

$$\mu_{\hat{p}} = p$$

$$\mu_{\hat{p}} = 0.6$$

The standard deviation of the sampling distribution (the standard error) is

$$\sigma_{\hat{p}} = \sqrt{\frac{0.6(1 - 0.6)}{25}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.6(0.4)}{25}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.24}{25}}$$

$$\sigma_{\hat{p}} \approx 0.0980$$

The question asks us for the probability that our sample proportion is within 5% of the population proportion $p = 60\%$. In other words, how likely is it that the sample proportion falls between 55% and 65%?

$$P(0.55 < \hat{p} < 0.65) \approx P\left(\frac{0.55 - 0.6}{0.0980} < z < \frac{0.65 - 0.6}{0.0980}\right)$$

$$P(0.55 < \hat{p} < 0.65) \approx P\left(-\frac{0.05}{0.0980} < z < \frac{0.05}{0.0980}\right)$$

$$P(0.55 < \hat{p} < 0.65) \approx P(-0.5102 < z < 0.5102)$$

$$P(0.55 < \hat{p} < 0.65) \approx P(-0.51 < z < 0.51)$$

In a z-table, a z-value of 0.51 gives 0.6950,

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549

and a value of -0.51 gives 0.3050.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121

Which means the probability under the normal curve of the sampling distribution of the sample proportion between these z-scores is

$$P(-0.51 < z < 0.51) \approx 0.6950 - 0.3050$$

$$P(-0.51 < z < 0.51) \approx 0.39$$

$$P(-0.51 < z < 0.51) \approx 39\,\%$$

Which means there's an approximately $39\,\%$ chance that our sample proportion will fall within $5\,\%$ of the pickleball player's claim.

6. A company reports that the proportion of its invoices that get paid on time is $p = 35\,\%$. A clerk on the Accounts Receivable team wants to verify this claim, so she takes a random sample of 80 invoices and records whether or not they were paid on time. How likely is it that her sample proportion will fall within $10\,\%$ of the company's claim?

Solution:

The sample size is $n = 80$ and $p = 0.35$, which means $1 - p = 0.65$, so

$$np = 80(0.35) = 28 \geq 5$$

$$n(1 - p) = 80(1 - 0.35) = 80(0.65) = 52 \geq 5$$

and we've verified normality. We were told in the question that the clerk's sample was random, and her sample $n = 80$ is less than $10\,\%$ of the total population of invoices issued by the company, so she's not violating the

10% rule. With the conditions for inference satisfied, she can calculate the mean and standard deviation of the sampling distribution of the sample proportion. The mean is

$$\mu_{\hat{p}} = p$$

$$\mu_{\hat{p}} = 0.35$$

The standard deviation of the sampling distribution (the standard error) is

$$\sigma_{\hat{p}} = \sqrt{\frac{0.35(1-0.35)}{80}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.35(0.65)}{80}}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.2275}{80}}$$

$$\sigma_{\hat{p}} = 0.053$$

The question asks us for the probability that the clerk's sample proportion is within 10% of the population proportion $p = 35\%$. In other words, how likely is it that the sample proportion falls between 25% and 45%?

$$P(0.25 < \hat{p} < 0.45) = P\left(\frac{0.25-0.35}{0.053} < z < \frac{0.45-0.35}{0.053}\right)$$

$$P(0.25 < \hat{p} < 0.45) = P\left(-\frac{0.1}{0.053} < z < \frac{0.1}{0.053}\right)$$

$$P(0.25 < \hat{p} < 0.45) \approx P(-1.89 < z < 1.89)$$

In a z-table, a z-value of 1.89 gives 0.9706,

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767

and a value of -1.89 gives 0.0294.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367

Which means the probability under the normal curve of the sampling distribution of the sample proportion between these z-scores is

$$P(-1.89 < z < 1.89) \approx 0.9706 - 0.0294$$

$$P(-1.89 < z < 1.89) \approx 0.9412$$

$$P(-1.89 < z < 1.89) \approx 94.12\,\%$$

Which means there's an approximately $94\,\%$ chance that the clerk's sample proportion will fall within $10\,\%$ of the company's claim.

THE STUDENT'S T-DISTRIBUTION

1. We take a random sample of size $n = 25$, and we want to be 99% confident about our results. What t-score will we find?

Solution:

If we look up $\text{df} = n - 1 = 25 - 1 = 24$, along with a 99% confidence level in the t-table, we find a t-score of $t = 2.797$.

	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**
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	**2.797**	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
	50%	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.8%**	**99.9%**
	Confidence level C									

2. We take a random sample of size $n = 18$, and we want to be 90% confident about our results. What t-score will we find?

Solution:

If we look up $\text{df} = n - 1 = 18 - 1 = 17$, along with a 90% confidence level in the t-table, we find a t-score of $t = 1.740$.

	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**
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.689	0.863	1.069	1.333	**1.740**	2.110	2.567	2.898	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
	50%	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.8%**	**99.9%**
	Confidence level C									

■ 3. We take a random sample of size $n = 8$, and we want to be 95% confident about our results. What t-score will we find?

Solution:

If we look up df $= n - 1 = 8 - 1 = 7$, along with a 95% confidence level in the t-table, we find a t-score of $t = 2.365$.

	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**
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	**2.365**	2.998	3.499	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
	50%	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.8%**	**99.9%**
	Confidence level C									

■ 4. We take a random sample of size $n = 14$, and our upper-tail probability will be 0.05. What t-score will we find?

Solution:

If we look up df $= n - 1 = 14 - 1 = 13$, along with an upper-tail probability of 0.05 in the t-table, we find a t-score of $t = 1.771$.

	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**
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.694	0.870	1.079	1.350	**1.771**	2.160	2.650	3.012	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
	50%	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.8%**	**99.9%**
	Confidence level C									

5. We take a random sample of size $n = 21$, and our upper-tail probability will be 0.001. What t-score will we find?

Solution:

If we look up df $= n - 1 = 21 - 1 = 20$, along with an upper-tail probability of 0.001 in the t-table, we find a t-score of $t = 3.552$.

	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**
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
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
	50%	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.8%**	**99.9%**
	Confidence level C									

6. We take a random sample of size $n = 3$, and our upper-tail probability will be 0.025. What t-score will we find?

Solution:

If we look up df $= n - 1 = 3 - 1 = 2$, along with an upper-tail probability of 0.025 in the t-table, we find a t-score of $t = 4.303$.

	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**
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.816	1.061	1.386	1.886	2.920	**4.303**	6.965	9.925	22.327	31.599
3	0.765	0.987	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
	50%	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.8%**	**99.9%**
	Confidence level C									

CONFIDENCE INTERVAL FOR THE MEAN

1. We 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 $\sigma = 350.2$. We 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. We take a sample of 30 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. So the confidence interval is given by

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

From the t-table, we find

	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**
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	**2.045**	2.462	2.756	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
	50%	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.8%**	**99.9%**
	Confidence level C									

Then the confidence interval is

$$(a, b) = 12 \pm 2.045 \cdot \frac{0.25}{\sqrt{30}}$$

$$(a, b) \approx 12 \pm 0.0933$$

$$(a, b) \approx (11.91, 12.09)$$

Based on the sample, we're 95% confident that the average bus trip from Kansas City to Denver takes between 11.91 and 12.09 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

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

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

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

$$(a,b)\approx(16.0499,17.9501)$$

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 we would get a better estimate of 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 we can see that we're adding and subtracting a smaller amount when the sample size is doubled. This would make the confidence interval narrower, which means we would get a better estimate of 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. Because population standard deviation is unknown, we'll use a t-test.

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 t-value associated with a 98% confidence interval, we realize that $\alpha/2 = 2\%/2 = 1\%$. So we'll look up the intersection of 0.01 and df $= 29$ in the body of the t-table. The t-value is $t = 2.462$.

	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**
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	**2.462**	2.756	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
	50%	**60%**	**70%**	**80%**	**90%**	**95%**	**98%**	**99%**	**99.8%**	**99.9%**
	Confidence level C									

So the confidence interval will be

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

$$(a, b) = 44.78 \pm 2.462 \cdot \frac{18.10}{\sqrt{30}}$$

$$(a, b) \approx 44.78 \pm 8.1359$$

$$(a, b) \approx (36.64, 52.92)$$

Based on the sample, we're 98% confident that the average amount spent on Easter baskets was between $\$36.64$ and $\$52.92$.

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 THE 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. We want to determine the proportion of teenagers who own their own cell phone. We 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. We want to see if this proportion holds at your own university. Find the minimum sample size we 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.

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

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

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

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

$$n = 2{,}662.56$$

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.

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

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

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

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

$$n \approx 96.78$$

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.

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

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

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

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

$$n \approx 138.30$$

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