Unit 5
Statistics –Revised
(20-25%)

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

F10 interpret and apply histograms

Elaboration – Instructional Strategies/Suggestions

F10 Students in this course should begin their study of statistics by reviewing ways of organizing and describing data. They should construct histograms (both with and without technology) and frequency polygons. When raw data is presented, students should be able to calculate the mean and median of the data. When a histogram is presented without the raw data, it is often not possible to calculate the mean and median precisely. When this occurs, students estimate the mean and the median of the data based on the histogram. Students should use frequency polygons to help them focus on the shape and spread (or dispersion) of the histogram. They also measure and interpret spread using standard deviation. In addition it will be necessary to determine whether or not data appears to be normally distributed based on a histogram and frequency polygon of the data.

Students will also examine various graphs and situations to describe the information presented. For example, they might determine the percentage of data that lies above or below a certain value.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

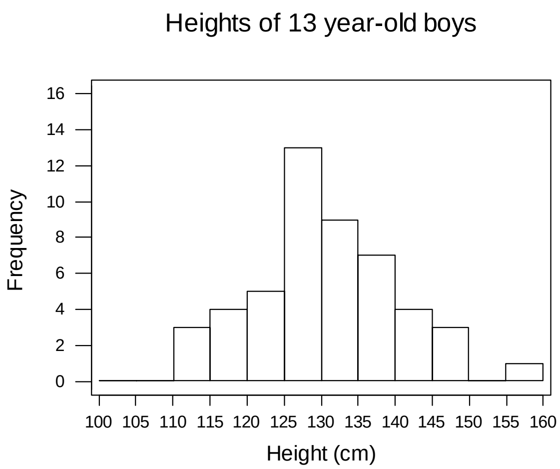
F10 *Activity*

F10.1 The Canadian Kennel Association did a study of litter size amongst certain breeds of dogs. Below are the results of the litter sizes last year of one particular breed, in one particular province:

3 6 5 6 5 5 7 5 7 6 6 6 4 6 5 6 4 3 5 6 8 3 6 9 2 5 7 9
4 5 8 2 5 7 6 6 7 7 8 4 5 6 5 9 10 3 4 5 11 2 7 9 5 3 7
6 7 4 5 8 9 3 6 5 7 5 6 4 8 7 3 9 10 4 5 12 5 8 9 5 6 8

a) Determine the mean and the median of the set of data.
b) Construct a histogram with bin-width 1 and a frequency polygon for this data.
c) Describe the shape of the frequency polygon.
d) What proportion of the litters had more than 6 pups? Between 3 and 8 pups?
e) What proportion of observations are at most 8 pups?
h) What proportion of litters contain between 5 and 10 (inclusive) offspring?

F10.2 The histogram below shows the heights of 13-year-old boys in a large urban school.



a) Draw the frequency polygon and describe its shape.
b) How many 13-year-old boys are in this school?
c) What do you think is the mean height of 13-year-old boys in this school?
d) What do you think is the median height of 13-year-old boys in this school?
e) What percentage of boys is between 130 cm and 150 cm?
f) What percentage of boys is less than 140 cm?
g) Give some examples of heights that are not typical of this group, and explain why you chose the heights you did.

(Continued on the next two-page spread ...)

Suggested Resources

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

F10 interpret and apply histograms

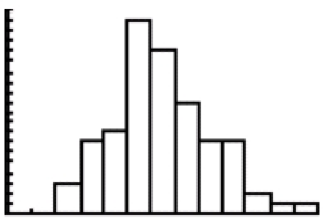
Statistics

Worthwhile Tasks for Instruction and/or Assessment

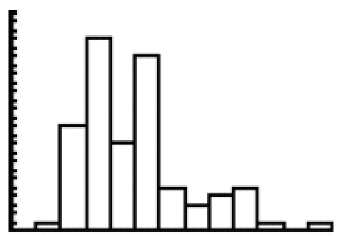
(... Continued from the previous two-page spread.)

F10.3 Which histograms appear to represent data that is approximately normally distributed? Explain your answer for each.

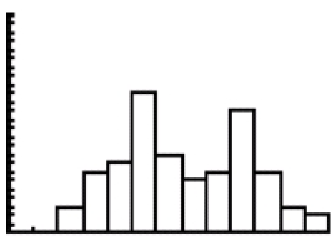
a)



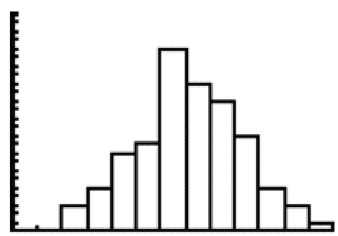
b)



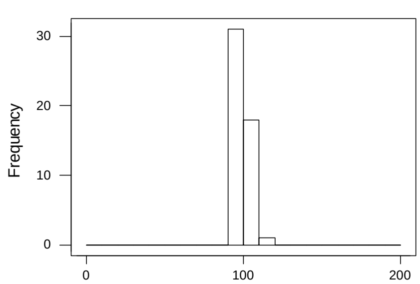
c)



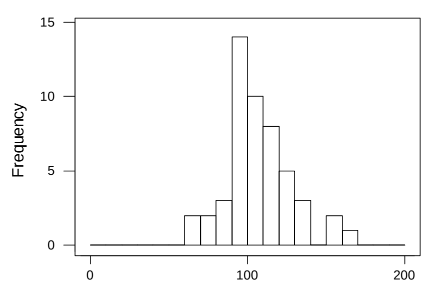
d)



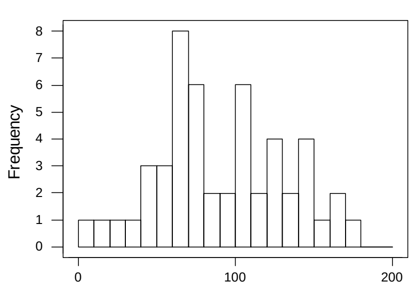
F10.4 Which of the following histograms has the largest standard deviation? the smallest? Explain your choices.



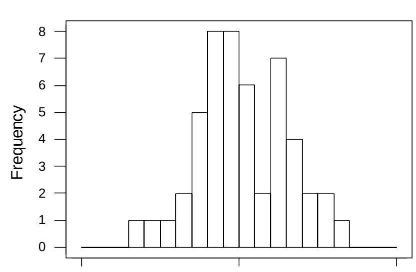
(a)



(b)



(c)



(d)

Suggested Resources

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

F9 demonstrate an understanding of the difference between sample standard deviation and population standard deviation

Elaboration – Instructional Strategies/Suggestions

F9 Before students can understand the difference between population standard deviation (σ_x) and sample standard deviation (S_x), they must first understand the difference between a population and a sample, as well as the difference between the sample mean and population mean. A population is the set of all the possible outcomes of interest to the investigator. A sample is a subset of the population. For example, considering all the grade 11 students in a particular school would represent the population of grade 11 students in that school. If we put all the names of the grade 11 students in a hat and select a subset of students by selecting names at random, we would have a sample. Further, if we are interested, for example, in the mean shoe size of grade 11 students in this school, calculating the mean based on the entire population data would be called the population mean, $\boldsymbol{\mu}$. Calculating the mean based on the sample data would be called the sample mean, $\bar{x}$. One important note is that we often find entire populations too large to work with for a number of reasons, such as cost and time; therefore we use random samples, which will hopefully be reflective of the entire population. If these samples are large enough (see Elaboration for F8/G3/F4 for discussion of sample sizes), we expect the value of the sample mean to be very close to the value of the population mean.

The calculation of standard deviation also depends on whether we are dealing with a population or a sample. If a set of data represents a population, then:

$$\sigma_x = \sqrt{\frac{(x_1-\mu)^2+(x_2-\mu)^2+...+(x_n-\mu)^2}{n}}$$, where $x_1, x_2, ..., x_n$ represent the data in the population.

However, when we are given a sample, we cannot find the value of the population standard deviation because we do not have the data set for the entire population. Therefore, we must use a value computed from the sample that is a good estimator of σ_x. We use a divisor of $n-1$ rather than n because, on average, the resulting value tends to be a bit closer to the true value of σ_x. That is:

$$S_x = \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+...+(x_n-\bar{x})^2}{n-1}}$$, where $x_1, x_2, ..., x_n$ represent the data in the sample.

The mathematical verification of this fact is beyond the scope of this course.

Students should be able to differentiate between when to use σ_x and S_x. While students should know how to calculate standard deviation both with and without technology, they should use technology for large data sets, as the calculation of standard deviation can be quite lengthy. Note that the TI-83 Plus will calculate both standard deviations.

As well, students should be able to use the mathematical symbols involved in describing the measurements in data:

- sample mean ($\bar{x}$)
- population mean (μ)
- sample standard deviation (S_x)
- population standard deviation (σ_x)

(Note: There is no difference in the calculation of $\bar{x}$ and μ for *finite* populations.)

Statistics

Worthwhile Tasks for Instruction and/or Assessment

F9 *Activity*

F9.1. The senate at a university with 12 000 students is interested in knowing how many hours per week students spend studying outside of class. Two hundred fifty students are selected at random and asked, "How many hours per week do you spend studying?"

a) What is the population of interest?
b) What group constitutes the sample?

F9.2. A biologist is researching bats and their ability to detect insects within close proximity of them. The following data represents the distances (cm) at which 11 different bats first detect a nearby insect:

62 23 27 56 52 34 42 40 68 45 83

a) Is this a sample or a population? Explain.
b) Without using the statistical capabilities of your calculator, determine the standard deviation.
c) Check your answer to (b) using technology.

F9.3. Mr. Sweeney has a small class of 15 math students. He is interested in knowing the mean and standard deviation of the class on their quiz. Their marks are given below:

82 76 65 78 81 90 52 93 50 89 70 85 59 60 52

a) Explain why this group represents a population and not a sample.
b) Calculate the mean and standard deviation of this data.
c) Create a set of data that has a smaller standard deviation and explain how you created it.

F9.4. The mathematics department at a small university expects all students enrolling in Mathematics 1000 to take an entry test. The exam is marked out of 20. The following shows the results of all 40 students who wrote the exam:

9 12 9 10 15 12 11 12 10 17 7 12 12 12 14 10 11 13 5 12
12 12 11 9 12 8 14 14 11 9 16 12 11 13 14 15 12 11 11 14

a) Construct a histogram and a frequency polygon.
b) Does the distribution appear roughly "bell-shaped"? Explain.
c) Does this data set represent a sample or a population? Explain
d) Using technology, find the mean and standard deviation.
e) How many scores fall within one standard deviation of the mean? Express this as a percentage of all the scores.

Suggested Resources

See the following website for more discussion on the use of n versus $n-1$ in the standard deviation formulas:

"Unbiased Estimators." *Herkimer's Hideaway*. http://herkimershideaway.org/writings/unbiae.htm

See the following website for a sample experiment using the standard deviation formulas:

"Re: A Note on Standard Deviation." *Math Central* [University of Regina]. http://mathcentral.uregina.ca/QQ/database/QQ.09.99/freeman2.html

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

A3 demonstrate an understanding of the application of random numbers to statistical sampling

F15 design and conduct surveys and/or simulate data collection to explore variability

F10 interpret and apply histograms

Elaboration – Instructional Strategies/Suggestions

A3/F15/F10 Introduce students to data creation through simulations. A simulation is the imitation of chance behaviour, based on a model that accurately reflects the experiment under consideration. The data generated in a simulation is always collected in a random fashion. Flipping a coin, drawing names from a hat, and utilizing random number generators are common methods used to generate data for simulations. For example, if we wanted to predict the average number of female children in a family with four children, we could use a coin. We could let “flipping a head” represent a male child and “flipping a tail” represent a female child. Then, we could flip the coin four times and count the number of tails. This would represent the number of female children in that family. If we repeated this process 100 times and found the average number of females per family, we would have conducted a valid simulation.

Likewise, we could use a random number generator. We could randomly generate an integer from 0 to 4 (inclusive) from an appropriate probability distribution to represent the number of female children in a family with four children. Generating 100 such numbers should provide us with a large enough data set to accurately predict the average number of female children in a family with four children.

Students can design and conduct simulations to model real-world phenomena. As well, students should realize that, because their data sets are randomly generated, they will usually generate data sets that are different from those of other students. They can organize their data into graphs and look for patterns that emerge.

It is also important that students distinguish between generating data in a simulation and drawing a sample from an actual population. In the example of female children above, the random number generator can be used to *imitate* surveying various families at random and recording the number of female children in each family. On the other hand, if we wished to take a sample of actual families and record the number of female children in each family, we might assign a number to each family in the population of interest and use a random number generator to select the families for the sample.

F15 When designing a simulation or survey, students should keep in mind that data must be created (in the case of a simulation) or collected (in the case of a survey) in a random fashion. Assumptions must also be stated, a trial must be defined, and the survey or simulation described. When randomly generating an integer from 0 to 4 (inclusive) to represent the number of female children in a family with four children, this would represent one trial of the experiment. This procedure would have to be repeated a large enough number of times to obtain a suitably reliable *simulated* sample. If each outcome from 0 to 4 in a trial is equally probable, then the assumption has been made that in reality each possible number of female children is equally probable and that there have been no outside factors that might influence the number of female children in a particular family.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

A3/F15/F10 *Activity*

A3/F15.1. Mytown High School produced 150 graduates last year. As part of a curriculum review, the government would like to ask 20 graduates of this school their perception of the value of the mathematics curriculum. Ask students to describe a method that the government might use to randomly select 20 students.

A3/F15.2. In a psychological study, subjects are shown 10 different images and asked to select one at random. There are a total of 100 subjects. Have students design a simulation that models this situation.

A3/F15/F10.1. A certain game of chance is based on randomly selecting three numbers from 0 to 99 and adding the numbers. A person wins if the resulting sum is a multiple of 5. Ask students to respond to the following:

a) Using technology, design and conduct a simulation that models this situation for 30 games. Record the number of wins.
b) Repeat part (a) 5 times. You should now have six numbers, with each number representing the number of wins in 30 games.
c) Collect results from other students until you have a data set of at least 50 numbers. Construct a histogram.
d) Does the histogram appear bell-shaped? Explain.

A3/F15/F10.2. Ask students to work in pairs using measuring tapes to measure and record their heights. Then ask students to measure the distance between their finger tips (which are pointing in opposite directions) when they hold their arms outstretched parallel to the floor.

a) Have students compute the ratio of their arm span to their height.
b) Ask students to create a distribution of the class results.
c) Ask students if the distribution is bell-shaped. Have them explain why they think their answer is correct. If students collected more ratios, what would be the effect on the distribution?
d) Ask students to calculate the mean and standard deviation.
e) Using their results, ask students to determine what percentage of all students in the school would have a ratio below one. Ask them why their sample may not be indicative of the population.

(Continued on the next two-page spread ...)

Suggested Resources

A free application (APP) for the TI-83 Plus called *ProbSim,* which performs simulations with dice, spinners, coins, cards, and random numbers, available under Statistics on the Texas Instruments educational website at <http://education.ti.com>.

A free program for Windows-based computers called *Winstats* performs simulation, creates histograms, scatter plots, and boxplots and calculates confidence intervals and many other advanced statistical computations.
It is available on the *Peanut Software Homepage* at <http://math.exeter.edu/rparris>.

Some *Winstats* tutorials may be found on the Newfoundland and Labrador Department of Education website under Professional Development Math Resources, <http://www.gov.nf.ca/edu/sp/mathres/softwaretutorials/softwaretutorials.htm>.

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

A3 demonstrate an understanding of the application of random numbers to statistical sampling

F15 design and conduct surveys and/or simulate data collection to explore variability

F10 interpret and apply histograms

Elaboration – Instructional Strategies/Suggestions

Statistics

Worthwhile Tasks for Instruction and/or Assessment	Suggested Resources
(... Continued from the previous two-page spread.) A3/F15/F10.3. In this activity, students will explore random sampling using letters from a Scrabble game. Give each group one complete set of letters and have them respond to the following prompts: a) Make a histogram of the point values for the population of the tiles. (Blank tiles represent zero points.) b) Is the histogram bell-shaped? Explain. c) Find μ . d) Randomly select 30 tiles, *with* replacement, from the Scrabble bag and find the mean score of the sample. After you have collected your sample of 30, put the tiles back in the bag. e) Repeat part (d) until you have at least 10 means. f) Collect other means from your classmates until you have 100 means. Create a histogram of your data collected in part (f). Is the histogram bell-shaped? Is this surprising? Why or why not? How does it compare to the histogram in part (a)? g) Find the mean of your data collected in part (f). How does it compare with the population mean?	

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

F8 apply characteristics of normal distributions

Elaboration – Instructional Strategies/Suggestions

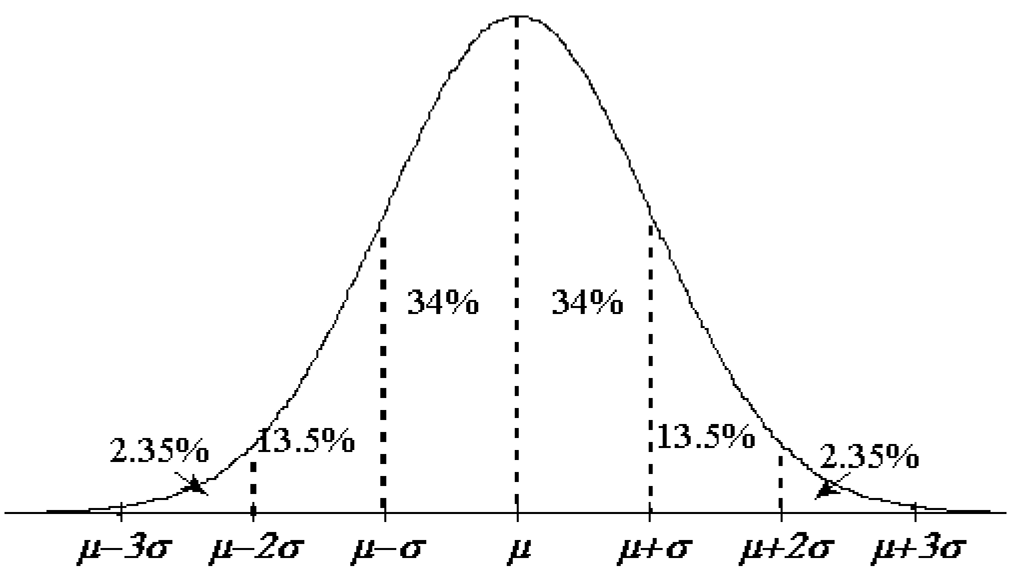
Students have studied properties of the normal distribution in a previous course. These properties may need to be reviewed. Students will need to know that a normal distribution

- is "bell-shaped"
- has a mean that is equal to the median
- is symmetrical and centred about the mean and median
- has most of its values clustered about the mean

Students should also know and understand the 68%-95%-99.7% rule:

- approximately 68%of the measurements are within 1 standard deviation of the mean
- approximately 95% of the measurements are within 2 standard deviations of the mean (for later work with confidence intervals, students will use the more precise value of 1.96 standard deviations from the mean)
- approximately 99.7% of the measurements are within 3 standard deviations of the mean

Students should be able to recognize a data set as being "approximately" normal when given a histogram. They can explore visual representations of different normal distributions and apply their knowledge of the normal distribution to real-world scenarios. They should also explore the symmetrical nature of the normal distribution:



These percentages represent the approximate percentages of the members of the population falling within each interval, and should be related back to work on frequency polygons and calculating percentages of population or sample members lying within certain classes on a histogram, e.g., refer to F10.1 and F10.2.

Note: In the diagram, μ and σ may be replaced with $\bar{x}$ and S_x if dealing with a sample that is normally distributed.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

F8 *Activity*

F8.1. A company is manufacturing hair dryers. The lifespan of a hairdryer is normally distributed with a mean of 6.5 years and a standard deviation of 1.5 years.

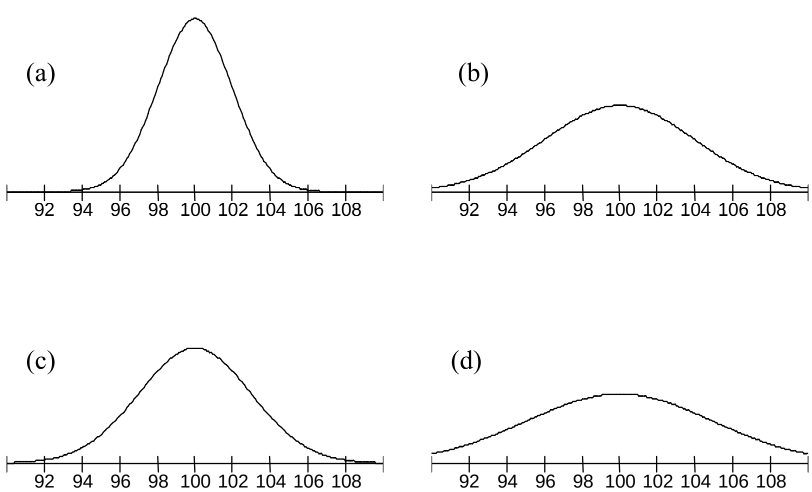
a) Draw a visual representation of this situation.
b) What percentage of hair dryers is expected to last between 5 and 8 years?
c) What percentage of hair dryers is expected to last between 3.5 and 8 years?
d) What percentage of hair dryers is expected to last more than 8 years?
e) The company decides to offer a one-year warranty. Does this seem like a good idea? Explain.

F8.2. A popular restaurant in Corner Brook does not take reservations, so there is usually a waiting time before customers are seated. The following data shows the waiting time (in minutes) of 100 randomly selected customers:

18	18	21	18	10	14	19	14	18	20	17	25	20	23	12	17	21	19	22	18
9	11	14	19	9	17	19	10	20	15	15	14	19	15	14	17	14	25	20	17
15	21	14	26	14	6	17	17	20	19	27	11	13	19	18	16	20	24	19	10
9	17	24	17	16	12	14	18	18	18	13	15	18	17	15	25	17	20	12	22
23	14	26	20	23	22	24	15	16	18	21	22	12	20	16	21	27	19	18	25

a) Is this a sample or a population? Explain.
b) Construct a histogram of the data. Does the data appear normally distributed? Explain.
c) Using technology, calculate the mean and standard deviation of the data.
d) What percentage of data lies within 1 standard deviation of the mean? 2 standard deviations? Do your percentages suggest that the data is normally distributed? Explain.

F8.3. Which normal distribution curve has the largest standard deviation? the smallest? Explain.



Suggested Resources

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

F14 distinguish between descriptive and inferential statistics

Elaboration – Instructional Strategies/Suggestions

Data, or numerical statistics derived from data, are essential for making many kinds of decisions in our lives. Statistical displays like bar graphs, stem-and-leaf plots, box plots, and histograms are ways in which we organize and present information. Descriptive statistics deal with organizing and summarizing data for effective presentation and for increased understanding of trends in the data.

Often the individuals or objects that are being studied are a sample that comes from a much larger population. The investigator or researcher may be interested in more information than just data summarization. Inferential statistics involves generalizing from a sample to the population from which it was selected. This often involves some risk, since some conclusions about the population might be reached on the basis of available, but incomplete information. The sample may possibly be unrepresentative of the population from which it came. So, an important part of the development of inference techniques would involve quantification of associated risks. For example, at most universities students are able to register for classes using a telephone. To assess the effectiveness of this system, one university has developed a set of questions to ask a sample of 150 randomly selected students who have used the system. This will result in a database of information. To make sense of the data and to describe student responses, it is desirable to summarize the data using various graphical displays. This would also make the results more accessible to others. In addition, inferential methods could be employed to draw various conclusions about the experiences of all students who used the system.

Some work on histograms from outcome F10 may be used as examples of descriptive statistics.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

F14 *Activity*

F14.1. The following data shows the number of shots on net that a minor hockey league team has taken in its last 40 games:

27 23 23 24 30 13 22 26 22 27 22 26 28 22 29 21 23 20 23 24
26 27 17 24 25 26 18 23 23 22 21 31 20 16 35 21 33 19 22 28

a) Using technology, calculate the mean, median, and standard deviation of the data.
b) Construct a histogram of the above data.
c) If you had to write a report to the team's general manager describing this data, what would you say?
d) If you were asked to predict how many shots you thought the team would take in the next game, what would be your response? Explain.
e) A friend tells you that he predicts that the team will take at least 30 shots on net in their next game. What would be your response? Why?
f) In which of the above questions were you using descriptive statistics? inferential statistics? Explain.

Suggested Resources

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

A3 demonstrate an understanding of the application of random numbers to statistical sampling

F2 identify bias in data collection, interpretation, and presentation

F19 demonstrate an understanding of the differences in the quality of sampling methods

Elaboration – Instructional Strategies/Suggestions

A3/F2/F19 To sample people in a mall is fast, cheap, and convenient. However, think about which people in the mall might be invited to participate. Often it is those who are well dressed, respectable looking, and friendly, because they seem easier to approach. The sample from malls may over-represent the middle class and retired, and under-represent the low income groups. Sampling techniques such as this one often produce data that is unrepresentative of the population. A sample is said to be biased if it systematically favours certain outcomes. Inferences about a population based on a sample that is biased are often inaccurate. It is preferable to collect unbiased samples. Even though unbiased samples will result in inferences that are at least slightly different from the population, unbiased samples are, on average, better representative of the population, and therefore, inferences based on unbiased samples tend to be more accurate.

Students should understand the difference between a biased and an unbiased sample. A unbiased sample is usually generated, at least in part, randomly. They will need to become be familiar with and be able to identify samples that tend to be unbiased:

- simple random sample—a sample that is selected from a population in a way that ensures that every different possible sample of the desired size has the same chance of being selected
- stratified random sample—a simple random sample selected from each of the given number of subpopulations, or strata
- cluster sample—a sample in which the population is divided into groups (or clusters), a cluster is randomly selected, and every member of that cluster is included in the sample
- systematic sample—a sample that is chosen according to a formula or rule

Students will also need to be able to identify samples that tend to be biased:

- convenience sample—a sample that uses results or data that are conveniently and readily obtained
- voluntary response sample—a sample where results are collected on a volunteer basis, such as a call-in poll or an online voting survey

Statistics

Worthwhile Tasks for Instruction and/or Assessment

A3/F2/F19 *Activity*

A3/F2/F19.1 During the previous calendar year, a city's small claims court processed 1562 cases. A legal researcher would like to select a random sample of 50 cases to obtain information regarding the average award in such cases. Assuming that the researcher wants his sample to be as unbiased as possible, describe two different methods the researcher could use to select his sample.

A3/F2/F19.2 A sociologist wants to know the opinions of employed adult women about government funding for day care. She obtains a list of the 520 members of a local business and professional women's club and mails a questionnaire to the first 100 people on the list; 63 surveys are returned.

a) What type of sample is this?
b) What are some sources of bias in her survey?
c) How would you improve her sampling method to reduce bias?

F2.1 Identify the sampling method used and identify possible sources of bias in each:

a) Shona stood by the mall entrance and questioned every 10th person.
b) Avril mailed questionnaires to all of the people who had rented videos at her store in the past three months.
c) Dominic randomly selected one of his hometown's eight skating arenas, and then surveyed all of the people who attended the next hockey game.
d) Ron visited several day-care centres and asked a few questions of children whom he judged to be typical four-year-olds.

A3/F2/F19.3 Marla stood by the front door of the local theatre immediately at the end of the movie, and asked every fifth person who left how much money they had spent at the popcorn counter. Of the 90 people asked, only 52 agreed to respond. Marla told the theatre management that the typical moviegoer spends $5.00 on treats before or during the movie.

a) What sampling method did she use?
b) Do you think these results are biased? Justify your reasoning.
c) Explain in a paragraph to Marla how she might have improved her results to be more representative of all moviegoers.

Suggested Resources

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

A3 demonstrate an understanding of the application of random numbers to statistical sampling

F8 apply characteristics of normal distributions

F15 design and conduct surveys and/or simulate data collection to explore sampling variability

Elaboration – Instructional Strategies/Suggestions

A3/F8/F15 Students have already been exposed to simulations and how simulations can be used to generate data. Students should now simulate sampling procedures from normal populations where the population mean (μ) and population standard deviation (σ) are known. Students can then compare sample means ($\bar{x}$) to μ .

For example, consider an IQ test where the scores are normally distributed with $\mu = 100$ and $\sigma = 15$. A random sample of 200 people is to be selected from this population. In order to simulate taking a random sample from this population, we need a method of generating numbers from a normal distribution with specified values of μ and σ for a specified number of trials. The randNorm(function on the TI-83 Plus is one method of generating such data. This function can be accessed by using the following keystrokes:

~ ~ ~ PRB – 6: randNorm(

The syntax for this command is:

randNorm(μ, σ ,number of trials)

Therefore, if a student wants to simulate taking a random sample of 200 people from the IQ test population given above, they should execute the following command:

randNorm(100,15,200)$\rightarrow L_1$ (Keystrokes ~ ~ ~ PRB - 6: randNorm(100,15,200) ¿ y À)

The "$\rightarrow L_1$" at the end of the command stores the list of 200 numbers in the first list, L_1 .

Students should then calculate the mean of their sample (see note below) and compare this to the population mean (100). This process should be repeated a number of times to explore sampling variability. Students should realize that μ remains constant while $\bar{x}$ varies. In later sections, students will learn how to use $\bar{x}$ to estimate the value of μ when μ is unknown.

Note: On the TI-83 Plus, it is possible to find the mean of a set of data without storing the data in a list. We can use the mean(command, which can be accessed using the following keystrokes:

y ... ~ ~ 3: mean(

For example, if we execute the command mean(randNorm(100,15,200)), the TI-83 Plus will generate a sample of size 200 from a normal distribution with $\mu = 100$ and $\sigma = 15$ and then return only the mean of this sample; it will not display the data in the simulated sample.

Note: While students are expected to be able to use graphing calculators and/or other technology for performing simulations, it is NOT expected that they memorize keystrokes and commands for the purposes of summative evaluation.

If available, computer software such as MINITAB can be used to carry out such simulations.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

A3/F8/F15 *Activity*

A3/F8/F15.1. The amount of soft drink in a particular brand of cola is normally distributed with $\mu = 355$ ml and $\sigma = 5.3$ ml.

a) Design a simulation to generate a sample of 100 soft drinks from this population.
b) Conduct your experiment and find the sample mean.
c) Compare the sample mean to the population mean. Do you think that þ is a good indicator of μ? Why or why not?
d) Repeat your simulation 3 more times and find the mean of each sample. Are your means the same? Explain.
e) If you did not know that $\mu = 355$ ml, would you be able to tell which value of þ was the closest to μ ? Explain.

A3/F8/F15.2. The lifespan of a certain brand of car is normally distributed with $\mu = 10$ years and $\sigma = 2$ years.

a) Design a simulation to generate the lifespan of 30 cars from this population.
b) Create a histogram of your data collected. Describe the shape of the histogram.
c) Conduct your simulation again, and generate the lifespan of 100 cars.
d) Create a histogram of your 100 lifespans. How does this histogram compare to the one you created in part (d)?
e) Based on your results to the above, how could you make your histogram appear more bell-shaped? Explain

A3/F8/F15.3. **Paper Activity Part 1:** Obtain seven different colours of 8" by 11" paper, e.g., one sheet each of yellow and white, three sheets each of pink and brown, six each of orange and blue, and nine sheets of green. Cut the paper into one-inch squares. Place each set of coloured paper into transparent containers such as plastic water bottles (each color in a different container) with the tops cut off, arranged to form an approximation of a “bell” curve (Note this is NOT truly a normal population.) Let each colour represent a number from 1 to 7, with green = 4. Discuss attributes such as the mean and standard deviation (dispersion) of the distribution.

Suggested Resources

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

G3 graph and interpret sampling distributions of the sample mean

F4 demonstrate an understanding of how sample size affects the variation in sample results

Elaboration – Instructional Strategies/Suggestions

G3/F4 Students should be introduced to sampling distributions of the sample mean. They will explore how the means of samples of size *n* are distributed if samples are repeatedly collected.

To create a good approximation of the sampling distribution of the sample mean, we often need the means from a large number of samples. Therefore, it may be beneficial to explore the sampling distribution of the sample mean as a whole-group activity. Consider the following example: The time taken for a certain reading test for 12-year-old children is normally distributed with a mean of 60 minutes and a standard deviation of 15 minutes. Figure 1 represents this normal distribution. Students could then simulate collecting a number of random samples of size 30 from this population and find the mean time for each sample. The mean times could then be collected as a class and a histogram constructed. (Note: The class should generate at least 100 samples, and the same scale for the horizontal axis that was used for the population distribution should be used.) A histogram based on 100 random samples is provided in Figure 2.

Population Distribution of Times

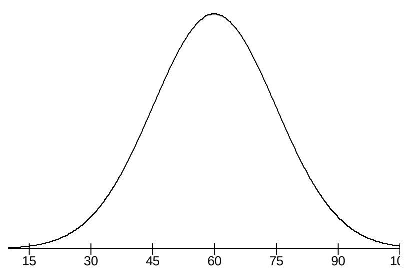


Figure 1

Distribution of 100 Sample Means (n = 30)

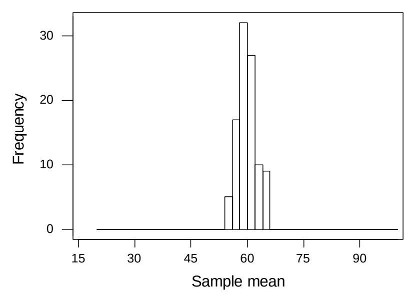


Figure 2

The histogram in Figure 2 is an approximation of the sampling distribution of the sample mean. Students should note that

- It appears approximately normal.
- The sampling distribution is centred at approximately 60 minutes, which is the population mean. (Calculating the mean of the data in the histogram should confirm this fact.)
- The histogram of the sample means has a smaller standard deviation. That is, there is less variability in the sampling distribution of the sample mean than in the population distribution.

Students should then explore the relationship between σ and the standard deviation of the data in their histogram. Students may need guidance to discover that the standard deviation of their simulated data should be approximately $\frac{\sigma}{\sqrt{n}}$, which in the above example is $\frac{15}{\sqrt{30}} \approx 2.74$.

Students can then repeat this exercise with a larger sample size (such as 60) to see the effect on the variability of the sampling distribution. A larger sample size will produce a more clustered histogram, which results in less variability.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

G3/F4 *Activity*

G3.1. An airline reports that the average weight of passenger baggage is 25 kg with a standard deviation of 8.2 kg.

a) Draw a visual representation of this population.
b) Design a simulation to collect a random sample of 40 weights from this population.
c) Conduct your simulation and find the mean of your sample.
d) Repeat part (c) until you have 10 sample means.
e) Collect results from other students until you have at least 100 means.
f) Construct a histogram of your data.
g) Compare your histogram in part (f) to your visual representation in part (a). What do they have in common? What is different?
h) Calculate the mean and standard deviation of your data. How do these values compare with μ and σ?

G3/F4.1. The age of university students in a certain province is normally distributed with an age of 24.1 years and a standard deviation of 2.3 years.

a) Describe how to create an approximate sampling distribution of the sample mean for a sample size of 50.
b) At what value would you expect your sampling distribution to be centred?
c) What would you predict the standard deviation of your sampling distribution to be?
d) If you were creating an approximate sampling distribution of the sample mean based on a sample size of 100, how would it be similar to the sampling distribution you created in part (c)? How would it be different?

G3.2. **Paper Activity Part 2 (continued from the previous Worthwhile Tasks):** Place the paper squares into a bag and shake vigorously. Have each student draw several samples of size 10 (see below—a larger sample size may be desired), ***replacing each sample before drawing the next,*** and record each sample mean. Then, combine class results until there are at least 100 sample means and construct a histogram. The class should make visual comparisons of the original population with the distribution of sample means, particularly the means and standard deviations, then calculate the mean of the sample means and compare to the population mean. Repeat with larger sample sizes. If the sample size is 30 or greater, students can also compare the value of $\frac{\sigma}{\sqrt{n}}$ to the standard deviation of the sampling distribution of the sample means. Discuss the hypothetical situation of obtaining an infinite number of sample means and the shape of the resulting distribution.

Suggested Resources

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

F8 apply characteristics of normal distributions

G3 graph and interpret sampling distributions of the sample mean

F4 demonstrate an understanding of how sample size affects the variation in sample results

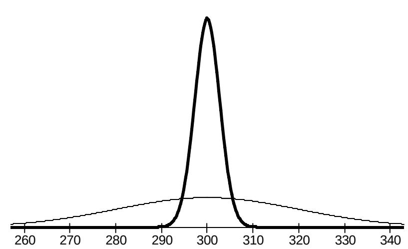
Elaboration – Instructional Strategies/Suggestions

F8/G3/F4 Students have explored approximate sampling distributions of the sample mean through simulations. However, because their histograms were based on a finite number of sample means, they could not create the actual sampling distribution of the sample mean. To describe the actual sampling distribution of the sample mean, we use the Central Limit Theorem:

- If samples of size *n* are drawn at random from any population with a finite mean and standard deviation, then the sampling distribution of the sample mean $\bar{x}$ is approximately normal when *n* is large (when $n \geq 30$).
- The mean of the sampling distribution is equal to the population mean ($\mu_{\bar{x}} = \mu$).
- The standard deviation of the sampling distribution is equivalent to the following: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

Students will need to understand and apply the Central Limit Theorem. Given a population where μ and σ are known, students should be able to determine the mean and standard deviation of the sampling distribution. Students should also understand that as the sample size increases, the variability in the sampling distribution decreases, and that the sampling distribution becomes more normal. This result is true *regardless* of the shape of the population distribution (for more on this, the online tutorial recommended under **Suggested Resources** on the next page can be accessed). Activity F8/G3/F4.3 on the next page can also be used to illustrate this point for students.

Students could be presented with examples such as the following: A company claims that one of its brands of cake has an average fat content of $\mu = 300$ grams with a standard deviation of 10 grams. Students could be asked to describe the sampling distribution of the sample mean if samples of size 50 were randomly and repeatedly taken from this population. Students could use the Central Limit Theorem to determine the mean of the sampling distribution ($\mu_{\bar{x}} = \mu = 300$ grams) and the standard deviation of the sampling distribution ($\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{50}} \approx 2.828$). Both the population distribution and the sampling distribution are graphed below (the sampling distribution is bolded):



Students should notice that the sampling distribution has less variability than the original population. Students could then be asked to describe the sampling distribution of the sample mean for samples of size 100. Again, using the Central Limit Theorem, students should determine that $\mu_{\bar{x}} = \mu = 300$ grams and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{100}} = 2$. This sampling distribution has less variability than the original population and that of the sampling distribution based on a sample size of 50.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

F8/G3/F4 *Activity*

F8/G3/F4.1. The base of live aspen trees in a national park is normally distributed with a mean of 34.5 cm and a standard deviation of 10.3 cm.

a) If samples of size 30 were repeatedly collected, what would be the mean of the sample means?

b) If samples of size 30 were repeatedly collected, what would be the standard deviation of the sample means?

c) How would your sampling distribution compare to (a) and (b) if samples of size 400 were repeatedly collected?

F8/G3/F4.2. The age of professional football players in a certain league is normally distributed with an average age of $\mu = 27.8$ and $\sigma = 3.5$ years. The following data shows the age of 40 randomly selected players from this league:

32.1	23.8	25.8	33.6	27.9	28.3	26.2	29.2	35.5	25.5
31.4	29.8	33.4	31.3	32.4	26.3	30.0	28.2	24.0	27.2
30.4	31.7	35.2	30.3	22.2	28.0	27.9	29.2	23.7	29.9
28.1	31.2	30.4	26.9	30.2	27.9	34.7	30.4	28.9	27.2

a) What is the sample mean? Is it close to the expected value? Explain.

b) If samples of size 40 were to be repeatedly collected, accurately predict the values of the mean and standard deviation of the sample means.

F8/G3/F4.3. Place 100 identical pieces of paper with the number 1 written on them in a clear plastic container such as a soft drink bottle with the top cut off. Repeat this for the numbers 2 through 5, using pieces of paper identical to those used in the first container. This will be the population, with $\mu = 3$ and $\sigma \approx 1.41$.

a) Construct a histogram for the distribution.

b) Place all the numbered pieces of paper in a bag and mix thoroughly. Randomly select a sample of size 30, without replacement, from the bag, and then calculate the sample mean and standard deviation. Compare these values to the corresponding population values.

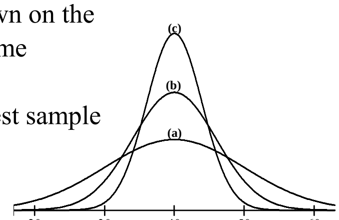
c) Repeat the procedure in b) until 100 such samples have been selected.

d) Construct a histogram for the sample means you obtained. How does its shape compare with that of the population? Is the standard deviation of the sampling distribution of sample means larger or smaller than the population standard deviation?

F8/G3/F4.4. Three sampling distributions of sample means are shown on the right. All three distributions were obtained from the same population. However, the sample sizes differ.

a) Which sampling distribution results from the largest sample size? the smallest? Explain.

b) What is the mean of the original population? How do you know?



Suggested Resources

For an online tutorial on the Central Limit Theorem, see the "Central Limit Theorem Statistics Workshops" on the Wadsworth website: <http://www.wadsworth.com/psychology_d/templates/student_resources/workshops/stat_workshp/cnt_lim_therm/cnt_lim_therm_01.html>.

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

F1 draw inferences about a population based on a sample

F11 determine, interpret, and apply confidence intervals

F20 demonstrate an understanding of how a confidence level affects a confidence interval

Elaboration – Instructional Strategies/Suggestions

F1/F11 Students should understand what confidence intervals are and how they can be used to make inferences about the population. Students should be able to identify and understand what is meant by a point estimate, an interval estimator, a confidence level, and a confidence interval. Initially, students can focus on interpreting and understanding confidence intervals, not calculating them.

While the population mean is fixed, it is often not known. A confidence interval is a method used to infer the population mean based on information collected from a sample. Students should understand that a confidence interval for the population mean is an interval of plausible values for the population mean. It is constructed so that the value of the population mean will likely be captured inside the interval with a chosen degree of confidence. In this course, students will work with three different confidence levels: 90%, 95%, and 99%. Students should also compare 90%, 95%, and 99% confidence intervals for the same data set to explore effect of the confidence level on the confidence interval. They should understand that a higher confidence level will result in a larger interval for a given sample size. However, if we increase the sample size, we can decrease the interval while maintaining the same level of confidence.

Discuss cases that may require different levels of confidence. For example, most polls regarding preferences for political parties are reported as accurate “19 times out of 20” (or with 95% confidence). On the other hand, the medical profession would require 99% or greater confidence, and likely a very large sample size, when examining the side effects of a new drug.

To help clarify this, the following example may be presented. A random sample of 100 people was selected, and their sleeping habits were studied. It was discovered that the mean time taken to fall asleep was $\bar{x} = 23.2$ minutes with a standard deviation of $S_x = 6.3$ minutes. From this information, the researchers conclude that they are 95% confident that the actual time it takes a person to fall asleep at night is between 22.0 minutes and 24.4 minutes. Students should understand that the researchers do not know the population mean or standard deviation because they studied a sample of only 100 people. Instead, the researchers will use the information to create a plausible range of the population mean. Students should identify that the point estimate is 23.3 minutes because this single number is used as a plausible value of the population mean. The interval estimator (or confidence interval) is $(22.0, 24.4)$ because this is the interval that the researchers believe contains the true population mean. However, because a sample will have different characteristics than a population, the researchers cannot be 100% confident that their interval actually contains the true population mean. Instead, the mathematical procedure used to create this confidence interval will entrap the population mean 95% of the time (for more elaboration on this subtle but important point, see the last paragraph below). Therefore, the confidence level is 95%. Students should not be initially concerned about how the confidence interval was derived. Instead, they should understand what a confidence interval is and what information about the population it gives.

Students could then be given the 90% and 99% confidence intervals for this set of data to explore the relationship between the confidence level and the confidence interval. The 90% and 99% confidence intervals are provided below:

- 90%—between 22.2 and 24.2 minutes
- 99%—between 21.6 and 24.8 minutes

It is tempting to say that there is a 95% chance that the population mean is between 22.0 minutes and 24.4 minutes. However, this is incorrect. The 95% refers to the percentage of all possible samples resulting in an interval that contains μ. That is, if we take sample after sample from the population and use each one individually to calculate a confidence interval, in the long run 95% of them will capture μ.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

F1/F11/F20 *Activity*

F1/F11/F20.1. A botanist collects a sample of 50 iris petals and measures the length of each. It is found that $\bar{x} = 5.55$ cm and $S_x = 0.57$ cm. He then reports that he is 95% confident that average petal length is between 5.39 cm and 5.71 cm.

a) Identify the point estimate, the confidence interval, and the confidence level.
b) Explain what information the confidence interval gives about the population of iris petal length.
c) How would the length of a 99% confidence interval be different from that of a 95% confidence interval?
d) If you did not know the point estimate but were still given that the confidence interval is between 5.39 cm and 5.71 cm, how could you determine the point estimate?

F11/F20.1. Explain what is meant by a 90%, a 95%, and a 99% confidence interval. How are these intervals similar? How are they different?

Suggested Resources

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

F1 draw inferences about a population based on a sample

F7 draw inferences from graphs, tables, and reports

F8 apply characteristics of normal distributions

F11 determine, interpret, and apply confidence intervals

F20 demonstrate an understanding of how a confidence level affects a confidence interval

F21 demonstrate an understanding of the role of the Central Limit Theorem in the development of confidence intervals

F22 distinguish between the calculation of confidence intervals for a known population mean versus an unknown population mean

G3 graph and interpret sampling distributions of the sample mean

Elaboration – Instructional Strategies/Suggestions

F1/F7/F8/F11/F20/F22/G3 Once students have an understanding of what a confidence interval is, they can determine and interpret confidence intervals from samples where the population mean and standard deviation are known. While this type of question is unrealistic (there is no need to calculate a plausible interval of μ if we already know the value of μ), it can help students further understand the meaning of a confidence interval.

The construction of a confidence interval is based on the Central Limit Theorem. If we construct a sampling distribution of the sample mean based on a sample of size *n*, we know that the distribution will be approximately normal with $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. To create a 95% confidence interval, we use the fact that approximately 95% of the data in a normal distribution lies within 1.96 standard deviations of the mean (see Figure 1). If we take a sample of size *n* and calculate $\bar{x}$, it has a 95% chance of lying within $1.96\sigma_{\bar{x}}$ of $\mu_{\bar{x}}$. Therefore, if we create the interval $\bar{x} \pm 1.96\sigma_{\bar{x}}$, we have a 95% chance of creating an interval that contains $\mu_{\bar{x}}$, which equals our original population mean (see Figure 2). Since there is a 5% chance of obtaining a sample that is beyond 1.96 standard deviations of the mean, there is a 5% chance that our confidence interval will not contain the population mean.

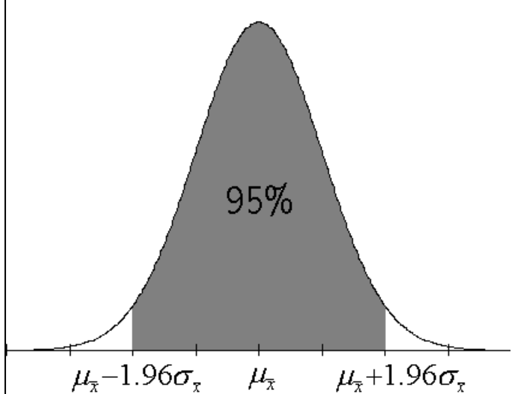


Figure 1

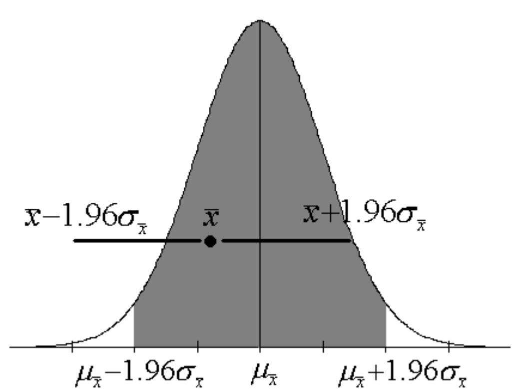


Figure 2

Students should determine and interpret 90%, 95%, and 99% confidence intervals from sample data where μ and σ are known.

(Continued on the next two-page spread ...)

Statistics

Worthwhile Tasks for Instruction and/or Assessment

F1/F7/F8/F11/F20/F22/G3 *Activity*

F1/F8/F11/F20/F22.1. The development time for a particular type of photographic printing paper is normally distributed with $\mu = 30$ minutes and $\sigma = 7.5$ minutes. A random sample of 50 developing times is collected, and the sample mean is determined to be $\bar{x} = 31.9$ minutes.

a) What is the point estimate?
b) If you were to determine the 90%, 95%, and 99% confidence intervals, which would be the smallest? the largest? Explain how you know.
c) Determine the 90%, 95%, and 99% confidence intervals for the mean developing time.
d) Does each confidence interval contain the population mean? Explain.
e) Another random sample, this time of 200 developing times, is collected, and the sample mean is $\bar{x} = 31.8$ minutes. Determine the 90%, 95%, and 99% confidence intervals based on this sample, then compare to the corresponding intervals constructed in part (c) for the sample of size 50. What do you notice?
f) Compare your 95% interval for sample size 200, with the 90% interval for sample size 50. Which is shorter? Explain why this happens.

F1/F7/F8/F11/F22.1. The mass of trout in a particular pond is normally distributed with $\mu = 445$ grams and $\sigma = 100$ g. The following represents the mass of 40 randomly selected trout:

446	407	447	380	400	511	467	257	613	530
599	359	489	451	352	710	322	410	633	520
624	519	427	503	547	275	519	505	303	396
408	308	358	623	426	402	347	443	460	588

a) What is the sample mean?
b) Determine the 90%, 95%, and 99% confidence intervals for the mean mass of the trout in this particular pond.
c) Another random sample of 100 trout was taken from the same pond and has the same sample mean as part (a). If a 95% confidence interval was calculated using this new sample, how would it differ from the 95% confidence interval calculated in part (b)?

F11/F20.1. A sample is taken from a normal population. Based on this sample, Lenna claims that she is 90% confident that the population mean is between 45.6 and 50.2. Based on the same sample, Phillip claims that he is 95% confident that the population mean is between 46.6 and 49.2. How do you know that someone has made a calculation error?

(Continued on the next two-page spread ...)

Suggested Resources

Winstats is a free program that has an excellent confidence interval demonstration. It is available on the *Peanut Software Homepage* at <http://math.exeter.edu/rparris>.

Tutorials for *Winstat*s may be found on the Newfoundland and Labrador Department of Education website under Professional Development Math Resources, <http://www.gov.nf.ca/edu/sp/mathres/softwaretutorials/softwaretutorials.htm>.

Applets demonstrating confidence intervals can be found at Rice University's Virtual Lab in Statistics, <http://www.ruf.rice.edu/~lane/stat_sim/conf_interval>, and on the University of South Carolina Department of Statistics website, <http://www.stat.sc.edu/~west/applets/ci.html>.

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

F1 draw inferences about a population based on a sample

F7 draw inferences from graphs, tables, and reports

F8 apply characteristics of normal distributions

F11 determine, interpret, and apply confidence intervals

F20 demonstrate an understanding of how a confidence level affects a confidence interval

F21 demonstrate an understanding of the role of the Central Limit Theorem in the development of confidence intervals

F22 distinguish between the calculation of confidence intervals for a known population mean versus an unknown population mean

G3 graph and interpret sampling distributions of the sample mean

Elaboration – Instructional Strategies/Suggestions

(... Continued from the previous two-page spread.)

Consider the following example: The automatic opening device of a military cargo parachute has been designed to open when the parachute is a certain distance above the ground. The opening altitude is normally distributed with $\mu = 250$ m and $\sigma = 15$ m. A random sample of 40 opening altitudes is taken, and it is determined that $\bar{x} = 246.83$ m. Students should be able to calculate the 95% confidence interval as follows:

$$\bar{x} \pm 1.96\sigma_{\bar{x}}$$
$$246.83 \pm 1.96\frac{\sigma}{\sqrt{n}}$$
$$246.83 \pm 1.96\frac{15}{\sqrt{40}}$$
$$\text{from } 242.18 \text{ to } 251.48$$

That is, we are 95% confident that the mean opening altitude is between 242.18 m and 251.48 m. Students should note that $\frac{\sigma}{\sqrt{n}}$ was substituted for $\sigma_{\bar{x}}$. This is a direct result of the Central Limit Theorem. Students should note that, in this instance, their confidence interval contains the population mean. This will happen for approximately 95% of our samples since this is a 95% confidence interval. Students could also calculate the 90% and 99% confidence intervals by understanding that

- 90% of the data in a normal distribution will lie within 1.645 standard deviations of the mean
- 99% of the data in a normal distribution will lie within 2.56 standard deviations of the mean

To calculate the 90% confidence interval, for example, students follow the same calculation as above, except they replace 1.96 with 1.645:

$$\bar{x} \pm 1.645\sigma_{\bar{x}}$$
$$246.83 \pm 1.645\frac{\sigma}{\sqrt{n}}$$
$$246.83 \pm 1.645\frac{15}{\sqrt{40}}$$
$$\text{from } 242.93 \text{ to } 250.73$$

Again, students should realize that this interval contains the population mean, and that this will happen for 90% of the samples since this is a 90% confidence interval. Students should also note that the 90% confidence interval is smaller than the 95% confidence interval, reaffirming the fact that higher confidence levels will result in larger confidence intervals for samples of the same size.

It is very important that students also consider the effect of sample size on the length of a confidence interval. While higher levels of confidence will produce longer confidence intervals when constructed from samples of the same size, using larger sample sizes will produce shorter confidence intervals. It is possible, for example, that a 90% confidence interval based on a sample of size 40 will be longer than a 95% confidence interval based on a sample of size 100 from the same population. Students could repeat the above example, but this time assuming a sample size of 100 with the same sample mean of $\bar{x} = 246.83$ m. They should compare the lengths of the 90%, 95%, and 99% confidence intervals with the corresponding intervals based on the smaller sample size.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

F1/F7/F8/F11/F20/F22/G3 *Activity*

(... Continued from the previous two-page spread.)

F1/F7/F8/F11/F22/G3.1. In the assembly line of the Canyonaro four-wheel-drive sport vehicle, the amount of time taken to install the exhaust system is normally distributed with $\mu = 7.5$ minutes and $\sigma = 1.7$ minutes.

a) Using technology, design and conduct a simulation that randomly selects a sample of 50 installation times.
b) Find the mean of your sample.
c) Using your sample, create a 90% confidence interval.
d) Repeat parts (a), (b), and (c) until you have 10 confidence intervals.
e) Collect confidence intervals from other students until you have at least 100 confidence intervals.
f) What percentage of your confidence intervals contains the population mean? Is this close to what you expected? Explain.
g) Repeat the simulation one more time, but this time for a sample of 100 installation times. Create a 90% confidence interval and compare its length to the ones you created in parts (a), (b), and (c). What do you notice? Why does this happen?

Suggested Resources

Statistics

Outcomes

SCO: By the end of Mathematics 2204-05, students will be expected to

F1 draw inferences about a population based on a sample

F2 identify bias in data collection, interpretation, and presentation

F7 draw inferences from graphs, tables, and reports

F11 determine, interpret, and apply confidence intervals

F22 distinguish between the calculation of confidence intervals for a known population mean versus an unknown population mean

Elaboration – Instructional Strategies/Suggestions

F1/F2/F7/F11/F22 Students should determine and interpret confidence intervals from large samples ($n \geq 30$) without knowing the values of μ and σ. They will need to approximate σ from their sample data. When a simple random sample of size $n \geq 30$ is taken from a finite population without replacement, and the sample size is no more than 5% of the population size, we can approximate σ by S_x. We can also approximate σ using S_x when dealing with infinite populations or with finite populations where sampling with replacement takes place.

Students should explore examples similar to the following: A sociologist is studying the length of courtship before marriage in Kyoto, Japan. A random sample of 56 couples was interviewed. It was found that the mean length of courtship was 4.3 years with a sample standard deviation of 1.0 years. Students could be asked to create a 99% confidence interval for the mean length of courtship in Kyoto. Students should realize that neither μ nor σ are known. In this case, we know that $\bar{x} = 4.3$ years and $S_x = 1.0$ years. However, since the sample size is greater than 30, and the sample is small compared to the number of couples in Kyoto, Japan, students should realize that they can approximate σ with S_x. They could then perform the following calculation to construct their confidence interval:

$$\bar{x} \pm 2.56\frac{S_x}{\sqrt{n}}$$

$$4.3 \pm 2.56\frac{1.0}{\sqrt{56}}$$

from 3.96 to 4.64

Students could then explain that they are 99% confident that the mean length of courtship in Kyoto, Japan, is between 3.96 years and 4.64 years. Students should also discuss the limitations of the data in this example. They should understand that their confidence interval is only applicable to Kyoto, Japan. They could not apply their confidence interval to a different part of the world. For example, they could not claim that this is true for all of Japan, as the length of courtship may be different in rural areas of the country. (Note: The shape of the distribution of courtship lengths is not important for the construction of the confidence interval since the Central Limit Theorem holds regardless of the shape of the parent distribution.)

Students should also understand what is meant by the term “margin of error.” The margin of error indicates the accuracy of the point estimate. For example, if an article reports that the mean waiting time in a bank line is 9.3 minutes with a margin of error of ± 2.1 minutes, and that these results are accurate 19 times out of 20, students should recognize that we are dealing with a 95% confidence interval (as stated by the “19 times out of 20”), and that the confidence interval is given by:

point estimate $\pm$ margin of error

9.3 ± 2.1

from 7.2 to 11.4

Students should determine the margin of error for different confidence intervals. The margin of error is calculated by the expression $z\frac{S_x}{\sqrt{n}}$ where z is

- 1.645 for a 90% confidence interval
- 1.96 for a 95% confidence interval
- 2.56 for a 99% confidence interval

Statistics

Worthwhile Tasks for Instruction and/or Assessment

F1/F2/F7/F11/F22 *Activity*

F1/F2/F11/F22.1. A random sample of 35 red pine trees was selected from a large forest containing 100 000 trees. The mean diameter was determined to be $\bar{x} = 25.3$ cm with $S_x = 3.6$ cm.

a) If we were to create a 95% confidence interval for the mean diameter of red pine trees in this forest, would we be allowed to approximate σ with S_x? Explain.

b) Create a 95% confidence interval and explain its meaning.

c) Can you make any conclusions about the mean diameter for all red pine trees in Canada? Why or why not?

F1/F2/F7/F11/F22.1. The following data shows a random sample of the ages of football players in a large European soccer league consisting of 1000 players:

26	24	25	36	26	32	31	34
32	27	32	23	24	29	30	30
29	33	25	32	24	25	28	22
22	24	23	33	32	31	26	28
32	25	22	28	25	29	25	27

a) What is the mean and standard deviation of the sample?

b) Determine a 99% confidence level for the mean age of soccer players in this league, and explain the meaning of this confidence interval.

c) What is the point estimate?

d) What is the margin of error of your confidence interval? Explain the meaning of the margin of error.

e) Based on this data, what, if anything, can you conclude about the mean age of soccer players in a Canadian soccer league? Explain.

F1/F7/F11/F22.1. A report claims that the average family income in a large city is \$32 000. It states the results are accurate 19 times out of 20 and have a margin of error of $\pm\$2\,500$.

a) What is the confidence level in this situation? Explain what it means.

b) What is the confidence interval?

c) Explain the meaning of the confidence interval.

Suggested Resources

Statistics

Outcomes

SCO: By the end of Mathematics 2205, students will be expected to

F16adv demonstrate an understanding of the difference between situations that involve a binomial experiment and those that do not

F18adv identify the characteristics of a binomial experiment

Elaboration – Instructional Strategies/Suggestions

The following material applies only to students in the advanced course.

F16adv/F18adv Students should also understand the difference between a population proportion (*p*) and a sample proportion $\left(\hat{p}\right)$. For example, if 60% of people in a certain community believe that smoking should be banned in public areas, then $p = .60$. If, in a sample of size 40 from this community, a researcher finds that 22 people believe in a public smoking ban, then $\hat{p} = \frac{\text{number of successes}}{\text{number of trials}} = \frac{22}{40} = 0.55$. Students should also recognize that probability is always expressed as a number between 0 and 1.

Students will explore binomial experiments. Students should be able to identify the characteristics of a binomial experiment:

- It consists of a fixed number of identical trials, *n*.
- Each trial can result in one of two outcomes labelled success (*S*) and failure (*F*).
- Outcomes of different trials are independent; that is, the outcome of one trial does not affect the outcome of any other trial.
- The probability that a trial results in *S* is equal to the population proportion (*p*) and is the same for each trial.

Students should be able to distinguish between situations that are binomial experiments and those that are not and justify their reasoning. For example, students could be given the following situation: a fair coin is flipped 200 times and the number of heads is recorded. Students should realize that this is a binomial experiment because

- There are 200 trials.
- Each trial has two possible outcomes: a head (success) and a tail (failure).
- the probability of obtaining a head for one trial does not influence the probability of obtaining a head for another trial.
- The probability of obtaining a head in each trial is the same: 0.50.

Students could also be given situations that are not binomial and explain why they are not binomial. For example, a biology student is taking a biology quiz consisting of 20 true/false questions. He knows the answers to 15 of the questions, but has to guess the other 5 answers. In this example, students should realize that this is not a binomial experiment because the probability of success is not the same for each trial. In some trials, the probability of success is 0.50 (when the student is guessing), and in other trials, the probability of success is 1 (when the student knows the answer).

Students should also understand that many surveys conducted in real-world situations do not represent true binomial experiments because the sample is collected without replacement from a finite population, which does not result in independent trials. Consider the following example: In a community with a population of 1000, 60% of the people believe that smoking should be banned in public areas. A researcher randomly selects 40 people from this community and asks them if they believe that smoking should be banned in public areas. The probability of success for the first trial is $\frac{600}{1000} = 0.60$. Students should realize that the first person interviewed will not be interviewed again. If on the first trial, the researcher selected a person who believes in banning smoking in public areas, then the probability of success on the second trial decreases to $\frac{599}{999} = .5996$ since the first person cannot be selected again. Even though there is little difference in the probability of success between the first and second trials, the fact that a difference exists means that this is not a true binomial experiment. However, students should recognize that as long as the sample is not too large in comparison with the population (as long as the population is at least 20 times bigger than the sample) the situation could be approximated by a binomial experiment.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

F16adv/F18adv *Activity*

F16/F18.1. Which of the following are binomial experiments? Which are not? Explain your reasoning.

a) A dice is rolled 600 times and the number of 3s is recorded.
b) In a town of 500 people, 50 people are asked, “Do you own two or more cars?”
c) A student is taking a true/false quiz consisting of 100 questions. The student decides not to read the questions and selects an answer (true or false) at random. The number of correct responses is recorded.
d) A student taking the same exam in part (c) has studied and knows the answers to 80 of the questions. He guesses at the remaining 20 questions in a random fashion. The number of correct responses is recorded.
e) Fifty people have entered a contest. Each person is asked to randomly select one of three boxes. The number of people who select each box is recorded.

F16/F18.2. Which of the following can be approximated by a binomial experiment?

a) In a city of 300 000 people, 100 people are randomly selected and asked, “Are you 18 years of age or older?”
b) In a town of 1000 people, 200 people are randomly selected and asked, “Do you own at least one cat?”
c) In a city of 40 000 people, 500 people are randomly selected for a taste test. People must decided between Cola A, Cola B, and Cola C.

F16/F18.3. In a large city of 100 000 people, it is known that 70% of the people use the public transit system at least three times a week. In a survey of 200 people, 136 people claim to use the public transit system at least three times a week.

a) Explain why this situation is not a binomial experiment.
b) Explain why we can approximate this situation with a binomial experiment.
c) What is p? $\hat{p}$? Explain how you determined each.

Suggested Resources

Statistics

Outcomes

SCO: By the end of Mathematics 2205, students will be expected to

F16adv demonstrate an understanding of the difference between situations that involve a binomial experiment and those that do not

A3 demonstrate an understanding of the application of random numbers to statistical sampling

F15 design and conduct surveys and simulate data collection to explore sampling variability

Elaboration – Instructional Strategies/Suggestions

The following material applies only to students in the advanced course.

F16adv/A3/F15 Students should imitate collecting data from binomial situations using simulations. They should recognize how to use resources such as dice, coins, spinners, random number tables, and technology to design and create such simulations.

For example, students could be told that 70% of a student population in a certain province takes a high school French course. Students could then be asked to design a simulation that would allow them to take a random sample of 50 students from this population and calculate the sample proportion of students who take French. Some possibilities include

- Create a spinner divided into two regions. One region contains 70% of the area and is labelled "Takes a French course." The remaining 30% is labelled "Does not take a French course." Students could use the spinner 50 times, record the number of times it lands in the first region, and calculate the sample proportion.
- Use a random number table or the **randInt(** function on the TI-83 Plus. Let 0, 1, 2, 3, 4, 5, 6 represent a student who takes a French course and 7, 8, 9 represent a student who does not take a French course. Students could generate 50 numbers, record the number of successes, and calculate the sample proportion.

Students should conduct simulations for given binomial experiments a number of times to understand variability in sampling. They should recognize that sampling from a binomial experiment will not always give the same sample proportion.

Note: The randBin(function on the TI-83 Plus or TI-84 Plus can be used to generate random data for a binomial experiment. This function can be accessed through the following keystrokes:

~ ~ ~ 7:randBin(

The syntax for this command is:

randBin(number of trials, probability of success, number of samples)

In the above population of high school students in a particular province, we could generate a sample proportion by using the command randBin(50,0.70). If we wanted to collect 100 sample proportions, we could use the command randBin(50,0.70,100).

Statistics

Worthwhile Tasks for Instruction and/or Assessment

F16adv/A3/F15 *Activity*

F16/A3/F15.1. For each of the following, design a simulation that would allow you to imitate collecting a random sample from the population.

a) Sixty-five percent of students at a large university take a calculus course. A random sample of 75 students is taken from this population and asked, “Have you taken a calculus course?”

b) Five out of six people in a large city believe that dogs should be on leashes in public areas. A random sample of 30 people from the city is selected and asked, “Do you believe that dogs should be on leashes in public areas?”

c) Sixty-eight percent of households in a certain community each have a DVD player. A random sample of 250 households is selected and an occupant is asked, “Do you have a DVD player in your house?”

d) Fifty percent of students in a particular high school are involved in extracurricular activities. A random sample of 92 students is selected and asked, “Are you involved in any extracurricular activity?”

F16/A3/F15.2. A team of eye surgeons has developed a technique for a risky eye operation to restore the sight of people blinded from a certain disease. In a population of 20 000 people who have had the surgery, 75% of the patients recovered their eyesight.

a) Explain why this situation can be approximated by a binomial experiment.

b) Design and conduct a simulation that mimics collection of a random sample of 60 people from this population. Calculate $\hat{p}$ from your sample.

c) Repeat part (a) until you have three sample proportions.

d) Are all of your sample proportions the same? Explain why you think this happens.

Project

F16/A3/F15.3. The principal of the school wishes to determine whether students prefer the current lunch period of 40 minutes or a new proposed lunch period of 30 minutes that would also allow the school day to close 10 minutes earlier in the afternoon. Design and conduct a survey that would allow you to collect a random sample of student preferences. Compare your value for the sample proportion with others in your class.

Suggested Resources

A application (APP) for the TI-83 Plus called *ProbSim* performs simulations with dice, spinners, coins, cards, and random numbers. It is available under Statistics on the *Texas Instruments* educational website at <http://education.ti.com>.

Winstats is another free program that can also perform simulations. It is available on the *Peanut Software Homepage* at <http://math.exeter.edu/rparris>.

Tutorials for *Winstats* may be found on the Newfoundland and Labrador Department of Education website under Professional Development Math Resources, <http://www.gov.nf.ca/edu/sp/mathres/softwaretutorials/softwaretutorials.htm>.

Statistics

Outcomes

SCO: By the end of Mathematics 2205, students will be expected to

F4 demonstrate an understanding of how sample size affects the variation in sample results

G3 graph and interpret sampling distributions of the sample proportion

Elaboration – Instructional Strategies/Suggestions

The following material applies only to students in the advanced course.

G3/F4 Students should be introduced to sampling distributions of the sample proportion. They should explore how the sample proportions of samples of size *n* are distributed if samples are repeatedly collected.

To create a good approximation of the sampling distribution of the sample proportion, we often need the sample proportions from a large number of samples. Therefore, it may be beneficial to explore the sampling distribution of the sample proportion as a whole-group activity. For example, in a particular province, 75% of all doctors have a private practice. Students could simulate collecting a number of random samples of size 50 from this population and determine $\hat{p}$ (the proportion of doctors in each sample who have a private practice). The sample proportions could then be collected as a class and a histogram constructed. (Note: The class should generate at least 100 samples.) A histogram based on 100 random samples for a sample size of 50 is provided in Figure 1. Students should then repeat the process for a larger sample size (such as 100). A histogram based on 100 random samples for a sample size of 100 is provided in Figure 2. (Note: The same horizontal scale should be used for both histograms.)

Figure 1 (n = 50)

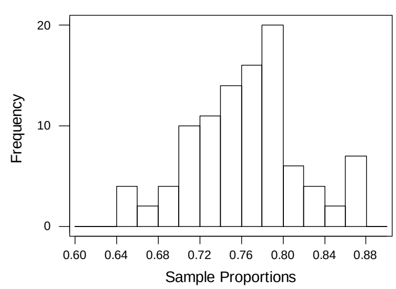
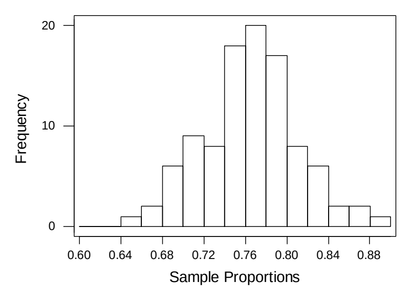


Figure 1 (n = 100)



Both histograms are approximate sampling distributions of the sample proportion. Students should note that

- The histograms appear approximately normal, and the one based on samples of size 100 is more normal than the one for samples of size 50.
- The sampling distributions are centred at approximately 0.75, which is the population proportion. (Calculating the mean of the data in the histograms should confirm this fact.)
- The histogram based on the sample size of 100 appears more clustered than the histogram based on 50 samples. That is, a larger sample size will produce a more clustered histogram, which results in less variability. (Calculating the standard deviation of both sets of data will confirm this fact.)

Students should then explore the relationship between the population proportion *p* and the standard deviation of the data in their histograms $\left(S_{\hat{p}}\right)$. Students may need guidance to discover that the standard deviation of their data can be approximated by $\sqrt{\frac{p(1-p)}{n}}$.

As a follow-up and for further investigation of these results, students can use technology to each generate 100 sample proportions for the above activity and create their own histograms to make comparisons with those of classmates. For the TI-83 Plus, the command would be randBin(50,.75,100) for sample size 50, and ranBin(100,.75,100) for sample size 100.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

F4/G3 *Activity*

F4/G3.1 Suppose that 77% of all households use coupons at grocery stores.

a) Design a simulation to imitate collection of a random sample of 60 households from this population.
b) Conduct your simulation and find the sample proportion of households that use coupons.
c) Repeat part (c) until you have 10 sample proportions.
d) Collect results from other students until you have at least 100 means.
e) Construct a histogram of your data. Describe the shape of your histogram.
f) Calculate the mean and standard deviation of your sample proportions. How could you have predicted these values using the population proportion?

F4/G3.2 Sixty percent of all students watch television in the morning before attending school.

a) Describe how to create an approximate sampling distribution of the sample proportion of students who watch television before school for a sample size of 50.
b) At what value would you expect your sampling distribution to be centred?
c) What would you predict the standard deviation of your sampling distribution to be?
d) If you were creating an approximate sampling distribution of the sample proportion based on a sample size of 200, how would it be similar to the sampling distribution you created in part (c)? How would it be different?

Suggested Resources

Statistics

Outcomes

SCO: By the end of Mathematics 2205, students will be expected to

F8 apply characteristics of normal distributions

G3 graph and interpret sampling distributions of the sample proportion

F4 demonstrate an understanding of how sample size affects the variation in sample results

Elaboration – Instructional Strategies/Suggestions

The following material applies only to students in the advanced course.

F8/G3/F4 Students have explored approximate sampling distributions of the sample proportion through simulations. However, because their histograms were based on a finite number of sample means, they could not create the actual sampling distribution of the sample proportion. The following results summarize the properties of the sampling distribution of the sample proportion where p is the population proportion:

- For random samples of size n where $np > 5$ and $n(1-p) > 5$, the sampling distribution of the sample proportion will be approximately normally distributed. (For the purposes of this course, these restrictions do not need to be mentioned to students.)
- The mean of the sampling distribution is equal to the population proportion $\left(\mu_{\hat{p}} = p\right)$.
- The standard deviation of the sampling distribution is equivalent to the following:
  $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$.

Students should understand and apply these results. Given a binomial situation where p is known, students should be able to determine the mean and standard deviation of the sampling distribution. They should understand that as the sample size increases, the variability in the sampling distribution decreases.

Students could be presented with examples such as the following: 60% of employees for a large company exercise during their lunch break. Students could be asked to describe the sampling distribution of the sample proportion if samples of size 30 were randomly and repeatedly taken from this population. Students could use the Central Limit Theorem to determine the mean of the sampling distribution ($\mu_{\hat{p}} = p = 0.30$ grams) and the standard deviation of the sampling distribution ($\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{0.30(0.70)}{30}} = 0.084$). Students could then describe the sampling distribution if samples of size 50 were randomly and repeated taken. Students could then compare the standard deviation of both sampling distributions to recognize that a larger sample size results in a lower standard deviation and less variability.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

F8/G3/F4 *Activity*

F8/G3/F4.1. In a large city, it is known that 15% of people have two or more jobs.

a) If random samples of size 150 were repeatedly taken from this population and a sampling distribution of the sample proportion of people with two or more jobs was created, what would be the shape of the distribution?

b) Where would the sampling distribution of $\hat{p}$ be centred?

c) What would be the standard deviation of the sampling distribution of $\hat{p}$?

F8/G3/F4.2. In a particular province, 50% of the voting public voted in the last election.

a) If random samples of 40 were repeatedly taken from this population and $\hat{p}$ was calculated for each sample, what would be the mean and the standard deviation of the sampling distribution of $\hat{p}$?

b) Design and conduct a simulation for this situation. Repeat your simulation until you have 10 sample proportions.

c) Collect other sample proportions from your classmates until you have at least 100 sample proportions.

d) Construct a histogram of your data in part (c). Describe its shape.

e) Calculate the mean and standard deviation of the data you collected in part (c). Are your results close to the expected values? Explain.

f) If random samples of size 100 were randomly taken and the histogram of the sample proportions was constructed, how would it compare to the histogram you created in part (d)?

Suggested Resources

Statistics

Outcomes

SCO: By the end of Mathematics 2205, students will be expected to

F11 determine, interpret, and apply confidence intervals

F1 draw inferences about a population based on a sample

F20 demonstrate an understanding of how a confidence level affects a confidence interval

F2 identify bias in data collection, interpretation, and presentation

F7 draw inferences from graphs, tables, and reports

Elaboration – Instructional Strategies/Suggestions

The following material applies only to students in the advanced course.

F11/F1/F20/F2/F7 Students have already explored confidence intervals to find plausible ranges for population means. Now, they will use confidence intervals to find plausible ranges for population proportions.

Students have been introduced to the terminology of confidence intervals in previous outcomes, so they should be familiar with the terms point estimate, interval estimator, confidence interval, margin of error, and confidence level. They can be exposed to examples similar to the following both as a review and to understand how these terms apply to confidence intervals for population proportions.

Example: An airline randomly selects 100 passengers and asks them if they were satisfied with their in-flight meal. In a report based on this survey, the airline claims that 34% of passengers are satisfied with the in-flight meals. Their results are accurate 19 times out of 20 with a margin of error of $\pm 9\%$. Students should be able to identify the point estimate (0.34), the interval estimator/confidence interval (from 0.25 to 0.43), the margin of error (± 0.09), and the confidence level (95%). Students should also realize that while they do not know the true value of *p*, they are 95% confident that it lies between 0.25 and 0.43 (or between 25% and 43%).

Students should also construct 90%, 95%, and 99% confidence intervals for population proportions. The construction of confidence intervals for the population proportion is similar to the construction of confidence intervals for the population mean. Since the sampling distribution for $\hat{p}$ is normally distributed with $\mu_{\hat{p}} = p$ and $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$, students should realize that a 95% confidence interval is constructed by

$$\begin{aligned} &\hat{p} \pm 1.96\sigma_{\hat{p}} \\ =\ &\hat{p} \pm 1.96\sqrt{\frac{p(1-p)}{n}} \end{aligned}$$

However, since we usually do not know the value of *p* (and therefore cannot calculate $\sigma_{\hat{p}}$), we approximate it with $\hat{p}$ when *n* is sufficiently large ($n \geq 30$). Therefore, when constructing 95% confidence intervals for sample sizes greater than or equal to 30, students will use:

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

(Constructing confidence intervals for sample sizes less than 30 are beyond the scope of this course.)

Students should also realize that to create a 90% confidence interval, we use $\hat{p} \pm 1.645\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, and for a 99% confidence interval, we use $\hat{p} \pm 2.56\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

(Continued on the next two-page spread ...)

Statistics

Worthwhile Tasks for Instruction and/or Assessment

Suggested Resources

F11/F1/F20/F2/F7 *Activity*

F11/F1/F20/F2/F7.1. In a national survey of 400 Canadians from the ages of 20 to 35, 37.5% of those interviewed claimed they exercise for at least four hours a week. The results were considered accurate within 4%, 9 times out of 10.

a) Are you dealing with a 90%, 95%, or 99% confidence interval? How do you know?
b) How many people in the survey claimed to exercise at least four hours a week?
c) What is the margin of error?
d) What is the confidence interval? Explain its meaning.
e) What are some limitations of this survey?
f) If the writers of the article created a 99% confidence interval based on this data, how would it be different? How would it be the same?
g) How would the confidence interval change if the sample size was increased to 500 but the sample proportion happened to remain the same?

F11/F1.1. A random sample of 500 first-year university students is obtained, and 355 of respondents claim to be attending the university of their first choice.

a) What is the sample proportion of students who claim to be attending the university of their first choice?
b) Construct a 90%, 95%, and 99% confidence interval for the population proportion of first-year students attending the university of their first choice.
c) If the sample size was decreased but the sample proportion happened to remain the same, how would your confidence intervals change?

F11/F1.2. Create a situation that would result in a 99% confidence interval from 73.5% to 86.5% based on a sample size of 200. Create at least three questions that could be asked about the situation and share them with a classmate. You may wish to ask questions about the sample proportion and population proportion, for example.

F11.1. A survey shows that 76% of high school students regularly attend dances. The survey was reported to be accurate to within 6.2%, 19 times out of 20. How many people were surveyed?

F11/F1.3. A fishing lodge in northern Alberta claims that 75% of its guests catch northern pike over 9.0 kg. A random sample of 83 guests indicated that 61 of them catch northern pike over 9.0 kg.

a) Create a 95% confidence interval for the population proportion of guests who caught northern pike over 9.0 kg.
b) Do you think the lodge's claim is correct? Explain.

(Continued on the next two-page spread ...)

Statistics

Outcomes

SCO: By the end of Mathematics 2205, students will be expected to

F11 determine, interpret, and apply confidence intervals

F1 draw inferences about a population based on a sample

F20 demonstrate an understanding of how a confidence level affects a confidence interval

F2 identify bias in data collection, interpretation, and presentation

F7 draw inferences from graphs, tables, and reports

Elaboration – Instructional Strategies/Suggestions

(... Continued from the previous two-page spread.)

The following material applies only to students in the advanced course.

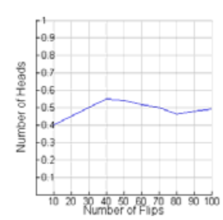
Students could be presented with examples similar to the following: A car rental company randomly selects 250 customers from Newfoundland and Labrador and asks them if they were satisfied with their last rental; 195 people claimed that they were satisfied. Students could be asked to calculate the sample proportion ($\hat{p} = 0.78$) and the 90%, 95%, and 99% confidence intervals; e.g., the 90% confidence interval is

$$\begin{aligned} &\hat{p} \pm 1.645\sqrt{\frac{\hat{p}\left(1-\hat{p}\right)}{n}} \\ =\ & 0.78 \pm 1.645\sqrt{\frac{0.78\left(1-0.78\right)}{250}} \\ =\ & 0.78 \pm 0.043 \\ & \text{from } 0.737 \text{ to } 0.823 \end{aligned}$$

Students should claim that they believe that the actual percentage of customers in Newfoundland and Labrador who were satisfied with their last car rental is between 73.7% and 82.3%. They should also claim that they are 90% confident of this. When calculating the 95% and 99% confidence intervals based on the same sample, students should recognize that a higher confidence level results in a larger interval. Students should also be aware of the limitations of such a survey. The car rental company surveys only randomly selected customers from Newfoundland and Labrador. Since they did not include customers from other provinces, they cannot apply this confidence interval to customers from other parts of Canada.

It is very important that students also consider the effect of sample size on the length of a confidence interval. While higher levels of confidence will produce longer confidence intervals when constructed from samples of the same size, using larger sample sizes will produce shorter confidence intervals. It is possible, for example, that a 90% confidence interval based on a sample of size 40 will be longer than a 95% confidence interval based on a sample of size 100 from the same population. Students could repeat the above example, but this time assuming a sample size of 500 with 390 claiming they were satisfied. They should compare the lengths of the 90%, 95%, and 99% confidence intervals with the corresponding intervals based on the smaller sample size.

If one were to flip a fair coin only once, then there is no chance that the sample proportion of heads would match the population proportion of $p = 0.5$. If it were flipped twice, while it is possible that one head would be obtained (giving $\hat{p} = 0.5 = p$), it is also very possible to obtain no heads ($\hat{p} = 0$) or two heads ($\hat{p} = 1$). However, as the number of repetitions is increased, it is generally the case that the sample proportion more consistently approximates the population proportion (at 10 flips, for example, one would usually get anywhere from 3 to 7 heads, meaning that $\hat{p}$ would range from 0.3 to 0.7, all of which provide better estimates of p than for two of the possible outcomes in the two-flip situation. For 50 flips, one would usually obtain somewhere between 20 and 30 heads, all outcomes that give values for $\hat{p}$ that are even more reasonable estimates of p). An interesting activity here might be to have students construct a broken line graph, similar to the one shown, based on an experiment with flipping coins. Students will see that the sample proportions more consistently approximate the population proportion as the number of flips is increased. This would also intuitively lead one to conclude that the larger sample sizes would usually allow for better estimates of the population proportion and therefore shorter confidence intervals than similar intervals based on smaller sample sizes.



Statistics

Worthwhile Tasks for Instruction and/or Assessment

(... Continued from the previous two-page spread.)

F11/F1/F20/F2/F7 *Activity*

F11/F1.4. A random sample of 267 Canadian doctors showed that 215 provided at least some charity care (i.e., treated people at no cost).

a) Let p represent the proportion of all Canadian doctors who provide some charity care. Find a point estimate for p.

b) Find a 99% confidence interval for p. Give a brief explanation of the meaning of your answer in the context of this problem.

c) Another random sample of 100 Canadian doctors showed that 79 of them provided at least some charity care. Find a 99% confidence for p and compare it to the one you found in (b). Which interval is of more use in drawing conclusions about the entire population of Canadian doctors and the proportion that provide at least some charity care? Explain, giving at least two reasons for your answer.

Suggested Resources

Statistics