Unit 4
Statistics
(20%)

Statistics

Outcomes

SCO: In this course, students will be expected to

F1 draw inferences about a population from a sample

F2 identify bias in data collection, interpretation, and presentation

Elaboration—Instructional Strategies/Suggestions

F1 Students might begin their study of statistics by reading and discussing some short articles that present results of statistical studies. They should discuss the different interpretations of the information that students might have, and any terminology that might cause confusion. Students should expect to see terms like survey, census, sample, population, and sample/population proportions or percentages.

F1/F2 A television survey might have been conducted to find the fraction of the public that watches a particular television program on a particular night and if it is watched by more women than men. A quality engineer must estimate what percentage of bottles rolling off an assembly line are defective. In both these situations, information is gathered about a large group of people or things. The expense of contacting every person or inspecting every bottle, not to mention the time it would take, is formidable. So information is gathered about only part of the group (a sample) in order to draw conclusions about the whole group (the population). Choosing a representative sample from a large and varied population can be a complex task. It is important to be clear about what population is to be described and exactly what is to be measured.

Areas of Concern

- How can a sample be chosen so that it is truly representative of the population?
- If a sample from a population differs from another sample from the same population, how confident can you be about predicting the true population percentage?
- Does the size of the sample make a difference?

F2 The process of sampling people at a mall is fast, cheap and convenient. But convenience samples often produce unrepresentative data. Think about people in the mall who might be asked. Often it is those who are well dressed, respectable looking, and friendly types, because they look easier to approach. Often the sample from malls will over-represent the middle class and retired, and under-represent the poor. When an error occurs due to bad sampling, the difference between the results obtained and the truth about the whole population is called bias.

Students should take every opportunity to discuss bias with respect not only to data collection, but its interpretation and presentation. These issues will be revisited on subsequent pages.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

Performance

F1

1) Some people say that statistics lie. State how confident you are on a scale of 1–5, 5 being very confident, with the sample results of the following situations:
   a) In March, the Gallup survey asked 1500 Nova Scotian adults this question: "Do you approve or disapprove of the way the Nova Scotia premier is handling his job?"
   b) Based on the survey, Gallup said that 56% of Nova Scotians approved of the way their premier is handling his job.
   c) After the same survey another newspaper reported that Gallup said 53% to 60% of Nova Scotians approved. They said they were 95% confident in their results.
   d) Shiniest toothpaste claims it's the best. Based on a survey of 10 dentists, 6 said Shiniest was tops.
   e) Based on the results of a survey of 100 music students at Swinging High, the school board will maintain the music program.
   f) Most Atlantic Canadians are opposed to nuclear power. This was the conclusion based on a survey conducted by phone in Lepreau, New Brunswick, site of a nuclear power plant.
   g) In a high school of 500 students, 10 students were surveyed and 6 said they thought the school year should be lengthened.

F2

2) Discuss how bias may affect your confidence from each of the above situations.

Journal

F1

3) a) Describe the difference between a survey and a census.
   b) Describe the difference between a sample proportion and a population proportion.
4) List 5 ways that bias might affect the outcomes of a survey conducted in your town.
5) You flip a coin 100 times and 62 heads turn up. Tell me what you think about my coin. How confident are you? Explain.

Suggested Resources

Garfunkel, Salomon, Consortium for Mathematics and its Applications (COMAP), *For All Practical Purposes*, W. H. Freeman and Company, NY, 1995

Landwehr, James M. et. al. *Exploring Surveys and Information from Samples, Quantitative Literacy Series*, Dale Seymour Publications, Palo Alto, CA, 1987

Statistics

Outcomes

SCO: In this course, students will be expected to

F1 draw inferences about a population from a sample

F15 design and conduct experiments/ surveys to explore sampling variability

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

Elaboration—Instructional Strategies/Suggestions

F1/F35 Students should understand that sampling is conducted to get information about a population. If chosen at random, the sampling is expected to represent the population. Your class may decide to ask a sample of your school population whether they watch a particular television program on a particular night. Gallup may decide to do the same survey by sampling the population of Atlantic Canada. The results of these samples from different populations are likely to vary. In fact, if repeated samples are taken from the same population, the results will vary from sample to sample. Can results from samples be trusted? How is trust affected by sample size? To explore this, students need to look more closely at sampling variability and consider how confident or certain they feel about the results. They can model this by taking samples from containers, flipping coins, rolling dice, flipping cards, using random number generators, and simulating by using technology.

Students should be able to make the link between proportions and probability. If 5 of 8 cubes in the bag are red, students should be able to infer that the probability of taking a red cube from the bag is 5/8 or 0.625, or that it will happen about 63% of the time.

Tossing coins is an easy way to start the exploration of sampling variability. Students would expect that since there are only two sides to a coin, heads would likely turn up one half of the time. Have students toss 10 coins and record the number of heads. Have them repeat this process many times. One thing that they should notice is how few times they will actually get 5 heads. Another would be that, after many trials, the results will cluster around the expected proportion (5 out of 10)(there will be more results of 4, 5, 6 than 1, 2, or 9, 10).

F15/F4 Increase the sample size to 40 (toss 40 coins many times) and revisit the proportions.

Students should notice that

- The probability of obtaining an exact particular result is greater for smaller samples. That is, you are more likely to get 5 heads when 10 coins are flipped than 50 heads when 100 coins are flipped.
- The probability of obtaining an approximate particular expected result is greater for larger samples. That is, you are more likely to get about 50 heads (45–55) in 100 tosses than about 5 heads (4, 5, or 6) in 10 tosses.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

Activity

F1/F15

1) a) It is assumed that the population proportion is 50% female. Sample the population. Select 4 persons at random and observe whether there are 0, 1, 2, 3, or 4 females. To simulate this

i) use a random device—4 coins, 4 spinners, 4 two-colour counters, or a random number generator...

ii) if coins are selected, toss 4 coins and count how many are heads (female) or use 4 spinners that have 2 colours, each representing the area

iii) calculate the sample proportion. Will this same sample proportion be obtained each time?

iv) repeat the experiment 40 times and calculate the 40 sample proportions

v) create a table for the data like the one shown

heads	sample proportion	tally	frequency	proportion of all trials
0	$\frac{0}{4} - 0.00$	|||	3	$\frac{3}{40} - 0.075$
1	$\frac{1}{4} - 0.25$	𝍸 |||	8	$\frac{8}{40} - 0.200$
2	$\frac{2}{4} - 0.50$	𝍸 𝍸 𝍸 |	16	$\frac{16}{40} - 0.400$
3	$\frac{3}{4} - 0.75$	𝍸 𝍸	10	$\frac{10}{40} - 0.250$
4	$\frac{4}{4} - 1.00$	|||	3	$\frac{3}{40} - 0.250$

Are the outcomes equally likely?

Make a chart or tree diagram to show how each toss of 4 coins, or spin on 4 spinners, results in a possible 16 outcomes (not 4) —FFFF, FFFM, FFMF, FMFF, ... to develop an understanding of equally likely

How many outcomes result in 0 females? 1 female? 2 females? 3 females? 4 females?

How accurate was the simulation of 40 trials?

b) Do the same thing now with 8 coins or spinners to model 8 persons being asked. First, spin or toss 10 times and complete a chart, then spin or toss 100 times and complete a chart. Then students should answer questions such as the following:

i) Are you more likely to get a sample proportion of exactly 0.5 if you toss 4 or 8 coins? [Answer: 4 coins]

ii) Are you more likely to get a sample proportion of between 0.25 and 0.75 if you toss 4 or 8 coins? [Answer: 8 coins]

iii) Are you more likely to get exactly 10 heads from tossing 20 coins or exactly 50 heads from tossing 100 coins?
[Answer: 10 from 20]

iv) Are you more likely to get a sample proportion of exactly 0.25 and 0.75 from tossing 20 or 100 coins? [Answer: 100]

Suggested Resources

Garfunkel, Salomon, Consortium for Mathematics and its Applications (COMAP), *For All Practical Purposes*, W. H. Freeman and Company, NY, 1995

Landwehr, Swift, Watkins, *Exploring Surveys and Information from Samples*, *Quantitative Literacy Series*, Dale Seymour Publications, 1987

Fathom, Dynamic Statistics Software, Key Curriculum Press, Palo alto, CA

Statistics

Outcomes

SCO: In this course, students will be expected to

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

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

Elaboration—Instructional Strategies/Suggestions

F4/A3 Students should approximate sampling distributions from populations for which the percentage of yeses is known, and study the variability that occurs through simulation. If, for example, students knew that 40% of the population of the town of Sackville answered yes when asked, "Would you agree to a 1% increase in town taxes to support the wild fowl park?" then they could take samples from the population to see how close their sample proportion of yeses would be to the 40%.

They might begin by taking small sample sizes of 5 or 10 and then trying some with large sample sizes of 40 or 80 to see the effect this would have on the distribution of the results. They should notice that as the sample size increases, the data would cluster more around the mean, decreasing the variation. If available, students could use technology to generate samples from known populations using various sample sizes and create histograms for each and note how the shape of the histograms change as the sample size increases. (For example, on the TI-83, they could use "randBin" found in the MATH PRB menu. They should use randBin because they are sampling from a population who can only respond "yes" or "no" (binary—only two choices.)

- randBin (10, 0.4, 100) would simulate ten (the 10) spinners with 40% of the area marked "yes" (0.4) being spun 100 times (the 100). This would simulate the situation above, i.e., asking 10 people from the known population that is 40% in favour of increasing taxes, and repeating it 100 times.

As the sample size increases from 10 to 20, to 40, the distribution should cluster more and more closely to the mean. This would become clear on the histograms of the distribution, as they would grow taller and skinnier as the sample sizes increased.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

Activity

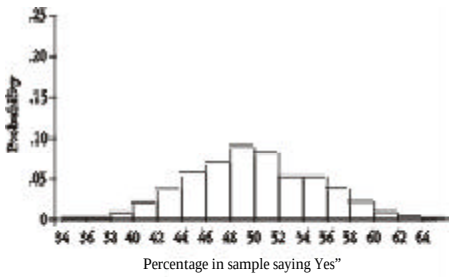
F4

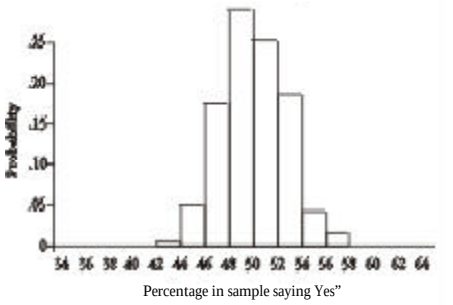
1) On a particular day in October, a census was taken of the entire grade 11 class at Your School. The results indicated that 62% of the 240 students answered "yes" to this question: "Should all grade 11 students take Physical Education 3 hours/week? Yes or No?"
   The school staff couldn't believe the results. They decided to sample the population, asking the same question.
   a) Describe a process they might use to simulate a sample size 5, with 50 repetitions.
   b) Conduct the simulation and display the results on a histogram.
   c) Repeat the process described above, but for sample size 10, with 50 repetitions, then sample size 50 with 50 repetitions.
   d) Compare the three histograms to explain how they are different.
   e) Describe how your confidence about the initial census results was strengthened or weakened by the results of your sampling.

Performance

F4

2) Displayed are two histograms showing the probability that a student would answer "yes" to a question. The first graph represents a sample size 100 and 1000 trials, the second, 1500 and 1000 trials.
   Describe the differences and explain why the differences are occurring.





Suggested Resources

Garfunkel, Salomon, Consortium for Mathematics and its Applications (COMAP), *For All Practical Purposes*, W. H. Freeman and Company, NY, 1995

Landwehr, Swift, Watkins, *Exploring Surveys and Information from Samples, Quantitative Literacy Series*, Dale Seymour Publications, Palo Alto, CA, 1987

www.education.ti.com (to download the APP for the TI-83 Plus called ProbSim.) This is a program that simulates spinners, dice, cards, etc.

Statistics

Outcomes

SCO: In this course, students will be expected to

F1 draw inferences about a population from a sample

F2 identify bias in data collection, interpretation, and presentation

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

Elaboration—Instructional Strategies/Suggestions

F1/F2 Students should become familiar with different kinds of sampling methods. The primary objective is for students to understand the differences in the sampling methods rather than memorizing the names by rote. Students should try to connect the names of the methods to the process used, in this way making sense of the name. The methods include

- simple random samples (This gives each member of a population an equal chance of being selected)
- cluster samples
- systematic samples
- convenience samples
- self-selected samples

Students should understand that randomizing procedures are required to reduce bias in data collection. The bias that is most easily identified results from people's own misgivings about including all groups of people in the sampling process. Other sources of bias in data collection include poorly worded questions, leading questions, untruthful responses, and unresponsive participants.

A3 An experiment or survey must be replicable, that is, the results should be more or less the same after many repetitions. Much of this depends on how the sample data is selected. Students who need an unbiased selection method should use a random number table or generator. Computers and calculators should be used for this task. For example, scientific calculators have built-in programs that generate random numbers between zero and one, such as the list that may appear: .531, .574, .072, .924, .490, .751, .436, .099. Each 3-digit decimal number between 0 and 1 has an equally likely chance of occurring. Random numbers can cluster, they just don't follow predictable or recurring patterns. If a student wants to interview 25 students from the grade 11 population in your school, she might multiply each of these random numbers by the total number of grade 11 students (212) in her school to find

$$.531 \times 212 = 112\text{th person on the list}$$
$$.531 \times 212 = 122\text{nd person on the list}$$
$$.072 \times 212 = 15\text{th person on the list}$$

Students can use the menu MATH PRB on a graphing calculator to select various random number generators. For example, to simulate the tossing of a six-sided die 50 times, students could select 50 integers from 1 to 6 by using the command randInt (1, 6, 50). To simulate yes/no situations, students can randomly sample using randBin (10, .4, 50) ("Bin" for "binary," or "only two possible answers"). This command would produce data simular to that if a student selected 50 random samples each of size 10 from a population where 40% are known to answer "yes" to a particular question in a survey.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

Performance

F1/F2/E3

1) A group of students wishes to take a survey of the 750 students in their school. Each student in the school has their own identification number and each student in the school is registered in one English course. Identify the following methods as random sampling, convenience sampling, cluster samples, or systematic sampling.
   a) Select 100 identification numbers using the random number generator of the school computer and interview the students selected.
   b) One teacher has four English classes, with a total of 120 students in the four classes. Interview all the students in those classes.
   c) Place each of the students who are conducting the survey at a different entrance to the school and select 1 student per minute until each surveyor has interviewed 20 students.
   d) Select every 7th person on the list of students who attend the school.
   e) Interview 120 students who have study periods at the same times as the students doing the interview.
2) What does a random sample mean?
3) Review the following question from a statistics quiz. This question contains 132 letters. Count the e's. Based on your count, what is the percentage of e's used in the English language? How confident are you in your response (on a scale of 1 to 10 where 10 is very confident)? Do you think that this is a biased sample of letters? Give a reason.
4) The following item appeared in a Saint John newspaper:
   "...is conducting a survey to determine the percentage of homes that have installed smoke detecting devices. Questionnaires will be mailed to a sample of homes. If you would like to participate in this survey, please contact..."
   Do you think that the results obtained from this survey will be biased? Give reasons.

Suggested Resources

Garfunkel, Salomon, Consortium for Mathematics and its Applications (COMAP), *For All Practical Purposes*, W.H. Freeman & Co., NY, 1995

Addenda series: *Data Analyses and Statistics, grades 9-12*, Burrill, Gail et al., NCTM, Reston, VA, 1992

Landwehr, Swift, Watkins, *Exploring Surveys and Information from Samples*, Quantitative Literacy Series, Dale Seymour Publications, Palo Alto, 1987

Statistics

Outcomes

SCO: In this course, students will be expected to

F5 organize and display information in various ways with and without technology

F13 calculate, analyse and interpret various statistics

F7 draw inferences from graphs and tables

F2 identify bias in data collection, interpretation, and presentation

Elaboration—Instructional Strategies/Suggestions

F5 Students will construct histograms, box plots, and other displays of data. They need to make decisions about which is the most appropriate display for the argument they are establishing or for the distribution they need to study.

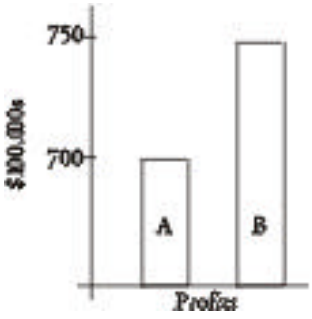
- A bar graph, circle graph, or a stem and leaf plot might be used to display test results in the grade 11 math class. Have students work in threes, each selecting one display method. Classmates should determine which graph displays the data in the most helpful or meaningful way.

Students should know that bar graphs are used to display discrete data, whereas histograms are intended for continuous data. Box plots are often used to compare variability in distributions. Students should note that the wider the box the more varied and widely distributed the data. Stem-and-leaf plots are like histograms, only they provide more detailed information.

F5/F13 Students should be encouraged to use appropriate statistical technology for calculating mean, median, and standard deviation (see page 92), and for creating displays (e.g., of histograms). Students will use histograms and/or frequency bar graphs to study distribution of data and to explore a distribution that is normal.

F7 Numerical facts are essential for making many kinds of decisions in our lives. Data is useful when we organize and present it so that it can be more easily interpreted. A few numbers computed from the data—averages, percents and the like—can be very helpful. Some numbers computed from the data lead to statistical inference—drawing conclusions and stating our confidence about the conclusions.

Students should investigate graphs, tables and displays of samples of data obtained in experiments and surveys. Students might come up with their own topics for a survey and survey the students in their class. They might then exchange displays of data, or post them on the classroom walls, from which whole class discussions can take place. They should be encouraged to explain what they know based on the information on the display. From the display, students should describe the population the sample represents and draw conclusions based on the sample.



F2 Students should analyse their displays for any bias that might be detected. Statisticians can alter the shape of a display in order to provide a picture that will help establish one particular side of an argument. Ask students how the graph to the right might mislead someone if they were working for company B and comparing their profits to those of company A.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

Performance

F5/F13/F7/F2

1) The Golf Problem:
The very first annual district golf tournament is presently underway, and you are the coach of your school team. The prize up for grabs is a TI-83 graphing calculator for each math student in the winning school, to be paid for by the Student Council of the losing teams' schools. Each school selects five golfers and two alternates for their team. There will be two 18 hole games of golf twice a month beginning in May, and ending in September, with each team entering 5 players to play. At the end of September the school team with the lowest accumulated score will be declared the winner, and the other schools will supply them with their prize.

Ronnie:

83	75	77	82	95	93	91	101	103	92
82	72	90	88	85	81	95	97	105	91
101	99	87	89	76	72	82	90	101	78
92	93	95	88	85	91	93	99	105	101

Bobbie:

68	89	101	67	107	110	98	89	72	108
71	102	88	92	95	87	69	111	96	68
78	82	67	86	92	90	103	96	107	75
91	99	105	101	65	87	86	92	91	101

Situation 1: In early August the accumulated scores show that any of the teams could go on and be the winner. But wait, your best golfer has developed a serious back problem and will not be able to play for the rest of this year. You must select an alternate to replace her. To assist you in your selection, the following tables represent the previous 20 golf scores recorded for each of your alternates while using the same golf course.
Analyse this information and select the alternate for the rest of the games this year. Present arguments and diagrams to support your decision.
Situation 2: There is only one game left in this year's tournament and one of your players cannot make it due to having to write exams at his school. Again, you have the two individuals above to choose from as a replacement. Here is the situation: one game left; your accumulated team score is 4 points higher than the leading school; you really want those graphing calculators! Present arguments and diagrams to support your decision.
(Note: this task may need to be done as a group activity. It contains a lot of reading which may make it challenging for some students.)

Suggested Resources

Garfunkel, Salomon, Consortium for Mathematics and its Applications (COMAP), *For All Practical Purposes*, W.H. Freeman & Co., NY, 1995

Addenda series: *Data Analyses and Statistics, grades 9–12*, Burrill, Gail et al., NCTM, reston, VA, 1992

Landwehr, Swift, Watkins, *Exploring Surveys and Information from Samples*, Quantitative Literacy Series, Dale Seymour Publications, Palo Alto, CA, 1987

Statistics

Outcomes

SCO: In this course, students will be expected to

F3 demonstrate an understanding of what can be inferred about a population by examining sample means and dispersions

F15 design and conduct experiments/ surveys to explore sampling variability

F5 organize and display information in various ways with and without technology

F13 calculate, analyse, and interpret various statistics

Elaboration—Instructional Strategies/Suggestions

F3 In studying the distribution of data (for more discussion on distribution, see page 92), students should understand that experimental results about a variable often differ from what results are expected. ("Expected" here simply refers to that result one would expect to see. For example, most people would expect to see about 50% of all coin tosses to turn up heads.) Students must decide how much deviation from the expected is reasonable before it can be concluded that something "strange" is happening. For example, if 60% of the coin flips turn up heads, would that be considered a "strange" result. How about 57%? Students will deepen their understanding of distributions and will use mean and standard deviation to describe the dispersion of data.

F15/F5 They might begin by examining data with a mean and median that are the same, but with displays that look quite different. By examining these data displayed in box plots, students will see a longer box and perhaps longer whiskers for one set of data than for another, even though their means and medians are the same. Lower and flatter histograms indicate that there is more variation in the data than a narrow high histogram where the data seem to cluster around the mean. This assumes, of course, a common scale for both histograms.

F3/F13 Students should discuss how dispersion of data can be measured. They should consider the range of values in the data. They might talk about the fact that a certain range of values might be expected, but that if the range was overly large some data might be considered quite suspicious. In discussing acceptable range, students might talk about a way to measure the dispersion of data. They may decide that dispersion can be described by some value arrived at by subtracting data values from the mean value. In other words, finding some average amount by which each value point is away from the mean value point.

Students should develop an understanding of how standard deviation is measured using manual calculations at first, then using technology. They should understand each step in the calculation and that higher values for deviation mean more dispersion. They should think of standard deviation as representing the average dispersion from the mean. With technology, students should be able to access values for single-variable statistics such as mean, median, range, standard deviation, and quartile values. They should be expected to use technology, if available, to find these single-variable statistics.

When introducing standard deviation, it might be helpful to give three sets of data and ask students to rate on a scale of 0 - 10, the amount of spread:

A: 7 7 7 7 7 B: 6 7 7 7 8 C: 0 4 7 10 14

This will help students get a feel for the meaning of standard deviation.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

Performance

F13/F15

1) The following data were gathered by a group of students performing an experiment to determine if amplitude has an effect on the period of a pendulum. The data (in seconds) represents the period of a pendulum, measured 20 times using a stop watch.

.58	.57	.51	.63	.61	.58	.59	.56	.55	.54
.54	.57	.59	.61	.53	.57	.56	.57	.58	.53

   a) Calculate the mean and standard deviation.
   b) Describe what you think is the period of the pendulum based on this experiment. Explain.
   c) Are you surprised by the different results that the students got within the same experiment? Explain.

F13/F3

2) This same experiment was conducted by 24 groups, each of which described the period as the mean of 20 trials. The following are the means.

.5624	.5705	.5680	.5581	.5713	.5615	.5496	.5564
.575	.5265	.5811	.5876	.5925	.5728	.5720	.5572
.5576	.5585	.5860	.5605	.5821	.5382	.6013	.5630

   a) Calculate the mean and standard deviation.
   b) How is this set of data different than the one in 1) above.

F5

3) a) Construct histograms for the experimental results in 1) and 2) above.
   b) Construct a box and whisker plot for each set of data.
   c) Describe how the box plots help you better understand the distributions.

Journal

F13

4) Explain why the differences from the mean are squared in the process of determining the standard deviation.
5) In the process of determining the standard deviation, each of the experimental results is subtracted from the mean. Why?
6) What does the standard deviation tell you?

Suggested Resources

Garfunkel, Salomon, Consortium for Mathematics and its Applications (COMAP), *For All Practical Purposes*, W.H. Freeman & Co., NY, 1995

Addenda: *Data Analyses and Statistics*, grades 9-12, Burrill, Gail et al., Reston, VA, NCTM, 1992

Statistics

Outcomes

SCO: In this course, students will be expected to

F1 draw inferences about a population from a sample

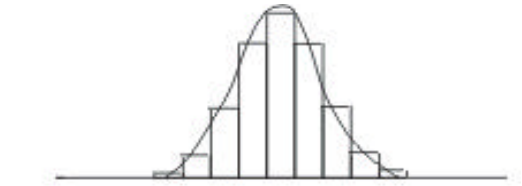
F3 demonstrate an understanding of what can be inferred about a population by examining sample means and dispersions

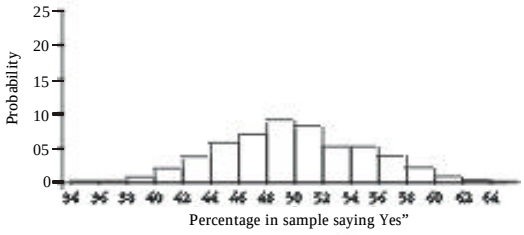
F4 demonstrate an understanding of how the size of a sample affects the variation in sample results

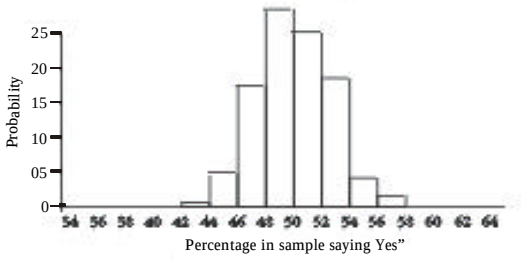
Elaboration—Instructional Strategies/Suggestions

F1/F3 In a previous course, students conducted multiple trials of certain experiments. They found that the measurement taken for each trial was not always the same. They explored this and came to an agreement that it is normal for this to happen. However, they wondered if some of the differences were normal or if they were caused by some error they were making. For example, when they attempted to measure the period of a pendulum (one complete swing), one time they might clock it at 1.80 seconds, the next time at 1.70 seconds, and the next at 1.85 seconds. In this course, students will formalize this notion that different measurements would be expected when performing the same experiment over and over again. What they need to formalize is how much deviation from the expected is reasonable before it can be concluded that the recorded measurement may have been affected by some other factor.

F1/F3/F4 When results vary, students need to question if the variation has occurred by chance or not. To do this, they should conduct many experiments of the same phenomena, pool all the results, and display them in histograms. For example, if a teacher had three or four math classes conducting these experiments and each class was made up of six to eight groups each collecting their own data, the teacher could have each group describe their data by finding the mean. The teacher could then collect all the means and form a histogram. Students would see that as the number of samples get larger and larger, the histogram would take on a shape that becomes more and more like a normal distribution; that is, as the number of trials increases, the distribution of sample means would seem to cluster more and more around the mean of all the data. It takes on a higher, narrower, more symmetrical appearance rather than a low and flat shape. Students should understand that a similar change in shape of the histogram is connected to sample size and that the sample size affects dispersion. As the sample size increases, the deviation of the data from the mean decreases.







Statistics

Worthwhile Tasks for Instruction and/or Assessment

Performance

F1

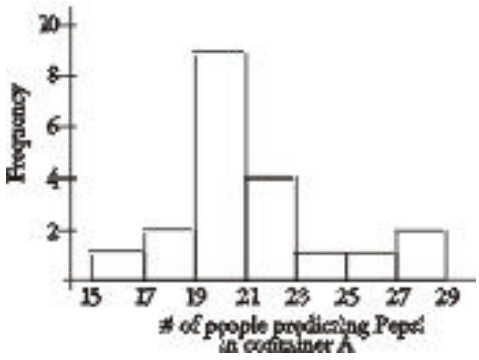
1) Barney and Martha set up a booth at their school's spring fair. They charged 25¢ for an opportunity to predict which cola was being tasted. Each participant would pay their 25¢ then taste two cups of liquid. They would then predict whether Container A or Container B was Pepsi. At the end of the day, Barney and Martha calculated their sample statistics. They had asked 125 people which container contained Pepsi. 42% of the people said container A was the Pepsi, with a standard deviation of 6.2%. Describe what you might hypothesize about the 1000 students at this school based on this sample.

F3

2) William wasn't satisfied with his hypothesis from question 1) above. He decided to conduct his own survey. He convinced 10 friends to help him. He asked each to conduct a similar survey at different locations around the school at noon time. Each was to collect 25 results, find the mean, and repeat this one more time. William then examined a distribution of the 20 sample means. Describe what you think he would see. How might it be different than what he saw in 1). Explain.

F4

3) The histogram displayed shows the results of a survey similar to that in question 2), only this time the sample size was 50, not 25. There are 20 sample means, each representing a sample size 50. Explain how increasing the sample size to 50 from 25 has affected the distribution.



Suggested Resources

Garfunkel, Salomon, Consortium for Mathematics and its Applications (COMAP), For All Practical Purposes, W.H. Freeman & Co., NY, 1995

Addenda: *Data Analyses and Statistics, grades 9–12*, Burrill, Gail et al., Reston, VA, NCTM, 1992

Statistics

Outcomes

SCO: In this course, students will be expected to

G3 graph sample distributions and interpret them using the language of probability

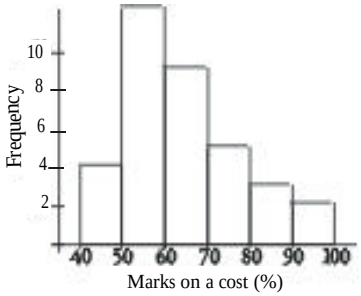
F13 calculate, analyse and interpret various statistics

F5 organize and display information in various ways with and without technology

F7 draw inferences from graphs and tables

Elaboration—Instructional Strategies/Suggestions

G3 Students will begin to interpret the frequencies of events occurring within a population using probability. The concept being developed here is that the relative area under any portion of a histogram corresponds with the probability of randomly choosing a member of the population with that characteristic.



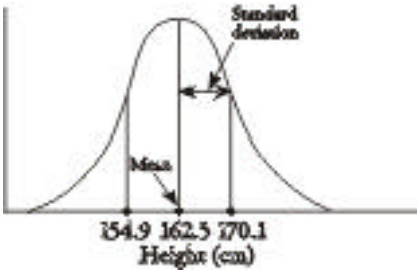
F13/F5/F7

- The probability of scoring between 40 and 50 on the test is determined by comparing in a ratio the area in the bar from 40-50, with the total area of all the bars. Class size is 35.

$$P(40-50)=\frac{4}{35} \text{ or } 0.114, \text{ or } 11\%$$

The probability that a particular student passed the test with a mark of 50% or better is determined using the ratio $\frac{31}{35}$ or 0.89, or 89%.

F7 Given a histogram of a distribution that is fairly symmetrical and approaches the shape of a normal curve, students could be asked what portion of the data is within one standard deviation of the mean. Using calculations like that above, students could calculate that the area under the curve within one standard deviation from the mean represents 68% of the data. Likewise, the 95% region represents the data within two standard deviation from the mean and 99.7% of the data lies within three standard deviations from the mean.



- Given data measuring heights of people with a mean height of 162.5 cm and standard deviation of 7.6 cm. A histogram could be constructed that approached the shape of a normal curve. Students could be asked what the probability is of having a height within one standard deviation of the mean. They would then approximate the area within the region below the curve between 154.9 and 170.1 cm, and express it as a ratio of the total area and estimate that to be 0.68 or 68%.

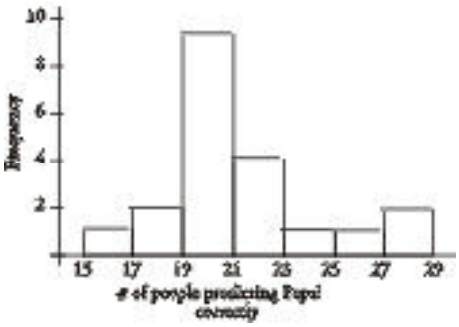
Statistics

Worthwhile Tasks for Instruction and/or Assessment

Performance

G3/F13/F5/F7

1) This frequency histogram represents the number of people who identified Pepsi in container A. It is based on 20 sample means from samples of size 50.



   a) Change this histogram to a probability bar graph representing the probability that a person would predict Pepsi in container A.
   b) From the probability bar graph what is the probability that from 19 to 21 people in every 50 would predict Pepsi in container A?
   c) What is the probability that 30 of 50 would predict Pepsi in container A? Explain.

F5/F13/G3/F7

2) Twelve of the 30 people in your math class are males. You have 5 tickets to a rock concert to give away to your classmates. Working as a team with other students, you are to draw 100 simple random samples of size five from this population (your class of 30). (Recommend that students use technology). Note: On the TI-83, use rand Int (1, 30, 5), where getting a number from 1 to 12 represents obtaining a male. Sample output:

rand Int (1, 30, 5)
{1, 30, 2, 16, 20}

Sample proportion is $\frac{2}{5} = 0.4$

   a) Complete the sample proportion of males in each sample.
   b) Make a probability bar graph of the situation.
   c) Describe the shape. In particular, does it look roughly normal?
   d) What is the average number of males in your 100 samples?
   e) You now give the five tickets to five students in the class. They were all females. Do you think there was any discrimination in giving the five tickets to the five females? Explain.

Suggested Resources

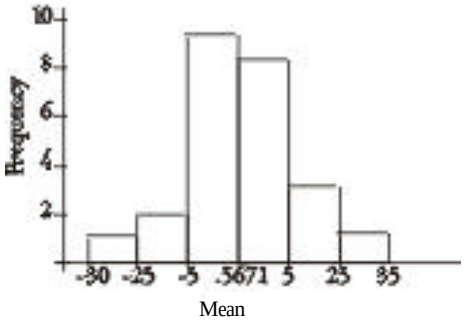
Statistics

Outcomes

SCO: In this course, students will be expected to

G3 graph sample distributions and interpret them using the language of probability

Elaboration—Instructional Strategies/Suggestions



G3 On the previous page it was discussed how students would calculate probabilities based on the amount of area under the curve versus the total area.

Students should formalize this with an activity where they would calculate the areas between first, second, and third standard deviations. They could do this by superimposing a grid over a normal curve and estimating the area within the stated regions. They should understand that in a normal situation 95% of all the data should fall within two standard deviations of the mean. (Students should remember that a standard deviation represents an average dispersion from the mean.) Students should now use this information about mean and standard deviations in normal curves to describe samples and populations. For example, from a census taken as fans entered the ball park at a baseball game, it was known that 80% of the fans said they wanted the opportunity to buy peanuts while watching the game. The standard deviation was 3%. Students should be able to say that if random samples of fans were asked their opinion about buying peanuts they would expect that the sample proportions of those in favour would be in the range of 74% to 86%, 95% of the time.

In studying the distribution of data, students should understand that experimental results about a variable often differ from expected results. They decide how much deviation from the expected is reasonable before it can be concluded that some other factor is affecting the outcomes.

When 95% of the sample measurements are within two standard deviations of the mean, then there is only a 5% chance that a randomly chosen measurement will differ from the mean by more than this, or only a 5% chance that the difference is accidental. When a measurement falls within two standard deviations of the mean, it is not considered significantly different from the mean.

Students should calculate means and standard deviations from various contexts to determine whether values differ sufficiently from the mean to be considered "significantly different."

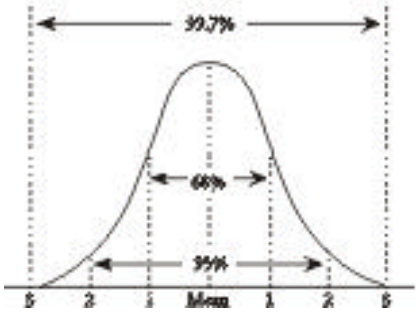
Statistics

Worthwhile Tasks for Instruction and/or Assessment

Activity

G3

1) On page 91, students were given problem 2) to address outcomes F13 and F3. Students calculated the mean and standard deviation for a distribution of 24 sample means. Each mean was the result of 20 trials. A histogram displays the distribution.



   a) Estimate the area on the histogram within (i) the first standard deviation; (ii) 2 standard deviations, (iii) 3 standard deviations.
   b) How close to normal would you say this distribution is? Explain.
   c) If a pendulum's period was measured (as below) how confident (scale 0–5, 5 being very confident) would you be of each coming from this distribution? Explain.
      i) .51
      ii) .53
      iii) .56
      iv) .59
      v) .60
      vi) .62

2) The scores on a Canadian math contest were normally distributed with a mean of 105 and a standard deviation of 25.
   a) 68% of the students had scores between what two values?
   b) 95% of the students had scores between what values?
   c) A student scored 130. What percentage of students scored better than that?
   d) How likely is it that a student scored 50 points? Explain.

Journal

3) The following is a letter to the editor written by Bruce Wark, School of Journalism, University of King's College, taken from March 1997, edition of the Halifax *Mail Star*. Re: Cameron MacKeen's March 12 report "Tories top Liberals, poll says." The article said the poll gave the Hamm tories the support of 32% of decided voters, while the Savage Liberals lagged with only 25%...A poll with a sample size of 400 has a standard deviation of $\pm 5\%$. That means, according to this poll, Tory support ranged from a high of 37% to a low of 27%. Liberal support ranged from 30% to 20%. It would be just as valid therefore, to report that the poll showed the Liberals ahead 30–27."
   a) Write a reaction to this letter. Do you agree with Mr. Wark's logic?
   b) Explain in your own words the logic in Mr. Wark's words in the last sentence.
   c) How else could the MacKeen report have announced the results for support?

Suggested Resources

Statistics

Outcomes

SCO: In this course, students will be expected to

F12 interpret normal curves and standard deviation to express levels of confidence

F17 design and conduct experiments/surveys, and interpret and communicate level of confidence

Elaboration—Instructional Strategies/Suggestions

F12 When students conduct surveys and polls and collect data from experiments, they need to be able to interpret these results and to use them to predict, with confidence, behaviours of the larger population. By studying sampling distribution, students should come to understand that the distribution of any random phenomena tends to be normal if it is averaged over a large number of independent repetitions. Students must be able to decide how much deviation from the expected is reasonable before it can be concluded that another variable is affecting the outcome. In other words, how far out on the tail of the normal curve will a result have to be in order to be considered different or unlikely to happen. Since a given measurement has a 68% chance of being within one standard deviation of the mean and a 95% chance of being within two standard deviation, any value further away from the mean (beyond two standard deviations) may result in being a measurement that makes a difference to the outcome.

The following is an example of using normal curve and standard deviations to express results with some degree of confidence.

F17

- Sammy White owns a “fish for trout” business on a small lake in a river system in southern Nova Scotia. In mid-April, he advertises that if you fish in his lake, you will catch trout between 21 cm and 30 cm. In fact, he guarantees that your catch will be in that size range or your money will be returned. During the month of April, Sammy closes his lake to fishing in order to conduct his survey of the trout population. Sammy has special trout traps set up in various parts of the lake. He spends most of his early mornings and late evenings canoeing from trap to trap to count and measure his trapped trout. He calculates the mean of the distribution to be 25.8 cm with a standard deviation of 2.1 cm. From these results (and a little bit of rounding) he feels he can be 95% confident that any fish taken from his lake during the next few weeks will measure between 21 cm and 30 cm. How can we use this information to support his claim?

Statistics

Worthwhile Tasks for Instruction and/or Assessment

Performance

F12

1) At a local cub rally, one event in which all 100 cubs participated was the racing car event. Each "cub" was given a block of wood, an axle, and four wheels. Each cub was expected to design a car for racing. The race consisted of releasing the cars on a sloped track and timing each car as it passed a particular spot. At the end of the day, the distribution of 100 times tended to be normal with a mean of 4.6 seconds and a standard deviation of 0.6.
   a) How many of the 100 times were within one standard deviation of the mean?
   b) Joe's time was 5.9 seconds. Is this unusually fast? Explain.
   c) If it took 3.5 seconds for a car to pass the spot, do you think that car was likely to win the event?
   d) If a car took 5.1 seconds, how would you describe its chances of winning? Explain.
   e) Cars with times 5.2 seconds or better made it to the final round. Estimate how many cars made it. Justify your estimate.
2) Suppose the following statement was made: 65% of high school students like music videos. Later, a random sample of 35 students finds that 18 like music videos. Is it fair to say that the statement is wrong? Why or why not?

F17

3) A committee at Yore School wants to provide a late bus so that students can stay at the school longer for academic and recreational activities. They want you to design and conduct a survey that would help them with their arguments when they present this idea to the school board. You are to design and conduct a yes/no survey and present the results, including the predicted population mean, standard deviation and your level of confidence, to the committee.

Suggested Resources

Statistics

Outcomes

SCO: In this course, students will be expected to

F17 design and conduct experiments/surveys, and interpret and communicate level of confidence

Elaboration—Instructional Strategies/Suggestions

F17 Students should use their understandings of relationships between samples and populations to conduct a survey and interpret the results of it. They should be expected to demonstrate their understanding by presenting, both orally and in writing, the methods they used, their displays, and their findings.

Students should design and conduct their own survey or poll using a large sample size. When students collect and analyse their own data, their appreciation of the principles and practice of survey techniques grows, and they are able to pull together what they have learned in a practical way.

One suggested approach might be to split the class into groups of four. Each group plans a questionnaire, including the specific wording of the questions and instructions. Each student should choose one yes/no question, but the group as a whole should determine the wording of the question. The group should then field test the questions by asking them to other students to make sure other students have no trouble answering yes or no. The group should then make any necessary revisions to the questions and decide on a sampling method. The group should collect data on all questions, but each student would prepare a report on his or her own question only.

Several evaluation schemes for survey projects are possible, but whatever the choice, make sure it is described to the students before they begin the project. One method is to give students a score from 0 to 4 in each of the following categories:

1. originality and independence
2. questionnaire design
3. sampling design
4. collection of data
5. report/Presentation

- A group of students agreed to survey people randomly selected at Hector's Quay in Pictou (near the ferry) on a Saturday. It was the responsibility of each of the four people in the group to ask 10 people the question at three different time periods in the day. This was the question: "Should Northumberland Ferries Ltd. discontinue its ferry service from Nova Scotia to Prince Edward Island?" After pooling their results, the group were 95% confident that that between 60% and 76% of the people in Pictou were against Northumberland Ferries Ltd. discontinuing their ferry service. Students should understand what this group did to be able to state their conclusion and whether, in fact, they were correct.
  Students created a histogram with their pooled results that approximated a normal curve with a mean of 68% and a standard deviation of 4.1. They could be confident that if they repeated this survey 20 times, 19 of the sample means would be within two standard deviations of 68% saying "no." In other words, roughly 60% to 76% of the people would say "no," 19 times out of 20.

Statistics

Worthwhile Tasks for Instruction and/or Assessment

Performance

F17

1) A high school wants to survey adults in the community to find out what percentage would attend the "spring-fling" event that the student council has in mind.
   a) Design and conduct a survey from which you can advise the student council with 95% confidence.
   b) Describe in detail the design and how you conducted the survey.
   c) Give your conclusion to the student council as a statement of what you found using 95% confidence.
2) On page 100, a survey is discussed that deals with the ferry service between Nova Scotia and Prince Edward Island. Do you think the results from that survey reflect the feelings of all the people in Nova Scotia? Atlantic Canada? How might you design and conduct the survey so as to get a less biassed result?

Suggested Resources