Unit 3: Day 15: Assessing Validity

Minds On: 5 Action: 50 Consolidate:20 Total=75 min	**Learning Goal**: • Interpret and assess statistics presented in the media. • Investigate interpretation by the media based on lack of knowledge of statistics. • Examine data collection techniques and analysis in the media.	**Materials** • BLM 3.15.1 to 3.15.5 • Acetates • Overhead projector

				Assessment Opportunities
	Minds On…	**Whole Class → Discussion** "There are lies, damned lies and statistics." --Mark Twain Discuss with students how statistics is generally accepted at face value. How skilled in numeracy and statistics are most journalists when they reference statistics in their articles? Is statistics part of the required training for journalists? Should it be? Why or why not? **Individual → Activity** Using BLM 3.15.1 students complete a checklist assessing their prior knowledge of using statistics. **Whole Class → Discussion** Using BLM 3.15.1 (Teacher Notes) review of the correct usage of statistics.		Stats Canada Website Diabetes Research Data Set : CCHS Cycle 3.1 Emphasize to the students that valid data for statistical analysis must come from some type of research. The data collection methodology must be unbiased.
	Action!	**Whole Class → Discussion** Teacher led discussion of how data and statistics are presented in the media **Students in Pairs → Activity** Description of activity Using BLM 3.15.5 students read the article and then complete the Statistics Use Checklist BLM 3.15.2 identifying any inappropriate use of statistics or misrepresentation of data. **Curriculum Expectations/Observation/Mental Note** Observe groups as they critique the article and complete the checklist.		
	Consolidate Debrief	**Whole Class → Discussion** Students share their findings using a Round Table discussion. Teachers may refer to BLM 3.15.5 (Teacher Notes) as students summarize the inappropriate use of statistics. Teacher summarizes the representation and misrepresentation of statistics in the media in general.		
	Exploration Application	**Home Activity or Further Classroom Consolidation** Using BLM 3.15.3, prepare a news article based on the diabetes data set made available or any Stats Canada data set. This article will then critiqued by another student using BLM 3.15.2 Checklist. **Process Expectations/Observations/Rubric:** Assess the students on the Getting It Write! activity using BLM 3.15.4 Rubric		

3.15.1: Your Best Guess!

TEST YOUR KNOWLEDGE OF USING STATISTICS

No.	Question	True	False
1	Numbers can stand alone without further explanation.		
2	Always check and verify the source of any statistics quoted.		
3	Statistics do not usually require interpretation.		
4	Always check for systemic bias in any survey results.		
5	Always check if data has been presented in the proper context.		
6	Accurate data generally provides valid evidence for a particular interpretation.		
7	The presenter's interpretation of statistics is generally correct.		
8	Confusing correlation with causation is a common error.		
9	True cause-and-effect relationships are easy to spot.		
10	Define your terms explicitly. (i.e. What is "average"?)		
11	Whenever possible, display the full range of data.		
12	Delete outliers. The user is not interested in seeing them.		
13	Knowing the margin of error is of no value to the user.		
14	The user is only interested in the graph, not the actual numbers.		
15	Whenever possible, include the mathematics and calculations behind the statistics – at least as an appendix.		
16	Through technology, the overall knowledge and understanding of statistics has increased.		
17	Pure data provides important, and useful, information and insight into relationships.		
18	Scatter plots are important sources of categorical data.		
19	Pie charts are one of the best means of presenting visual information in a form directly accessible to the user.		
20	The Law of Large Numbers is also known as The Law of Averages.		

3.15.1: Your Best Guess! (Teacher Notes)

TEST YOUR KNOWLEDGE OF USING STATISTICS

No.	Question	True	False
1	Numbers can stand alone without further explanation.		x
2	Always check and verify the source of any statistics quoted.	x	
3	Statistics do not usually require interpretation.		x
4	Always check for systemic bias in any survey results.	x	
5	Always check if data has been presented in the proper context.	x	
6	Accurate data generally provides valid evidence for a particular interpretation.		x
7	The presenter's interpretation of statistics is generally correct.		x
8	Confusing correlation with causation is a common error.	x	
9	True cause-and-effect relationships are easy to spot.		x
10	Define your terms explicitly. (i.e. What is "average"?)	x	
11	Whenever possible, display the full range of data.	x	
12	Delete outliers. The user is not interested in seeing them.		x
13	Knowing the margin of error is of no value to the user.		x
14	The user is only interested in the graph, not the actual numbers.		x
15	Whenever possible, include the mathematics and calculations behind the statistics – at least as an appendix.	x	
16	Through technology, the overall knowledge and understanding of statistics has increased.	x	x
17	Pure data provides important, and useful, information and insight into relationships.		x
18	Scatter plots are important sources of categorical data.		x
19	Pie charts are one of the best means of presenting visual information in a form directly accessible to the user.		x
20	The Law of Large Numbers is also known as The Law of Averages.	x	

3.15.2: Statistics Use Checklist

Reviewer:_____Date:_____

Item Reviewed:_____

Item	Yes	No	N/A	Comment
DATA				
Has all relevant data been included?				
Is the data interpreted correctly?				
Has the original data source been referenced?				
Has the data been presented in context?				
SURVEY RESULTS				
Have the sponsor(s) of the survey been identified?				
Has the population and sample size been included?				
Has any possible sampling bias been stated?				
Is the sample representative of the population?				
Does the survey results support the conclusions reached?				
REPRESENTATION OF DATA				
Have the data ranges been truncated or reduced?				
Are the axes scaled proportionally?				
Has the start/stop points of the time series been adjusted to support the point of the article?				
Have outliers been removed?				
ANALYSIS				
Has correlation been confused with causation?				
Has accidental association been confused with causation?				
Has the margin of error been included?				
MATHEMATICAL CALCULATIONS				
Are there any obvious mathematical errors?				
Has percentage been confused with percentage points?				
Has the base value been established for percentage references?				

3.15.3: Getting It Write!

Using the Diabetes Data Set supplied by your teacher or any data set available through Stats Canada, write a newspaper article referencing a variety of statistics and numerical data. Your article should use the data, and your analysis of the statistics, to support a conclusion or hypothesis that you have made.

HINTS:

1. Use BLM 3.15.2 as a guide as to the appropriate use of statistics in your article.

2. The following website: http://www.notrain-nogain.org/Train/Res/Num/NUM.asp provides a variety of hints and guidelines as to how numeracy and statistics should be referenced in a newspaper article.

Once you have completed your article, give it to one of the other students in the class who will complete BLM 3.15.3 Statistics Use Checklist to provide feedback to you on your article.

Hand in both your article and the completed Checklist. Note: It is important to obtain the feedback from someone else but you do not have to change your article based on their feedback.

Every student must submit their own article and every student must have reviewed at least one other student's article and completed a Checklist.

Your mark will be based on the article that you submit.

3.15.4: Getting It Write! Rubric

Reasoning and Proving				
Criteria	Level 1	Level 2	Level 3	Level 4
Making inferences, conclusions and justifications	Justification of the answer presented has a limited connection to the problem solving process and models presented	Justification of the answer presented has some connection to the problem solving process and models presented	Justification of the answer presented has a direct connection to the problem solving process and models presented	Justification of the answer presented has a direct connection to the problem solving process and models presented, with evidence of reflection
Connecting				
Criteria	Level 1	Level 2	Level 3	Level 4
Making connections among mathematical concepts and procedures	Makes weak connections	Makes simple connections	Makes appropriate connections	Makes strong connections
Communicating				
Criteria	Level 1	Level 2	Level 3	Level 4
Degree of clarity in explanations and justifications in reporting	Explanations and justifications are partially understandable	Explanations and justifications are understandable by me, but would likely be unclear to others	Explanations and justifications are clear for a range of audiences	Explanations and justifications are particularly clear and detailed

3.15.5: Numbers That Lie!

INSTRUCTIONS:

1. In your groups, read over the attached article that was prepared with diabetes research data from the Canadian Community Health Survey (CCHS) Cycle 3.1. Complete the Statistics Use Checklist BLM 3.15.2 to identify numerical and statistical errors that you find within the article.

2. Refer to BLM 3.15.8 to see the actual source data that the numbers where taken from.

3. Be warned! Some of the errors are obvious but others may be somewhat tricky.

4. **DO NOT ACCEPT THE PREMISE OR STATISTICS PRESENTED IN THIS ARTICLE AT FACE VALUE. VALIDATE AND CHECK EVERYTHING!**

Is Canadian Diabetes Research Money Paying Off?

Diabetes is a serious disease that affects a large number of Canadians. In Canada roughly $3000.00 per capita is spent annually. This is half of what is spent in the United States. Since the total population of Canada is 33,390,141 we are spending roughly $100 billion annually on health services. Are we getting our money's worth?

A statistical analysis of data provided by the Canadian Community Health Services (CCHS) Cycle 3.1 would tend to suggest otherwise. Table 1.1 below summarizes some of the data from this source:

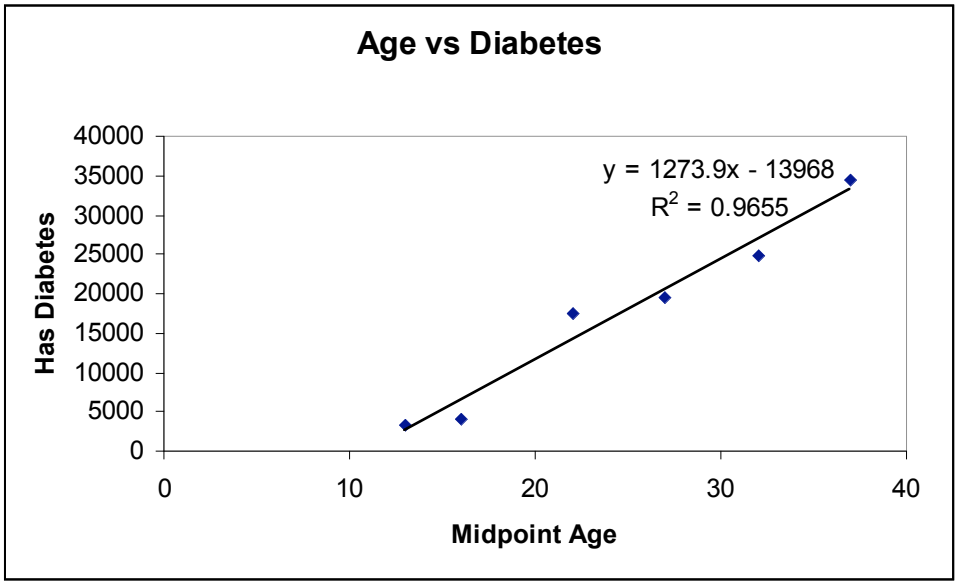
Table 1.1: Diabetes Compared to Age

Midpoint Age	Age Group	Has Diabetes? YES
13	12 TO 14 YEARS	3363
16	15 TO 17 YEARS	4001
22	20 TO 24 YEARS	17546
27	25 TO 29 YEARS	19412
32	30 TO 34 YEARS	24818
37	35 TO 39 YEARS	34323

3.15.5: Numbers That Lie! (continued)

Is Canadian Diabetes Research Money Paying Off? (Cont'd)

By completing a scatter plot and calculating standard deviation, variance and the regression line we obtain the graph below:



This has a variance of 0.9655 and a correlation coefficient of 0.9826. Since this graph demonstrates a very strong degree of positive correlation we are very confident in suggesting, from the data, that increasing age causes the onset of diabetes.

Using the regression line formula, we can estimate the number of Canadians who would have diabetes at ages 57, 67 and 77.

Age	Number with Diabetes	Percentage of Population
57	58,644	0.00176%
67	71,383	0.00214%
77	84,122	0.00252%

In 2007 it is estimated that approximately $13 billion will be spent on diabetes treatment, research and all factors associated with diabetes. This is approximately 10% of the annual money that we spend on health services in Canada. In other words, we are spending roughly $60,000 annually on diabetes patients at ages 57, 67 and 77. This is approximately 20 times the per capita rate. Collectively, they account for less than 1 % of the population. By age 90 there are still over 117,000 persons with diabetes.

It is clear, from the data, that the money would be better spent on prevention rather than treatment and research. The numbers don't lie!

3.15.5: Numbers That Lie! (continued)

Source Data: CCHS Cycle 3.1

AGE GROUPS

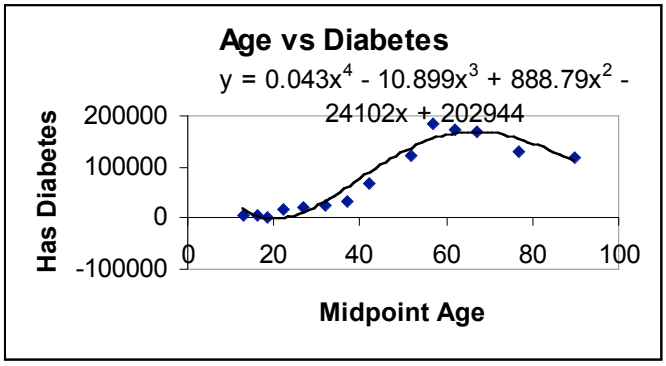
			Has Diabetes?	
Midpoint Age	Age Group	Total	YES	NO
13	12 TO 14 YEARS	1235378	3363	1231764
16	15 TO 17 YEARS	1304374	4001	1299971
18.5	18 TO 19 YEARS	813009	2263	809595
22	20 TO 24 YEARS	2239016	17546	2221077
27	25 TO 29 YEARS	2103064	19412	2083062
32	30 TO 34 YEARS	2079630	24818	2054112
37	35 TO 39 YEARS	2282458	34323	2247042
42	40 TO 44 YEARS	2803793	69378	2730245
52	50 TO 54 YEARS	2251247	123803	2124130
57	55 TO 59 YEARS	1973801	185760	1785570
62	60 TO 64 YEARS	1580369	174674	1403611
67	65 TO 69 YEARS	1213621	170572	1041598
77	75 TO 79 YEARS	807900	128520	678899
90	80 YEARS OR OLDER	875354	117377	757090

3.15.5: Numbers That Lie! (Teacher Notes)

Item	Yes	No	N/A	Comment
DATA				
Has all relevant data been included?		x		Ages 42 to 90 were excluded.
Is the data interpreted correctly?		x		Improper analysis
Has the original data source been referenced?	x			Although it was truncated
Has the data been presented in context?		x		Not all data shown
SURVEY RESULTS				
Have the sponsor(s) of the survey been identified?	x			
Has the population and sample size been included?		x		
Has any possible sampling bias been stated?		x		
Is the sample representative of the population?			x	Sampling methods can be reviewed on CCHS website
Does the survey results support the conclusions reached?		x		
REPRESENTATION OF DATA				
Have the data ranges been truncated or reduced?	x			Data ranges truncated
Are the axes scaled proportionally?	x			
Has the start/stop points of the time series been adjusted to support the point of the article?	x			
Have outliers been removed?	x			18.5 removed as outlier
ANALYSIS				
Has correlation been confused with causation?	x			
Has accidental association been confused with causation?		x		
Has the margin of error been included?		x		
MATHEMATICAL CALCULATIONS				
Are there any obvious mathematical errors?	x			Several. See notes below.
Has percentage been confused with percentage points?		x		
Has the base value been established for percentage references?	x			

3.15.5: Numbers That Lie! (Teacher Notes) (continued)

When all the data is included, the actual scatter plot appears below:



When all the data points are considered, the line of best fit is non-linear. In the article, through data truncation and removing data points, we only examined the linear portion of the graph.

The numbers quoted: population of Canada: 33,390,141, annual expenditure related to diabetes: $13 billion, per capita health care spending: $3,000 are accurate. The calculations shown at the end of the article are sheer nonsense, however. Why only consider spending for the age groups selected? The $13 billion represents spending over **all** diabetes patients. The number of patients quoted for age 90 is taken from the original data set and was **NOT** derived from the regression line equation quoted in the article – as it was implied.

The number of diabetes patients quoted for ages 57, 67 and 77 do not reflect the actual number of diabetes patients. The values should have been taken from the original source, not extrapolated beyond the range of data provided.