

Lesson: Misuses of Data

Eighth Grade Objective: 4.03 Identify misuses of statistical and numerical data.

Lesson:

Data and statistics are frequently manipulated to serve the purpose of the person or company presenting the material. For example, a frequently used statistic states that “4 out of 5 dentists with a preference prefer Brand A toothpaste”. The commercial that uses the statistic obviously is trying to sell their product, however, they neglect to mention how the survey was conducted or how many dentists didn’t have a preference.

It is also common that people use the word “average” to suit their needs. Always keep in mind that “average” can mean multiple things and that you should be wary when hearing that term.

What is misleading about the following situations?

1. In city X, there have been 42 robberies in the past year. In city Y, there have been 18 robberies. City Y is much safer than city X.

We don’t know how many people live in each city. It is quite possible that city Y is safer than city X, however, if the population of city X is 100,000 people and the population of city Y is 1000 people, city X is actually safer.

Consider city X: 42 robberies out of 100,000 people, that’s 0.042%.

Consider city Y: 18 robberies out of 1000 people, that’s 1.8%.

In this instance, city X is safer.

The data presented, in this case, is incomplete. It is important to know how many people live in each city.

2. The mean math test scores in Mrs. Allen’s class is 82%. The mean math test scores in Mrs. Bowden’s class is 86%. The students in Mrs. Bowden’s class are smarter than the students in Mrs. Allen’s class.

Not necessarily. Sample possible scores:

Mrs. Allen’s class: 82, 82, 82, 82, 82, 82, 82, 82, 82, 82, 82, 82, 82, 82

Mrs. Bowden’s class: 80, 80, 80, 80, 80, 80, 80, 80, 80, 80, 95, 95, 100, 100, 100

Mrs. Bowden’s class has 5 students who did better than the students in Mrs. Allen’s class, but the other 10, 2/3 of the class, did worse. Without knowing the spread of the data, the distribution of numbers, it is impossible to tell the exact factors that caused one mean to be higher than the other. It is possible that one set of data was more spread out, but had more or higher outliers, raising the mean.

3. On Tuesday during the lunch hour, the local news station asked viewers to call in and answer the following: “Do you think more taxes should be collected to pay for public school programs?” On the evening news, they state: “85% of the community does NOT want more taxes to pay for public school programs.”

This survey is misleading on several levels, the first being that the survey is self-selected. That means that only people who have the time, were watching the news station at noon, and have a strong opinion are going to respond. Another issue: the question was posed during the school day when most teachers and working parents would be unable to respond. These are the people who are most likely to vote that they would be for more taxes to fund public school programs.

4. Create a two sets of employee salaries for the following company. The first set should represent the statement, assuming it is completely accurate. The second set of data should assume the statement is misleading.

“A small company of 7 employees is looking for a motivated, hard-worker to join their firm. Average salary is \$35,000. Apply soon! Job will fill quickly!”

There are infinite solutions to this problem, an example is shown below:

Data set 1: 35,000, 35,000, 35,000, 35,000, 35,000, 35,000, 35,000

Data set 2: 20,000, 20,000, 20,000, 20,000, 20,000, 20,000, 125,000

Try these on your own:

1. Create two sets of data in which the first set has a higher mean and the second set has a higher median.

2. What could be done to improve the following survey:

John is trying to determine how much students would be willing to pay for juice boxes in the cafeteria. He asks every 2nd person in the lunch line during lunch on Tuesday.

3. Town M has a graduation rate of 97%. Town P has a graduation rate of 85%. More students graduate from Town M than from Town P.

Create two sets of population data for the scenario above. One that makes the statement true, the other that makes the statement false.

Check your answers:

1. There are infinite solutions to this problem, one of which is:

Set 1: 1, 2, 3, 4, 5; the mean is 3 and the median is 3

Set 2: 1, 1, 4, 4, 4; the mean is 2.8 and the median is 4

2. While asking every second person isn't a bad strategy, asking in the lunch line does not survey all students who may or may not purchase a juice box based on the price. Students who bring their lunches were not surveyed. John could instead ask every 2nd student as they enter the cafeteria.

3. There are many solutions to this problem, one of which is:

True: Town M has 100 people, 97% graduate, therefore 97 students graduate.
Town P has 100 people, 85% graduate, therefore 85 students graduate.

False: Town M has 100,000 people, 97% graduate, therefore 97,000 graduate.
Town P has 1,000,000 people, 85% graduate, therefore 850,000 graduate.

Quiz Yourself:

1. Create two sets of data in which the first set has a mean that is at least 10 points higher than the second and the second set has a larger range than the first set.
2. Shana is surveying her neighbors, asking each when the community pool should open. She goes door to door at 3 pm, immediately after getting home from school. The overwhelming majority of people said the pool should open as soon as possible. Is there bias in her survey?
3. The local pet supply store emailed a survey to the community asking them how much they would be willing to pay to have their dogs groomed. Seventy-five percent of the replies said they would pay \$50 or more. Is there bias in this survey?
4. School A has a 98% pass rate on the EOG's. School B has a 90% pass rate on the EOG's. More students at school A passed the EOG compared to students at school B.

Give two sets of population data, one that makes the above statement true, and one that makes it false.

Check your answers:

1. Set 1: 50, 55, 60, 65, 70 (Mean 60, Range 20)
Set 2: 30, 30, 30, 30, 100 (Mean 44, Range 70)
There are infinite solutions to this problem.

2. Shana does try to get each neighbor's opinion, but it is unlikely that each neighbor will be home when she arrives home from school. It is more likely that someone of school age will answer the door and complete Shana's survey. It is likely that adult community members are underrepresented in her survey.

3. Yes. First of all, it is unlikely that every member of the community has email. Secondly, this is a voluntary survey. The store cannot be sure that the respondents even

actually have dogs. Also, unless the potential respondent feels strongly about the question being asked, it is unlikely they will respond.

4. School A: 100 students, 98% pass, 98 students pass.
School B: 100 students, 90% pass, 90 students pass.

School A: 400 students, 98% pass, 382 students pass
School B: 500 students, 90% pass, 450 students pass