

Average part measurements = Summarize, represent and interpret data on a single count or measurement variable.

Program Task: Develop a Quality Control Plan.

Program Associated Vocabulary:
AVERAGE, OUTLIER, SAMPLING, SPC

Program Formulas and Procedures:
Preparing and reading X-bar charts and R-charts involves calculating the range and average of a sample number of parts.

Sometimes one sample part's size varies greatly from others in the sampling. That one "bad" part in a group of several "good" parts is most likely caused by a non-repeatable, unique factor that will not happen on a regular basis. This "bad" part is called an outlier.

If this part is included in the sampling, SPC data will give an inaccurate representation of the overall machining operation. The decision is usually made to remove that part from the sample, or remove it and replace it with another.

Example:
Three samplings are listed. Note the effect of the outlier ("bad" part) on the average from sampling 2.

S ₁	S ₂	S ₃
.250	.251	.249
.251	.205	.249
.252	.252	.251
.252	.253	.250
.249	.250	.252

$$S_1 = (.250 + .251 + .252 + .252 + .249) \div 5 = .2508$$

$$S_2 = (.251 + .205 + .252 + .253 + .250) \div 5 = .2412$$

$$S_3 = (.249 + .249 + .251 + .250 + .252) \div 5 = .2502$$

Removing the outlier of .205 from sampling 2 actually gives a better representation of the operation:

$$S_2 = (.251 + .252 + .253 + .250) \div 4 = .2515$$

PA Core Standard: CC.2.4.HS.B.1

Description: Summarize, represent and interpret data on a single count or measurement variable.

Math Associated Vocabulary:
MEAN, MEDIAN, MODE, OUTLIER

Formulas and Procedures:

Outlier	An extreme value in a set of data which is much higher or lower than the other numbers.
Mean (Average)	The average of set of data that is calculated by dividing the sum of the data by the number of items in the set.
Median	The middle value when data are arranged in numeric order or the average of the two middle numbers when the set has an even number of data.
Mode	The value that occurs most frequently in a set of data.

Measures of central tendency are mean, median and mode. Outliers affect the mean value of the data but have little effect on the median and mode of a given set of data.

Example:
A student receives a zero on a quiz and subsequently has the following scores:

0, 70, 70, 80, 85, 90, 90, 90, 95, 100

Outlier: 0

Mean:

$$\frac{0 + 70 + 70 + 80 + 85 + 90 + 90 + 90 + 95 + 100}{10} = 77$$

Median: since the data set has 10 values, there are two middle numbers, so one must find the mean of these two values, 85 and 90.

$$\frac{85 + 90}{2} = 87.5$$

Mode: The score 90 occurs more frequently than the other values (three times), so 90 is the mode.

Receiving a zero on a quiz significantly affects a student's mean, or average. Notice that the outlier had a small effect on the median and mode of the data.

It should be noted that because outliers affect the mean and have little effect on the median, the median is often used to describe "average" income. Often, one hears that the median income for a group is a certain value. Mean is not typically used because outliers, people who make significantly more or make no money at all, affect this measure.

Instructor's Script – Comparing and Contrasting

The Machine Tool Technology examples show real life situations where data can be affected by an outlier. Whether you are talking about bowling scores, swimming times, test scores, sets of foul shots, or hours of sleep, outliers change the data. The data value it affects the most is the mean. That is why sometimes the median or the mode is the best representation of a set of data. Outliers can be either much higher or much lower than the mean score. In the machining field, a value too high or too low would be considered a “bad” part, since accuracy is extremely important.

Common Mistakes Made By Students

Calculator error when finding the mean: Students often forget to use parenthesis when finding the mean of a data set. For instance, to find the average of 40 and 50, parenthesis must be used for the sum before dividing by two. Students often enter $40 + 50/2$, which yields an answer of 65 instead of entering $(40 + 50)/2$ which yields the correct answer of 45.

Changing the divisor: When determining how an outlier affects the mean of a data set, the student must find the mean with the outlier, then find the mean again once the outlier is removed. Removing the outlier decreases the number of data by one and therefore you must decrease the divisor. For instance, when you find the mean of 0, 10, 10, 12, 12, you must divide the sum by 5, but when you remove the outlier of 0, you must then divide by 4.

When calculating the median, students must list the data need in numerical order.

Finding the median of an even set of data: Finding the median or middle number, of a set of data is simple when there is an odd number of data. When there is an even number, there are two middle numbers, and these numbers must be averaged to obtain the median. For instance, the median of 1, 1, 2, 3, 3 is 2 because 2 is the middle number. If the data set is 1, 2, 3, 3, then 2 and 3 are the middle numbers and must be averaged to obtain the median of 2.5.

CTE Instructor's Extended Discussion

Since averaging can be used in countless machining applications, recognizing how an outlier can distort an average is an important concept.

Remember that a value is only an outlier if it occurs infrequently and in a non-repeatable way. If large variations become more and more frequent, there is probably a reason that can be identified and steps taken to remedy the situation.

Problems	Career and Technical Math Concepts	Solutions												
1. Find the average with and without the outlier for the sampling sizes of .505, .508, .495, and .470.														
2. Your custom chopper machine shop produced the following numbers of custom motorcycle swingarms over the last six months. How many would you project to machine monthly in the future? <table border="1" style="margin: 10px auto; border-collapse: collapse;"> <tr> <td style="padding: 5px;">Jan</td> <td style="padding: 5px;">Feb</td> <td style="padding: 5px;">Mar</td> </tr> <tr> <td style="padding: 5px; text-align: center;">6</td> <td style="padding: 5px; text-align: center;">8</td> <td style="padding: 5px; text-align: center;">7</td> </tr> <tr> <td style="padding: 5px;">Apr</td> <td style="padding: 5px;">May</td> <td style="padding: 5px;">Jun</td> </tr> <tr> <td style="padding: 5px; text-align: center;">6</td> <td style="padding: 5px; text-align: center;">12</td> <td style="padding: 5px; text-align: center;">9</td> </tr> </table>	Jan	Feb	Mar	6	8	7	Apr	May	Jun	6	12	9		
Jan	Feb	Mar												
6	8	7												
Apr	May	Jun												
6	12	9												
3. An apprentice cut material for die blocks that you are going to machine on the VMC. What size should you plan for when writing your program if the blocks were cut at the following sizes? 1.66, 1.51, and 1.67														

Problems	Related, Generic Math Concepts	Solutions
4. Sally earned scores of 60, 65, 65, and 80 on 4 tests. How would scoring 100 on a fifth test affect the mean?		
5. Tom recorded his daily caloric intake for 5 days. The results were as follows: 2500, 2600, 2600, 2400, and 3900. How would removing the outlier affect the mean, median and mode of the data?		
6. Angela recorded the number of hours she spent watching TV for one week. The results were as follows: 6, 2, 2, 1.5, 3, 2.5, 2. How would removing the outlier affect the mean, median, and mode of the data?		

Problems	PA Core Math Look	Solutions
7. Which of the following measures of central tendency does an outlier affect the most? a) Mean b) Median c) Mode		
8. Which measure of central tendency would best depict the following data: 10, 200, 200, 300, 325, 350 and 400? a) Mean b) Median c) Mode		
9. How would removing the outlier affect the mean of the following data: 1200, 2400, 2400, 2500 and 9000?		

Problems	Career and Technical Math Concepts	Solutions												
1. Find the average with and without the outlier for the sampling sizes of .505, .508, .495, and .470.		Average w/outlier = $(.505 + .508 + .495 + .470) \div 4 = .4945$ Average w/o outlier = $(.505 + .508 + .495) \div 3 = .5027$												
2. Your custom chopper machine shop produced the following numbers of custom motorcycle swingarms over the last six months. How many would you project to machine monthly in the future? <table border="1" data-bbox="289 449 576 604"> <tr> <td>Jan</td> <td>Feb</td> <td>Mar</td> </tr> <tr> <td>6</td> <td>8</td> <td>7</td> </tr> <tr> <td>Apr</td> <td>May</td> <td>Jun</td> </tr> <tr> <td>6</td> <td>12</td> <td>9</td> </tr> </table>	Jan	Feb	Mar	6	8	7	Apr	May	Jun	6	12	9		Remove the high month count of 12 or you could expect more work than is realistic. Average = $(6 + 8 + 7 + 6 + 9) \div 5 = 7.2$ You should be able to expect to machine 7 swingarms monthly.
Jan	Feb	Mar												
6	8	7												
Apr	May	Jun												
6	12	9												
3. An apprentice cut material for die blocks that you are going to machine on the VMC. What size should you plan for when writing your program if the blocks were cut at the following sizes? 1.66, 1.51, and 1.67		Average w / outlier = $\frac{(1.66+1.67+1.51)}{3} = 1.613$ Average w / o outlier = $\frac{(1.66+1.67)}{2} = 1.665$ It would be better to use the average without the outlier to allow for the larger material size to prevent cutter breakage.												
Problems	Related, Generic Math Concepts	Solutions												
4. Sally earned scores of 60, 65, 65, and 80 on 4 tests. How would scoring 100 on a fifth test affect the mean?		Initial Mean = $\frac{60 + 65 + 65 + 80}{4} = 67.5$ Mean with outlier = $\frac{60 + 65 + 65 + 80 + 100}{5} = 74$												
5. Tom recorded his daily caloric intake for 5 days. The results were as follows: 2500, 2600, 2600, 2400, and 3900. How would removing the outlier affect the mean, median and mode of the data?		The mean would decrease from 2800 to 2525. The median would decrease from 2600 to 2550. The mode would remain constant at 2600.												
6. Angela recorded the number of hours she spent watching TV for one week. The results were as follows: 6, 2, 2, 1.5, 3, 2.5, 2. How would removing the outlier affect the mean, median, and mode of the data?		The mean would decrease from 2.71 to 2.17 The median would remain constant at 2. The mode would remain constant at 2.												
Problems	PA Core Math Look	Solutions												
7. Which of the following measures of central tendency does an outlier affect the most? a) Mean b) Median c) Mode		a) Mean												
8. Which measure of central tendency would best depict the following data: 10, 200, 200, 300, 325, 350 and 400? a) Mean b) Median c) Mode		Median, because the outlier of 10 would make the average lower, and the mode of 200 would represent the lower data. The median is 300.												
9. How would removing the outlier affect the mean of the following data: 1200, 2400, 2400, 2500 and 9000?		The mean would decrease from 3,500 to 2,125.												