Thomas B. Fox

Transformations on Data Sets and Their Effects on Descriptive Statistics

The development of concepts in statistics and functions is an important part of the school mathematics curriculum. Also important is the formulation and verification of mathematical conjectures (National Council of Teachers of Mathematics 2000). This activity asks students to examine the effects on the descriptive statistics of a data set that has undergone either a translation or a scale change. They make conjectures relative to the effects on the statistics of a transformation on a data set. Students then defend their conjectures and deductively verify several of them.

The activity centers on important content in the statistics curriculum, an often-neglected topic in school mathematics, and helps make an often underused connection between mathematics and statistics. Statistics, such as the mean, median, and variance, are all functions. One can perform transformations on them as on any function, and they behave as do other functions. This activity can help reinforce the notion of basic descriptive statistics as functions and can help make this important connection between functions in general and statistical functions.

The first time I taught this concept I was somewhat disappointed with the results. It was not that the topic was not important or that the majority of my high school students did not learn the material. My dissatisfaction came from the lack of student engagement as a result of the largely teacher-centered manner in which I taught the subject. This experience led me to rethink my teaching and to construct this activity. When I used the revised activity, my students' level of engagement increased, and I was pleasantly surprised with what my students could learn more independently.

TEACHER'S GUIDE

This activity is accessible to students who have studied basic descriptive statistics. Developing skill in calculating descriptive statistics using paper and pencil is not a goal here. If students have not previously studied transformations, this activity may serve as a context to help introduce the topic to them. Those students who have studied the effects of transformations on functions can particularly profit from it.

The materials required to complete this lesson are minimal and include a metric tape measure (about 3 to 4 meters long) and some sort of spreadsheet technology. Basic familiarity with either a spreadsheet or a graphing calculator with a statistics editor is very helpful. Technology use in this activity frees students from the tedious calculations that would accompany the study of this topic if

This department is designed to provide in reproducible formats activities for students in grades 7–12. The material may be reproduced by classroom teachers for use in their own classes. Readers who have developed successful classroom activities are encouraged to submit manuscripts, in a format similar to the "Activities" already published, to the journal editor for review. Of particular interest are activities focusing on the Council's curriculum standards, its expanded concept of basic skills, problem solving and applications, and the uses of calculators and computers. Please send submissions to "Activities" by accessing **mt.msubmit.net**.

Edited by ***Gene Potter***
Hazelwood West High School (retired), Hazelwood, MO 63031

technology were not used. For the various operations on the graphing calculator required by this activity, the teacher may choose to model the appropriate keystrokes, depending on students' experience. I use the TI-83 throughout the article.

Students begin the activity by forming groups of approximately four. Each group receives a copy of the activity sheets and a tape measure.

Each student collects data for the different variables (height, wrist circumference, neck circumference, and arm length), shares them with the other group members, and records the data in the appropriate column in the chart on **sheet 1.**

Each student is responsible for entering the groups' data for one of the variables into the first column of his or her calculator's statistics editor. Students use the calculator to find descriptive statistics for their variable and then place them in the appropriate column in the table on **sheet 1.** Student answers here will vary. Since the calculator may not be able to find all of the necessary statistics, the mode, IQR, and variance may need to be computed by hand. However, the standard deviation, first and third quartiles, and the maximum and minimum are typically supplied by the calculator, which simplifies this task.

At this point in the lesson, I like to have my class as a whole examine more generally the idea of translating data sets and connect this idea to a concrete example. This is also an ideal time to introduce the notion of descriptive statistics as a multivariate function. I begin by considering the idea of the mean as a multivariate function

$$f(x_1, x_2, x_3, \ldots, x_n) = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

and by reviewing the general notion of a translation of a function. See **figure 1.**

The average function maps all the data points onto one number—its arithmetic average. A translation of a set of data $A = \{x_1, x_2, x_3, \ldots, x_n\}$ is a transformation that maps each x_i to $x_i + h$, where h is a nonzero constant. T is a translation if and only if $T: X \rightarrow X + h$ or $T(X) = X + h$. When translated by a factor of h, set A looks like $\{x_1 + h, x_2 + h, x_3 + h, \ldots, x_n + h\}$. The number h is called the *translation factor* and the value $X + h$ is called the *image of X.* A translation of the average function on the set $\{x_1, x_2, x_3, \ldots, x_n\}$ looks like this:

$$\text{avg}(x_1 + h, x_2 + h, x_3 + h, x_4 + h, \ldots, x_n + h)$$
$$= \frac{(x_1 + h) + (x_2 + h) + (x_3 + h) + (x_4 + h) + \cdots + (x_n + h)}{n}.$$

The same translation applied to the range function

$$\text{range } (x_1, x_2, x_3, x_4, x_5, \ldots, x_n) = \text{MAX}(x_1) - \text{MIN}(x_1)$$

would be

$$\text{range } (x_1 + h, x_2 + h, x_3 + h, x_4 + h, x_5 + h, \ldots, x_n + h)$$
$$= \text{MAX}(x_1 + h) - \text{MIN}(x_1 + h)$$

Fig. 1 Translation of a function

We also review the relationship between the graph of $y = f(x - h)$ relative to the graph of $y = f(x)$. After affirming the graph of $f(x - h)$ is h units to the right of $f(x)$, I ask students to use their understanding of translations of functions and of the meaning of the various statistical functions to make predictions regarding how different statistical functions would behave under a translation. Students record these predictions in the table at the top of **sheet 2.**

Continuing our class discussion, students consider the following scenario:

> Suppose we are translating the data points in a data set by adding the constant 2 onto each of the data points. Based on our discussion of translations of functions, how far (or how many units) might you expect the average function to map this translated set? Would this shift be to the left or to the right? Answer: 2 units to the right.

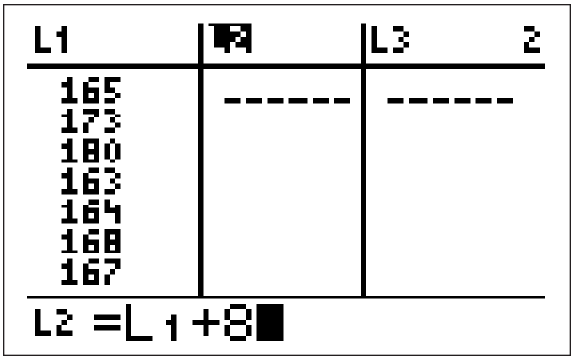


Fig. 2 TI-83 view screen set up to correctly translate a set of height data

I then ask the students to consider a similar scenario in which they add the constant of –0.5 onto each set. In this instance, the shift would be 0.5 units to the left.

After considering these scenarios, students use their calculators to translate the data sets they collected. Students do not need to translate each data point in their data set, since their statistics editor/spreadsheet can be programmed to do this task. The view screen should look something like the one shown in **figure 2.** An example of a translated height data set can be found in **figure 3.**

Using their graphing calculators, students then calculate the descriptive statistics of the translated data set. Student answers here will vary, depending on the data collected. Once students find the descriptive statistics for their data sets, they share the information with their group and log the data in the appropriate column in the chart at the bottom of sheet 1.

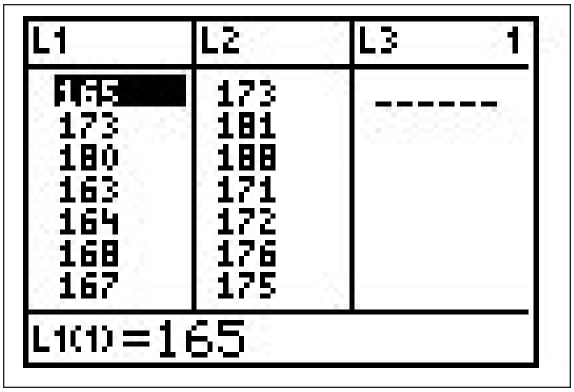


Fig. 3 An example of a set of height data that has been translated

By inspecting their table at the bottom of sheet 1, students can compare their descriptive statistics for each of the variables for their original data set with those for the corresponding translated data set, look for patterns, and formulate conjectures regarding what happens in general for each of the descriptive statistics when the data set is translated. Students note their conjectures in the appropriate chart at the bottom of sheet 2. They should find that the mean, median, mode, Q_1, and Q_3 are shifted by 8 units, while the measures of variation are unaffected. These effects are summarized in **table 1.**

Students can test some of the conjectures by constructing on their calculator a box plot of their data for their original data set and for their translated data set for each of their variables. Using these box plots, students test their conjectures and reformulate them, if necessary, for the median, first and third quartile, IQR, and range. If students have trouble verifying their conjectures pictorially using the box plots, the teacher might encourage them to use the TRACE feature of their calculator to find important values on the box plot. **Figure 4** shows an example of box plots for a set of height data (original and translated data sets).

TABLE 1

Effects on Descriptive Statistics of a Translation or Scale Change on a Data Set

	Translation of h Units	Scale Change by Factor of A
Mean	shifted by h	$A \cdot$ (Mean)
1st Quartile	shifted by h	$A \cdot$ (1st Quartile)
Median	shifted by h	$A \cdot$ (Median)
3rd Quartile	shifted by h	$A \cdot$ (3rd Quartile)
Mode	shifted by h	$A \cdot$ (Mode)
Variance	none	$A^2 \cdot$ (Variance)
Standard Deviation	none	$\lvert A \rvert \cdot$ (Stand. Dev.)
Range	none	$\lvert A \rvert \cdot$ (Range)
IQR	none	$\lvert A \rvert \cdot$ (IQR)

At this point in the activity, students share their conjectures with the whole class, helping students further test the conjectures by examining what other groups have found. If different groups have constructed different conjectures, which does not usually happen in this portion of the activity, the whole class can examine the statistics of a few particular data sets gathered by the students until progress toward a consensus is made.

To conclude this portion of the activity, I ask students to compare their initial predictions (found in the table at the top of sheet 2) to their final conjectures for each of the descriptive statistics (found in the table at the bottom of sheet 2). As a class, we discuss why students think that they differed.

Some students may be confused that the addition of h to x for a function causes in general a left shift by h, yet the mean and median are shifted to the right by h. To help students understand that this really is not a contradiction, teachers can ask students to investigate the behavior of the line $y = x$ under similar circumstances. For example, they can ask students to graph on their calculators $f(x) = x$ and then $f(x + 2) = x + 2$. Students discover that the graph of $y = x$ is shifted 2 units to the left to produce the graph of $f(x + 2)$, as would be expected. However, the y values of $f(x+2)$ are 2 units *higher* relative to that of $f(x)$. Thus the resulting graph is a translation of $f(x)$ up by 2. Other values of h yield similar results. For this same reason the mean and the median are increased by h; there is no contradiction.

The next part of the activity examines the effects on descriptive statistics of a data set after a scale change has been performed on it. Transforming the data sets from metric to English system units serves as the context.

To begin this portion of the activity, I introduce the idea of a scale change by comparing scale changes and translations and by introducing a scale change with a scale factor of $m = 100$ on a small data set, {1.8 m, 1.75 m, 1.85 m}, in which the

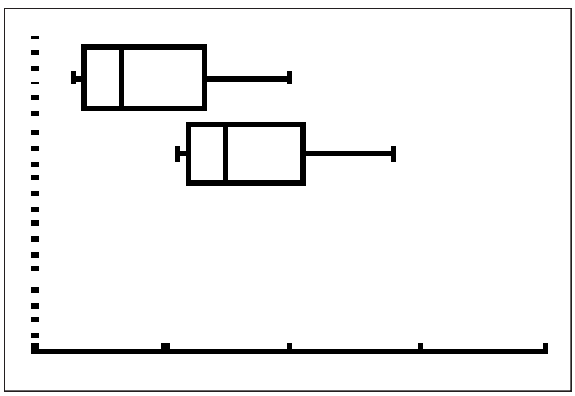


Fig. 4 An example of box plots for a set of height data—original (above) and translated (below) data sets

A scale change of a data set A = $\{x_1, x_2, x_3, \ldots, x_n\}$ is a transformation that maps each x_i to mx_i, where m is some nonzero constant. S is a scale change if and only if $S : X \to mX$ or $S(X) = mX$. $S(A) = \{mx_1, mx_2, mx_3, \ldots, mx_n\}$. The number m is called the *scale factor of the scale change,* and the value mX is called the *image of X.* Applying such a scale change to the average function, results in

$$\text{avg}(mx_1, mx_2, mx_3, mx_4, \ldots, mx_n) = \left(\frac{mx_1 + mx_2 + mx_3 + mx_4 + \cdots + mx_n}{n}\right).$$

Fig. 5 A scale change of a data set

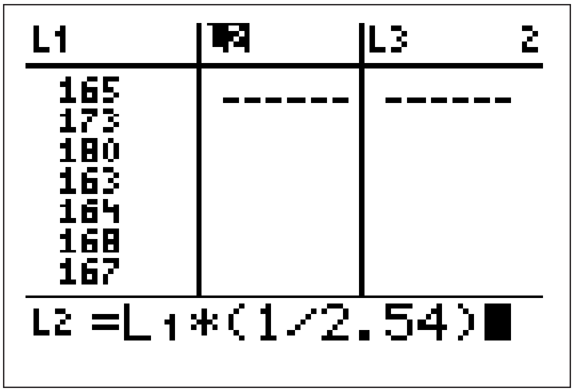


Fig. 6 TI-83 view screen set up to correctly scale change a set of height data

units of the data points are converted from meters to centimeters: {180 cm, 175 cm, 185 cm}. I then connect these examples to the more general idea of a scale change on a data set. See **figure 5.**

Students consider another example of scale changes on a data set, tying in some of the scale change notation and vocabulary just discussed.

> Suppose that you wanted to perform a scale change on a set of height data {1.8 m, 1.75m, 1.85m}. The scale change on this set would be $S : X \to 1000X$ or $S(X) = 1000X$. The image of the original data set under this scale change would be {1800, 1750, 1850}. What would the units of measure for the transformed data be? [Answer: millimeters]

After this example, I ask students to use their understanding of scale changes on functions and of the meaning of the various statistical functions to make predictions regarding how different statistical functions behave under a scale change. These predictions should be recorded in the table at the top of sheet 3. Students should then discover the scale factor from centimeters to inches that will be used in this activity

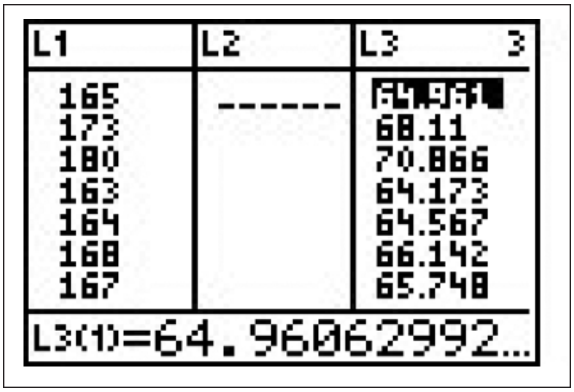


Fig. 7 TI-83 view screen containing a set of scale-changed height data

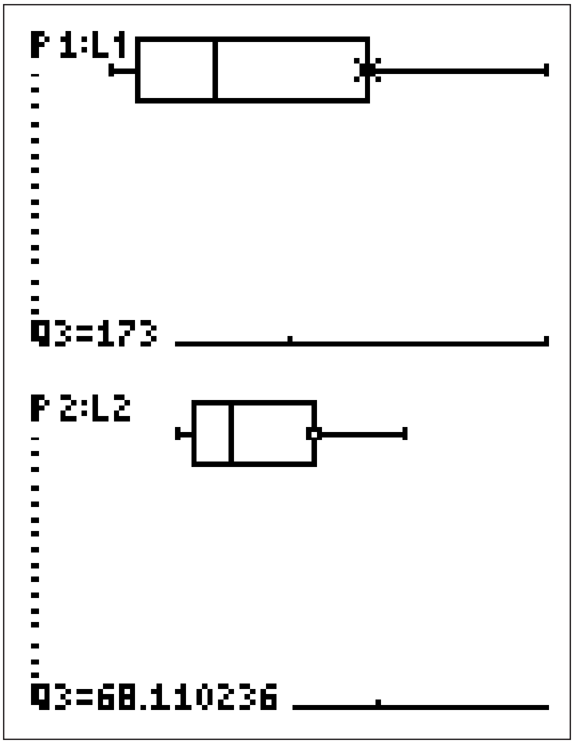


Fig. 8 An example of box plots for a set of scale changed height data—original (metric) and scale-changed (English system) data sets

by making conversions of several measurements from centimeters to inches and by looking for a pattern.

Once the scale factor for conversion from centimeter to inches has been discovered, students perform this scale change on their original data set. They can use the statistics editor to perform this scale change in a manner similar to the translation portion of the activity. An example of a TI-83 view screen set up to scale change a set of height data correctly can be found in **figure 6.** After students perform the scale change, their display looks something like the one shown in **figure 7.**

Students are now ready to calculate descriptive statistics for their data sets in English system units. This calculation is done as on sheet 2. Students then share their results with their group and place them in the correct column in the chart found on sheet 1. Student answers will vary.

At this point in the activity, students formulate their conjectures about the effects of the scale change on the descriptive statistics. Students should be careful to compare the statistics for their original and scale changed data sets (and not the translated data sets) and record their conjectures in the chart on the bottom of **sheet 3.** Correct answers can be found in **table 1.**

To test some of their conjectures, students use their calculators to construct box plots of their data for their original data set and for their scale-changed data set for several of the variables. Using these box plots, students test and reformulate, if necessary, their scale change conjectures for the median, first and third quartile, IQR, and range. An example of box plots for a set of height data can be found in **figure 8.**

Once the box plots are completed, a whole-group discussion of students' conjectures can take place. This discussion may give students cause to question some of their conjectures. The class may even reach consensus on a few conjectures that are "close" but not entirely accurate. This consensus is especially true for the different measures of dispersion. **Sheet 4** is designed to challenge students' potentially inaccurate conjectures and gives them an opportunity to reformulate any incorrect conjectures they may have formulated. Teachers can ask students to complete this portion of sheet 4 and then discuss the questions as a group.

To conclude this portion of the activity, students compare their initial predictions (found in the table at the top of sheet 3) to their final conjectures for each of the descriptive statistics (found the table at the bottom of sheet 3). As a class, we discuss briefly why students think their final conjectures and initial predictions differed. Students then verify their conjectures.

Complete proofs are shown in the **appendix.**

FOLLOW-UP ACTIVITIES

Up to this point in the activity students have only informally discussed their conjectures. Students who show appropriate readiness at this point can more formally verify their conjectures for the average function for both the translation and scale change. The arguments involving the mean are largely algebraic in nature, involving factoring, and should be accessible to most students who have successfully completed first year algebra.

SOLUTIONS

Sheets 1 and 2

Student answers will vary according to data collected. The correct answers for the effects on the descriptive statistics on a translated data set can be found in **table 1.**

Sheet 3

1. $\dfrac{1}{2.54} \approx 0.394$

2. and 5. Student conjectures may vary. See **table 1** for correct responses.

Sheet 4

1. Examples of box plots for a set of height data (original and scale changed data sets) can be found in **figure 8.**

2. Yes, because they are all measures of central tendency. When the same number is added to each data point, the middle of the data set, the most common data point, and the average will all be increased by the same amount.

3. Yes, because they are all measures of dispersion of the data set. A translated data set is no more disperse than the original data set.

REFERENCES

National Council of Teachers of Mathematics (NCTM). *Principles and Standards for School Mathematics*. Reston, VA: NCTM, 2000.

APPENDIX: COMPLETE PROOFS

Proof of average under a translation

Let $\{x_1, x_2, x_3, \ldots, x_n\}$ be a data set subject to a translation, which is translated by h units. The average (avg) of the original data set is:

$$\begin{aligned}\text{avg}(x_1, x_2, x_3, x_4, \ldots, x_n) &= \frac{x_1 + x_2 + x_3 + x_4 + \cdots + x_n}{n} \\ &= \bar{x}\end{aligned}$$

The average of the translated data set is:

$$\begin{aligned}&\text{avg}(x_1 + h, x_2 + h, x_3 + h, x_4 + h, \ldots, x_n + h) \\ &\quad = \frac{(x_1 + h) + (x_2 + h) + (x_3 + h) + (x_4 + h) + \cdots + (x_n + h)}{n} \\ &\quad = \frac{x_1 + x_2 + x_3 + x_4 + \cdots + x_n + nh}{n} \\ &\quad = \frac{x_1 + x_2 + x_3 + x_4 + \cdots + x_n}{n} + h \\ &\quad = \bar{x} + h\end{aligned}$$

Proof of average under a scale change

Let $\{x_1, x_2, x_3, \ldots, x_n\}$ be a data set subject to a scale change, with a scale factor of m. The average of the original data set is

$$\begin{aligned}\text{avg}(x_1, x_2, x_3, x_4, \ldots, x_n) &= \frac{x_1 + x_2 + x_3 + x_4 + \cdots + x_n}{n} \\ &= \bar{x}\end{aligned}$$

The average of the scale-changed data set is

$$
\begin{aligned}
&\operatorname{avg}(mx_1, mx_2, mx_3, mx_4, \ldots, mx_n) \\
&\quad = \left(\frac{mx_1 + mx_2 + mx_3 + mx_4 + \cdots + mx_n}{n}\right) \\
&\quad = m\left(\frac{x_1 + x_2 + x_3 + x_4 + \ldots + x_n}{n}\right) \\
&\quad = m\bar{x},
\end{aligned}
$$

which is what we needed to prove. ∞

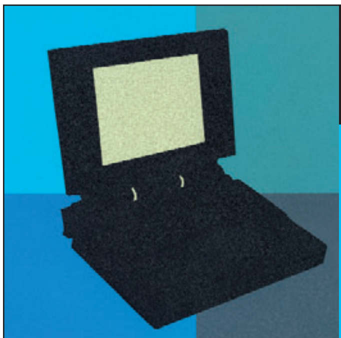
Activity sheets follow.

THOMAS FOX, fox@cl.uh.edu, is a former secondary school teacher who teaches preservice and in-service teachers at the University of Houston–Clear Lake. His interests include the acquisition of mathematics content knowledge by preservice teachers as well as issues in the teaching of calculus.

You Think That You Have Problems?

Then we would like to hear from you!

The Editorial Panel is planning a special commemorative poster of 100 problems in conjunction with the 100th anniversary of the *Mathematics Teacher*. Be a part of the 100th anniversary celebration by submitting your favorite problems from past *Mathematics Teacher* calendars. Problems addressing a wide variety of topics and levels of difficulty are needed to appeal to a large audience. Please send the month, calendar date, and year that the problem appeared in the *Mathematics Teacher*. Send submissions by March 31, 2006, to Judith Covington at mathprobs@lsus.edu.



Save Time: Submit Your Manuscripts Online

NCTM is simplifying the manuscript submission process. The old process involved writing, mailing several copies of the manuscript to headquarters, then waiting while the arduous task of reviewing and revising occurred. The new streamlined process will allow you to write, submit your material online, and then receive confirmation and updates as often as you like by accessing your own Web page. It will save both time and mounds of paper and allow you, the author, to get your approved manuscript in print much faster.

How to Start

Step-by-step information is given for submitting manuscripts to *Mathematics Teacher* (*MT*), *Mathematics Teaching in the Middle School* (*MTMS*), and/or *Teaching Children Mathematics* (*TCM*).

1. Enter one of these sites, keyed to either *MT, MTMS,* or *TCM:*

mt.msubmit.net **mtms.msubmit.net** **tcm.msubmit.net**

2. Log in and enter a password.
3. Click New Users: Please register here. Once you register for one journal, that information will be in the system, under that same name. However, if you write for more than one journal, you must log in to each journal's site separately.
4. Click Author Instructions under **Author Tasks**. Once you have read the step-by-step instructions and your manuscript is complete, you are ready for the last step.
5. Click Submit Manuscripts.

You will upload your manuscript file to this site, and the process will begin. A red arrow will alert you to any information required or action you need to take. An acknowledgment e-mail will be sent to you, explaining how to log back into the system and check the status of your article at any time. If you have questions, clicking on **HELP** in the top bar will direct you to various question marks—noted as ?—spread throughout the site. If you have additional questions not answered by the site, e-mail Sandy Berger, managing editor, at sberger@nctm.org.

We look forward to receiving your online submissions, and thank you again for writing.

Descriptive Statistics and Translations and Scale Changes

Data Collection

1. Collect the following data from each member in your group: height, circumference of wrist, circumference of neck, and arm span (from fingertip to fingertip). Enter the data (in centimeters) into this chart for these four variables:

	Height (cm)	Wrist (cm)	Neck (cm)	Arm (cm)
Names				

2. Calculate the statistics for each of these four variables: height, wrist circumference, neck circumference, and arm length. Place your results in the column marked "Original."

	Height			Wrist			Neck			Arm		
	Original	Translated	Scaled	Original	Translated	Scaled	Original	Translated	Scaled	Original	Translated	Scaled
Mean												
Median												
Mode												
Variance												
Standard Deviation												
Range												
1st Quartile												
3rd Quartile												
IQR												

Descriptive Statistics and Translations and Scale Changes

Translations

Suppose that in a five-year period each group member's height increases by 8 cm, their wrist circumference increases by 1 cm, their neck circumference increases by 2 cm, and their arm span increases by 4 cm.

1. If you add a constant *h* to each data point (a translation), what happens to the descriptive statistics on that data set? Predict what effect such a translation will have on the descriptive statistics and enter your predictions in the space provided in the top chart on this sheet.

2. Translate your data set by adding the appropriate constant and recalculate the same descriptive statistics as you did for your original data set. The calculator/spreadsheet can help you do the calculations. Share your results with those in your group, and enter them into the appropriate "translated" columns in the chart at the bottom of sheet 1.

3. Compare the statistics in the "translated" column with those of the data set in the "original" column for each of the four variables. What do you see happening? Are they the same? Are they changed? If so, how are they changed? Is the change related to the number you added to each piece of data? Conjecture what would happen in general for each of these statistics when the data set is translated. Summarize your conjectures in the bottom chart on this sheet.

Mean	
Median	
Mode	
Variance	
Standard Deviation	
Range	
1st Quartile	
3rd Quartile	
IQR	

4. Construct a box plot for your assigned data variable for both the original data set and the translated data set. Compare the two box plots. Use the TRACE feature of your calculator to locate values on your box plot.
   Do your box plots confirm your conjectures in question 4 regarding the median? _______
   the first and third quartiles? _______
   the IQR?_______
   the range? _______
   Share this with the others in your group. Re-examine and reformulate any conjectures that do not agree and place any new conjectures in the chart.

Mean	
Median	
Mode	
Variance	
Standard Deviation	
Range	
1st Quartile	
3rd Quartile	
IQR	

Descriptive Statistics and Translations and Scale Changes

Scale Changes

Suppose we want to make unit conversions on our original data, converting our height, arm span, neck circumference, and wrist circumference from centimeters to inches.

1. Convert several measurements in centimeters to inches. Compare the measurements and look for a pattern. Each measurement in inches is what multiple of the corresponding measurement in centimeters? ________ This number will be our scale factor for the data that we collected and placed on sheet 1.

2. Predict what effect a scale change on a data set will have on the following descriptive statistics. Place your predictions in the top chart.

Mean	
Median	
Mode	
Variance	
Standard Deviation	
Range	
1st Quartile	
3rd Quartile	
IQR	

3. Your graphing calculator can perform this scale change on your original data set. Place the scale-changed data set in the third column of your statistics editor/spreadsheet.

4. Calculate the same descriptive statistics for your scale changed data set. Share your statistics with your group and enter them into the "scaled" column for each of the variables in the chart at the bottom of sheet 1.

5. Compare the statistics in the "scaled" column with those of the data set in the "original" column for each of the four variables. What do you see happening? Are they the same? Are they changed? If so, how are they changed? Is the change related to the number by which you multiplied each piece of data? Conjecture what would happen in general for each of these statistics when the data set undergoes a scale change. Summarize your conjectures in the bottom chart.

   If you multiply each data point in a set by a factor *m*, what happens to the statistics in the table?

Mean	
Median	
Mode	
Variance	
Standard Deviation	
Range	
1st Quartile	
3rd Quartile	
IQR	

Descriptive Statistics and Translations and Scale Changes

Sheet 4

1. Construct a box plot for your assigned data variable for both the original data set and the scale changed data set. Compare the two box plots. Use the TRACE feature of you calculator to locate values on your box plot.
   Do your box plots confirm your conjectures with regard to the median? _______
   the first and third quartiles? _______
   the IQR?_______
   the range? _______
   Share this with the others in your group. Re-examine and re-formulate any conjectures that do not agree and place any new conjectures in the second chart on sheet 2.