Issues in Teaching the Comparison of Variability to Non-Statistics Students

Lorraine Garrett and John C. Nash
University of Ottawa

Journal of Statistics Education Volume 9, Number 2 (2001)

Copyright © 2001 by Lorraine Garrett and John C. Nash, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Homoskedasticity; Teaching statistics; Test; Variability.

Abstract

One of the main themes of statistics courses is to teach about variability, as well as location. This is especially important for non-statistics students, who often overlook variability. We consider particularly the problem of comparing variability among \( k \) samples (\( k > 2 \)) that are not necessarily drawn from Gaussian populations. This can also be viewed as testing for homoskedasticity of samples. We examine tools for this problem from the perspective of their suitability for inclusion in elementary statistics courses for students of non-mathematical subjects. The ideas are illustrated by an example that arose in a student project.

1. Introduction

Much effort is needed in service courses in elementary statistics to help students to appreciate and to be able to deal with the concept of variability, which is central to the subject. Having achieved some success in convincing students that variability is important, it is embarrassing to note that the traditional test for equality of variance, the standard \( F \)-ratio test (for example, in Aczel 1996, pp. 341-349), handles only two samples, and the parent populations must be assumed independent and Gaussian.

Moreover, modern statistical software makes it straightforward to check whether data conform to a Gaussian distribution, using tools such as the normal probability plot. Thus students are well-positioned to challenge the use of tests and other tools designed for Gaussian populations. Some are also astute enough to recognize that two-sample tests are not directly suitable for \( k \)-sample situations.

In a class project, one of us wanted to compare textbook prices across different faculties. We provide these data in Table 1 and a Minitab script for reading them in Appendix 1. Though it is clearly important to good statistical practice to specify the data definitions and the protocol for gathering the samples of textbook prices, these details are not central to the discussion here and will be omitted. We note that the samples in the example are of unequal size. From graphical displays, some of the samples appear non-Gaussian, and the variability is apparently different across faculties. How can one decide if the population variability is really different based on the sample data? From the point of view of the instructor, the issue is one of finding appropriate tools that can be taught to and used by non-mathematical students in an introductory statistics course. This note considers some possibilities.
Table 1. Textbook Prices in Dollars for Eight Faculties

| a   | b   | c   | d   | e   | f   | g   | h   |
|-----|-----|-----|-----|-----|-----|-----|-----|
| 76.95 | 24.5 | 30.4 | 47.1 | 54.95 | 94.95 | 54.95 | 81.95 |
| 99.95 | 34.95 | 76.95 | 99.95 | 26.95 | 48.5 | 78.95 | 31.95 |
| 84.95 | 18.8 | 93.95 | 34.95 | 23.15 | 79.7 | 73.5 | 61.95 |
| 67.95 | 29.95 | 103.95 | 65.95 | 79.95 | 86.95 | 32.95 | 72.95 |
| 79.95 | 20 | 98.5 | 11 | 24 | 79.7 | 36 | 83.95 |
| 71.95 | 40.95 | 55 | 81.95 | 24.5 | 112.15 | 56.55 | 35 |
| 30.95 | 40.95 | 84.95 | 85.95 | 17.99 | 42.55 | 68.95 | 84.95 |
| 30.95 | 37.95 | 90.95 | 76.95 | 78.95 | 104.95 | 23.95 | 31.95 |
| 41.95 | 50.95 | 99.95 | 21.05 | 25.95 | 96.7 | 26.95 | 98.95 |
| 79.95 | 29.95 | 95 | 19.95 | 20.95 | 104.95 | 65.2 | 29.95 |
| 93.95 | 36.95 | 115.5 | 27 | 79.95 | 86.65 | 81.2 | 94.95 |
| 39.95 | 40.5 | 94.95 | 27.5 | 39.95 | 94.95 | 32.95 | 80.95 |
| 108.95 | 44.95 | 112.95 | 27.5 | 65.95 | 102.95 | 21.75 | 25.95 |
| 57.95 | 32.95 | 82.75 | 66 | 33.5 | 76.95 | 59.95 | 66.95 |
| 77.95 | 53.95 | 104.95 | 63.5 | 39.95 | 62.3 | 32.95 | 44.95 |
| 27.95 | 52.95 | 100.8 | 69.95 | 15.95 | 95.95 | * | 63.95 |
| 61.95 | 39.95 | 76.95 | 26.5 | 70.95 | 78.5 | * | 78.95 |
| 51.95 | 27.8 | 73.05 | 101.95 | 39.95 | 89.95 | * | 32.6 |
| 88.95 | 19.95 | 92.95 | 89.95 | 56.5 | 100.95 | * | 84.95 |
| 80.95 | 14.5 | 98.95 | 59.95 | 27 | 104.95 | * | 56.95 |
| 100.95 | 75 | 90.95 | 78.95 | 59.95 | 108.95 | * | 71.95 |
| 16 | 20.25 | 122.95 | 26.95 | 76.95 | 78.5 | * | 98.95 |
| 95 | 42.5 | 112.95 | 32.5 | 60.95 | 61.95 | * | 89.95 |
| 86.95 | 59 | 86.95 | 16 | 85 | 50.5 | * | 84.95 |
| 58.95 | 78.95 | 104.95 | 69.95 | 55.95 | 53.95 | * | 32.95 |
| 89.5 | 59.95 | 99.95 | 37.95 | 69.55 | 88.5 | * | 83.95 |
The statistical problem of interest is, we believe, one that should attract attention. Our example concerns the comparison of the variability of prices, a topic of interest for consumers, vendors, and regulators. In quality management, the variability across samples or batches in many types of processes is often as important as the differences in level. In the process of developing this paper, we noted that (1) no "business statistics" textbooks that we could find addressed this problem, and (2) few statistics books, business or otherwise, index a test for two or more variances.

Madansky (1988) uses the term "test for homoskedasticity," but this terminology is also not generally applied. Thus novices in the field may have some difficulty in finding suitable information on the topic. Nevertheless, from the works we cite and the references therein, a number of tools have been developed and studied. (We note that different names are sometimes used for similar techniques.) We will be content to use such results and will not attempt to develop new methods.

Though statistical tests usually concern the variance, we will use variability as a general term to cover any measure of spread, since the traditionally used variance or standard deviation may not be suitable to our data or situation. Moreover, we are willing to consider graphical and similar tools that provide support for decision making with less rigour than hypothesis tests.

To summarize, we want to consider what existing and well-documented tools are suitable, in the context of a course in statistics for non-mathematical students, for comparing the variability of \( k \) samples (\( k > 2 \)), possibly of unequal size, where some of the parent populations may be non-Gaussian.

2. Background

Textbooks for elementary applied statistics courses provide little guidance about this problem. Indeed, though the "analysis of variance" is a prominent topic in both courses and textbooks, we see surprisingly few actual comparative analyses of the variability of samples. Most elementary textbooks present only the Fisher \( F \)-ratio test for Gaussian populations.

More advanced monographs give some pointers. For example, Bradley (1968) suggests applying a nonparametric test of location to distances from median or mean. However, Bradley gives a number of provisos about such tests, for example, the Siegel-Tukey test (Bradley 1968, p. 118). Students find such caveats confusing; professors find that they take a lot of time and effort to communicate. A somewhat different treatment, using traditional tests, but discussing transformations to deal with non-normality, is given by Neter, Wasserman, and Kutner (1990, pp. 614-623).

Even when data are drawn from Gaussian populations, the Fisher test compares just two samples at a time. A multiple comparison test similar to the Fisher test is that of Bartlett (see Snedecor and Cochran 1967, p. 296; Madansky 1988, p. 61 ff.; or Conover, Johnson, and Johnson 1981). It also requires Gaussian populations.

The sensitivity of the Fisher and Bartlett tests to non-normality is well-known but bears underlining. For example, Hoel (1971, p. 273) states, "Unfortunately the preceding test is not reliable if \( X \) and \( Y \) do not possess normal distributions." Snedecor and Cochran (1967, p. 298) are more precise: "Unfortunately, both Bartlett's test and this test (an unequal sample version of Bartlett's test) are sensitive to non-normality in the data, particularly kurtosis."

The research literature on what we will call the \( k \)-sample variability comparison problem offers some help. The monograph by Madansky (1988) and the paper by Conover et al. (1981) present several approaches. More recently, Lim and Loh (1996) performed a number of simulation experiments to compare a range of variance equality tests, extending Loh's (1987) simulation study of a modified Levene (1960) method, in particular the variant of Brown and Forsythe (1974). Our challenge is to adapt such results to the capabilities and level of the introductory service course in statistics, and to do so in such a way that the content of the course remains balanced.
3. Approaches and Their Merits

We have considered three main themes:

1. The transformation of our data so that tools that are already part of the student repertoire may be applied;
2. The use of resampling methods;
3. The adaptation of two-sample approaches to the $k$-sample problem.

Of these approaches, the last is not, in our opinion, suitable for teaching to novices. Though courses often include mention of the use of the Tukey pairwise comparisons test (Aczel 1996, p. 383), the foundation of such approaches is not discussed, largely because it involves subtle and detailed thinking (Neter et al. 1990, p. 579). This impinges on the use of resampling statistics, since the major use, in our opinion, of bootstrap and other resampling methods is to compute more reliable measures of the variances of the $k$ populations, which must then be compared.

Resampling has become popular with the availability of cheap computing power. However, the term "bootstrap" appears only once in a Journal of Statistics Education title through March 2001, while "resampling" does not appear (though it is mentioned in a sub-heading of "Teaching Bits" in at least one issue). Moreover, though we suspect most university and college teachers would like to introduce some aspects of resampling into applied statistics courses, few have chosen to do so in the introductory courses. The most active and long-standing proponent of their application to teaching has been Simon (1969 and later articles and books), who is associated with a software package (Resampling Stats) for this purpose.

Our reticence in proposing resampling methods for the current problem arises from the following arguments:

1. There are many details involved in applying resampling methods to our problem. Lim and Loh (1996) require almost half a page of tightly printed "recipe" to describe the general structure of the bootstrap procedure they used.
2. There are many choices in applying resampling methods. Should we use the bootstrap or jackknife? How many samples are "enough"? Have we used the correct method for sampling?
3. Making these choices often implies the imposition of assumptions that may or may not be suitable to the situation at hand.
4. Do the resulting pseudo-data or the resampling statistics provide more insight into the comparison of the variability than the original data?

None of these objections is applicable to advanced or even to intermediate statistics courses, whether for statistics or non-statistics students. Indeed, one textbook we have used for a course that immediately follows the introductory one is Hamilton (1992), which has an excellent discussion of resampling in Appendix 2. However, they are serious obstacles in the introductory service course.

4. Transformations

Transformation of our data so that we can use available tools to solve the problem at hand is a standard and traditional tool in applied mathematics and statistics. It is an appealing approach, since it shows students that we can recycle our intellectual capital and increase the efficiency of learning.

In the introductory service course, students generally have limited skills with mathematical functions. Our experience is that we need to review log, exp, and their relationship to $a^b$, i.e., the power function, as well as square, cube, and other roots if and when such functions are needed. (By design, we try to avoid situations that need them, but we may want to reconsider this choice in the light of the present discussion.) Thus, the traditional Box-Cox transformation to attempt to render data Gaussian would not be appropriate for most introductory level courses. (It is a topic in intermediate courses, including our own.) However, for students who have seen a number of manipulations of data, it may be appropriate to show an example of the Box-Cox transformation, especially in a case study and as a topic that is "not on the final."
The present dataset, however, for which we present boxplots in Figure 1, is not suitable for transformations of the Box-Cox kind because the sub-samples appear to have different distributional shapes, even from the boxplots. Stem-and-leaf diagrams or other distributional plots make this clear. For those students who have been shown the tool, the normal probability plot could even be used. We note that one can view the normal probability plot as a graph of a transformation of the ranks of data versus the data themselves.

More directly useful to us are transformations of the data that convert variability to level. Two transformations of particular interest are

\[ y_{1i} = \text{abs}(x_i - \text{mean}(x)) \]

and

\[ y_{2i} = \text{abs}(x_i - \text{median}(x)), \]

that is, the absolute deviations from the mean or median. These transformations allow a number of possible tools to be used to assess the variability, which is now given by measures of location or level of the transformed data. Later we consider further transformations to attempt to symmetrize the new \( y_1 \) or \( y_2 \) data.
Finally, many nonparametric methods use ranks rather than the raw data, then transform the ranks by various scoring schemes so that tests based on probability calculations using common distributions may be made. Such transformations are similar to the calculations used for normal probability or other quantile plots that elementary statistics students may already have seen, though few will comprehend them well.

5. Graphical Tools to Display Variability

The most obvious tool to display variability is the multiple boxplot. In Figure 1 we showed the boxplots of the data themselves. Figure 2 displays the absolute deviations from the medians. We have included the "notches," that is, the approximate 95% confidence intervals recommended by McGill, Tukey, and Larsen (1978) and Velleman and Hoaglin (1981). This gives us a visual method for comparing the variability. Groups for which the boxplot notch intervals do not overlap are likely different in variability. (Here we encounter once again the multiple comparison issue.) We note that Minitab appears to offer notched boxplots only in the "obsolete" character version of graphs. (The Minitab macro MEDBOX.MAC draws notched boxplots of the absolute deviations from group medians using the character graphics format.) A different restriction was noted with Stata (version 5 or earlier), in that the maximum number of groups is six. A rather old version of Systat produced boxplots similar to those here, but the JPEG file did not reproduce as cleanly as that from the most recently available stable download of R.

Figure 2. Notched Boxplots of the Absolute Deviations From Group Medians of the Textbook Price Data. Drawn with R, version 1.010. Once again, Faculty g has a box that does not cover the notches.
6. Traditional Tests Robust to Non-Normality

It is well-established (see Madansky 1988 or Conover et al. 1981) that the traditional analysis of variance (ANOVA) for comparing means is quite robust to non-normality of the samples. We therefore wish to find a way to use this to compare variability. The absolute deviations from group medians provide a set of distances whose means can be compared by a one-way ANOVA. This is the central idea of the Levene test (Conover et al. 1981, Loh 1987, Lim and Loh 1996, Hines and O'Hara Hines 2000). Absolute deviations from means were used in the original Levene (1960) test, but Conover et al. found that deviations from medians are preferable. Moreover, Conover et al. (1981) suggest that the use of the square roots of the absolute deviations from medians does not result in great benefit. This could, however, be a useful subject for a student project, given the following statement by Cleveland (1993, p. 51):

The square root transformation is used because absolute residuals are almost always severely skewed toward large values, and the square root often removes the asymmetry.

The reader can see that the square root transformation improves the symmetry of the data (Figure 3).

Figure 3. Boxplots of Square Roots of the Absolute Deviations From Group Medians of the Book Price Data.

On the other hand, the log transformation makes things rather more skewed (Figure 4). Worse, several points cannot even be drawn because of zeros in the deviation data. (R gave some warning messages.)
Tests of the Levene type can be accommodated well in a statistics course that includes one-way ANOVA, as many do. Though some bootstrap versions of this test appear to have a few advantages in the simulation study of Lim and Loh (1996), the original test still does quite well, especially if the sub-sample (i.e., group) sizes are not too small. "Small" for Lim and Loh was five, and our view is that students should be encouraged to avoid sample sizes smaller than 10. We note the choice of Levene-type (Brown-Forsythe) tests in Stata (Cleves 2000), using deviations from mean, median, and trimmed mean. The Minitab macro LEVENE.MAC carries out the Levene deviation-from-mean and deviation-from-median tests.

The computation of the deviations from the medians is potentially messy, but is certainly not difficult. Tools such as the Minitab macros DEVMEAN.MAC and DEVMED.MAC that we provide in Appendix 2 allow much of the tedium to be avoided. Furthermore, Tukey paired comparisons allow us to decide which groups are different, in addition to the decision that at least two groups have different variability.

7. Nonparametric Test Approaches

In courses where tests based on ranks have not been introduced, a "new" test that uses such principles is not appropriate. However, when students have already had some exposure to the ideas of using ranks in place of data, we can suggest the Fligner-Killeen tests.
Some elementary service courses include rank-based tests such as the Wilcoxon or Mann-Whitney tests, so that methods of this type could be considered. Madansky (1988) suggests two similar normal scores techniques under the title of the Fligner-Killeen tests. Hollander and Wolfe (1999) present a similar approach that arrives at somewhat different p-values under the name of the van der Waerden or normal scores method.

The Fligner-Killeen tests (as well as their cousins in Hollander and Wolfe) are based once again on the absolute deviations from group medians. Now, however, we want to pool all these deviations and rank them from smallest to largest. We then transform the ranks, labelled $i$, to scores

$$\text{score}(i) = \Phi^{-1} \left( 0.5 + \frac{0.5i}{n+1} \right),$$

where $n$ is the size of the total sample (i.e., the sum of the group sample sizes) and $\Phi^{-1}()$ is the inverse of the cumulative standard normal distribution. That is, if

$$1 - \alpha = \Phi(z) = \text{area under standard normal curve from } -\infty \text{ to } z,$$

then

$$z = \Phi^{-1}(1 - \alpha).$$

We can then compute a variance of the scores of all observations and compare this to the within-group variance of the scores using the Fisher $F$ test. See Madansky (1988, p. 65) for details, or consult the macro FKTEST.MAC in Appendix 2.

Using deviations from medians or means, we could also carry out the Kruskal-Wallis test, but note that the Kruskal-Wallis test assumes similar distribution shapes for each group.

While the Fligner-Killeen and similar tests are not particularly difficult to implement and use, it is our opinion that they introduce too many new concepts for appropriate use in an elementary statistics course. We have noted that rank-based tests are novel enough. The further complication of scores and then the distribution of a relatively complicated function of these scores is too much to introduce. Moreover, a test of homogeneity of the variances will not tell us where the differences lie. We will, however, note where such tests agree with the other methods we recommend.

8. Spread Versus Level

Having decided that the samples appear to be from populations with neither the same distributional shape nor the same variability, there is the possibility that variability in textbook prices is somehow related to price level. That is, we may be concerned that the variability is proportional to the level. Building on our transformations, we can plot spread versus level (or location). See, for example, Cleveland (1993, p. 50 ff). Such graphs almost always involve a (further) transformation of the data. Cleveland recommends that the square root of the absolute deviation from the median be used as the measure of spread. In the present example, we have prepared such graphs from both the raw and log data by computing the square roots of the absolute deviations from group medians and graphing them against the group medians. The Minitab macro SPRLEVGR.MAC prepares a fitted line plot with these data, which not only draws the scatterplot but adds a simple regression line whose slope shows whether spread is increasing or decreasing with level. We recommend presenting such a graph only after simple regression has been covered, and generally would do so only if there is a reason to do so, such as the case that prompted this paper.

For the textbook price example, plotted in Figure 5, the spread seems to be roughly constant with level, assuming that the square root transformation of the distance from the median is appropriate.
9. Choice and Illustration of Techniques

The techniques that we consider appropriate for teaching comparison of variability for samples drawn from populations that are possibly non-Gaussian are

- Multiple boxplots of absolute deviations from group medians and
- One-way ANOVA of absolute deviations from group medians (Levene median test).

We do not regard the Fligner-Killeen or other rank-based tests to be appropriate as a regular topic in an introductory course, but they could be shown to interested students. Similarly, if the course includes ANOVA and nonparametric statistics, the Kruskal-Wallis procedure could be similarly presented, including its application here to deviations from group medians, but we need to mention the assumption of similarly shaped distributions. We do not feel it appropriate to examine students on these topics, however.

In the process of preparing this paper, we recognized that transformations of data should be given more prominence, since they are part of so many statistical methods or these methods can be presented via transformations. We believe that it may be worthwhile to place more emphasis on transformations in the introductory service course, possibly linking coverage to material on functions in typical introductory mathematics courses where appropriate. However, such emphasis is only warranted if we have examples that
show the utility of transformations. Indeed, in the service course, each topic should be well-illustrated with practical examples.

For our textbook price data, all the methods suggest that variability in textbook price differs among faculties. First, the notched boxplots (Figure 2) show that groups a and d differ from groups b and c, but that all overlap the remaining four groups.

The one-way ANOVA for the Levene median $F$ test (Table 2) gives a $p$-value of just over .01 for the hypothesis of equal variances in all groups. The 95% confidence intervals for the group means in the Minitab ANOVA output show differences only for groups a and b and groups b and d. Tukey paired comparisons paint a similar picture.

Minitab (here we are using the Student Edition for Windows, Version 12) allows ANOVA to be carried out either by listing the individual variables or by providing a concatenated (or stacked) set of data in a single variable along with an index variable that specifies the group membership. This latter method allows Tukey and other comparisons to be computed. However, we note that the usual commands to produce the stacked variable with data create a simple numerical index and that this must be recoded to give the faculty labels. Such "how to" details related to software are a frequent source of student frustration and require careful classroom presentation.

### Table 2. Minitab Output for Levene Test for Textbook Price Data

| Source   | DF | SS   | MS   | F          | P        |
|----------|----|------|------|------------|----------|
| faculty  | 7  | 3649 | 521  | 2.68       | 0.011    |
| Error    | 287| 55778| 194  |            |          |
| Total    | 294| 59428|      |            |          |

| Level | N   | Mean | StDev | Individual 95% CIs For Mean Based on Pooled StDev |
|-------|-----|------|-------|---------------------------------------------------|
| a     | 52  | 23.75| 14.09 | (--------*--------)                               |
| b     | 47  | 13.81| 11.33 | (--------*-------)                                |
| c     | 42  | 16.34| 16.94 | (--------*-------)                                |
| d     | 28  | 24.05| 15.26 | (--------*-------)                                |
| e     | 26  | 20.61| 9.32  | (--------*-------)                                |
| f     | 55  | 17.63| 15.20 | (--------*-------)                                |
| g     | 15  | 18.45| 10.31 | (--------*-------)                                |
| h     | 30  | 19.91| 13.84 | (--------*-------)                                |

Pooled StDev = 13.94

Tukey’s pairwise comparisons

| Family error rate = 0.0500 |
| Individual error rate = 0.00264 |
| Critical value = 4.29 |

Intervals for (column level mean) - (row level mean)

|         | a   | b   | c   | d   | e   | f   | g   |
|---------|-----|-----|-----|-----|-----|-----|-----|
|         |     | 1.43|     |     |     |     |     |
| b       | 18.46|     |     |     |     |     |     |
| c       | -1.36| -11.51|     |     |     |     |     |
|         | 16.19| 6.45   |     |     |     |     |     |
| d       | -10.21| -20.33| -18.02|     |     |     |     |
|         | 9.62 | -0.14| 2.61   |     |     |     |     |
The Fligner-Killeen tests give *p*-values of 0.0117 and 0.0105 for the hypothesis of equal variance. The *p*-value of the normal scores test, as computed by StatXact 4 for Windows, is 0.0033 using an asymptotic approximation and 0.0026 using a Monte-Carlo estimate.

The Kruskal-Wallis test gave the output in Table 3 with a very small *p*-value for equality of the medians of the deviation data. (StatXact gave equivalent results.) The output suggests that the mean ranks of groups a and d are the most elevated from the group mean ranking, while those of groups b and c are the most reduced from this mean ranking. We should, in using this procedure, consider whether the boxplots of the deviation data (Figure 2) allow us to accept similarly shaped distributions for all groups, as there are clearly some differences in symmetry and outliers. This may account for the small *p*-value in comparison to the Levene and Fligner-Killeen approaches, which are quite similar to each other.

### Table 3. Minitab Output for Kruskal-Wallis Test on Absolute Deviations From Group Medians for Textbook Price Data

| idx | N  | Median | Ave Rank | Z  |
|-----|----|--------|----------|----|
| 1   | 52 | 25.00  | 179.3    | 2.92|
| 2   | 47 | 11.00  | 115.1    | -2.89|
| 3   | 42 | 10.73  | 120.4    | -2.27|
| 4   | 28 | 22.03  | 178.1    | 1.96|
| 5   | 26 | 21.80  | 170.6    | 1.42|
| 6   | 55 | 14.30  | 135.3    | -1.22|
| 7   | 15 | 22.00  | 156.3    | 0.39|
| 8   | 30 | 14.00  | 155.4    | 0.50|
| Overall | 295 |        | 148.0    |    |

H = 25.32  DF = 7  P = 0.001  
H = 25.33  DF = 7  P = 0.001 (adjusted for ties)

### 10. Conclusion

Given the availability of quite modest computational tools, we believe that techniques for comparing the variability of *k* > 2 samples can be taught in an elementary statistics course. If ANOVA is not part of the course,
then multiple notched boxplots based on absolute deviations from group medians are simple and effective. One-way ANOVA on these data, with the addition of Tukey paired comparisons or the graphical display of confidence intervals for the means, allows a reasonable test along with additional insight as to the origin of the non-homogeneity of the variability.

As we have noted, the theme of transformation of data is one that is important in statistics:

- To allow better analysis of data properties, such as variability in the present example;
- To permit graphs to be drawn that show such properties; or
- To cause the distribution of the transformed data to be such that available tools can be used to analyze the data.

While students in introductory courses are unlikely to appreciate this generality and the importance of transformations, those with reasonable mathematical skills -- who we caution are a minority in our business statistics classes -- could benefit from carrying out an investigation of transformations on a dataset similar to the example presented here. Given that introductory courses such as our own present the normal probability plot as well as histograms, boxplots, and stem-and-leaf diagrams, and that these tools are readily available within software such as Minitab, this could make a good student project that is challenging, but doable. If students are not self-starters, a case study approach could be used where there is a structured set of exercises, possibly even using pre-written scripts to prepare graphs.

**Acknowledgments**

We are grateful for personal or e-mail discussions with a number of colleagues while refining this paper: Paul Velleman, Richard Goldstein, Raoul Lepage, Colin Chalmers, Alan Hutson, Terry Flynn, Tim Auton, and John Haywood. The original class project that motivated this paper was carried out in collaboration with students Christopher Charron and Tatiana Botchoukova.

**Appendices**

- **Appendix 1: Minitab Script to Load the Book Price Data**
- **Appendix 2: Minitab Macros to Perform Some of the Calculations**

**References**

Aczel, A. (1996), *Complete Business Statistics* (3rd ed.), Chicago: Richard D. Irwin.

Bradley, J. V. (1968), *Distribution-Free Statistical Tests*, Englewood Cliffs, NJ: Prentice-Hall.

Brown, M. B., and Forsythe A. B. (1974), "Robust Tests for the Equality of Variances," *Journal of the American Statistical Association*, 69, 364-387; Correction (1974), 69, 840.

Cleveland, W. S. (1993), *Visualizing Data*, Summit, NJ: Hobart Press.

Cleves, M. (2000), "Robust Tests for the Equality of Variances Update to Stata 6," *Stata Technical Bulletin*, STB-53, January, 17-18.
Conover, W. J., Johnson, M. E., and Johnson, M. M. (1981), "A Comparative Study of Tests for Homogeneity of Variances, With Applications to Outer Continental Shelf Bidding Data," *Technometrics*, 23(4), 351-361.

Hamilton, L. C. (1992), *Regression With Graphics: A Second Course in Applied Statistics*, Belmont, CA: Wadsworth.

Hines, W. G. S., and O'Hara Hines, R. J. (2000), "Increased Power With Modified Forms of the Levene (Med) Test for Heterogeneity of Variance," *Biometrics*, 56, 451-454.

Hoel, P. G. (1971), *Introduction to Mathematical Statistics*, New York: Wiley.

Hollander, M., and Wolfe, D. A. (1999), *Nonparametric Statistical Methods* (2nd ed.), New York: Wiley.

Levene, H. (1960), "Robust Tests for Equality of Variances," in *Contributions to Probability and Statistics*, ed. I. Olkin, Palo Alto, CA: Stanford University Press, pp. 278-292.

Lim T.-S., and Loh, W.-Y. (1996), "A Comparison of Tests of Equality of Variances," *Computational Statistics and Data Analysis*, 22(3), 287-301.

Loh, W.-Y. (1987), "Some Modifications of Levene's Test of Variance Homogeneity," *Journal of Statistical Computation and Simulation*, 28, 213-226.

Madansky, A. (1988), *Prescriptions for Working Statisticians*, New York: Springer-Verlag.

McGill, R., Tukey, J. W., and Larsen, W. A. (1978), "Variations of Boxplots," *The American Statistician*, 32, 12-16.

Neter, J., Wasserman, W., and Kutner, M. H. (1990), *Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs* (3rd ed.), Homewood, IL: Irwin.

Simon, J. L., and Holmes, A. (1969), "A Really New Way to Teach Probability and Statistics," *The Mathematics Teacher*, LXII, April, 283-288.

Snedecor, G. W., and Cochran, W. G. (1967), *Statistical Methods* (6th ed.), Ames, IA: The Iowa State University Press.

Velleman, P. F., and Hoaglin, D. C. (1981), *Applications, Basics and Computing of Exploratory Data Analysis*, Belmont, CA: Duxbury.

Lorraine Garrett
8.5 Range Road
Ottawa, Ontario, K1N 8J3, Canada
lgarrett@omnilogic.com

John C. Nash
Faculty of Administration
University of Ottawa
136 Jean-Jacques Lussier Private
Ottawa, Ontario, K1N 6N5, Canada
jcnash@uottawa.ca
