Using Van Valen's Procedure In Business Research To Assess Consistent Differences In Multidimensional Variability In Two Or More Groups

Mark L. Berenson, Montclair State University
Kimberly Killmer Hollister, Montclair State University

ABSTRACT

Much business research involves comparisons in two or more groups on many dimensions. This paper primarily focuses on demonstrating and providing guidance as to how researchers should approach a multivariate analysis in the comparison of sets of corresponding characteristics in two or more independent groups. In particular, this paper demonstrates the utility of a simple but not widely known procedure developed by Van Valen (1978) that should be employed to test for the significance of differences in overall variability in the sets of corresponding characteristics in two or more groups, a test that enjoys much statistical power in detecting significant subtle group differences when the set of characteristics in one group consistently demonstrates greater variability than the corresponding set of characteristics in the other group(s).

INTRODUCTION

*M*uch business research involves comparisons in two or more groups on many dimensions. All too often, however, researchers attempt an analysis of differences in the groups by assessing one characteristic at a time and fail to take advantage of a more global multivariate assessment in the groups that encompass the joint associations among each group's multi characteristics.

Although assessment of differences in two or more groups usually involves comparisons of central tendencies (i.e., the means of the set of corresponding characteristics in each group), there are also occasions when interest primarily concerns comparisons of patterns of dispersion (i.e., the variability around the means of the set of corresponding characteristics in each group). Sometimes, such analysis is an "end in itself" because the researcher is primarily concerned with differences in variability in two or more groups; in other situations, it is a "means to an end." An investigation of differences in variability may be conducted in an attempt to uncover some significant distinction in two or more multivariate distributions that show no significant differences in central tendencies or a researcher may first decide to assess potential differences in overall variability in the groups before embarking on an analysis of possible differences in their central tendencies.

This paper primarily focuses on demonstrating and providing guidance as to how researchers should approach a multivariate analysis in the comparison of sets of corresponding characteristics in two or more independent groups. In particular, based on the findings of Manly (1986a, 1986b, 1988), when assessing significant differences in overall variability in the corresponding multiple characteristics between two groups, the simple but not widely known Van Valen (1978) procedure should be used if the researcher can either *a priori* hypothesize that the variability in the corresponding set of characteristics is consistently greater in one of the groups or, following a exploratory investigation of the data, the researcher knows this to be true. In such situations the Van Valen procedure enjoys high power to detect such significant differences in overall variability. On the other hand, if a researcher has no reason to make such an assertion *a priori,* or, if through a preliminary descriptive analysis the

researcher finds one or more characteristics for which the differences in the variability is greater in one group while observing that for at least another characteristic the differences in the variability is greater in the other group, the Levene procedure (1960) should be used for detecting significant differences in overall variability.
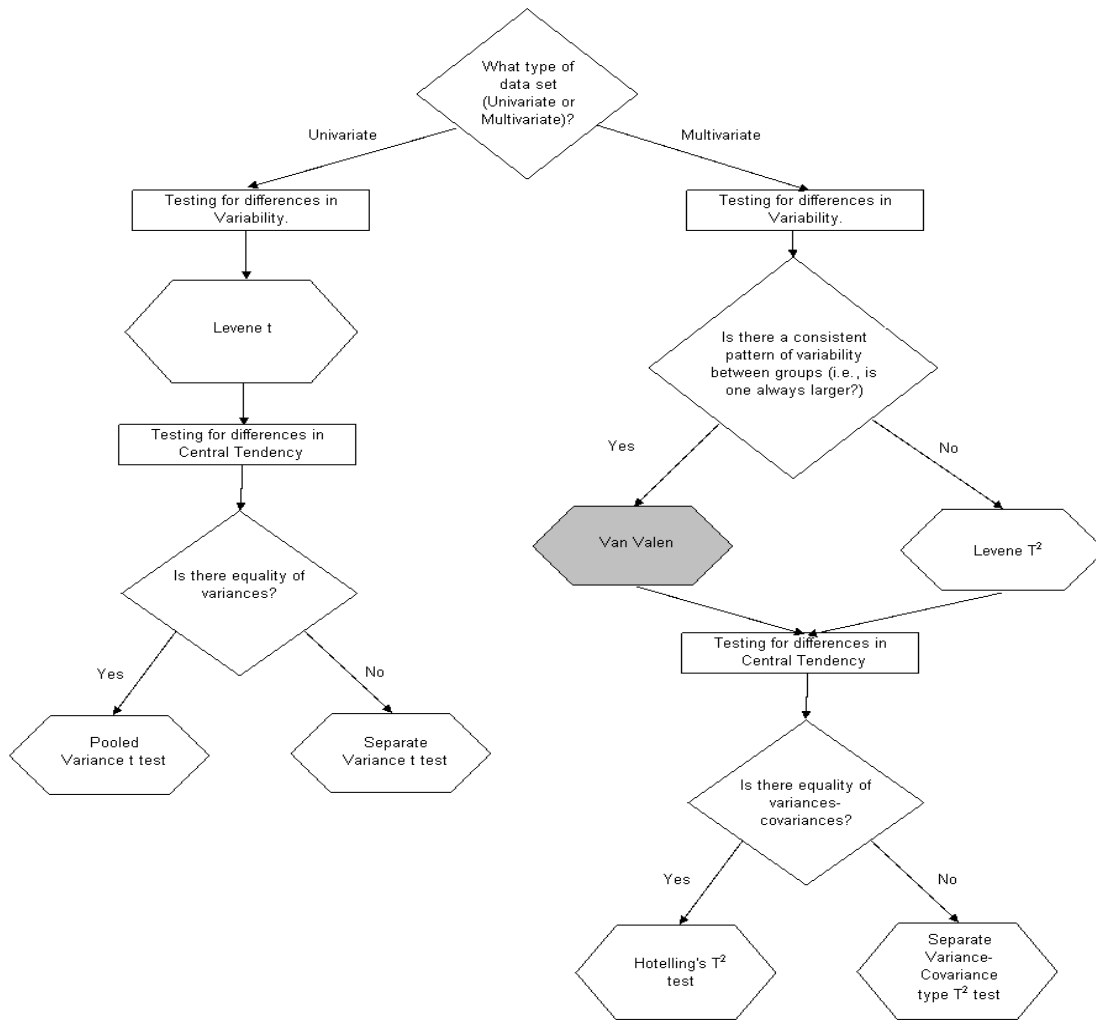
Figure 1: Flowchart of Appropriate Test Selection for Two Independent Groups of Data

BACKGROUND

When assessing differences in two or more independent groups it is imperative that the researcher employ the correct tools for the situation at hand, i.e., those which appropriately correspond to *a priori* hypotheses regarding particular parameters and/or those which are appropriate for the particular data sets being studied. Failure to do so results in a less than optimal analysis and reduces the opportunity for detecting significant differences when they truly exist.

Figure 1 provides a guide for appropriate test selection for situations involving two independent groups of numeric data. (The corresponding more general "distribution-free" or nonparametric procedures (Daniel, 1990) are not considered here). When comparing two independent groups of numerical data on but one dimension (i.e., a

univariate analysis), the researcher has some choice in test selection, depending on the particular hypothesis of interest and the shape of the underlying population distributions from which the samples are obtained. If the objective is to assess differences in central tendency, either the pooled *t* or separate variance *t* tests can be employed to compare differences in means when the assumption of at least approximate normality in the two populations from which the respective samples were drawn can be met, or when a normalizing transformation on the sampled data can be made (Berenson et al., 1983). If the objective is to assess differences in dispersion in two groups, Manly (1986b) recommends that the Levene procedure should be selected over Fisher's *F* test involving the ratio of two sample variances because the *F* test is unreliable. It confounds lack of normality in the data with lack of homogeneity of the variances. The Levene test is robust to lack of normality in the data.

As displayed in Figure 1 for the comparable multivariate situations containing numeric data, either Hotelling's T^2 test or a "separate variance-covariance T^2 type" adjustment are used for assessing differences in means of the corresponding sets of characteristics between the two groups (Hotelling, 1931, 1951; Kim, 1992; Yao, 1965; Lattin et al, 2003) and an extension of the aforementioned Levene procedure can be employed for comparing differences in the variances in the corresponding sets of characteristics between two groups (Manly, 1986b). When the differences are consistent, however, and the entire multivariable set indicates more spread in one of the groups than in the other, the Van Valen procedure is more powerful in capturing what the Levene procedure misses as just a set of subtle differences. And it is this latter situation that is the focus of this paper.

THE PROBLEM AND ITS IMPORTANCE

Although Hotelling's T^2 test – the appropriate procedure for comparing differences in means in the corresponding sets of characteristics between two independent groups – has been available for three-quarters of a century, all too often, however, when faced with two samples of multidimensional data researchers inappropriately employ a set of univariate *t* tests. But since there is at least some correlation among the multidimensional set of characteristics being studied, the multiple set of univariate *t* tests both violate and liberalize the stated α level of significance and lead the researcher to erroneously conclude that a significance difference exists between the two groups for at least one of the characteristics when in reality such differences do not surface.

As an example, when evaluating differences in the means of a characteristic from two independent groups, the level of significance α, traditionally selected as .05, represents the probability of a "type I" error, i.e., drawing the erroneous conclusion that the means in the two groups differ significantly when in reality they do not. Its complement, $1-\alpha$, traditionally .95, represents the probability of a correct decision, i.e., asserting there is no evidence of a significant difference in the two group means when in actuality there is none. However, when multiple characteristics are observed in the two independent groups and the researcher makes individual tests on the same data sets the probability of a "type I" error increases. Suppose, for example, four characteristics on the same sample subjects were completely independent of each other and there were no real differences in the means in any of the corresponding characteristics across the two groups. When performing individual *t* tests on each of the four characteristics, the probability of making all four correct decisions is:

$$(1-\alpha)\times(1-\alpha)\times(1-\alpha)\times(1-\alpha)=(1-\alpha)^4=(.95)^4=.8145$$

Thus, when there are no real differences in the means of the corresponding characteristics, the probability of making at least one incorrect decision when performing four individual tests is:

$$1-(1-\alpha)^4=1-(.95)^4=.1855$$

It must be realized that if a researcher does not account for and "adjust" the individual α levels of significance in the analysis, erroneous conclusions may be drawn and the overall study findings may not be optimal. If the overall α level of significance desired is the traditional .05, the adjustment suggested here is to make the four

individual tests with an $\alpha/4 = .0125$ level of significance so that the probability of making at least one incorrect decision is:

$$1-(1-\alpha)^4 = 1-(.9875)^4 = .0491$$

Of course, in a multivariate analysis we do not expect the characteristics to be completely independent; however the suggested adjustment still substantially reduces the likelihood of an incorrect conclusion.

The Hotelling T^2 test, however, enables the researcher to draw a global conclusion about differences in means in the corresponding sets of characteristics between the two groups while preserving the stated α level of significance and, if deemed significant, allows for *a posteriori* comparisons to determine which of the individual means differ over the two groups or which combinations of means differ over the two groups.

Like the univariate *t* test, Hotelling's T^2 procedure is deemed robust to violations in the assumptions of (multivariate) normality and equality of the variances-covariances if the two multidimensional samples are coming from populations that are not too skewed, if the sample sizes are not too small and if the sample sizes are equal or approximately the same (Manly, 1986b, Lattin et al., 2003). A failure to reject the null hypothesis of equality of means in the corresponding sets of characteristics between the two populations typically leads to a belief that if there is no evidence of a difference in central tendency the two populations are basically the same.

But multivariate distributions can differ in variability or shape, not just central tendency. Thus, the question of interest sometimes is not whether there is a difference in the means in the corresponding sets of characteristics between the two groups but rather is there evidence that the set of characteristics in one of the groups shows consistent and significantly greater spread around their respective means than do the corresponding characteristics in the other group. It is in such situations that the little known Van Valen's procedure provides optimal opportunities for uncovering differences in variability. It is important to note that the Van Valen procedure can be used whether or not there are differences in central tendency.

THE STUDY: COMPARING HOUSING CHARACTERISTICS IN DENVILLE AND SPARTA – TWO NEW JERSEY COMMUNITIES

To illustrate how Van Valen's procedure can be used to study differences in overall variability between two groups we collected real estate data on single-family houses listed for sale in two New Jersey towns, Denville and Sparta Township, using the public multiple listing service site for Northwest New Jersey (GSMLS, 2007). From the list of all properties for sale in the two towns on July 13, 2007, a random sample of 24 homes was selected from Denville and a random sample of 27 homes was selected from Sparta.

To address the multivariate nature of the real estate market, data on property size (in acres), total number of rooms, list price, and house age (in years) were collected for each of the 51 properties included in the study.

The two towns have many similarities but we would expect the real estate market in Sparta to be more variable than the market in Denville. We examine some characteristic differences in geography, demographics, economics, and transportation between the two towns.

	Denville, New Jersey	Sparta Twp., New Jersey
Population	15,825	18,080
Housing Units	6,178	6,590
Land Area (Sq miles)	12.11	37.39
Population Density (per sq. mile)	1,307.1	483.5
Housing Density (per sq. mile)	510.3	176.2

Table 1: Overall Characteristics of Denville and Sparta (Sperling 2007)

Table 1 shows overall characteristics of both towns. Although Sparta and Denville do not differ substantially in terms of their overall population and housing units, there is a large difference in the overall land area. As a result, the population density in Denville is 2.70 times that of Sparta and the housing density is 2.90 times that of Sparta. The difference in land area between the two towns and the resulting housing density differences contribute to the increased variability in the housing market in Sparta.

	Denville, New Jersey	Sparta Twp., New Jersey
Income per Capita	$44,513	$42,580
Household Income	$89,623	$102,925
Commute Time, minutes	31	42.3
Commute Mode Auto - alone	85.5%	82.8%
Cost of Living Index	158.5	148.2
Median Population Age	41.1	38.6
Household Size	2.6	2.9

Table 2: Income, Age, Household, and Other Data for Denville and Sparta (Sperling, 2007)

Table 2 shows income and other household characteristics for Denville and Sparta. Although household income in Sparta is higher than that in Denville, the income per capita is actually higher in Denville because of its lower average household size. Another notable difference between the two towns is the average commute time; average commute time in Sparta is 36.5% longer than in Denville. And it is interesting to note that the average commute in both towns is longer than the US average of 27.4 minutes (Sperling, 2007). A slightly higher percentage of Denville residents commute by automobile (85.5% in Denville versus 82.8% in Sparta). Also, the median population age in Denville is higher than that of Sparta, as is the cost of living higher in Denville than in Sparta. The longer commute time in Sparta contributes to increased variability in its housing market.

	Denville, New Jersey	Sparta Twp., New Jersey
Median Home Age	42	33
Percentage of Homes Owned	82.8%	85.0%
Housing Vacant	3.0%	5.5%
Property Tax Rate per $1,000 assessed value	$20.12	$26.41

Table 3: Home Ownership and Occupancy Data for Denville and Sparta (Sperling 2007)

Home ownership and occupancy data for Denville and Sparta are shown in Table 3. The median age of homes in Denville is 9 years older than that of Sparta. Rates of home ownership are similar in the two towns; however, the proportion of houses vacant in Sparta is higher than that of Denville. The property tax rate in Sparta is also higher than the rate in Denville. The higher proportion of vacant homes in Sparta and a lower median home age (resulting from new developments) contribute to the increased variability of the housing market in Sparta. Geographically, Denville has better access to highways and mass transportation and is approximately 20 minutes closer to New York City (a destination for many commuters in the Northern New Jersey area).

	Denville, New Jersey	Sparta Twp., New Jersey
School Expenditure	$8,382	$8,936
Pupil/Teacher Ratio	16.9	14.9
% Going to 4yr College	67.4%	73.2%
% Going to 2yr College	23.1%	19.9%
High School Grads	92.5%	94.6%
SAT Average	1571	1586

Table 4: Education Data for Denville and Sparta (Sperling 2007, NJDOE 2005-2006)

Home	Acres	Rms	Price	Age	Town	Zacres	Zrooms	Zprice	Zage	AbsdZac	AbsdZro	AbsdZpr	AbsdZag	dVanvalen
1	0.15	6	299.90	61	D	-0.578	-0.959	-1.393	0.517	0.121	0.644	1.218	0.202	1.398
2	0.22	7	319.90	65	D	-0.294	-0.315	-1.150	0.719	0.162	0.000	0.975	0.404	1.067
3	0.30	6	327.50	67	D	0.029	-0.959	-1.057	0.819	0.486	0.644	0.882	0.505	1.297
4	0.10	5	365.00	77	D	-0.780	-1.602	-0.600	1.324	0.324	1.287	0.425	1.009	1.721
5	0.17	5	368.00	57	D	-0.497	-1.602	-0.563	0.315	0.040	1.287	0.389	0.000	1.345
6	0.12	7	375.00	22	D	-0.699	-0.315	-0.478	-1.452	0.243	0.000	0.303	1.767	1.809
7	0.17	7	379.00	44	D	-0.497	-0.315	-0.429	-0.341	0.040	0.000	0.255	0.656	0.705
8	0.17	7	394.90	44	D	-0.497	-0.315	-0.236	-0.341	0.040	0.000	0.061	0.656	0.660
9	0.21	9	395.00	50	D	-0.335	0.972	-0.234	-0.039	0.121	1.287	0.060	0.353	1.341
10	0.23	6	399.00	67	D	-0.254	-0.959	-0.186	0.819	0.202	0.644	0.011	0.505	0.843
11	0.73	7	399.00	69	D	1.770	-0.315	-0.186	0.920	2.226	0.000	0.011	0.606	2.307
12	0.17	6	399.90	57	D	-0.497	-0.959	-0.175	0.315	0.040	0.644	0.000	0.000	0.645
13	0.17	7	399.90	29	D	-0.497	-0.315	-0.175	-1.099	0.040	0.000	0.000	1.413	1.414
14	0.19	7	399.90	65	D	-0.416	-0.315	-0.175	0.719	0.040	0.000	0.000	0.404	0.406
15	0.20	7	399.99	57	D	-0.375	-0.315	-0.174	0.315	0.081	0.000	0.001	0.000	0.081
16	0.11	8	409.90	20	D	-0.740	0.328	-0.053	-1.553	0.283	0.643	0.122	1.868	1.999
17	0.14	10	420.00	31	D	-0.618	1.615	0.070	-0.998	0.162	1.930	0.245	1.312	2.353
18	0.17	7	424.90	57	D	-0.497	-0.315	0.130	0.315	0.040	0.000	0.305	0.000	0.307
19	0.38	6	429.00	77	D	0.353	-0.959	0.180	1.324	0.809	0.644	0.355	1.009	1.488
20	0.17	8	469.00	39	D	-0.497	0.328	0.667	-0.594	0.040	0.643	0.842	0.909	1.396
21	0.25	10	499.90	45	D	-0.173	1.615	1.044	-0.291	0.283	1.930	1.218	0.606	2.379
22	0.50	8	499.90	50	D	0.839	0.328	1.044	-0.039	1.295	0.643	1.218	0.353	1.924
23	0.92	8	499.90	12	D	2.538	0.328	1.044	-1.957	2.995	0.643	1.218	2.271	4.003
24	0.62	8	499.99	67	D	1.324	0.328	1.045	0.819	1.781	0.643	1.220	0.505	2.308
25	0.16	5	259.00	58	S	-0.537	-1.602	-1.892	0.365	0.364	1.931	1.717	0.202	2.617
26	0.27	7	315.00	69	S	-0.092	-0.315	-1.209	0.920	0.081	0.644	1.034	0.757	1.437
27	0.11	6	325.00	78	S	-0.740	-0.959	-1.087	1.375	0.567	1.287	0.913	1.211	2.068
28	0.33	6	329.00	81	S	0.151	-0.959	-1.039	1.526	0.324	1.287	0.864	1.363	2.089
29	0.18	6	340.00	43	S	-0.456	-0.959	-0.905	-0.392	0.283	1.287	0.730	0.555	1.605
30	0.25	7	340.00	32	S	-0.173	-0.315	-0.905	-0.947	0.000	0.644	0.730	1.110	1.476
31	0.30	5	349.90	48	S	0.029	-1.602	-0.784	-0.140	0.202	1.931	0.609	0.303	2.057
32	0.17	8	355.00	50	S	-0.497	0.328	-0.722	-0.039	0.324	0.000	0.547	0.202	0.667
33	0.42	10	359.00	47	S	0.515	1.615	-0.673	-0.190	0.688	1.287	0.498	0.353	1.582
34	0.15	6	368.50	70	S	-0.578	-0.959	-0.557	0.971	0.405	1.287	0.383	0.808	1.618
35	0.18	7	369.90	56	S	-0.456	-0.315	-0.540	0.264	0.283	0.644	0.366	0.101	0.799
36	0.42	9	387.00	65	S	0.515	0.972	-0.332	0.719	0.688	0.643	0.157	0.555	1.105
37	0.60	11	389.90	40	S	1.243	2.259	-0.297	-0.543	1.416	1.930	0.122	0.707	2.499
38	0.21	9	399.90	34	S	-0.335	0.972	-0.175	-0.846	0.162	0.643	0.000	1.009	1.208
39	0.22	9	419.90	79	S	-0.294	0.972	0.069	1.425	0.121	0.643	0.244	1.262	1.442
40	0.14	8	424.50	57	S	-0.618	0.328	0.125	0.315	0.445	0.000	0.300	0.151	0.558
41	1.57	9	425.00	31	S	5.169	0.972	0.131	-0.998	5.342	0.643	0.306	1.161	5.513
42	0.33	8	439.90	23	S	0.151	0.328	0.313	-1.401	0.324	0.000	0.487	1.565	1.671
43	0.18	8	449.00	75	S	-0.456	0.328	0.424	1.223	0.283	0.000	0.598	1.060	1.245
44	0.21	7	449.90	50	S	-0.335	-0.315	0.435	-0.039	0.162	0.644	0.609	0.202	0.923
45	0.20	6	469.90	55	S	-0.375	-0.959	0.678	0.214	0.202	1.287	0.853	0.050	1.558
46	0.27	11	489.90	29	S	-0.092	2.259	0.922	-1.099	0.081	1.930	1.097	1.262	2.555
47	0.13	7	499.00	54	S	-0.659	-0.315	1.033	0.163	0.486	0.644	1.208	0.000	1.452
48	0.25	8	519.00	2	S	-0.173	0.328	1.276	-2.461	0.000	0.000	1.451	2.625	2.999
49	0.30	10	549.00	75	S	0.029	1.615	1.642	1.223	0.202	1.287	1.817	1.060	2.474
50	0.29	7	579.90	58	S	-0.011	-0.315	2.019	0.365	0.162	0.644	2.193	0.202	2.300
51	0.53	10	749.90	1	S	0.960	1.615	4.090	-2.512	1.133	1.287	4.265	2.675	5.318

Table 5: Data on Property Size (in Acres), Number of Rooms, List Price and Age (in Years) for 24 Single-Family Homes in Denville and 27 Single-Family Homes in Sparta.

Education data for Denville and Sparta are depicted in Table 4. Overall, the two towns do not differ much for all factors listed; however Sparta's education data is marginally better than that of Denville. Sparta schools have a slightly higher percentage of students attending two or four-year schools (93.1% in Sparta compared with 90.5% in Denville). The Average SAT score in Sparta is also slightly higher than in Denville and Sparta boasts a slightly higher proportion of students graduating from high school.   Sparta also has higher per student expenditures and reports a lower student to teacher ratio.

The real estate market has a high degree of variability. Investors/potential home buyers often narrow their searches to one or two towns based on overall characteristics such as location, education, and demographics. Many do not consider the level of variability within their reduced set of choices or within each of the towns under consideration. Once narrowing their decision geographically the next decisive factor is price of homes and their features. A buyer who understands the variability of the market within an area is at an advantage to other buyers. The next section applies Van Valen's procedure to compare variability in housing characteristics between the two towns.

APPLYING VAN VALEN'S PROCEDURE TO COMPARE VARIABILITY IN HOUSING CHARACTERISTICS IN TWO COMMUNITIES

Table 5 displays the real estate data collected in Denville (houses numbered 1 through 24 in column 1) and in Sparta (houses numbered 25 through 51 in column 1). In addition to the measured numerical characteristics on property size (in acres), number of rooms, list price and house age (columns 2-5), the corresponding standardized values for these characteristics are also displayed (columns 7-10), as are the corresponding sets of   *absolute deviations* between the standardized values and their respective standardized medians (columns 11-14). Based on the latter four columns, the final column of the table provides the list of Van Valen "distance" statistics *d* necessary for applying the Van Valen procedure.

The means and standard deviations for each of the four measured housing characteristics in both towns as well as in the combined study group are summarized in Table 6. The data show that for all four characteristics under consideration, the standard deviation in Sparta exceeds that of Denville, indicating that Van Valen's procedure should be used to study overall differences in variability in the two communities.

Summary Statistics	Acres	Rooms	Price	Age
Denville (24 Houses)				
• **Mean**	0.273	7.167	407.266	51.208
• **Standard Deviation**	0.211	1.308	55.345	18.137
Sparta (27 Houses)				
• **Mean**	0.310	7.778	420.441	50.370
• **Standard Deviation**	0.278	1.717	100.786	21.528
Combined (51 Houses)				
• **Mean**	0.293	7.490	414.241	50.765
• **Standard Deviation**	0.247	1.554	82.069	19.812

Table 6: Means and Standard Deviations for Four Housing Characteristics in Denville and Sparta.

Table 7 provides a listing of descriptive summary measures for the four standardized housing characteristics in the two towns.

Summary Statistics	Zacres	Zrooms	Zprice	Zage
Denville (24 Houses)				
• **Mean**	-0.079	-0.208	-0.085	0.022
• **Median**	-0.457	-0.315	-0.175	0.315
• **Standard Deviation**	0.852	0.842	0.674	0.916
Sparta (27 Houses)				
• **Mean**	0.070	0.185	0.076	-0.020
• **Median**	-0.173	0.328	-0.175	0.163
• **Standard Deviation**	1.127	1.105	1.228	1.087

Table 7: Descriptive Summary Measures for Four Standardized Housing Characteristics in Denville and in Sparta.

The question that must be addressed is whether or not there is any evidence of a significant difference in the overall variability in the corresponding four housing characteristics between properties listed for sale in Denville and those listed for sale in Sparta. Given that the descriptive summary data in Tables 6 and 7 support the *a priori* conjecture that if the null hypothesis of equality of variances for the set of four characteristics between Denville and Sparta is rejected it would lead to the conclusion that there is evidence of a significantly consistent pattern of greater variability in the set of four characteristics in Sparta. The Van Valen procedure is most appropriate to address this question.

Using a step-by-step approach for developing Van Valen's procedure we note:

1. The data from a sample of n_1 = 24 properties listed for sale in Denville containing a set of measurements on four numerical characteristics (acres, rooms, price and age) and a sample of size n_2 = 27 properties listed for sale in Sparta containing a corresponding set of measurements on the same four characteristics are presented in columns 2 through 5 in Table 5.

2. The means and the standard deviations for each of the four characteristics from the sample of n_1 = 24 properties in Denville and from the sample of n_2 = 27 properties in Sparta are displayed in Table 6. The overall means and overall standard deviations for each of the four characteristics based on the combined study group of 51 properties are also displayed in Table 6. For each of the four characteristics, the original measurements are standardized by obtaining the difference between each measurement and its particular characteristic's overall mean and dividing this result by the particular characteristic's overall standard deviation. The results are displayed in columns 7 through 10 of Table 5.

    To demonstrate the calculation of the standardized value for house number 1 in Denville with respect to property size (in acres), take the difference between 0.15 (i.e., the entry for house number 1 in column 2 of Table 5) and 0.293 (i.e., the overall mean for acres listed in Table 6) and divide this result by 0.247 (i.e., the overall standard deviation for acres listed in Table 6). The resulting standardized value of the measurement of the property size in acres for house number 1 is – 0.578 and this is listed in column 7 of Table 5.

3. From the n_1 = 24 standardized measurements in Denville, we then obtain standardized medians for each of the four characteristics. Similarly, for the n_2 = 27 standardized measurements in Sparta, we obtain the standardized medians for each of the corresponding four characteristics. These are presented in Table 7.

4. For each of the four characteristics in Denville, a set of n_1 = 24 *absolute deviations* between the standardized measurements in columns 7 through 10 of Table 5 and their respective standardized medians from Table 7 are obtained. Similarly, for each of the corresponding four characteristics in Sparta, a set of n_2 = 27 *absolute deviations* between the standardized measurements in columns 7 through 10 of Table 5

and their respective standardized medians listed in Table 7 are obtained. The results are listed in columns 11 through 14 of Table 5.

To demonstrate the calculation of the *absolute deviation* of a standardized measurement and its particular group standardized median for house number 1 in Denville with respect to property size (in acres), take the absolute difference between –0.578 (i.e., the value computed in step 3 and presented in column 7 of Table 5) and –0.457 (i.e., the standardized median for acres in Denville displayed in Table 7). The result, 0.121, is displayed in column 11 of Table 5.

5. For the $n_1 = 24$ properties in Denville, Van Valen's "distance" statistic d is computed by taking the square root of the summation of the squared *absolute deviations* over all four characteristics. That is, from columns 11 through 14 in Table 5, for each property in Denville the listed *absolute deviation* values are squared, summed, and the square root of the result is then obtained. This same process is repeated for each of the $n_2 = 27$ properties in Sparta. The "distance" statistic d is listed in column 15 of Table 5 for all 51 properties.

    To demonstrate the calculation of the "distance" statistic d for house number 1 in Denville, we square the four *absolute deviation* values shown in columns 11 through 14 and sum the results. The "distance" statistic d is the square root of these results. Thus, for house number 1 in Denville,

    $$d = \sqrt{(0.121)^2 + (0.644)^2 + (1.218)^2 + (0.202)^2} = 1.398.$$

6. Using the Van Valen "distance" statistic d as newly derived "measurements," the two sample group means $\overline{d}_1$ and $\overline{d}_2$ and the two sample group variances S_1^2 and S_2^2 are obtained. The results are displayed in the Excel printout of Table 8.

7. To test the null hypothesis of homogeneity of spread in the four housing characteristics in Denville and Sparta against the alternative that the set of four variances in Sparta show consistent and significantly greater variation than the corresponding set of four variances in Denville a one-tailed pooled-variance t test is performed. The results are shown in Table 8.

8. Using a level of significance of .05 (one-tailed), the pooled-variance t test indicates a significant difference in overall variability in the two groups. The t statistic is -1.68045 and the p-value is .0496. The variances in the set of four housing characteristics in Sparta are consistently and significantly larger than the variances in the corresponding set of housing characteristics in Denville. Thus, there is greater overall variability in Sparta with respect to acres, number of rooms, list price and age of single-family homes than there is in Denville -- even though, as will be observed in the next section, the means in the sets of four housing characteristics between the two New Jersey communities are not significantly different.

Van Valen procedure t test results

	Denville	*Sparta*
Mean	1.466513	1.957089
Variance	0.741563	1.38474
Observations	24	27
Pooled Variance	1.082841	
Hypothesized Mean Difference	0	
Df	49	
t Stat	-1.68045	
p value of Van Valen procedure	0.049617	
t Critical one-tail	-1.676551	

Table 8: Van Valen's Procedure for Comparing Overall Variability
between housing characteristics in Denville and Sparta.

OTHER POSSIBLE ANALYSES OF THE DENVILLE AND SPARTA HOUSING CHARACTERISTICS

Differences in the Means: A typical concern of most researchers is whether or not there is evidence of a difference in the means of the set of characteristics studied in both groups. To test the hypothesis of no difference in the means in the set of four housing characteristics observed in Denville and Sparta, either Hotelling's T^2 test or a "separate variance-covariance T^2 type" adjustment can be used (Kim, 1992, Yao, 1965). Hotelling's T^2 test is robust to violations in multivariate normality and to moderate violations in homogeneity of variances-covariances in the set of characteristics being studied, particularly when the two group samples sizes are equal or nearly equal (Manly, 1986b; Lattin et al., 2003). Thus Hotelling's T^2 test could be used to compare the means of the set of housing characteristics after concluding, via Van Valen's procedure, that there are differences in overall variability.

Comparing the means of the set of four housing characteristics, Hotelling's T^2 test fails to reject the null hypothesis at the .05 level of significance. The *p*-value is .717. No differences between the corresponding sets of means are observed. If such comparisons were the only question of interest, the researcher would miss the opportunity to learn via the Van Valen procedure, that there are significant differences between the two communities in terms of overall variability in the set of housing characteristics. Interestingly, the application of the four univariate *t* tests, whether inappropriately unadjusted or appropriately adjusted for the correct level of α, yields the same conclusions – no evidence of significant differences in the means for property size (in acres), number of rooms, list price or age (in years) – in the two communities. The respective *p*-values are .602, .163, .572 and .882. Although not observed here, Manly (1986b) points out that a cumulative or overall effect could be uncovered via the Hotelling's T^2 test with its initially correct level of α but that individual tests using the adjusted level of α may not lead to such conclusions for the particular characteristics being compared and, on occasion, the reverse will occur – the individual test will declare a significant difference in the means of a particular characteristic while the difference is "washed out" in the overall, global Hotelling's T^2 test approach.

Differences in the Variances: As observed in Figure 1, there are two recommended procedures that may be used to assess possible differences in overall variability for multivariate data. The Van Valen procedure emphasized in this paper is most powerful when there is a consistent directional difference in the variances in the set of characteristics, i.e., the variances in the characteristics in one group are greater than the corresponding set of variances in the characteristics of the other group. On the other hand, if more general differences in overall variability in the set of characteristics between two groups are of interest, Van Valen's procedure should not be used. For these more general situations an extension of the aforementioned Levene procedure (1960) used in univariate analysis can be employed (Manly, 1986b). The extension applies Hotelling's T^2 test on the sets of absolute differences between the standardized measurements and their particular group characteristic standardized medians to determine if there are significant differences in overall variability. Manly argues that this extended Levene procedure would enjoy much power if the differences in variability in the characteristics are more general, with some characteristics having variances in group 1 significantly larger than that in group 2 and other characteristics demonstrating the reverse effect.

For the real estate study, hypothesizing that the variation in each of the characteristics will be greater in Sparta than in Denville, the application of four univariate one-tailed Levene *t* tests yields the same conclusion in all four situations – adjusted for the correct level of α there is no evidence of significant differences in the variances for property size (in acres), number of rooms, list price or age (in years) – in the two communities. The respective Levene *t* test *p*-values are .423, .036, 020 and .275. Each of these four *p*-values exceeds the adjusted .0125 level of significance.

Note, however, that if these tests were inappropriately performed at the unadjusted .05 level of significance, a researcher employing the Levene tests would erroneously draw the conclusion that there was evidence that the variability in the number of rooms and in the list price of Sparta homes was significantly greater in Denville.

Employing Manly's (1986b) suggested multivariate extension of the Levene procedure (i.e., Levene T^2 in Figure 1) results in a *p* value of .211. At the traditional .05 level of significance, the Levene T^2 test fails to find significant differences in overall variability and thus fails to uncover the effects that the more powerful Van Valen procedure found.

DISCUSSION

Van Valen (1978) developed his procedure based on a review of a study conducted by Bumphus (1898) which compared a set of measurements on five charactersistics of 21 female sparrows that survived a fierce winter storm with those of 28 female sparrows that perished as a result of the storm. Although there were no differences in the means of the sets of five characteristics from the two groups of sparrows, there was a significant difference in the overall variability in these five characteristics over the two groups. Darwinian theory suggested that the more homogeneous sparrows would be the survivors. In both his text and in an article, Manly (1986b, 1988) advocates for the use of Van Valen's procedure to demonstrate that two groups may differ consistently and significantly with respect to overall variability in a set of measured characteristics but not necessarily differ in their central tendencies. Manly demonstrates that under such conditions the Van Valen procedure enjoys high statistical power in detecting differences in overall variability. He then demonstrates with the Bumphus sparrow data that Hotelling's T^2 test did not uncover significant differences in the means of the corresponding sets of five characteristics over the two groups and, further, none of the five inappropriate individual *t*-tests show any real differences in the means of the measured characteristics – even without controlling for the proper α level of significance that would then make the finding of such significant differences more difficult. Manly then shows that when repeatedly employing Levene's *t* test (1960) only one of the five measured characteristics demonstrated significant differences in variability between the surviving and perishing groups of female sparrows but did not reinforce his earlier comment that such five individual analyses were inappropriate without first adjusting for the α level of significance. Such an adjustment to α would have resulted in no significant differences in any of the five measured characteristics.

Manly (1986b) describes an appropriate usage of the Levene procedure extended to a multivariate setting. For each of the five measured characteristics he obtains a set of absolute differences in the standardized values from their respective standardized group medians and then performs Hotelling's T^2 test on these absolute differences. No significant difference in overall variability between the two groups was uncovered. Manly then argues that this Levene T^2 procedure would have more power than the Van Valen procedure to uncover differences in variability in the measured multiple characteristics if consistent order was not preserved. On the other hand, when there is consistent order in the variability between the two groups with respect to the multiple measured characteristics, as was observed in the Bumphus sparrow data, Van Valen's procedure is preferred.

In a recent study comparing student learning under two hybrid environments, Hollister and Berenson (2007) found results which are completely in concordance with the above comments made by Manly regarding the analyses of the Bumphus sparrow data. In the current application comparing the four housing characteristics in Denville and Sparta, the results once again illustrate Manly's textbook example and are in agreement with what he described in the analysis of the Bumphus sparrow data.

IMPLICATIONS: EXTENSION TO COMPARISONS IN THREE OR MORE GROUPS

Figure 2 provides a guide for appropriate test selection for situations involving three or more independent groups of numeric data, depending on the particular hypothesis of interest and the shape of the underlying population distributions from which the samples are obtained.

When comparing three or more groups of numerical univariate data the *classical* ANOVA (i.e., Analysis of Variance) *F* test can be used for assessing differences in central tendency when the variability in the groups are assumed to be approximately equal. The separate variance *F* test can be used when that assumption cannot be met (Berenson et al., 1983). Both require that samples are drawn from populations whose underlying distributions are approximately normal. Tests developed by Bartlett, Hartley, Cochran and by Levene have been employed to assess differences in dispersion. However, empirical research has shown that the first three procedures confound

homoscedasticity with non-normality so that in general Levene's procedure (1960) is much preferred (Milliken and Johnson, 1984; Berenson et al., 2006).

When assessing differences in means in the corresponding sets of characteristics among three or more independent groups, the *classical* MANOVA (i.e., Multivariate Analysis of Variance) with the "likelihood ratio" test statistic ϕ is used in lieu of Hotelling's test statistic T^2 if differences in the variances-covariances are only moderate. On the other hand, of major differences in the variance-covariances are found, a separate variance-covariance type ϕ test can be used. When comparing differences in the variability in the corresponding sets of characteristics among three or more groups, extensions of either Levene's or Van Valen's procedures (Manly, 1986b) should be used.

By analogy to the two group multidimensional situation, if the differences in spread in the set of variances are consistent across the groups in an ordered sequence (i.e., a "trend" effect) or, if the set of variances in at least one of the groups was consistently larger (or smaller) than the sets of variances in the other groups (i.e., an "end gap" effect), Manly (1986b) advocates for an extension of the Van Valen procedure that would employ the one-way ANOVA *F* test on the computed *d* statistics in lieu of the pooled-variance *t* test used for comparing two groups.
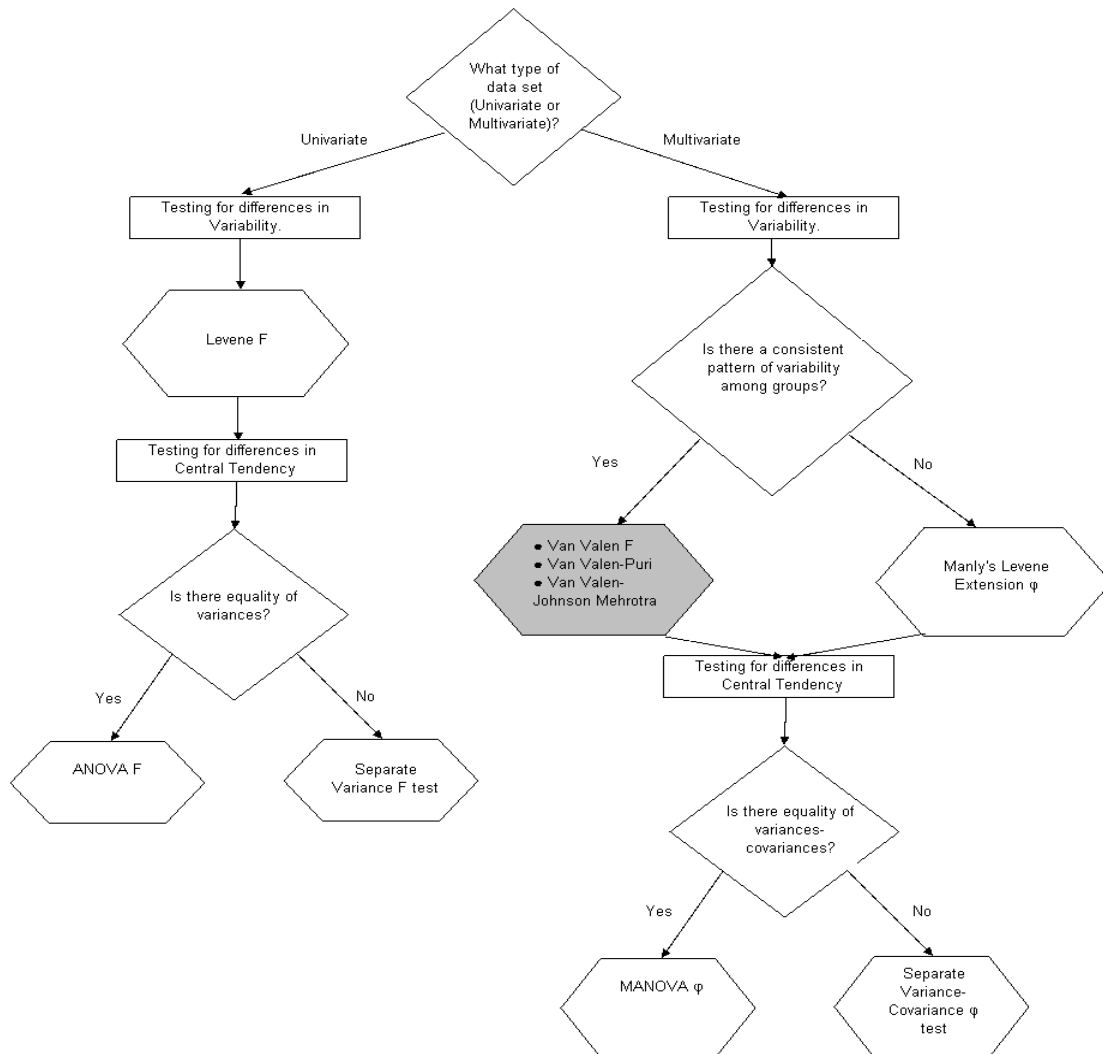

Figure 2: Flowchart of Appropriate Test Selection for Three or More Independent Groups of Data

SUGGESTIONS TO IMPROVE THE VAN VALEN PROCEDURE WITH MULTIVARIATE DATA IN THREE OR MORE GROUPS

Given that the Van Valen procedure comparing two groups is one-tailed, Manly fails to mention that the corresponding one-way ANOVA *F* test would need to be evaluated using a level of significance of 2α in order for the power of the procedure to be consistent with that from the use of the pooled *t*-test comparing only two groups. That being said, even further improvements to the Van Valen procedure would be expected if, instead of using the more general, omnibus one-way ANOVA *F* test, a procedure designed for evaluating ordered alternatives were selected. That is, if, in advance, a researcher were to hypothesize the direction of the trend in the variances of the set of numerical characteristics across the multiple groups, then the very powerful *normal scores* procedure developed by Puri (1965) could be employed to assess the significance of those group differences (see Figure 2). When the groups have the same sample sizes the Puri test statistic can be stated simply (Berenson, 1982, 1986 and see Statistical Appendix). When the sample sizes differ from group to group, however, the Puri test statistic formula is unfortunately much more complex.

On the other hand, if, in advance, the researcher hypothesized an "end gap" effect in the variances in the set of characteristics among the groups, further improvements to the Van Valen procedure would be expected over the suggested use of the one-way ANOVA *F* test by employing the very powerful *normal scores* procedure developed by Johnson and Mehrotra (1971). (See Figure 2). When the groups have the same sample sizes, the Johnson-Mehrotra test statistic can be written simply (Berenson, 1982 and see Statistical Appendix). When the sample sizes differ from group to group, the more generalized version of the Johnson-Mehrotra test statistic based on an extension developed by Schaafsma and Smid (1966) could be employed.

SUMMARY

The Van Valen procedure used in our analysis of the real estate housing market study between Denville and Sparta revealed that there are consistent and statistically significant differences in overall variability between the two communities. Significant differences, however, were not found by using the Levene procedure. Our example shows that when one group has consistently higher variability across all data characteristics, the Van Valen procedure has higher power to detect differences between groups.

There are many potential business research applications where differences in variability in two or more groups are of interest that could benefit from the application of the Van Valen procedure. Examples include market research survey analysis, pharmaceutical studies, and financial data analysis.

Another avenue for application of the Van Valen procedure is in business education research. As mentioned in the previous section, the authors used the Van Valen procedure to study consistent differences in overall variability in student performance between two hybrid sections of a course. There are numerous studies comparing student performance between traditional and on-line/hybrid sections of courses (Anstine and Skidmore, 2005; Liu, 2006; Weber and Lennon, 2007); authors in these studies typically compare group differences in central tendency and ignore differences in variability; the Levene and Van Valen procedures would be applicable to explore overall variability differences across groups.

DIRECTIONS OF FURTHER RESEARCH

A Monte Carlo simulation study enabling a comparison of empirical α and power levels among Manly's suggested extension of the Levene process using MANOVA, Manly's suggested use of the ANOVA *F* test on the computed Van Valen *d* statistics and our suggested Puri and Johnson-Mehrotra tests on the computed Van Valen *d* statistics is in development. Controls are imposed for number of characteristics, number of groups, and varying but equal sample sizes within groups. Also controlled are the levels of significance, the underlying multivariate populations being sampled (normal versus "mixed")with four conditions of variability (equal variability across groups, unequal but inconsistent differences in variability across groups, and two forms of unequal but consistent variability across groups (i.e., "trend" versus "end gap").

This simulation study is intended not only to corroborate the findings of Manly in his previously reported simulations (1986a) but further to assess improvements over Manly's suggested one-way ANOVA *F* test extension to the Van Valen procedure through the use of approaches that *a priori* hypothesize the consistent directions in variability that the sets of measured characteristics would display over the groups.

REFERENCES

1. Abelson, R.P., and Tukey, J.W. (1963) Efficient utilization of non-numerical information in quantitative analysis: General theory and the case of simple order. *Annals of Mathematical Statistics* 34, pp. 1347-1369.
2. Anstine, J. and Skidmore, M. (2005) A Small Sample Study of Traditional and Online Courses with Sample Selection Adjustment. *Journal of Economic Education*, 36, pp. 107-128.
3. Berenson, M.L. (1982) A comparison of several *k* sample tests for ordered alternatives in completely randomized designs. *Psychometrika* 47, pp. 265-279. (Erratum in *Psychometrika* 47, pp. 535-539).
4. Berenson, M.L., Levine, D.M., and Goldstein, M. (1983) *Intermediate Statistical Methods and Applications: A Computer Package Approach.* Prentice-Hall, Englewood Cliffs, NJ.
5. Berenson, M.L. (1986) Puri's expected normal scores test. In *Encyclopedia of Statistical Sciences* 7 (Read, C.B., ed.), John Wiley, New York, pp.364-366.
6. Berenson, M.L., Levine, D.M., and Krehbiel, T.C. (2006) *Basic Business Statistics: Concepts and Applications*, 10th ed. Prentice-Hall, Upper Saddle River, NJ.
7. Bumphus, H.C. (1898) The elimination of the unfit a illustrated by the introduced sparrow, *Passer domesticus*. *Biological Lectures, Marine Biology Laboratory*. Woods Hole, MA, pp. 209-226.
8. Daniel, W.W. (1990) *Applied Nonparametric Statistics, 2nd ed.* PWS-Kent, Boston.
9. Fisher, R.A., and Yates, F. (1963) *Statistical Tables for Biological, Agricultural, and Medical Research*, 6th ed. Hafner, New York
10. GSMLS (2007). Garden State Multiple Listing Service, Friday, July 13, 2007. http://www.gsmls.com
11. Hollister, K.K. and Berenson, M.L. (2007) Proctored vs. Un-proctored Exams: Studying the Impact of Exam Environment on Student Performance. *Informs Transactions on Education* .Currently Under Review.
12. Hotelling, H. (1931) The generalization of Student's ratio. *Annals of Mathematical Statistics* 2, pp. 360-378.
13. Hotelling, H. (1951) A generalized T-test and measure of multivariate dispersion. Proceedings of the Second Berkeley Symposium of Mathematical Statistics and Probability 2, pp. 23-41.
14. Johnson, R.A., and Mehrotra, K.G. (1971) Some *c*-sample nonparametric tests for ordered alternatives. *University of Wisconsin Technical Report No. 263*.
15. Kim, S. (1992) A practical solution to the multivariate Behrens-Fisher problem. *Biometrika* 79, pp. 171-176.
16. Lattin, J., Carroll, J.D., and Green, P.E. (2003) *Analyzing Multivariate Data*. Brooks/Cole, Pacific Grove, CA.
17. Levene, H. (1960) Robust tests for equality of variance. In *Contributions to Probability and Statistics* (Olkin, I., Ghurge, S.G., Hoeffding, W., Madow, W.G., and Mann, H.B., eds.), Stanford University Press, Palo Alto, CA, pp. 278-292.
18. Liu, Y. (2006) Effects of Online Intstruction vs. Traditional Instruction on Students' Learning. *International Journal of Instructional Technology & Distance Learning*, 2, 57-64.
19. Manly, B.F.J. (1986a) Testing for stabilizing selection. In *Pacific Statistical Congress* (Francis, I.S., Manly, B.F.J., and Lam, F.C., eds.), North-Holland, Amsterdam, pp. 339-344.
20. Manly, B.F.J. (1986b) *Multivariate Statistical Methods: A Primer.* Chapman and Hall, London.
21. Manly, B.F.J. (1988) Van Valen's test. In *Encyclopedia of Statistical Sciences* 9 (Read, C.B., ed.), John Wiley, New York, pp.462-465.
22. Milliken, G. A. and Johnson, D.E. (1984). Analysis of Messy Data, *Volume 1: Designed Experiments*, Van Nostrand Reinhold, New York.
23. New Jersey, Department of Education (2005-2006). New Jersey School Report Card 2005-2006, http://education.state.nj.us

24.   Puri, M.L. (1965) Some distribution-free *k*-sample rank tests of homogeneity against ordered alternatives. *Communications on Pure and Applied Mathematics* 18, pp. 51-63.
25.   Schaafsma, W., and Smid, L.J. (1966) Most stringent somewhat most powerful tests against alternatives restricted by number of linear inequalities. *Annals of Mathematical Statistics* 37, pp. 1161-1172.
26.   Sperling (2007). Sperling's Best Places. http://www.bestplaces.net
27.   Van Valen, L. (1978) The statistics of variation. *Evolutionary Theory* 4, pp. 33-43. (Erratum in *Evolutionary Theory* 4, p. 202).
28.   Weber J. and Lennon R. (2007) Multi-Course Comparison of Traditional versus Web-based Course Delivery Systems, *The Journal of Educators Online, 4(2).*
29.   Yao, Y. (1965) An approximate degrees of freedom solution to the multivariate Behrens-Fisher problem. *Biometrika* 52, pp. 139-147.

STATISTICAL APPENDIX

Van Valen Procedure- The following is a step-by-step approach for developing Van Valen's procedure:

1.   Obtain a sample of size n_j observations from multivariate population *i* that contains a set of numerical measurements $(X_{ij1}, X_{ij2},..., X_{ijp}$ where $i = 1$, n_j and $j = 1,g$) on each of *p* characteristics.   Thus, X_{ijk} represents the *i*th observation in the *j*th sample group for variable *k*, where $i = 1, n_j$, $j = 1, g$ and $k = 1, p.$

2.   Compute the means $(\overline{X}_{.j1}, \overline{X}_{.j2},..., \overline{X}_{.jp})$ and the standard deviations $(S_{.j1}, S_{.j2},..., S_{.jp})$ for the set of *p* characteristics from the sample of n_j observations in group *j* where $j = 1,g$. For each of the *p* characteristics being studied combine the $n = n_1 + n_2 + \cdots + n_g$ measurements and compute the overall means $(\overline{X}_{.1}, \overline{X}_{.2},..., \overline{X}_{.p})$ and the overall standard deviations $(S_{.1}, S_{.2},..., S_{.p})$. Then, for each of the *p* characteristics, standardize the *n* measurements so that $Z_{ijk} = \dfrac{(X_{ijk} - \overline{X}_{.k})}{S_{.k}}$.

3.   Using the n_j standardized measurements in group *j*, obtain the set of standardized medians $(M_{.j1}, M_{.j2},..., M_{.jp})$ for each of the *p* characteristics.

4.   For each of the *p* characteristics (k=1,*p*), obtain the *absolute deviations* between the standardized measurements and their respective group medians (*j*=1,*g*) so that
$$A_{ijk} = |Z_{ijk} - M_{.jk}|$$

5.   For each of the $n = n_1 + n_2 + \cdots + n_g$ observations in the study compute Van Valen's "distance" statistic
$$d_{ij} = \sqrt{\sum_{k=1}^{p} A_{ijk}^2}$$

6.   If two groups are being compared (i.e., *g*=2), using the Van Valen "distance" statistic d_{ij} as newly derived "measurements," compute the two sample group means $\overline{d}_1$ and $\overline{d}_2$ and the two sample group variances S_1^2 and S_2^2 .

7.   To test the null hypothesis of homogeneity of spread for the *p* characteristics over the two groups against the alternative that the set of *p* variances in one of the groups (say 2) shows significantly greater variation than the corresponding set of *p* variances in the other group (say 1), perform a one-tailed pooled-variance *t* test.

8.   Using an α alpha level of significance, reject the null hypothesis if $t < t_{\alpha(n_1+n_2-2)}$ with $n_1 + n_2 - 2$ degrees of freedom (one-tailed) and conclude there is consistent and significantly greater variability in the p characteristics from group 2 than in group 1.

9.   If three or more groups are being compared (i.e., $g \geq 3$), follow steps 1 through 5 and in step 6 compute the g sample group means $\bar{d}_j$ and g sample group variances S_j^2 where $j = 1,g$. In steps 7 and 8, to test the null hypothesis of homogeneity of spread for the p characteristics over the g groups against the alternative that the set of p variances in at least one of the groups shows significantly greater (or less) variation than the corresponding sets of p variances in the other groups, replace the one-tailed pooled-variance t test by the one-way ANOVA F test and, for a level of significance 2α, reject the null hypothesis if $F > F_{2\alpha(g-1,n-g)}$ where g is the number of groups and $n = n_1 + n_2 + \cdots + n_g$.

<u>Puri Test</u> - The following is a step-by-step approach for developing Puri's V statistic to be used in lieu of the ANOVA F test statistic as part of the Van Valen procedure if three or more groups are being compared and a consistent "trend effect" is hypothesized:

     Follow steps 1 through 5 of the Van Valen procedure for all g groups and the first part of step 9.

1.   To test the null hypothesis of homogeneity of spread for the p characteristics over the g groups against the alternative that the set of p variances in at least one of the groups shows significantly greater (or less) variation than the corresponding sets of p variances in the other groups (i.e., a "trend" effect), replace the one-way ANOVA F test suggested by Manly in Step 9 by Puri's procedure if the g group samples sizes are equal.

2.   The Puri test statistic V is given by $V = \sum_{j=1}^{g} l_j \bar{e}_{.j}$ where $l_j = n \sum_{j=1}^{g} (2j - g - 1)$, the linear contrast coefficients and $\bar{e}_{.j}$ is the mean of the expected normal scores assigned to group j. The expected normal scores e_{ij} are the "expected normal order statistics" first presented in 1938 by Fisher and Yates (1963) in an attempt to transform a set of non-normal data into one of normal form. The e_{ij} represent the expected Z values for each of the ordered measurements d_{ij} when taken from a standard normal distribution.

3.   The Puri test statistic V is approximately normally distributed and the null hypothesis can be rejected at the

$$\alpha \text{ level of significance if } V > Z_\alpha \left[\left(\frac{n' \sum_{i=1}^{n'} \sum_{j=1}^{g} e_{ij}^2}{n-1} \right) \left(\sum_{j=1}^{g} l_j^2 \right) \right]^{1/2}$$ where e_{ij}^2 is the square of the expected

normal score assigned to the ith observation in the jth group, n' is the common sample size and Z_α is the upper-tailed critical value from the standardized normal distribution.

<u>Johnson-Mehrotra Test</u> - The following is a step-by-step approach for developing the Johnson-Mehrotra $T*$ statistic to be used in lieu of the ANOVA F test statistic as part of the Van Valen procedure if three or more groups are being compared and a consistent "end gap effect" is hypothesized:

1.   Follow steps 1 through 5 and the first part of step 9 of the Van Valen procedure for all g groups.

2.   To test the null hypothesis of homogeneity of spread for the p characteristics over the g groups against the alternative that the set of p variances in one of the groups shows significantly greater (or less) variation than the corresponding sets of p variances in all the other groups (i.e., an "end gap" effect), replace the one-

way ANOVA *F* test suggested by Manly by the Johnson-Mehrotra procedure if the *g* group samples sizes are equal.

3. The Johnson-Mehrotra test statistic $T*$ is given by $T* = \sum_{j=1}^{g} m_j \bar{e}_{.j}$ where $m_j = \left[(j-1)[1-((j-1)/g)] \right]^{1/2} - \left[j(1-(j/k)) \right]^{1/2}$ are the "optimal contrast" coefficients given by Abelson and Tukey (1963) and $\bar{e}_{.j}$ is the mean of the expected normal scores assigned to group *j*. The expected normal scores e_{ij} are defined in step 3 of the Puri test.

4. The Johnson-Mehrotra test statistic $T*$ is approximately normally distributed and the null hypothesis can be rejected at the α level of significance if $T* > Z_\alpha \left[\left(\frac{\sum_{i=1}^{n'} \sum_{j=1}^{g} e_{ij}^2}{n'(n-1)} \right) \left(\sum_{j=1}^{g} m_j^2 \right) \right]^{1/2}$ where e_{ij}^2 is the square of the expected normal score assigned to the *i*th observation in the *j*th group, n' is the common sample size and Z_α is the upper-tailed critical value from the standardized normal distribution.

<u>Levene Procedure</u> - The following is a step-by-step approach for developing Levene's procedure:

1. Follow steps 1 through 4 as in the Van Valen procedure.
2. For univariate data (i.e., *p* = 1; only one characteristic being measured) with two independent groups, the Levene statistic *L* is obtained by performing a pooled-variance *t* test on the *absolute deviation* measurements A_{ij} in the two groups.
3. To test the null hypothesis of equal variances in the two groups, either a one-tailed test or two-tailed test may be used, depending on whether the researcher *a priori* hypothesizes the direction of the difference expected.
4. For univariate data with three or more groups, the Levene statistic *L* is obtained by performing a one-way ANOVA *F* test on the *absolute deviation* measurements A_{ij} in the *g* groups.
5. To test the null hypothesis of equal variances in the *g* groups, the Levene procedure is more general and is unable to specify directional differences in the variances. The null hypothesis may be rejected at the α level of significance if $L > F_{\alpha(g-1,n-g)}$.
6. For multidimensional data in two independent groups, the Levene statistic *L* is obtained by performing Hotelling's T^2 test on the *absolute deviation* measurements A_{ijk} in the two groups.
7. To test the null hypothesis of homogeneity of spread for the *p* characteristics over the two groups against the alternative that the set of *p* variances in one of the groups is significantly different from the corresponding set of *p* variances in the other group the null hypothesis is rejected at the α level of significance if $L > F_{\alpha(p,n_1+n_2-p-1)}$. Note that with multidimensional data the Levene procedure is more general than the Van Valen procedure and, unlike the latter, is therefore unable to specify consistent directional differences in variances in the set of characteristics across the two groups.
8. For multidimensional data in three or more groups, Manly suggests that the Levene statistic *L* be obtained by performing a *classical* MANOVA with the "likelihood ratio" test statistic ϕ on the *absolute deviation* measurements A_{ijk} in the *g* groups.
9. To test the null hypothesis of homogeneity of spread for the *p* characteristics over the *g* groups against the alternative that the set of *p* variances in one of the groups is significantly different from the corresponding

set or sets of p variances in at least one of the other groups the null hypothesis is rejected at the α level of significance if $L > \chi^2_{\alpha(p(g-1))}$, the critical value from a χ^2 distribution with $p(g-1)$ degrees of freedom. Note again that with multidimensional data the Levene procedure is more general than the Van Valen procedure and, unlike the latter, is therefore unable to specify consistent directional differences in variances in the set of characteristics across the g groups.

NOTES