FRACTAL GRADING SYSTEM AND ITS USE IN THE PHILIPPINES

Rowena A. Decena and Fe Sharon C. Tubal
Northwestern Mindanao State College of Science and Technology
Labuyo, Tangub City

ABSTRACT

This paper proposes a new measure of "typical performance" and examines its behavior in comparison with the traditional method of grading. The new measure is based on the concept of a fractal statistics. When one of the test scores become extremely low or extremely high, the values obtained by using the sample mean will be adversely affected. This type of grading system is not fair to the student because he is unduly penalized for a single bad performance. Likewise, he is also unduly rewarded for a single lucky performance. In order to correct this situation, we attempt to control the amount of "reward" a student obtains for a single lucky performance. That is, we do not put the same weight on a single high score as the rest of the scores, but we, instead, adjust this reward as a function of his consistency in class. The study made use of a simulation design using Monte Carlo methods. Outlier probability models are assumed so that student scores are generated from these models. Results revealed that a fractal grading system rationalizes the weights placed on examination scores as basis for the final marks given to students. The system uses the "worst" performance as base score and adds a 'reward" component based on his consistency in performance as measured by his fractal dimension. In contrast, the traditional grading system ignores the "reward" component and computes the final marks of students based on an averaging or smoothing process.

Keywords: fractal statistics, traditional grading system, fractal grading system, smoothing

Introduction

Current grading systems used in the Philippines and elsewhere in the world make use of the mean and standard deviation to describe a student's typical or average performance in class. This grading system is defined as the method utilized in measuring the student's typical performance in order to come up with his final mark. This practice is anchored on the assumption that test scores are normally distributed so that the mean is a good measure of central tendency (Tukey, 1977; Santos et al. 2007). Recent developments, however, have shown that test scores are more likely skewed to the right with very high variance making the mean a spurious measure of typical performance (Huber, 1986; Padua, 2013). The sample mean is easily distorted by a single extreme observation; hence, using it as a measure of typical performance may not be advisable. This paper proposes a new measure of “typical performance” and examines its behavior in comparison with the traditional method of grading. The new measure is based on the concept of a fractal statistics.

There are many factors that affect students’ scores that would cause the values to fluctuate erratically (Santos et al., 2007). The students’ psychological state at the time he is taking an examination could either make his score high or low. Personality factors such as stress and fatigue may lead to poor performance while school factors such as conducive learning environment may precipitate excellent performance. Rubio (2014) identified personal-demographic, social, economic and psychological factors that influence the students’ scores in a standardized entrance examination given by a University. She concludes that students’ scores, generally, tended to be low with the occurrence of sporadic extremely high scores attributed to their different characteristics. Thus, extreme observations will distort the overall picture of a group’s performance (inter-student effect) as well as the picture of a typical student’s performance (intra-student effect). The use of the mean to describe a typical performance value for each student is, thus, compromised in reality.

Tukey (1975), Huber (1986), Andrews (1987), Padua (1989) and others have long recognized the sensitivity of the mean to the effect of outliers. The term “robustness” describes the ability of a statistical measure to down weight the effect of extreme values. The median or the 50th percentile is more robust that the mean in that it takes 50% of the scores which are extremely high or low before it breaks down completely. Alternatives to the sample mean had been proposed to protect against outliers. Stigler (1979) proposed putting small weights to extreme values and called the estimates an L-estimator of location; Huber (1986) proposed a different estimator based on minimizing a distance function called an M-estimator; Sen (1987) suggested replacing the scores by their rank values and called the estimator as R-estimator.

The importance of having a stable measure of typical performance cannot be over-emphasized. Since the beginning of school systems, grades have always remained a debatable issue. Grades tend to brand and stereotype students and such stereotyping has a profound impact on the future of these students not to mention the effect of such to their self-esteem.

Conceptual Framework

The study is anchored on Benoit Mandelbrot's (1982) exposition of the existence of continuous yet non-differentiable curves. These are curves that are highly erratic and irregular such that they do not possess slopes at every point. Such curves are called a fractal curve. Instead of using the derivative (standard deviation) to describe the "smoothness" of a curve, Mandelbrot (1982) proposed a measure of ruggedness called the fractal dimension (λ). Padua (2013) suggested the formula:

$$(2.1) \qquad \lambda = 1 + \frac{n}{\sum_{i=1}^{n} \log\left(\frac{Xi}{\theta}\right)}$$

where X_1, X_2,……, X_n are the scores and θ is the minimum of the scores.

Let X_1, X_2,……., X_n be the test scores of a student. The usual way to describe the students' typical performance is:

$$(2.2) \qquad \text{Typical Performance} = \bar{X} = \frac{\sum_{i=1}^{n} Xi}{n}$$

$$\text{Standard Deviation} = \sqrt{\frac{\sum_{i=1}^{n} (\mathrm{Xi} - \bar{X})^2}{n-1}}$$

When one of the test scores become extremely low, the values obtained in (2.2) will be adversely affected. This type of grading system is not fair to the student because he is unduly penalized for a single bad performance. Likewise, he is unduly rewarded for a single lucky performance.

In order to correct this situation, we attempt to control the amount of "reward" a student obtains for a single lucky performance. That is, we do not put the same weight on a single high score as the rest of the scores, but we, instead, adjust this reward as a function of his consistency in class. Alternatively, we propose the measure:

$$(2.3) \qquad \text{Fractal Performance Measure} = \text{min score} + \text{reward}$$

where :

Reward = $\frac{Max\{X_1,\ X^2 \ldots, Xn\}}{\lambda}$

Ruggedness Measure of Performance = λ

Here, Max $\{X_1, X_2, \ldots\ldots, X_n\}$ is the highest test score of the student and λ is the fractal dimension computed from (2.1). The grade obtained in (2.3) favors the student in the sense that we take his "worst" performance as the base measure and then reward him accordingly based on his "best performance". A student who is consistent in his performance will obtain a larger reward, while a student who is erratic in his performance will obtain a smaller reward.

In order to compare (2.2) and (2.3), we examine the stability of the two measures and compare to determine which measure gives a more stable consistent estimate. A theoretical true performance value φ is assured and then we compute:

(2.4) Stability Measure = Mean-Square Error (MSE) = $\frac{\sum(\hat{x}-\varphi\}^2}{N-1}$

where $\hat{x}$ is either (2.2) or (2.3). The smaller the mean-square error (MSE) is, the more stable the measure is.

Research Design and Methods

The study made use of a simulation design using Monte Carlo methods. An outlier probability model is assumed so that student scores are generated from this model. The expected value of the estimators is considered the true parameter ϕ to be estimated. Each student is assumed to have four (4) test scores X_1, X_2, X_3 and X_4 representing test scores for the prelims, midterms, semi-finals, and finals. The probability distributions used are:

Probability Model 1: Moderate Outlier Model

Value	Probability
70	0.45
75	0.25
85	0.20
90	0.10
Total	1.00

Probability Model 2: Extreme High Outlier Model

The Tukey's contaminated normal model with e = .25 is used. The model is given as follows:

$$F(x) = (1-e)\ N(70,5) + e\ N(95,5)$$

where N(70,5) is a normal distribution with mean 70 and standard deviation 5 contaminated with a normal distribution N(95,5).

The theoretical expected value E(Estimator)= ϕ is used in the computation of the MSE. The mean-square error (MSE) is computed over 100 runs repeated 1000 times.

Intra-student performance is also analyzed by simulation. A theoretical student is given a battery of 100 tests equi-spaced in time. The test scores fluctuate erratically, but a definite upward trend is provided. The fractal average and the traditional average are then computed for this particular student and the results are subsequently discussed. The purpose of this particular simulation is to demonstrate when fractal grading systems are most applicable and beneficial to use in an actual classroom setting.

The simulation experiments are generally set up to determine instances in real classroom situations where traditional grading and fractal grading systems are realistic and practical.

Results and Discussion

1. Results under the First Probability Model: Moderate Outlier Model

Figures 1 and 2 show the histograms of the Fractal grades and Traditional grades respectively:

Figure 1. Histogram of the Fractal Grades of 100 students

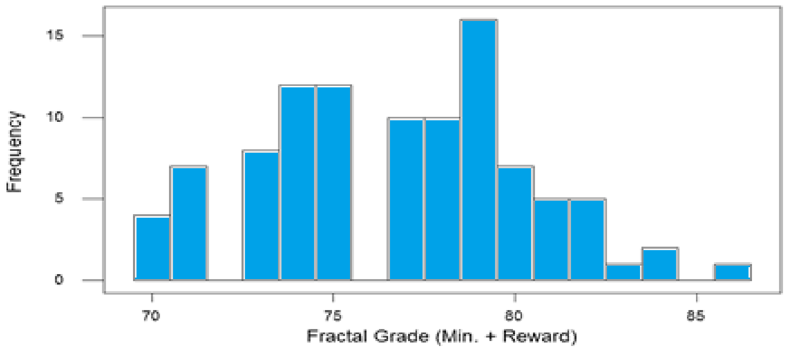
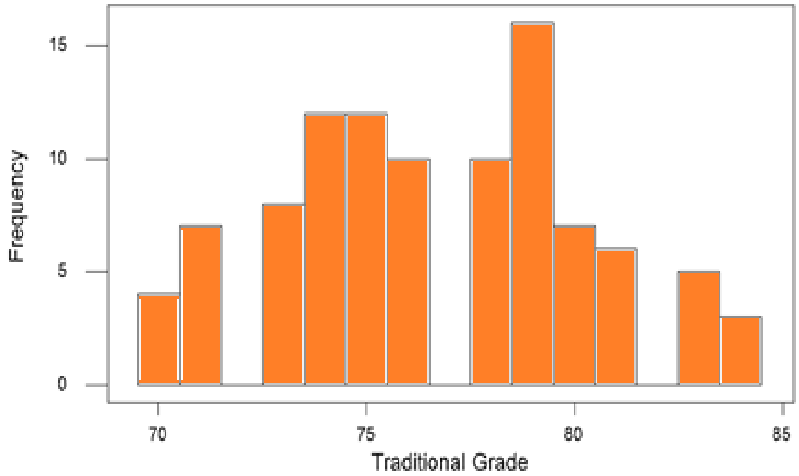


Figure 2. Histogram of the Traditional Grades of 100 students



Figures 1 and 2 show that the two types of grading systems produce an almost identical probability distribution of grades. We verified this by finding the correlation between the two types of grades. Table 1 shows the results of the correlation test:

Table 1. Correlation of the Traditional and Fractal Grades

GRADING TYPE	PEARSON CORRELATION
Fractal Grading System	0.984
Traditional Grading System	P value = 0.000

The very high correlation value of 0.984 which is statistically significant beyond the 0.01 probability level confirms the assertion that the two grading systems are almost perfectly matched under the first probability model.

Table 2, on the other hand, provides the mean-squared errors for the two types of grading systems:

Table 2. Mean-Squared Errors of the Traditional and Fractal Grading System

Variable	N	Mean
Fractal Grading	1000	.013130
Traditional Grading	1000	.01258

Tabular values show that the fractal grading system is slightly less stable than the traditional grading under the first probability model.

2. Results under the High Outlier Model

Table 3 shows the ten (10) of the typical fractal and traditional grades under the high outlier model. We note at once that the traditional grading system tended to give slightly lower grades to students than the fractal grading system. Thus, even if all ten (10) students exhibited higher grades in the final examination demonstrating extra effort on the part of the students, the traditional grading system discounts such efforts by putting the same weight to all the examinations. On the other hand, the fractal grading system recognizes such an effort and rewards the students for such.

Table 3. Typical Fractal and Traditional Grades under the High Outlier Models

Fractal Grade	Traditional Grade
81.84147	81.17285
74.32004	73.71628
76.90475	76.151
77.7833	77.01443
76.17418	75.59028
78.53137	77.0223
74.27637	73.29138
76.92358	75.42485
75.10301	73.9884
74.8237	74.16465

Table 4 shows the mean-squared errors over 1000 runs of 100 students:

Table 4. Mean-Squared Errors under Fractal Grading and Traditional Grading

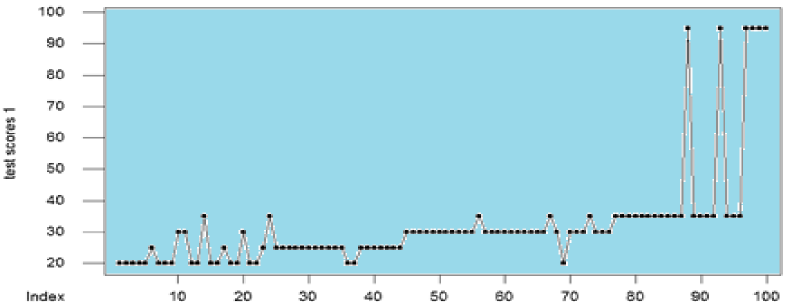
Variable	N	Mean
Fractal Grading	1000	.0489
Traditional Grading	1000	.0487

Tabular values show that in terms of mean-squared errors, the two systems of grading are comparable under the high outlier model.

3. Intra-Student Comparison

A student is given 100 tests on a given subject. His theoretical performance is displayed as a time series graph in fig. 3:

Figure 3: Test Performance of a Student over Time



The time series plot clearly demonstrates that the student's performance is generally improving over time with episodes of backslides. Erratic performances are noted in the time intervals 10 to 30 and then again between 90 and 100. Given this picture, how should the student's typical performance be assessed? Table 5 shows the fractal grade and the traditional grade respectively, for this particular student.

Table 5: Fractal Grade and Traditional Grade of Student A

GRADE TYPE	GRADE	STD. ERROR
Fractal Grade	46.8899	2.41
Traditional Grade	32.1000	1.68

The traditional grade or the mean of the test scores appears to be insensitive to the upward trend of the student's performance, whereas the fractal grade does take into account the fact that the student had reached the maximum performance score towards the end of the battery of tests. It would then appear that the fractal grading system is more psychologically acceptable for students performing in the manner depicted above.

Discussion

- The traditional and fractal grading systems are consistent with each other, in the sense that "good" and "poor" performance are quantitatively differentiated by the two systems. The correlation between the two grading systems is generally high and statistically significant.
- The traditional grading system is more stable than the fractal grading system when the scores are normally distributed. However, the fractal grading system competes with the traditional grading system in the event that the scores are widely dispersed and skewed to the right.
- The fractal grading system has the advantage that the student's performance is assessed based on a reward system that incorporates his consistency in his class performance. On the other hand, the traditional grading system puts equal weight to extreme high scores as the other typical scores of the student.
- The psychological advantage of the fractal grading system is that the student is made aware of the importance of consistency in his academic performance while the traditional grading system ignores such phenomenon altogether.

Conclusion

A fractal grading system rationalizes the weights placed on examination scores as basis for the final marks given to students. The system uses the "worst" performance as base score and adds a 'reward" component based on his consistency in performance as measured by his fractal dimension. In contrast, the traditional grading system ignores the "reward" component and computes the final marks of students based on an averaging or smoothing process.

References:

Andrews, J (1987). *On Robust Statistical Functionals*. (Journal of the American Statistical Association, Vol. _, No. 32)

Huber, P. (1986). *M-Estimators of Location Parameters*. (Journal of the American Statistical Association, Vol. 38, No.28)

Mandelbrot, B. (1982). The Fractal Geometry of Nature. W.H. Freeman and Co. Retrieved from http://ordinatous.com/pdf/The_Fractal_Geometry_of_Nature.pdf

Padua, R. N. (1989). *A Generalized Asymptotic Theory of Multivariate L-Estimate*. (87th Proceedings of the International Statistical Institute, Paris, France).

Padua, R (2013). *Fractal Statistical Inference*. (The Recoletos Multidisciplinary Journal of Research, Vol. 3, No. 1, CHED-JAS-A2)

Santos, R. and R. Padua (2007). *Educational Measurement and Evaluation*. (Diwa Publishing Co., Manila, Philippines)

Stigler, M. (1979). *Robust L-Estimators of Location Parameters*. (Journal of the American Statistical Association, Vol. 27, No.22)

Tukey, J.W. (1975). Mathematics and the picturing of data. In Proc. International Congress of Mathematicians Vancouver 1974. Canadian Mathematics Congress, Montreal. 2 523-531.

Tukey, J.W. (1977). Exploratory data analysis. Pearson, London