

Addressing Negative Trends in Statistics and Testing

Q. E. Igabari, P. A. U. Ossai
Department of Guidance and Counselling,
(Measurement and Evaluation),
Delta State University, Abraka.

ABSTRACT

Certain situations in the realm of statistics and testing are currently exhibiting unfavorable patterns. When students or experts come across these adverse trends, they can leverage foundational mathematical concepts to gain a better understanding. In this comprehensive review, we delve into several crucial aspects, including skewness in test scores, response patterns in testing, score transformation techniques, and the utilization of coding methods for calculating standard deviation. Throughout this analysis, potential sources of confusion are identified and elucidated, paving the way for actionable recommendations to address these intricacies effectively.

ARTICLE INFO

Article History
Received: March, 2023
Received in revised form: May, 2023
Accepted: May, 2023
Published online: June, 2023

KEYWORDS

Skewness, Response Pattern, Coding,
Transformation, Scores

INTRODUCTION

It is common place to see situation that appear unusual. In other words, there are situations that depart from normality. Some of these situations are encountered in statistics and measurement. This review examined skewness in test scores, response pattern in testing, transformation of scores and coding method of standard deviation and scoring of questionnaire skewness in test scores.

Skewness in Testing Scores

Skewness is a measure of distributional a symmetry (Hippel, 2010). According to the author, skewness describes which side of a

distribution has a longer tail. Similarly, National Teachers' Institute (NTI, 2000), as cited by Ossai (2016), viewed skewness as the degree of asymmetry of a distribution. Ossai (2023), Okwuije and Orluwene (2012) as well as Gravetter and wallnua (2004) noted that skewness is either positive or negative. If the long tail is on the right, the skewness is positive or rightward 'but if it is on the left, the skewness is negative or leftward (Hippel, 2010).

Okwuije and Orluwene (2012) as well as Ossai (2021) and Ossai (2016) showed the diagram of positive skenwness and negative skewness as follows:

Fig. 1 (Positively Skewed)

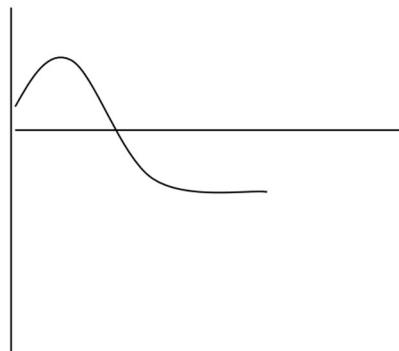
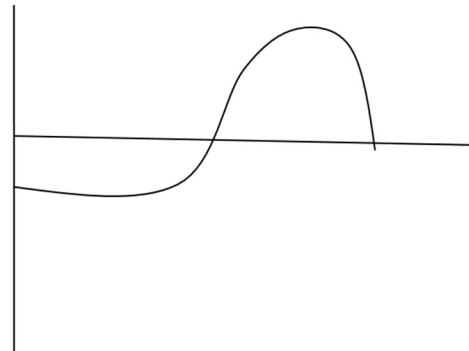


Fig. 2 (Negatively Skewed)



*Corresponding author: Ossai, P. A. U.

✉ aqbadobipeter@gmail.com

Department of Guidance and Counselling, Delta State University, Abraka.

© 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved



The possible reasons advanced by Ossai (2016) for positive skewness and negative skewness in the classroom situation are difficult test questions and simple test items respectively. According to the author, difficult test questions give rise to test scores that are low while easy test items produce scores that are high. For positively skewed distribution, the mode and the median are less than the mean while the mean and the median are less than the mode in a negatively skewed distribution. It is tempting to associate high scores with positive skewness and relate and relate low scores with negative skewness, given that a positive phenomenon is desirable while negative scenario is undesirable. Unfortunately, the reverse is the case. Thus, there appears to be a confusion. Individuals, therefore, appear to misinterpret the concepts of skewness as relate to test scores. Teachers are, therefore, advised to watch out for where the longer tail is while explaining the two concepts.

Examinees' Response Pattern Test Items

Examinees' detailed responses to test items (objective) are shown in item analysis data. Item analysis indicates the item that is simple, difficult and which one fails to discriminate well (Oji, 2011). Two basic concept are involved, namely item difficulty index and item discrimination power. Item difficulty index (also known as p-value) shows how easy or difficulty an item is, based on the value that ranges on 0.0_ and 1.0 derived from pupils correct responses (Bisi & Embong, 2018, as cited by Vincent, et al. 2020). It is the proportion of individuals who correctly answer an item (Oji, 2011). In item difficulty index, the higher the p-value, the easier the item and vice versa (Vincent & Shanmigam, 2020). On the other hand, item discrimination power, D, is an index that shows how well an item differentiates between examinees who have greater or less command of the knowledge of the specific content (Egbule, 2008). In discrimination index, the value ranges from -1 to +1, indicating that +1 implies perfect discrimination, -1 showing that more pupils from the lower group got the item right, and 0 indicating that there is no discrimination (Bichi & Embong, 2018). As against

difficulty index the item discrimination index is directly proportion to the difficulty of the item. That is the higher the index, the more difficult the item is and vice versa.

The formula for calculating the item difficulty index is $P = \frac{R}{T} \times \frac{100}{1}$ where R= percentage passes in an item, T = number of testees involved in the items analysis.

The discrimination power is given as $D = \frac{R_u - R_l}{1/2(T)}$

Where D = item discrimination power, R_u = the number of upper 27%. that got the item right R_l = number of lower 27%. that got the item right t = number of Testees used in the item analysis (Oji, 2011)

The purpose of item analysis according to Anastasi and Urbina (2007) teacher-made tests, is to identify deficiencies either in the test or in the teaching; therefore, class discussion of questionable items can diagnose the problem.

Below is a hypothetical case of an item analysis of a class of 36 students where the upper group and the lower group have 12 students each, the middle group having been discarded. In the hypothetical case, the item and option are; which of the following objects has five sides? The options are A) Hexagon B) quadrilateral C) Pentagon D) Heptagon

Item analysis is thus presented.

Alternatives	A	B	C*	D
Upper 12 pupils	1	0	11	0
Lower 12 pupils	3	2	4	3

Referring to the formula already given, the item difficulty index is 67% or 0.63 while the item discrimination index is 0.58 suggesting that the distracters are plausible or effective. Note that the key in the item has been asterisked, that is option C.

A negative scenario can arise from the hypothetical case given. Suppose the item analysis was presented thus:

Alternatives	A	B	C*	D
Upper 12 pupils	1	1	5	5
Lower 12 pupils	2	3	6	1

Given that C* is the key, the item difficulty and discrimination indices are 0.46 and –

*Corresponding author: Ossai, P. A. U.

✉ aqbadobipeter@gmail.com

Department of Guidance and Counselling, Delta State University, Abraka.

© 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved



0.08 respectively. It is obvious that more testees from the lower group got the item right compared to the upper group. Besides, option D (a distracter) attracted more testees from the upper group. This is consistent with Ossai (2019) who reported that implausible or ineffective distracters occur in testing when more testees from the low achieving group get an item right or a distracter attracts more testees from the upper group. Situations like this are abnormal; they necessitate discussion between the subject teacher and the learners so as to diagnose the problem

The situation becomes complicated when student resort to guessing, (Ossai, 2015) in which an examiner may apply correction for guessing as noted by Ossai and Enwefa (2020). This excludes the score obtained through guessing.

Standard Scores

According to Best and Kahn (2007), standard scores is a method of expressing any score in a distribution in terms of its distance from the mean in standard deviation units. Thus raw scores remain meaningless until they are expressed in or transformed into standard scores. A researcher that uses raw scores for comparison creates a false basis for comparison.

Converting raw scores into standard scores has some merits. Adeyemi (2011) pointed out that raw score conversions are (a) used in comparing two or more distributions and (b) used to compare an individual's level of performance on different occasions. The commonly reported means of transformation of score are percentile ranks, Z-score and T-score. Although percentile ranks was not treated here, suffice it to say that it compromises the property of equal intervals.

Given that $Z = \frac{X - \bar{X}}{SD}$, Z-score, X = raw score, \bar{X} = mean, SD = standard deviation, Z-score is given as $X - \text{minus } \bar{X}$ divided by the standard deviation (Adeyemi, 2011; Egbule, 2008). In a hypothetical case by Egbule (2008) where a student had a score of 80% in physics

testing with a mean of 75% and a standard deviation of 10, the Z-score is 0.5. This can be interpreted that the individual with a raw score of 75% has a score that is 0.5 standard deviations above the mean. However, in a different hypothetical case given by Oji (2011) where a score in a mathematics test is 94 and the mean and raw scores are 98 and 12 respectively, the Z-score is -0.33. The interpretation is that the individual with a score of 94 has performed 0.33 standard deviations below the mean. Thus raw score above the mean have a positive value of Z-score. Conversely, those below the mean have negative Z-score values.

As responses to negative question pose challenges, (Odili, 2023) so it is with the interpretation of Z-score values that are negative. It is possible, for instance, for an individual to interpret a Z-score value of -0.33 as "the individual with a raw score of 94 performed -0.33 standard deviation below the mean. This is where the knowledge of integers and the number line come in. Any number to the right of zero is positive, but any number to the left is negative. Thus, three units to the left of zero is not -3 units below zero but 3 units below zero. This negative scenario is capable of causing confusion when interpreting Z-scores with negative values. Therefore, teachers need to adopt strategies of presenting the idea to students. The T-score is not associated with negative values or fractions. These are the demerits of Z-score (Oji, 2011; Best & Kahn, 2007).

Standard Deviations Using Coding Method

Best and Kahn (2007) posited that deviation method and raw score methods can be used in computing standard deviation. The former is associated with negative values. Determining where to start and how to handle the negative values on the deviation column become challenging. Consider a hypothetical case from Okorodudu (2012),

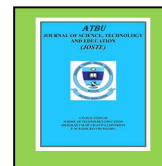
Class Internal	f	d	fd	fd ²
45 - 49	2	4	+8	32
40 - 44	5	3	+15	45

*Corresponding author: Ossai, P. A. U.

✉ aqbadobipeter@gmail.com

Department of Guidance and Counselling, Delta State University, Abraka.

© 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved



35 – 39	7	2	+14	28
30 – 34	9	1	+9	9
25 – 29	13	0	0	0
20 – 24	8	-1	-8	+8
15 – 19	5	-2	-10	+20
10 - 14	1	-3	-3	+9
$\Sigma f=50$				

It is better to start the deviation from the class interval that contains the observation at the middle (in this case it is 25 -29). The classes below this interval have negative deviations while those above have positive deviation. The next step is to balance the negative values and positive values on the deviation column to get Σfd . In this case it is. Note that there are fifty observations.

Determining the starting point (or zero interval) appears challenging. Although the same result is achieved when one decides to start with any class interval, it is advisable to start with the interval that contains the middle score or observation. The number line can then be used to explain the values on the deviation column.

CONCLUSION

This review has shown that students confuse positive skewness with negative skewness. In each of the cases, where the longer tail points is important. Teachers need to take note.

Sometimes examinees response patterns show that some distracters attract more individuals from the group of high achievers. Also, the correct option is selected by more testees from the lower group. These are unusual situations that a test constructor should pay attention to.

It is common knowledge for students to misinterpret Z-scores with negative values. The interpretation should not be done using "minus below the mean"; that is, the word "minus" should be dropped in the interpretation.

Raw score and coding methods are used in finding the standard deviation of a distribution of scores. The latter is associated with positive and negative integers. The starting point should be carefully located; also the teacher may use the ideas of the number line in presenting the concept

REFERENCES

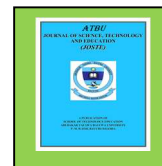
- Adeyemi, T. O. (2011). The effective use of standard scores for research in educational management. *Research Journal of mathematics and statistics*, 3(3), 91-96.
- Agbadobi Uloku, Ossai Peter; Otue Emuobor, Efeludu Patricia (2023). Students' Mathematics Achievement: Is Location A Factor? *ATBU Journal of Science, Technology and Education*, 11(2), pp. 143-146, ISSN 2277-0011. Available at: <http://www.atbuftejoste.net/index.php/joste/article/view/1800>. Date accessed: 04 Aug. 2023.
- Agbadobi Uloku, Ossai Peter; Maureen, Junosa Amainuren (2023). Predictive Validity of the Unified Tertiary Matriculation Examination (UTME) on First-Year Undergraduate GPA. *ATBU Journal of Science, Technology and Education*, 11(2), p. 382-386, ISSN 2277-0011. Available at: <http://www.atbuftejoste.net/index.php/joste/article/view/1824>. Date accessed: 04 Aug. 2023.
- Agbadobi Uloku, Ossai Peter (2023). Assessing post-graduate Students' Ability to Match Levels of Measurement with the Appropriate Measures of Central Tendency. *ATBU Journal of Science, Technology and Education*, 11(2), pp. 127-133, ISSN 2277-0011. Available at: <http://www.atbuftejoste.net/index.php/joste/article/view/1798>. Date accessed: 04 Aug. 2023.

*Corresponding author: Ossai, P. A. U.

✉ agbadobipeter@gmail.com

Department of Guidance and Counselling, Delta State University, Abraka.

© 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved



- Anastasi, A. & Urbina, S. (2007). *Psychological testing*. Prentice Hall of India.
- Au, O. P. Relationship between Students' scores on Plain Questions and Worded Problems in Mathematics.
- Best, J. W. & Kahn, J. V. (2007). *Research in education*. Prentice Hall of India.
- Bichi, A. A. & Embong, I. (2018). Evaluating the quality of Islamic civilization questions. *Asian People Journal*, 1, 93-109.
- Egbule, J. F. (2008). *Fundamentals of test measurement and evaluation*. White and White Publishers.
- Gravetter, F. J. & Wollnau, L. B. (2004). *Statistics for Behavioural Science*. Thompson Wadsworth.
- Hippel, P. (2010). Skewness. *International Encyclopedia of statistical Science*. Springer.
- National Teacher Institute (2000). Education. National Teacher's Institute.
- Odili, J. N. (2023). Managing, item bias in educational assessment in pluralistic society for testing and national integration, 94th inaugural lecture, Delsu.
- Oji, J. O. (2011). *Educational measurement and evaluation*. Lincel Publishers.
- Okorodudu, R. I. (2012). *Understanding educational and psychological measurement and evaluation*. University Printing Press.
- Ossai, P. A. (2022) Maintaining Integrity in Nigeria's Academic Environment through Students' Evaluation. Maintaining Integrity in Nigeria's Academic Environment, 413.
- Ossai, P. A. U. (2019). Examiner's perceived impact of multiple choice test items on students. *Benue State Journal of Education*, 19(3), 35-44.
- Ossai, P. A. U. (2020). Examining Relationship and Differences between Student Teachers' First Supervision and Second Supervision Scores.
- Ossai, P. A. U. (2013). Using continuous assessment to curb bullying in Schools. *International Journal of Advancement in Educational Methods and Management*, 3(3), 94-100.
- Ossai, P. A. (2015). Gender differences in guessing tendencies in mathematics. *British Journal of Education*, 3(8), 78-85.
- Ossai, P. A. U. (2019). Gender variation in undergraduate students' achievement in English language. *Journal of Science Technology & Education*, 7(3), 130-136.
- Ossai, P. A. U., & Enwefa, C. (2020). Undergraduate students' achievement in GST 101 using corrected scores and uncorrected scores. *Journal of Education and Social Research*, 10(2), 126-133.
- Ossai, P. A. U. (2014). Dimensionality of West African Senior School Certificate Examination in Mathematics Objective Test Items. *An Unpublished Ph. D. Thesis, Delta State University, Abraka, Nigeria*.
- Ossai, P. A. U. (2014). Reporting affective scores in school assessment: Implications to counselling. *Delta Journal of Guidance and counselling*, 2(1), 24-31.
- Ossai, P. A. U. (2016). Relationship between Students' Scores on Research Methods and Statistics, and Undergraduate Project Scores. *Journal of Education and Practice*, 7(8), 75-78.
- Ossai, P. A. U. (2016). Test scores and skewness: implications to examiners and students. *Journal of Educational Research and Policies*, 11(2), 156-160.
- OSSAI, P. A. U. Means of Admission and Sex as Determinants of Students' Achievement in the use of English and Library (GST 101).
- Ukwuije, R. P. I & Orluwene, W. W. (2012). *Peanut educational statistics*. Chadi Printing Press.
- Vincent, W. & Shaninugam, S. K. S. (2020). The role of classical test theory to determine the quality of classroom teaching of test items. *PEDAGOGIA: Journal Pendidikan*, 9(1), 5-14.

*Corresponding author: Ossai, P. A. U.

✉ aqbadobipeter@gmail.com

Department of Guidance and Counselling, Delta State University, Abraka.

© 2022 Faculty of Technology Education, ATBU Bauchi. All rights reserved