

# Blood oxygenation among healthy adults: Recurrence Plots analysis and quantification

Gennady Chuiko<sup>a</sup>, Yevhen Darnapuk<sup>a</sup>, Olga Dvornik<sup>a</sup> and Olga Yaremchuk<sup>a</sup>

<sup>a</sup>*Petro Mohyla Black Sea National University, 68 Desantnykiv, 10, 54003, Mykolaiv, Ukraine*

## Abstract

Authors have used Recurrence Plots analysis for a published dataset of blood oxygenation for healthy adults. They offered a new way of the patients' sorting regarding the subsets using the Recurrence Plots and Recurrence Ratio for normal blood oxygenation (SpO<sub>2</sub>) trials. The Recurrence Plot of a SpO<sub>2</sub> record turned out a quite individual portrait of a patient. We found three subsets, from which the entire dataset consists. They vary with levels SpO<sub>2</sub> and its recurrence ratios. The weaker subset shows the lowest levels of oxygenation and low Recurrence of trials, hence the highest their variability. Perhaps, the subset is riskiest for COVID-19, which mostly accompanied itself by hypoxemia. The subset is about 22 % of the full population. The most vital subset is the opposite, their recurrence ratios are highest, as well as the level of oxygen saturation. This subset also is about 22 % of the population. The rest majority (56 %) of healthy adults have quite a high level of oxygenation with moderate Recurrence and variability. All differences among subsets concerning recurrence ratios ranges turned out statistically significant.

## Keywords

Blood oxygenation, SpO<sub>2</sub>, Recurrence Plots, Recurrence Ratio, COVID-19, Variability

## 1. Introduction

Coronavirus disease 2019 (COVID-19) forces the studies of reasons and the provenance of accompanying hypoxemia. Hypoxemia, or sharp oxygen deficiency in the arterial blood, was pointed out in the recent studies of COVID-19 as a severe mortality factor [1, 2, 3, 4]. An element of blood red cells, hemoglobin, is responsible for foe blood oxygen saturation. Each molecule of hemoglobin can capture and delivery to the need place up to four oxygen molecules. Oxygen saturations less than 92 % are associated with significant adverse events in outpatients with pneumonia [5]. Some authors set this critical threshold even higher, up to 95 % [4].

Pulse oximetry (SpO<sub>2</sub> data collection) is a routine medical, non-invasive rapid measurement using small gadgets with a finger clip. There is also a more laborious and invasive method of the saturation measuring by gas sensors introduced live inside arteries. Such data have a bit other notation (SaO<sub>2</sub>). The divergences between these measures are minor as a rule [6]. The reason is that pulse oximeters are calibrated mostly via the direct SaO<sub>2</sub> data.

---

*ICT&ES-2020: Information-Communication Technologies & Embedded Systems, November 12, 2020, Mykolaiv, Ukraine*

✉ gennchuiko@gmail.com (G. Chuiko); yevhen.darnapuk@chmnu.edu.ua (Y. Darnapuk);

olga.dvornik@chmnu.edu.ua (O. Dvornik); olga.yaremchuk.77@ukr.net (O. Yaremchuk)

🆔 0000-0001-5590-9404 (G. Chuiko); 0000-0002-7099-5344 (Y. Darnapuk); 0000-0002-4545-1599 (O. Dvornik); 0000-0002-0891-4216 (O. Yaremchuk)



© 2020 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

In [7, 8] was presented new SpO<sub>2</sub> dataset for 36 healthy adults. The first analysis of these data [7, 9] showed:

1. Relative small variability of data within each record. Especially it touches the short-time variability descriptors (SD1). The long-time variability of the oxygen saturation (SD2) does not exceed 1.1 %, while the short-time variability was less than 0.2 % [9].
2. High Recurrence of some trials. They could be repeated hundreds and even over a thousand times. Therefore Poincare plots turned out strongly clusterized.
3. Significantly non-Gaussian distribution for trials within every record.

The high repeatability and low variability of the blood oxygen saturation records hint that this process is close to almost stable and hard-changing one. Poincare discovered the Recurrence as a fundamental property of the conservative dynamical systems yet in 19 century. However, the only three last decade it is in use as the modern computerized method of investigations [10, 11]. Mostly it is realized via Recurrence Plots (RPs) and their quantification analysis (QRA) [11].

Here we will be keeping in this way, studying the RPs for records of the dataset [7, 8]. Besides, we are going to consider the main QRA factor for them: that is the Recurrence Ratio (RR). Such is the goal of this report.

## 2. Dataset and Methods of processing

The dataset [7, 8] includes 17 males and 19 females, in the age range (19 - 66) years. The patients' body mass index (BMI) was in the range (18.5 - 28.4), which was close to the recognized norm. The majority of participants were non-smokers (28 persons); the rest was either smokers (3) or ex-smokers (5). The duration of every personal record of the blood oxygen saturation was about an hour with the sampling rate equal to 1 Hz (one measuring each second). Note, that one can get all dataset from the source [8].

The peculiar "rectangular" shape of the records [7, 8, 9] advises us a the filtering of them by Haar wavelets as shape fittest ones. We have carried out such filtering for all records with two-fold downsampling. We got the low-frequency (LF) and high-frequency (HF) parts of each record as a result of such filtering. The first of them presents the filtered signal while the second one – noises. Further, we dealt mostly with filtered LF parts of records.

Statistical Shapiro and Wilk's W-tests, which we were applying to all 36 LF parts of records, showed the non-Gaussian nature of them with the probability equal to 0.99. The matter is mainly in the excess kurtosis and massive outliers within the records. Hence, such known variability descriptors as the ranges, or even interquartile ranges, look initially as unreliable for the dataset [7, 8]. So, one should pay attention to other descriptors. Maybe these descriptors should be found outside of statistics.

Let pay attention to the high Recurrence of trials within SpO<sub>2</sub> records. Consider only one example of records (the original code is 080217B [7, 8]). The filtered LF part of this record comprises 1793 values, but only 34 of them are unique. These unique trials were repeated from once up to 1443 times. The most recurred saturation was equal to 98.6 %. No one can now be surprised by the highly clusterized Poincare plots for records of this dataset [9].

Recurrence is a property opposite to variability in some sense. As intensive is the Recurrence so smaller is the variability and reversely. The similarity matrices describe the Recurrence in a series of trials mathematically [10, 11]. Let consider a series, the terms of which are indexed by independent indexes  $one \leq i, j \leq N$ . Then each pair of series' terms  $(i, j)$  corresponds to the matrix element of the square similarity matrix. This element is either equal to 1 if the absolute value of the difference between the pair's terms less than some given threshold, or equal to zero in the opposite case.

The similarity matrix has dimensioned  $(N \times N)$  depending on the series length. For extensive records, like in our case, it may be a big matrix, required the computer methods for processing and building of RPs. Now, we can understand why the RPs have developed together with computers [10].

The Recurrence Plot (RP) is the image of the similarity matrix [10, 11]. Let warn the reader that there are two ways of numeration for the rows and columns of this matrix. The ordinary matrix view suggests the numeration from the left upper corner of the matrix to down and right. The enumeration starts from the lower-left corner to up and right in another way [10, 11]. Here we will be keeping on the matrix view.

The quantification of RPs (QRA) suggests many quantitative factors for the similarity matrix [11]. The most useful and straightforward of them seems to be the Recurrence Ratio (RR). This factor is the ratio between the number of matrix elements, which are equal to 1, and the total number of matrix elements. If the similarity matrix is considering as an image, then RR meets the mean intensity of this image. There are existing also some rules for the visual qualitative analysis of RPs [11]. We will be using these rules below for built RPs.

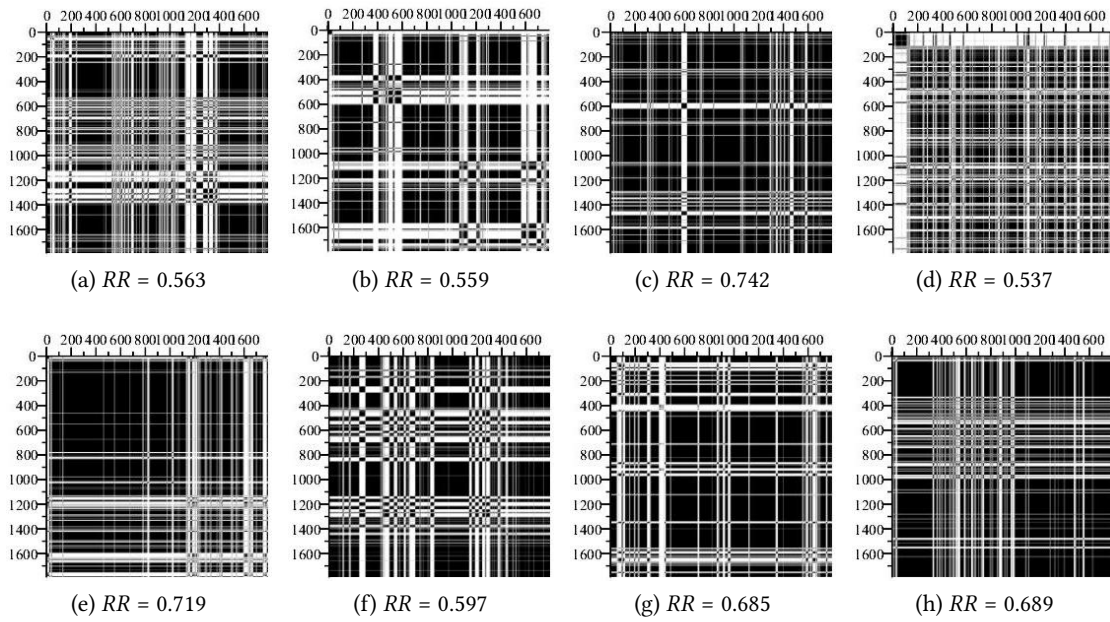
An RP enables us to investigate the multi-dimensional phase space trajectory through a two-dimensional and visual representation of its recurrences [11]. The recurrence threshold value mentioned above is a critical thing for the RPs building Authors [11] pointed out that recurrence threshold selection is a trade-off. On one hand, we intend to a threshold as small as possible, but on the other hand, a sufficient number of recurrences need a higher threshold. We choose the recurrence threshold equal to 0.1 %, which is about equal to the short-time variabilities of records [9]. This value also roughly meets the standard deviation of HF parts of records.

### 3. Results

#### 3.1. RPs for the vital subset

Those people from the dataset [7, 8], who showed the highest RR in the range (0.537 – 0.742), were included in the vital subset. It was eight persons or about 22 % of the population. The mentioned above Shapiro and Wilk's W-test shows the distribution of RR values close to the standard (Gaussian) one with the probability 0.99. Gaussian distribution without any outliers turned out as fit also for the other two subsets. Fig. 1 shows the collection of RPs for this subset.

The mean value for RR is equal to 0.636, with the standard deviation equal to 0.081 for this subset. Thus, the Recurrence is relatively high and concentrated within the higher range. That means that the variability of the oxygen is lowest for these people. The oxygen saturation process is stable enough and hard-changing.



**Figure 1:** Recurrence Plots for the persons from the more vital subset. Note that all RPs are unique by their textures.

However, this process is significantly non-stationary, which confirms by inhomogeneity of plots. Periodic patterns (block-like textures) show that the process can have characteristic cyclic ties. Dark rectangles are signs that some states do not change or change slowly for some time intervals (that is so-called laminar states) [11].

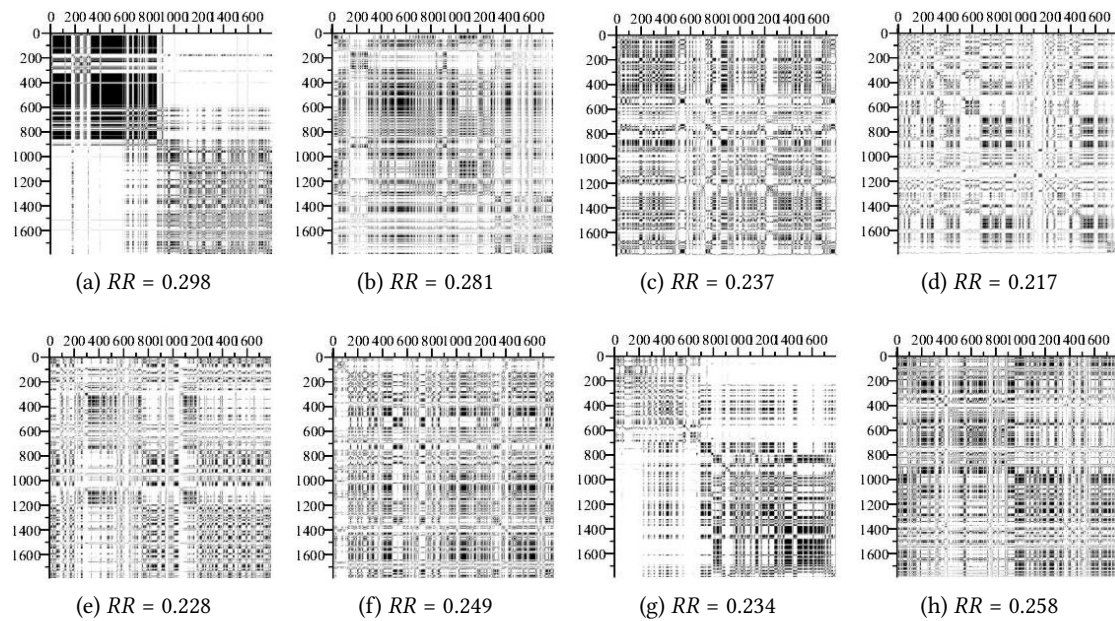
### 3.2. RPs for the riskier subset

People from the riskier subset have RR values from other, palpably lower ranges (0.217 – 0.298) with a mean value equal to 0.250, and the standard deviation about 0.027. Participants demonstrate significantly higher variability and much fewer recurrence ratios. Fig. 2 presents their RPs. The difference of mean intensities of plots is quite evident if one compares Fig. 1 and Fig. 2.

Thus, the oxygen saturation process is here less stable and easier to change. That is why we called this subset riskier to COVID-19. Besides, the saturation process for the people from the subset is non-stationary, though the cyclic ties and laminar states are less expressed as for the previous subset. Pay attention to the "fading" of corners for some RPs of Fig. 2. That is an additional quantitative sign of a non-stationary process [11].

### 3.3. The main subset

This subset comprises 20 participants or 56 % of the total population. The range of RR is (0.327 – 0.568) with a mean equal to 0.430 and a standard deviation of about 0.065. Thus this range



**Figure 2:** Recurrence Plots for the persons from the riskier subset.

lightly overlays with the previous vital subset, while the border with the riskiest subset looks clearer.

The participants demonstrate middle variabilities and recurrences. The conclusions made above for smaller subsets remain valid for the main subset. We mean here the Gaussian distribution, unique textures of RPs, non-stationarity, and presence of cyclic ties and laminar states in the blood oxygen saturation process.

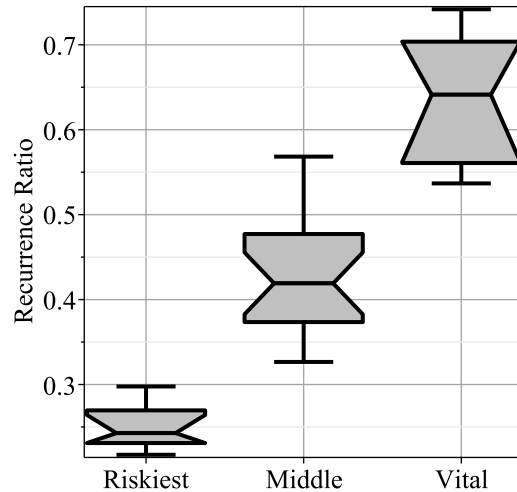
Fig. 3 presents the statistical box-and-whisker plot for RR ranges of all three subsets. The height of boxes reflects the interquartile ranges (the distance between upper and lower quartiles). The "whiskers" shows the ranges of each subset. The lines inside the boxes are the medians. Statistical Two-Sample T-test confirms the statistical significance of the differences between means of RRs for all possible pairs of subsets with the probability equal to 0.99.

Therefore, the intervals of Recurrence Ratios pointed out for each of subsets are a relatively reliable tool for the separation of patients, concerning subsets. Some uncertainty may arise on the border between the upper-middle and the lower-vital subgroup, but the riskiest one is entirely separable via the RR measuring.

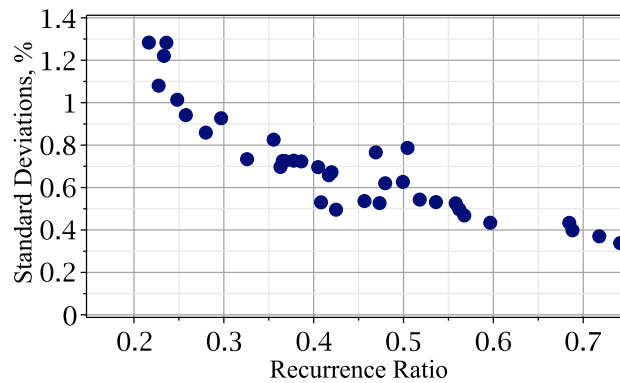
### 3.4. The negative correlation between RRs and standard deviations (variabilities)

Let consider the bond between the RRs and standard deviations for each record of the dataset [7, 8]. Here we reckon the standard deviation as the most reliable indicator of the total variability of each record [12]. Fig. 4 shows this correlation.

The correlation showed by Fig. 4 is quite vital that the correlation coefficient is equal to about



**Figure 3:** Statistical box-and-whisker plot for Recurrence Ratios intervals, which characterizes each of subsets. Note that RR data are free of outliers in contrast to the dataset [7, 8].



**Figure 4:** The negative correlation between Standards Deviations and Recurrence Ratios. As higher is the Recurrence as lower is the variability (the Standard Deviation). Thus the vital subgroup demonstrates the lowest variabilities while the riskiest has the highest ones.

-0.88. Table 1 show some other correlation coefficients.

Note that we have to accept as essential non-zero correlation any correlation coefficients exceed the critical value equal to  $\pm 0.39$  if we take confidence on the level 0.99 and use well-known Student's critical tables. Thus, all coefficients in Table 1 are statistically significant. The sole positive correlation of Table 1 tells us that as higher is the Recurrence as higher is and oxygen saturation of a patient. Although this correlation is moderate enough, the connection sounds as essential for us.

It is visible that the correlation coefficients are lower for the ranges and interquartile ranges than for standard deviations. We can explain this fact by numerous outliers in the dataset [7, 8]. They have a noise-like effect on the veracity of these ranges, what was said above.

**Table 1**

The correlation coefficients between Recurrence Ratios (RR) and some statistical indicators of the dataset variability [7, 8]

RR and Standard Deviations	RR and interquartile ranges	RR and modes (most probable values of the oxygen saturations)	RR and ranges
-0.88	-0.83	+0.73	-0.61

## 4. Discussion and conclusions

Processes and systems, which are ubiquitous, also in medicine and physiology besides, mostly are non-linear and non-stationary ones. They usually exist in noisy mediums. Of course, all of that is conditions breaking to use the arsenal of powerful linear methods and tools [11]. The reader could see how the non-Gaussian nature of the signals, especially the presence of numerous outliers, inhibit the use of statistical tools (ranges and interquartile ranges analysis) or the Poincare plots technique concerning the dataset.

The RPs and QRA are relatively new methods of non-linear dynamics, permitting the study of such processes and systems successfully [10, 11]. We have used here only one of the many qualifications within QRA [10, 11]. However, even that was enough for the straight separation of the dataset on three subsets. Quantification of Recurrence Plots can be continued and looks promising.

It is worthy to point out that RPs are individual portraits of the patients. We did not find even two identical among all of them. The qualitative analysis of their textures, much more profound, than in this paper, also looks like a coming way.

The correlations between Recurrence Ratios and statistical parameters, which shows Table 1, permit us a few conclusions:

1. Recurrence and statistical descriptors of variability, such as standard deviation, ranges, and interquartile ranges [12, 13], have the essential negative correlations. It means, as higher is Recurrence as lower is the variability, as we assumed it above.
2. The vital subset, for instance, has not only the lowest variabilities but and highest oxygen saturation levels. The riskiest subset, reversely, have not only much highest variability, but and the lowest oxygenation. It follows from the positive correlation in Table 1 between Recurrence and the most probable values of oxygenation (modes).
3. The Recurrence Ratios, as the measure for the variability, are free from dataset drawbacks because they have Gaussian distribution and have no outliers. Therefore, they are more reliable, as we believe.

We suppose that healthy adults, incoming to either the vital subset or in riskiest one, should have different chances of arising and deepness, if it has begun, of the hypoxemia accompanying COVID-19.

## Acknowledgments

This report is a part of the research project entitled "Development of hardware and software complex for non-invasive monitoring of blood pressure and heart rate of dual purpose" (registration number 0120U101266) Ukrainian Ministry of education and science financially support this project, and the authors are grateful for that.

## References

- [1] Asociacion Argentina de Medicina Hiperbárica e Investigacion, Covid-19: Hypoxia, inflammation, and immune response, 2020. URL: <https://storage.googleapis.com/wp-aamhei/2020/03/Hypoxia-and-COVID-19-AAMHEI.pdf>.
- [2] S. Rezaie, Covid-19 hypoxemia: A better and still safe way, 2020. URL: <https://rebelem.com/covid-19-hypoxemia-a-better-and-still-safe-way/>.
- [3] K. B. Kashani, Hypoxia in covid-19: Sign of severity or cause for poor outcomes, *Mayo Clin Proc.* 95 (2020) 1094–1096. doi:10.1016/j.mayocp.2020.04.021.
- [4] J. Xie, N. Covassin, Z. Fan, P. Singh, W. Gao, G. Li, T. Kara, V. K. Somers, Association between hypoxemia and mortality in patients with covid-19, *Mayo Clin Proc.* 95 (2020) 1138–1147. doi:10.1016/j.mayocp.2020.04.006.
- [5] S. R. Majumdar, D. T. Eurich, J.-M. Gamble, A. Senthilselvan, T. J. Marrie, Oxygen saturations less than 92 % are associated with major adverse events in outpatients with pneumonia: A population-based cohort study, *Clinical Infectious Diseases* 52 (2011) 325–331. doi:10.1093/cid/ciq076.
- [6] A. Jubran, Pulse oximetry, in: M. Pinsky, L. Brochard, G. Hedenstierna, M. Antonelli (Eds.), *Applied Physiology in Intensive Care Medicine 1*, Springer, Berlin, Heidelberg, Berlin, 2012, pp. 51–54. doi:10.1007/978-3-642-28270-6\_12.
- [7] A. S. Bhogal, A. R. Mani, Pattern analysis of oxygen saturation variability in healthy individuals: Entropy of pulse oximetry signals carries information about mean oxygen saturation, *Front. Physiol.* 8 (2017). doi:10.3389/fphys.2017.00555.
- [8] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, H. E. Stanley, Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals, *Circulation* 101 (2000) e215–e220. doi:10.1161/01.CIR.101.23.e215.
- [9] G. Chuiko, O. Dvornik, Y. Darnapuk, Y. Krainyk, Oxygen saturation variability: Healthy adults, in: *Proceedings of the 2019 XIth International Scientific and Practical Conference on Electronics and Information Technologies, ELIT '2019, IEEE, Lviv, Ukraine, 2019*, pp. 72–75. doi:10.1109/ELIT.2019.8892319.
- [10] N. A. Marwan, A historical review of recurrence plots, *Eur. Phys. J. Spec. Top.* 164 (2008) 3–12. doi:10.1140/epjst/e2008-00829-1.
- [11] N. Marwan, C. L. Webber, Mathematical and computational foundations of recurrence quantifications, in: J. C. Webber, N. Marwan (Eds.), *Recurrence Quantification Analysis. Understanding Complex Systems, Understanding Complex Systems*, Springer, Cham, 2015, pp. 3–43. doi:10.1007/978-3-319-07155-8\_1.



- [12] J. Frost, Measures of variability: Range, interquartile range, variance, and standard deviation, 2019. URL: <https://statisticsbyjim.com/basics/variability-range-interquartile-variance-standard-deviation>.
- [13] A. Field, Discovering Statistics Using SPSS, 3rd. ed., SAGE, Los Angeles-London-New Delhi-Singapore-Vashington DC, 2009. URL: [https://www.academia.edu/24632540/Discovering\\_Statistics\\_Using\\_SPSS\\_Introducing\\_Statistical\\_Method\\_3rd\\_edition\\_2\\_](https://www.academia.edu/24632540/Discovering_Statistics_Using_SPSS_Introducing_Statistical_Method_3rd_edition_2_).