UNTREF
UNIVERSIDAD NACIONAL
DE TRES DE FEBRERO

Histogram preparation and calculation of normalized kurtosis

Denise Romaris

National University of Tres de Febrero, Sound Engineering, Caseros, Buenos Aires, Argentina
dnic.91@gmail.com

Ignacio Maldonado

Universidad Nacional de Tres de Febrero, Sound Engineering, Caseros, Buenos Aires, Argentina
ignaciomberaza@gmail.com

This report presents the preparation of a histogram from an environmental noise measurement and a measurement of a rotating equipment. At the same time, the distribution obtained and its difference in relation to a normal distribution is analyzed from the calculation of the kurtosis and the Kolmogorov-Smirnov test.

Keywords: histogram, normal distribution, kurtosis.

1. Introduction

An analysis for an environmental signal, a rotating equipment signal and the sum of both is presented in this report. The results are shown in the form of a histogram. The normalized kurtosis is then calculated for the three records. The level of similarity of the obtained distributions and the normal distribution is analyzed using the Kolmogorov-Smirnov (K-S) test and the variations of the results are analyzed based on the number of bins.

A histogram is a graphical representation of a variable in the form of bars, which has two types of variables: the frequency of the values and the values themselves. In this case, the equivalent sound pressure level is analyzed for 20 minutes and the histogram shows how frequent each range of measured levels is.

In turn, kurtosis measures how steep or flattened a curve or distribution is. Therefore, it serves to characterize a probability distribution of a random variable, indicating the degree of concentration of the values around the mean. Therefore, the higher the degree of kurtosis, the steeper the shape of the curve and the lower the degree, the more uniform the distribution. According to its value, three types of distributions are obtained:

- If this coefficient is null, the distribution is said to be normal (similar to the normal Gaussian distribution) and is called *meso-kurtic*.
- If the coefficient is positive, the distribution is called *leptokurtic*, there is a greater concentration of data around the mean.
- If the coefficient is negative, the distribution is called *platykurtic* and there is a lower concentration of data around the mean.

In addition, to verify whether or not the distributions obtained from the samples follow a normal distribution, the K-S statistical test is used. This allows measuring the degree of agreement between the distribution of a set of data and a specific theoretical distribution. The test result is represented by the letter Z, which is calculated from the largest difference between the theoretical and observed cumulative distribution functions. For this purpose, a hypothesis test is performed assuming that both samples have the same distribution against the alternative hypothesis that they are different.

In such a test, two hypotheses are put forward: a null hypothesis and an alternative hypothesis. Which one prevails depends on the t e s t result. In t u r n , for large samples there is the
Table 1 presents these already calculated critical values. α (%). The

Table 1.1: Tabulated critical values.

α	1%	5%	10%
C_α	0,819	0,895	1,035
$k(n)$	$\sqrt{n}-0,01+0,85/\sqrt{n}$		

This value gives the largest tolerable absolute difference for a certain level of significance. For this report, the value corresponding to 5 % is used and from this value and k(n) for a normal distribution, where n is the number of samples, we have
value and $k(n)$ for a normal distribution, where n is the number of samples, it is obtained that that:

if $Z > Z_\alpha \Rightarrow$ *The null hypothesis* (H_0) *is rejected.*

if $Z < Z_\alpha \Rightarrow$ *There is not enough evidence to reject* H_0

Where Z is the result of the Kolmogorov-Smirnov test and $Z_\alpha$is the critical value divided $k(n)$.

2. Procedure

2.1 Recording of environmental signal and signal from a rotating equipment.

For the recording of the signals, a time of approximately 20 minutes is measured in a 15 m bath[2]. An AEG L7FBR169L washing machine with a load of 3kg is used as rotating equipment. The equipment setup includes a Rode NT2-A microphone, a Lexicon Lambda audio board, a Macbook Air computer and REW V5.19 software. The
The microphone is calibrated from its sensitivity, as detailed in Report 1. In turn, the Z-weighting curve is used since the low-frequency content in the noise measurement of a washing machine is significant. The graphs of both records can be found in Appendix 6.1.

2.2 Preparation of the histogram and calculation of kurtosis

The Excel program was used for the analysis of the results. They are placed from the file .txt file provided by the REW program, the decibel values of both measurements are placed, the maximum and minimum values are calculated, the difference between them, and this difference is divided by the number of bins desired for the histogram analysis. For the sum of the measured signals, Equation (1) is used.

$$L_{eq} = 10\ log\left(\sum_{i=1}^{n} 10^{L_{(e)(q)i}/10}\right) \qquad (1)$$

L_{eqi}being the values of $L_{eq}$of the ambient noise measurements and of the rotating equipment over time.

The range and frequency table is assembled by selecting the minimum Leq value and then an array that sums the difference between the minimum and maximum divided by the number of bins. Then the number of values between the respective limits is calculated for each range. To analyze the values, select the Leq values from the Data tab and select Data analysis. Among the results of this function is the kurtosis. Finally, the variation of the results in the histograms as a function of the number of bins selected is analyzed for each case.

2.3 Kolmogorov-Smirnov test

The test is performed for the three signals obtained with the IBM SPSS Statistics 25 program, in order to obtain a more precise analysis of the distributions they represent. The theoretical distribution with which the signals are compared is that of a normal distribution.

3. Results

3.1 Preparation of the histogram

The histograms of the ambient noise measurement, of the rotating equipment and the sum of both signals are presented below. For the ambient noise histogram (Figure 3.1.1) an asymmetric left-tailed distribution is observed. Although this can be verified visually, it is also corroborated by taking into account the shift between the mean and the median. In this case, the mean is 53.4 dBZ and the median 53.2 dBZ, resulting in a difference of +0.2 dBZ, which indicates this type of asymmetry. This is to be expected due to the characteristics of the measurand. That is, being an ambient noise inside a home, it is expected that the level is presented in a constant way with the exception of some few rises in the level, as can be seen in the histogram.

Figure 3.1.2 shows the histogram corresponding to the measurement of a rotating equipment, which in this case is a washing machine operating with a load of 3 kg. Unlike that obtained for ambient noise, in this case the values present a distribution that is not concentrated in a few ranges of values. This can be attributed to the dynamics in terms of rotation speed, being of low revolutions for washing and increasing its speed for spinning.

As for the skewness of the distribution it is of negative or left-tailed type, the difference between the mean and the median being -0.34 dBZ.

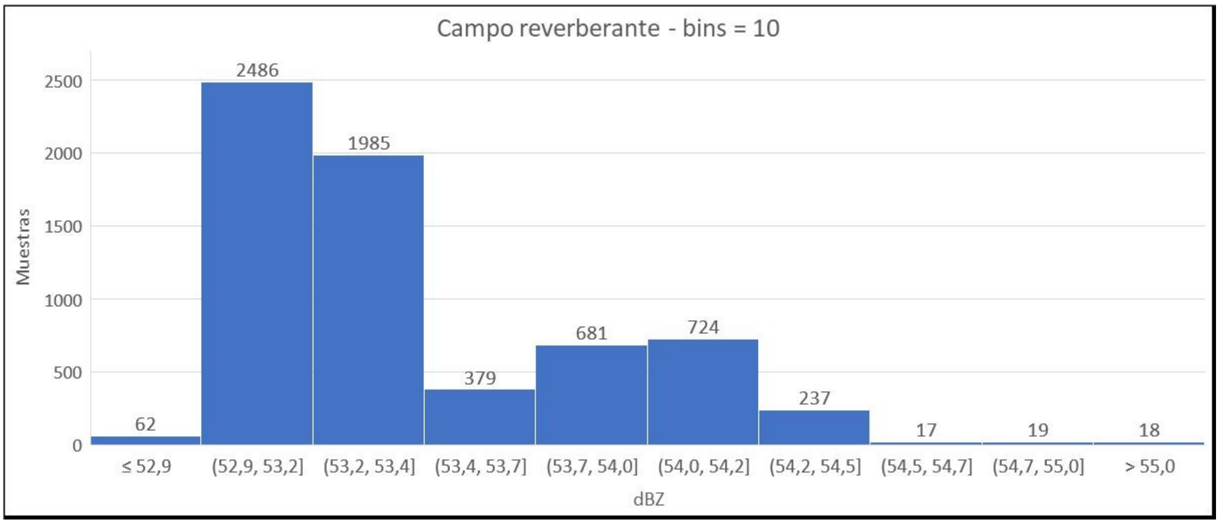

Figure 3.1.1: Histogram of 10 bins of the ambient noise measurement.

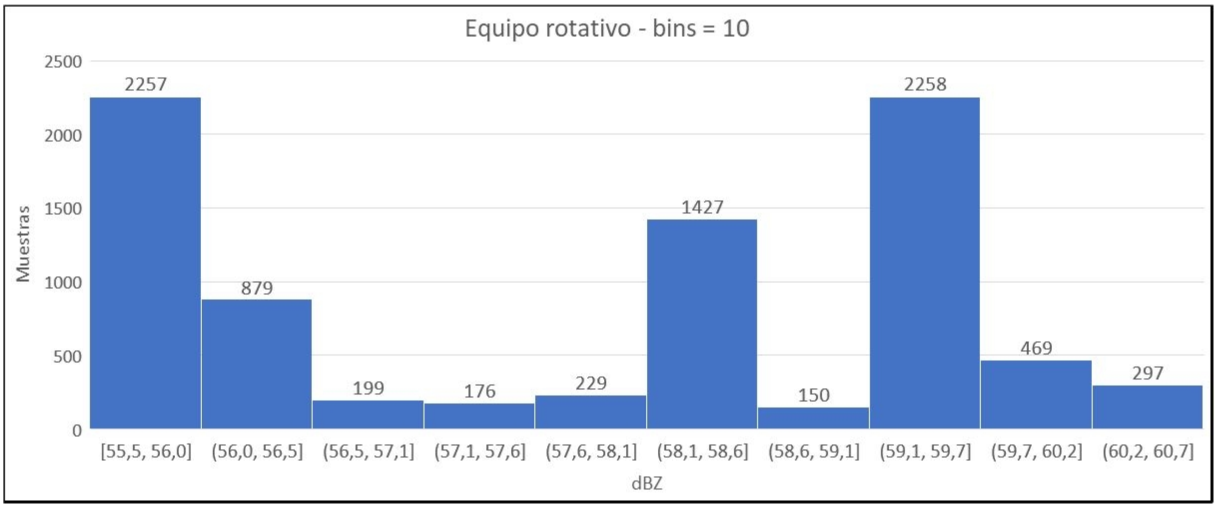

Figure 3.1.2: 10-bin histogram of the rotating equipment measurement.

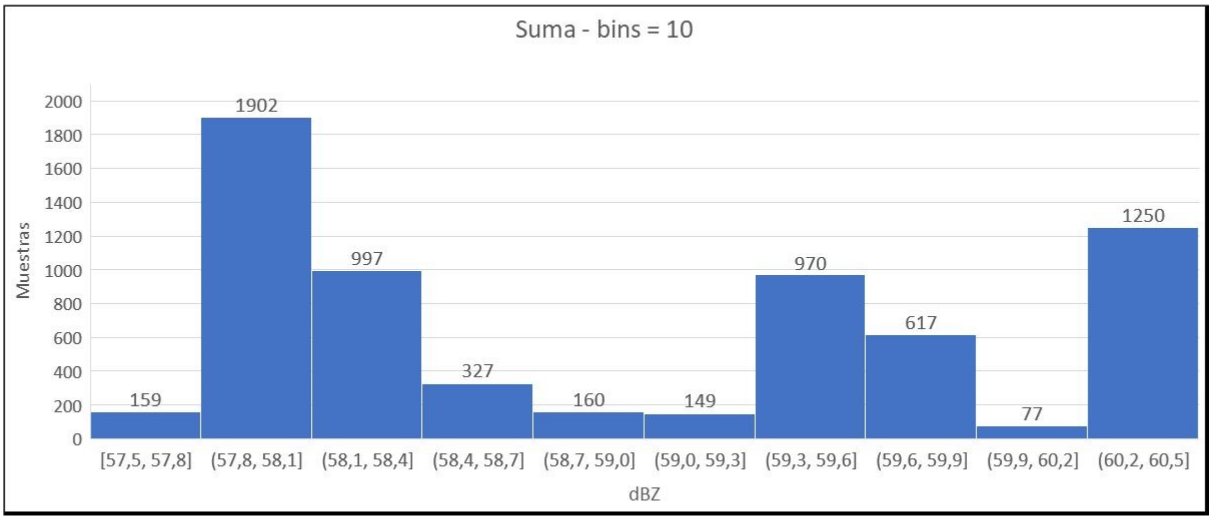

Figure 3.1.3: 10-bin histogram of the sum of signals.

For the third case of analysis, the sum of the two previous signals is obtained which is represented in Figure 3.1.3. As a first observation, we have a distribution that is more similar to that of Figure 3.1.2 corresponding to the rotating equipment. This

This result is to be expected since the ambient noise measurement is of constant character and lower level, which to the total sum contributes an increase only in the left part of the distribution. The latter is reflected in the difference of +0.4 dBZ between the mean and the median, which is also inverted with respect to that obtained for the rotating equipment measurement. Although the asymmetry seen from this difference is of a positive nature (like that obtained in the ambient noise measurement), this value is not sufficient to analyze the distribution obtained. Therefore, in the next section the calculation of the kurtosis is performed, which reflects a more consistent value for the analysis and comparison of the distributions.

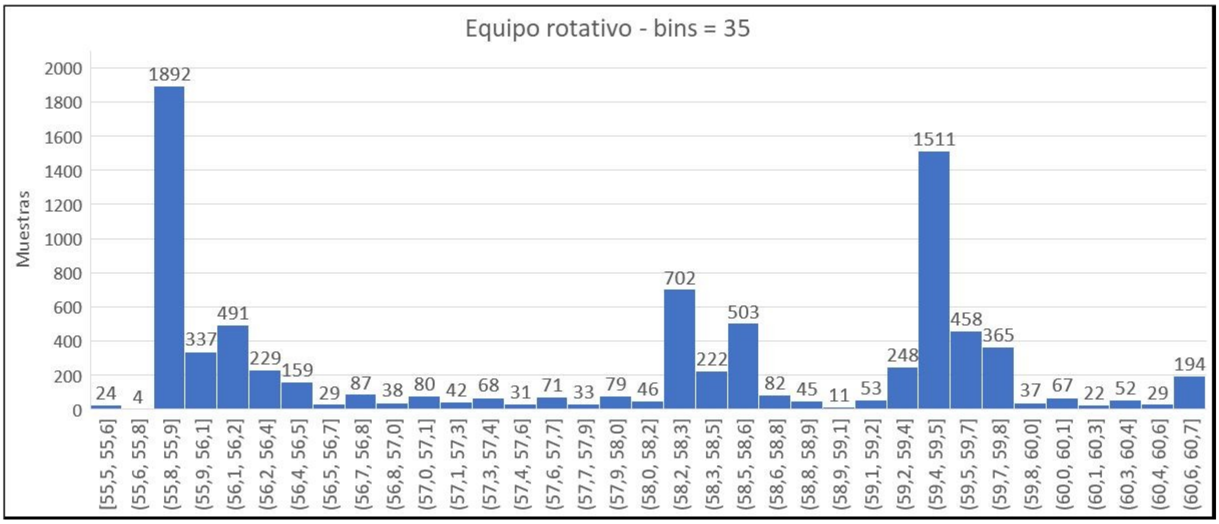

Figure 3.1.4: Histogram of 35 bins of the rotating equipment.

Finally, Figure 3.1.4 shows the histogram representing the rotating equipment but with a larger number of bins. As expected, the resolution for this case is higher, which allows a more precise visualization of its distribution. In addition, both at the upper and lower end, values with very few samples are observed. These could represent outliers. On the one hand, the lower level ones could be attributed to a short duration pause during washing. On the other hand, the higher level outliers could be caused by the addition of some kind of random sound source external to the measurement, or even by some malfunction of the appliance that causes an increase in the level during a short period of time. Appendix 6.2 shows the histograms with more bins for the remaining signals.

3.2 Calculation of kurtosis

Table 2 shows the results of the data analysis.

Table 3.2.1: Analysis of the L_{eqZ}values of the measurements.

	Ambient noise	**Rotating equipment**	**Signal summation**
Average [dBZ] 53.40	53,40	57,86	58,90
Median [dBZ] 53.20	53,20	58,20	58,50
Standard deviation [dBZ] 0.43	0,43	1,61	0,96
Kurtosis	**1,63**	**-1,59**	**-1,45**
Range [dBZ] 3.90	3,90	5,20	3,01
Minimum [dBZ] 51.30	51,30	55,50	57,47

Maximum [dBZ] 55.20	55,20	60,70	60,48
Samples	6.608	8.341	6.608

As already established, there are three possible distributions from the calculation of the kurtosis. In the case of ambient noise we have a positive value, i.e. a *leptokurtic* distribution. According to the definition, this distribution concentrates its values around the mean, which in this case is 53.4 dBZ. Taking into account this value, this phenomenon can be verified by observing the histogram in figure 3.1.1. On the contrary, both for the rotating equipment and the sum of both signals, a negative or *platykurtic* kurtosis is obtained. By definition, this indicates a lower concentration around the mean, which is verifiable by observing their respective histograms, for which the mean is around the center of the graph. Whereas most of the samples are located towards the extremes, away from the mean. In turn, this also verifies that the sum of both signals has the characteristics of that of the rotating equipment, as described in the previous section.

The kurtosis shows, for all cases, a considerable departure from a normal distribution. In order to corroborate this observation, the results obtained for each test are shown below.

3.4 Kolmogorov-Smirnov test

First, the hypotheses are posed:

- *Null hypothesis: the measurements follow a normal distribution .*
- *Alternate hypothesis: the measurements do not follow a normal distribution.*

For a confidence interval of 95%, the null hypothesis is accepted if the K-S test result, Z is less than $Z_\alpha$for each case. According to Table 3.4.1 for the three signals a higher value of Z is obtained, therefore, the null hypothesis is rejected. It is determined for the three cases that none of them follows a normal distribution.

Table 3.4.1: K-S test for the three signals.

	Ambient noise *n= 6608*	**Rotating equipment** *n= 8341*	**Sum of signals** *n= 6608*
Z	**0,246523**	**0,191754**	**0,181256**
$C_{(\alpha)}$ (95%)	**0,895**	**0,895**	**0,895**
$k(n)$	**81,29**	**91,35**	**81,29**
$Z_\alpha = C_\alpha / k(n)$	**0,011**	**0,0098**	**0,011**

Finally, in appendix 6.2 the histogram of the sum of signals compared with a theoretical normal distribution curve is observed, being in evidence the results obtained. The same is observed for the remaining signals.

4. Conclusions

Considering each measurement, the ambient noise follows a leptokurtic distribution, with most of its samples around 53 dBZ, with small variations between 51.30 dBZ and 55.20 dBZ, being asymmetric towards the higher values. These may be due to noises coming from outside the passenger compartment. As for the measurement of the rotating equipment, it shows peaks in three different ranges, due to the pauses of the washing machine, in addition to its speed changes and even to some potential failure. This histogram with a platicurtic distribution does not resemble a normal distribution, although if each of these three peaks is analyzed separately, a similarity could be found, as shown in Figure 6.2.4. The analysis of this sum of signals leads to the same conclusions as when analyzing the measurement of the rotating equipment. For all cases, as the number of bins in the histogram increases, outliers are displayed.

From all this, it was possible to verify with the use of statistical processing such as kurtosis or the K-S test, the characteristics observed from the results obtained. From all the results obtained, the main question that arises refers to the reason why none of the distributions obtained resembles a normal distribution. This could be attributed to the existence of sources of errors introduced in the measurements that lead to different results.

5. References

[1] https://economipedia.com/definiciones/prueba-de-kolmogorov-smirnoff-k-s.html

[2]https://nephi.unice.fr/ftp/users/francour/Cours_Stat_4.pdf

[3]http://catarina.udlap.mx/u_dl_a/tales/documentos/lat/ramirez_r_pa/apendiceC.pdf

[4] Mark Denny. More physics in the laundromat (2010).

[5] https://ocw.unican.es/pluginfile.php/1301/course/section/1594/12-Oscilaciones.pdf

[6] R. N. Garces. Influence of thermal noise on JMC67 integrators. Master Thesis, Faculty of Engineering Sciences, Universidad Tecnológica Acústica de Argentina (2005).

[7] P. L. Kuman, F. A. Robledo, G. F. Sanabria, and Y. S. Kim. S. Kim. Efficient calculation of DFT using non-invertible quaternions. Technical Report No 2561-08, National Institute Corporation, Canelones, Uruguay (2008).

6. Annex

6.1 Photos and graphs of the measurements

The following Figures show the results of the temporal measurements provided by the REW program.

Figure 6.1.1: Color reference for the interpretation of the results.

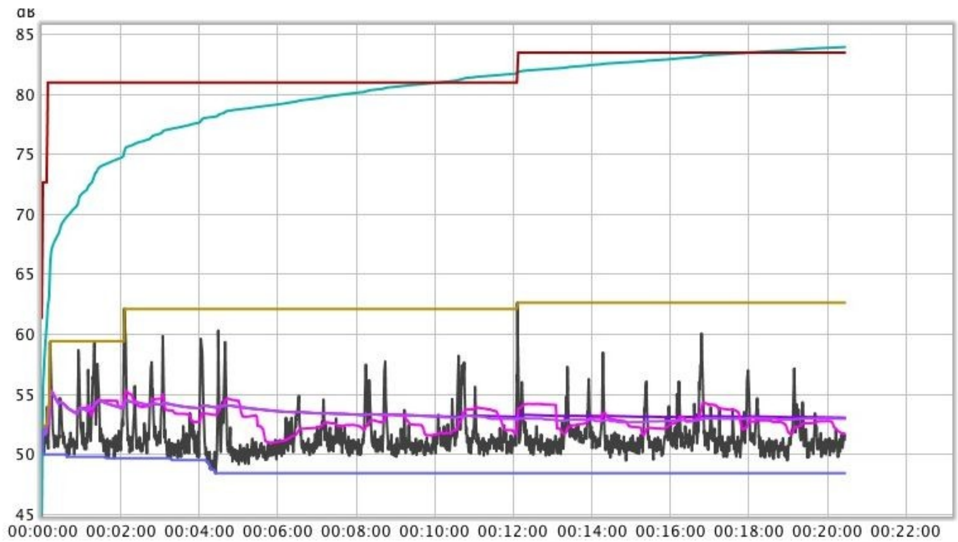

Figure 6.1.2: Ambient noise measurement in a 15 m bath^2.

Figure 6.1.3: Noise measurement of rotating equipment in 15 m^2bath.

Figure 6.1.4: Photo of ambient noise measurement in 15 m^2bath.

Figure 6.1.5: Photo of AEG L7FBR169L washing machine used as rotating equipment.

6.2 Histograms

The different histograms calculated with different with a higher number of bins and also comparative histograms with respect to a theoretical normal distribution are shown below:

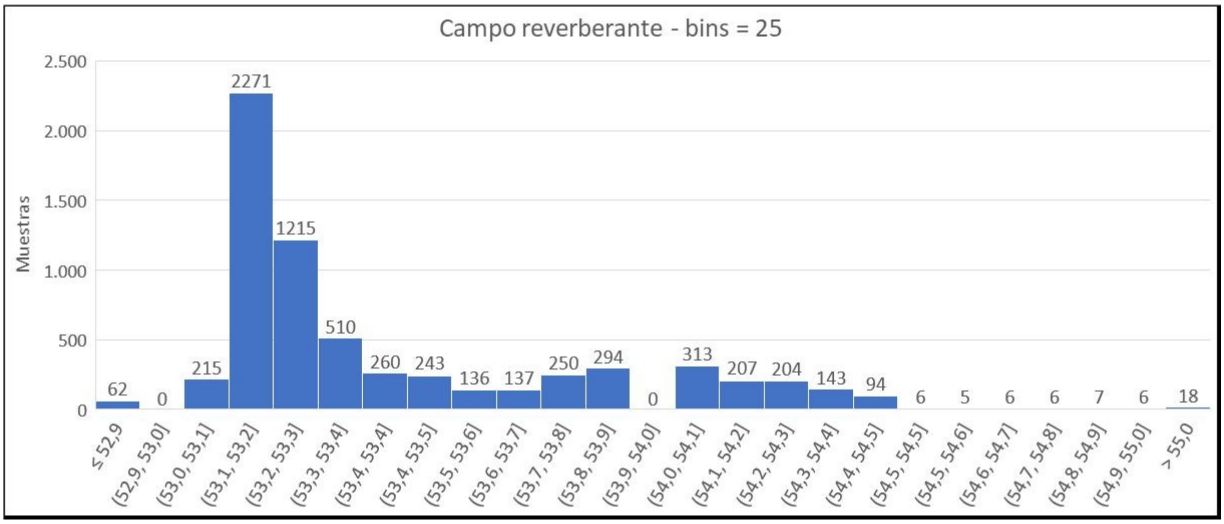

Figure 6.2.1: Histogram of 35 bins of ambient noise.

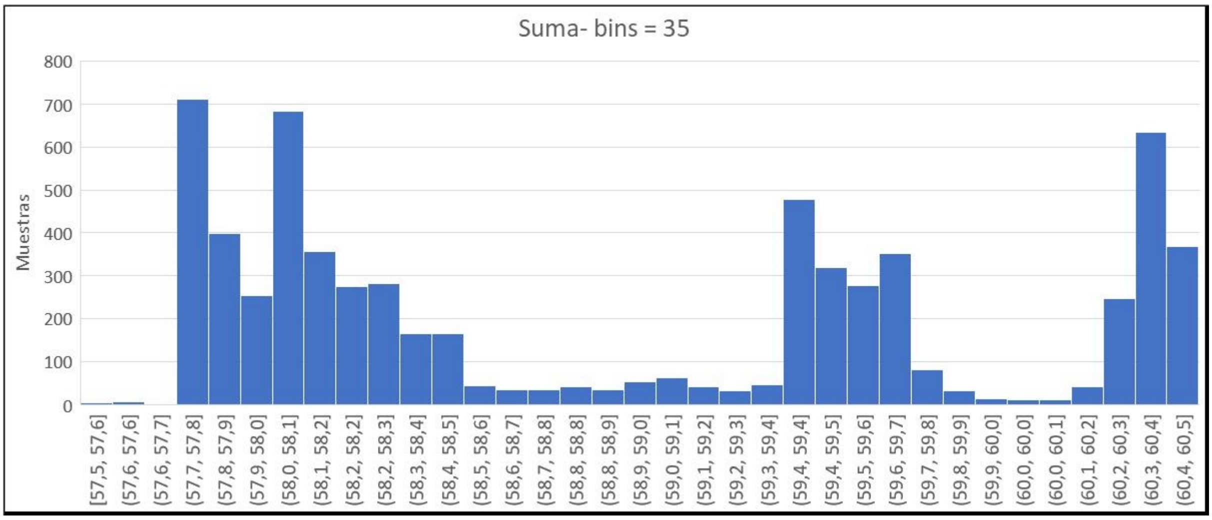

Figure 6.2.2: Histogram of 35 bins of the sum of signals.

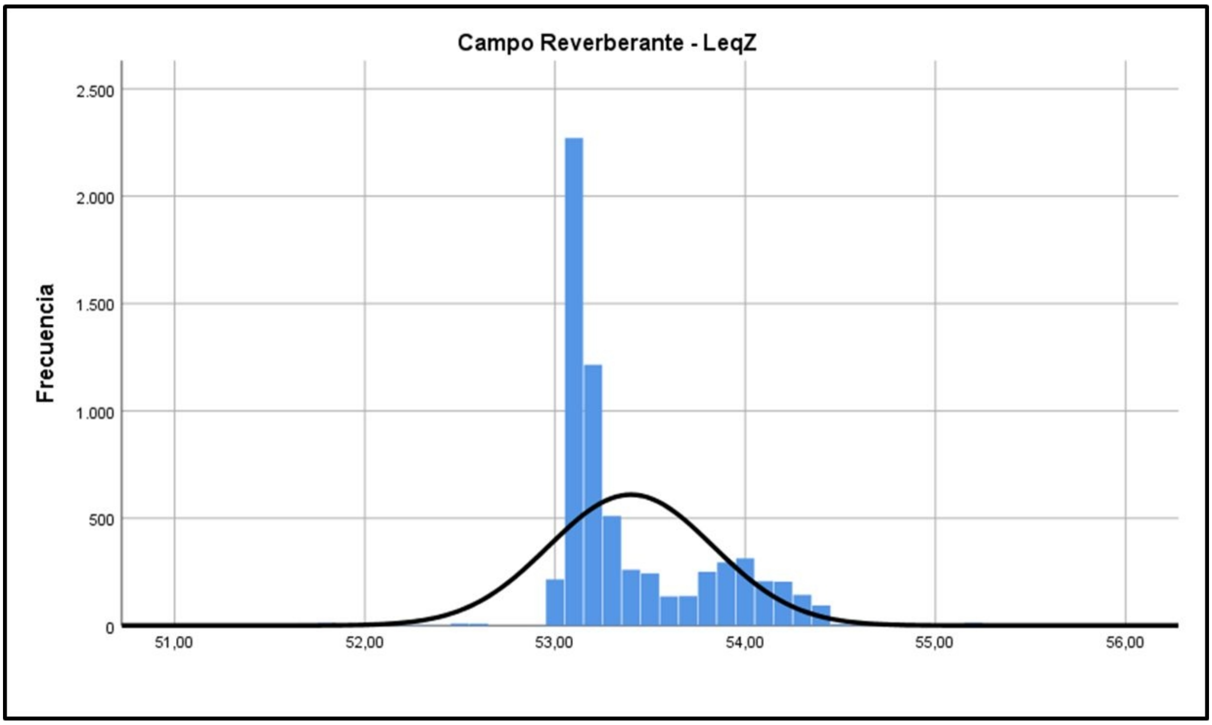

Figure 6.2.3: Histogram of ambient noise vs. theoretical normal distribution curve.

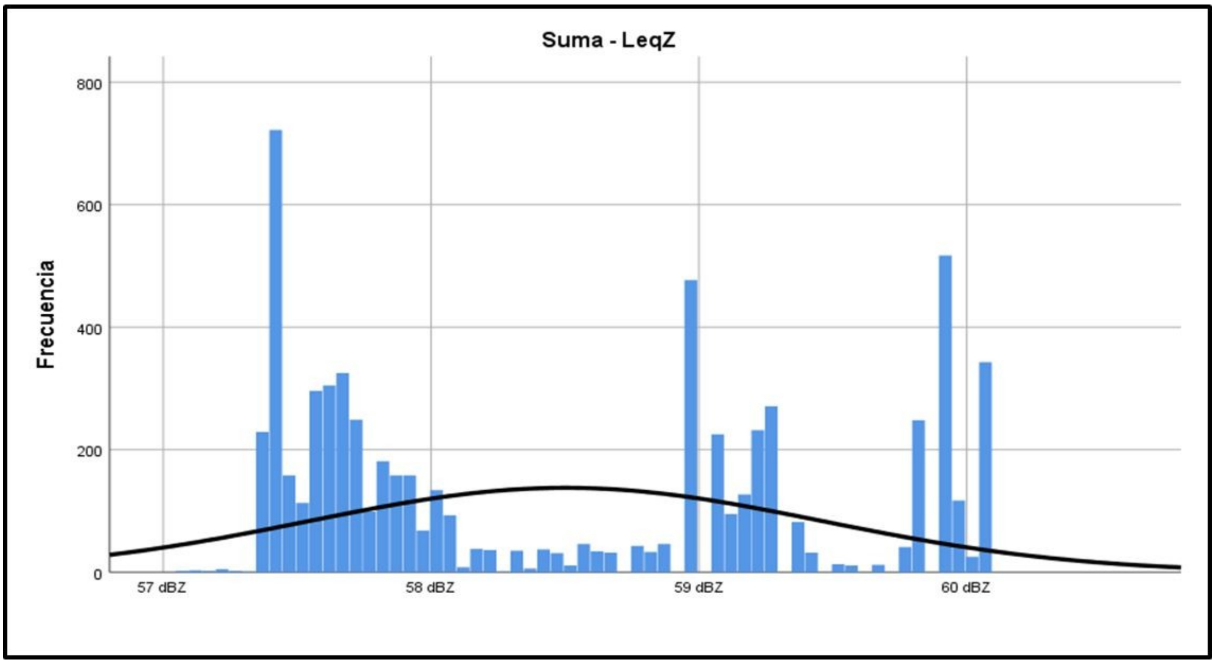

Figure 6.2.4: Histogram of the sum of signals vs. theoretical normal distribution curve.

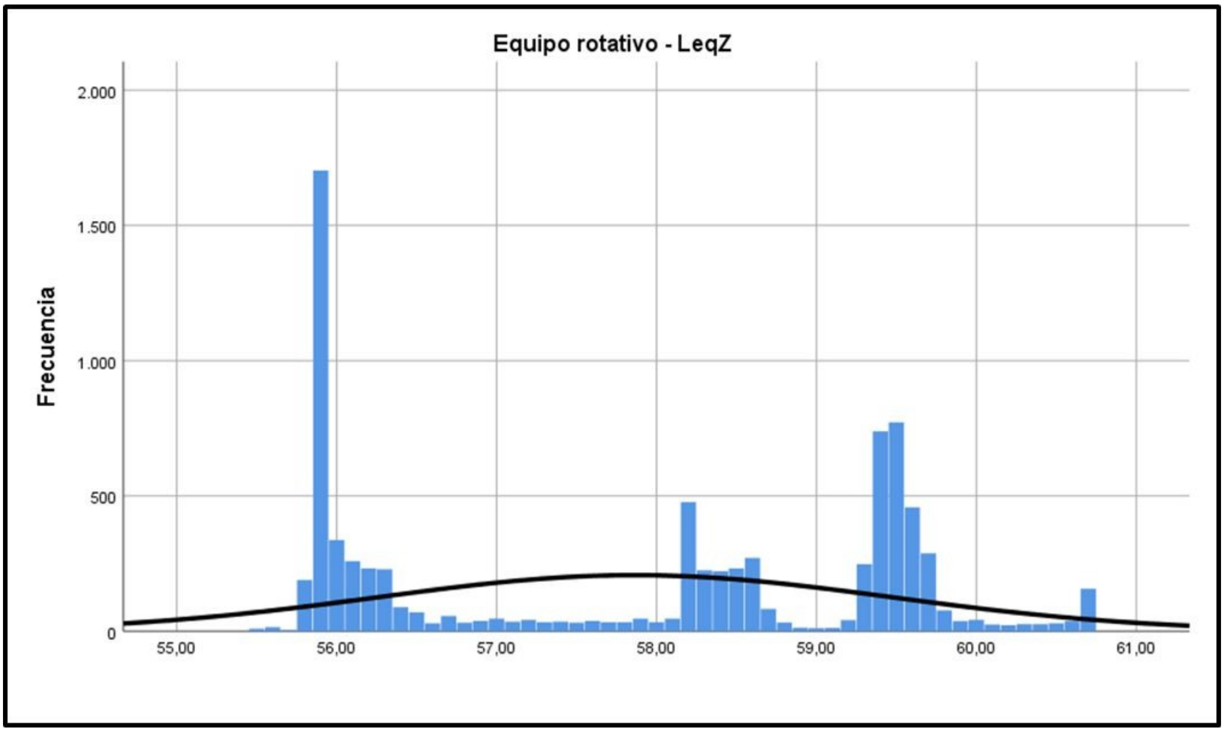

Figure 6.2.5: Histogram of rotating equipment vs. theoretical normal distribution curve.