Acoustics'08 Paris
June 29-July 4, 2008
www.acoustics08-paris.org

A proposal for doing a touch of anova with noise levels

Michel Maurin

INRETS-LTE, case 24, 69 675 Bron cedex, 69 500 Bron, France
maurin@inrets.fr

Because noise levels are logarithms of additive variables, the usual numerical processings are not suitable for them, (here 60 + 60 ≠ 120 !). One has to deal with them differently, following another rules which may be called the 'logic of levels'. Of course the question is correctly resolved with the energetic mean, but it fails as soon as one has to deal with variances and covariances implying levels in dB. Some solutions may be developed which are necessarily approximate compromise between the logic of levels and the usual arithmetical logic. Here we introduce a new statistic, the h-dispersion which has many technical properties of variance and fortunately takes account of the logarithmic status of data. This new compromise may be used for levels in the case of design of experiments such as a one-way design for acoustical comparisons (before and after an operation for instance), and also, partially, in the case of two-way designs. However this approximation is an improvement in relation to crude ANOVA on levels.

1 Introduction

In environmental acoustics, the principle variables with which we are working are levels, noise, pressure, power, … of the general form $L_s = 10 \log s/s_0$ [5], and they are respective logarithms of other magnitudes. Of course they are also numerical data, but as logarithms they can not follow the same calculations' rules than with ordinary numbers ; very logically they have to respect other specific rules and their set may be called "the logic of levels", [7, 8, 9]. The question is often raised when one has to deal with statistical calculations and processings which follow the arithmetic rules by default, and one has to develop compromises between arithmetic and logarithmic logics of calculations.

Previously we introduced an "equivalent variance" as an order 2 (statistical) moment convenient and useful in regressions techniques ; here we propose another statistic for levels L_i which has some analogous properties with the variance, and which may be used for the populations' comparisons as we usually do in design of experiments and analysis of variance, but here between levels' populations.

2 Some recalls on statistics and logic of levels

2.1 About the mean of levels

The mean of levels is not veritably problematic because one deals straightaway (and fairly and physically) with the energetic mean $10 \log\{1/I \sum_i 10^{Li/10}\}$ of several levels L_i and the arithmetic mean of additive levels powers $10^{Li/10}$ (under acoustical independancy). This is the equivalent level, and also in mathematics the h-mean $h^{-1}\{1/I \sum_i h(y_i)\}$ of y_i data for every bijective application h ; the arithmetical mean belongs to the transformed space of y_i and the h-mean returns in the y_i space with the inverse transformation h^{-1}, [3].

In acoustics we obtain the equivalent level with the applications $h(L) = 10^{L/10}$ and $h^{-1}(x) = 10 \log x$, and Aczél notes that this special h-mean has the "translativity" property, [1]. That is to say that whenever levels are translated of the same quantity δ, the corresponding equivalent level $10 \log \{1/I \sum_i 10^{(L_i+\delta)/10}\} = 10 \log \{1/I \sum_i 10^{L_i/10}\} + \delta$ also, as an usual arithmetical mean.

2.2 About the variance

The question is raised more severely when one looks for a dispersion statistic because there is the unfair mean levels $1/I \sum_i L_i$ in the variance (as for every centered moment of order ≥ 2) ; then the variance and all the covariances do not formally agree with the logarithmic status of levels.

2.3 The transform of variables in statistics, the equivalent variance

A first way to reconcile both arithmetical and logarithmic logics may be attempted with the change of variables as is generally done in statistics. One considers an X variable and a transformed one Y = g(X) with a regular (continuous, derivable, …) transformation h. Then knowing the first X moments $E(X) = m_x$ and $var(X) = \sigma_x^2$ we may have an approximation for the same moments of the transformed variable $E(Y) = m_y$ and $var(Y) = \sigma_y^2$ with the help of the technical expansions, [4] :

$E(Y) = m_y = g(m_x) + \sigma_x^2/2\, g''(m_x) + \dots$,

and $\sigma_y^2 = \sigma_x^2\, g'^2(m_x) + \dots$.

Concerning the logic of levels we take the notation Y = f(X), and f(x) = 10 log x.

a) Every time X variable is additive, the statistics m_x and σ_x^2 are meaningful and the other m_y and σ_y^2 are not. But we note that the first term of the E(Y) expansion is the f^{-1}-mean for Y data, here one uses the arithmetical mean for the additive $f^{-1}(Y)$, and after that the f^{-1}-mean may be considered as a central tendency statistic in the Y space, an equivalent mean $m_{eqy} = f(m_x)$.

b) For the variance, by analogy, one takes the first member $\sigma_x^2\, f'^2(m_x)$ of the σ_y^2 expansion in the Y space. This "equivalent variance" $\sigma_{eqy}^2 = \sigma_x^2\, f'^2(m_x)$ comes from m_x and σ_x^2 statistics ; and with the inverse function $h = f^{-1}$ we note it is also the usual variance of the $h(y_i)/h'(m_{eqy})$ data because here $m_{eqy} = f(m_x)$ and more generally $h'[f(m_x)]\, f'(m_x) = 1$.

In acoustics Y_i data are the levels L_i and transformed variables $x_i = 10^{L_i/10}$ are additive powers. One has f(x) = M ln x with M = 10/ln10, $f'(x) = M/x$ and $h(y) = 10^{y/10}$, and finally $\sigma_x^2 = 1/n \sum 10^{L_i/5} - (1/n \sum 10^{L_i/10})^2 = 1/n \sum 10^{L_i/5}$ -

$10^{Leq/5} = 10^{Leq/5}\{1/n \sum 10^{(L_i-Leq)/5} - 1\}$, $f'(m_x) = M\, 10^{-Leq/10}$. It results that the expression $\sigma_{eqL}^2 = M^2 \{1/n \sum 10^{(L_i-Leq)/5} - 1\}$ is depending on levels, and as a variance it is invariant by any δ translation on levels, (see the Leq translativity, [1]).

3 Another para-variance

The equivalent variance is useful for regressions [8, 9], but not for other classical fields where variance is implied, for instance in the analysis of variance and the related designs of experiments.

3.1 h-mean and h-dispersion

We come back here to general data x_i (not acoustics). Whenever the h function is monotonous and convex one obtains the classical inequality $h^{-1}\{1/I \sum_i h(x_i)\} \geq 1/I \sum_i x_i$, [3], then the difference $h^{-1}\{1/I \sum_i h(x_i)\} - 1/I \sum_i x_i$ between h-mean and arithmetical mean is positive (or zero). It is a first common property with the variance.

One may also show that this difference is only depending on $\Delta x_i = 1/I \sum_j x_j - x_i$ if and only if the h function is an exponential function e^{cx}, [10]. This means that only in this special case the positive difference $h^{-1}\{1/I \sum_i h(x_i)\} - 1/I \sum_i x_i$ has a second common property with the usual variance $\sigma^2 = 1/I \sum_i \Delta x_i^2$. For these two reasons this difference is called an "h-dispersion" noted h-disp(x_i), with the detailed expression h-disp(x_i) $= 1/c \ln (1/I \sum_i e^{c\Delta x_i})$. Note that this name is sligthly ambiguous because the h-dispersion is defined only for the special h exponential functions e^{cx}, but it is like that by semantic continuity with the more general h-mean.

3.2 A property for mixed populations

Lastly the h-dispersion has another very interesting common property with the variance, it concerns the mixing of statistical sub-populations, [10].

a) Classically, faced with a mixed population of data x_{pr}, with a p index for the sub-populations p = 1…P, and an r index for repeat, r =1…$n_p \geq 2$, one has first the means and variances for each p population, and second the additive decomposition of the total variance

total variance = mean of variances(p) + variance of means(p).

The variances(p) are the respective internal variances of each sub-population and their mean is called the intra-variance, while the variance of means(p) is called the inter-variance and it is due to the true differences between all the sub-populations. Then the total variance has two very separate parts coming from different origins, the intra part and the inter part.

b) one observes a quite similar property for the total h-dispersion when one mixes several sub-populations, coming from the internal h-dispersions(p) and also the respective h-means(p), [10] :

total h-dispersion = mean of h-dispersions(p) + h-dispersion of h-means(p).

Following this, globally, the first term of this additive decomposition of the total h-dispersion is a h-dispersion intra, with expression h-disp intra = mean of h-dispersions(p) ; and the second term is a h-disp.inter, with expression h-disp.inter = h-dispersion of h-means(p).

4 Another compromise between arithmetical rules and logic of levels

Until now all the considerations in § 3 are algebraic and statistical, but they find an immediate echo among acoustics because of the logarithm and because an equivalent level is an h-mean with the exponential function $10^{y/10} = e^{y/M}$. Of course we find Leq(L_i) $\geq 1/I \sum_i L_i$ again, and we pose h-disp(L_i) = Leq(L_i) - $1/I \sum_i L_i$ as an acoustical h-dispersion for levels. For previous reasons it may be considered as another compromise between the arithmetic calculations and the logic of levels, with a non artificial introduction of the two respective means.

4.1 A levels population's comparison with the acoustical h-dispersion

Here we examine the simple case of two sub-populations of levels, P = 2, the first one with n_1 levels L_i and a second one with n_2 levels G_j. For instance it may be the case for the comparison between acoustical situations on the same site, before and after any noticeable change of traffic management, ….

a) the definitions give us
h-disp(pop1) = h-disp(L_i) = Leq(L_i) - $1/n_1 \sum_i L_i$,
and h-disp(pop2) = h-disp(G_j) = Leq(G_j) - $1/n_2 \sum_j G_j$

b) then we obtain

h-disp intra = n_1/n_+ {Leq(L_i) - $1/n_1 \sum_i L_i$} + n_2/n_+ {Leq(G_j) - $1/n_2 \sum_j G_j$} = $1/n_+$ {n_1 Leq(L_i) + n_2 Leq(G_j)} - $1/n_+$ [$\sum_i L_i + \sum_j G_j$] ,

c) h-disp inter = Leq{Leq(L_i), Leq(G_j)} - $1/n_+$ [n_1 Leq(L_i) + n_2 Leq(G_j)] ,

and finally the additive decomposition for the total h-dispersion with the two parts h-disp inter + h-disp intra. Then we obtain the easy possibility to compare the two, and to note which is the most important one.

4.2 An acoustical effect's characterization

We may go beyond, and with the help of the h-disp.inter we may characterize the acoustical effect there is between both the {L_i} and {G_i} sub-populations.

If we suppose this effect is due to a collective translation which may be represented by the Leq difference δ = Leq(G_j) - Leq(L_i), (positive, otherwise Leq(L_i) - Leq(G_j)), one obtains

h-disp inter = Leq{Leq(L_i), Leq(L_i) + δ} - Leq(L_i) - $\delta\, n_2/n_+$

$= 10 \log[1/n_+ (n_1\ 10^{Leq(Li)/10} + n_2\ 10^{Leq(Li)/10}\ 10^{\delta/10})]$ - $Leq(L_i) - \delta\ n_2/n_+ = Leq(L_i) + 10 \log(1 + n_2/n_+\ (10^{\delta/10}-1))$ - $Leq(L_i) - \delta\ n_2/n_+ = 10 \log(1 + n_2/n_+\ (10^{\delta/10}-1)) - \delta\ n_2/n_+$ or finally $M \ln(1 + n_2/n_+ (e^{\delta/M}-1)) - \delta\ n_2/n_+$.

Then the inter h-dispersion between the two sub-populations is a simple function of the δ Leq difference. It may be seen as a trite result, but it comes from the algebraical construction of a compromise of two logics (arithmetic and levels ones) which is not at all trite, and the resulting additive decomposition in two parts of the total h-dispersion, an internal sub-population one - the h-disp.intra because each sub-population has its own dispersion, and the inter h-dispersion which is due to the true acoustical effect between sub-populations.

Moreover $f_{n1,n2}(\delta) = M \ln(1 + n_2/n_+ (e^{\delta/M}-1)) - \delta\ n_2/n_+$ is defined for $\delta \geq 0$, the function is zero at the origin, growing and convex on R^+ and alike to $\delta\ n_1/n_+ + M \ln n_2/n_+$ when δ tends towards $+\infty$.

5 A random term and statistical imprecision

The previous calulations do not account for the presence of the random terms which practically affect any measure. Here we try to introduce some of them in accordance with the logic of levels cares.

5.1 A random term (approximate)

a) let a signal s_i corresponding to the measured magnitude s and the random term ε_i , the classic model is $s_i = s + \varepsilon_i$ with an additive formulation for the perturbation ;

b) for the level $L_i = 10 \log\ s_i/s_0$ one has to consider the errored level $L = 10 \log\ (s+\varepsilon_i)/s_0$; and here the general logarithm properties give $\ln (z + \varepsilon) = \ln z + \ln (1+ \varepsilon/z)$ # $\ln z + \varepsilon/z$ for the rather weak perturbations such as ε/z is small in comparison with the unit.

In the case of acoustics it yields $10 \log (s+\varepsilon_i)/s_0 = 10 \log [s/s_0 (1+ \varepsilon_i/s)] = 10 \log s/s_0 + M \ln (1+\varepsilon_i/s)$ # $10 \log s/s_0 + M\ \varepsilon_i/s$.

The following result is that when a signal is affected by an additive and rather weak perturbation in relative terms, its level $10 \log s/s_0$ which is (of course) logarithmic and not additive is in spite of this affected by an additive errored dimensionless term ε_{Li} following the approximate relation $M\ \varepsilon$ # $s\ \varepsilon_{Li}$, [6].

c) and here as one deals with levels and not signals themselves, this is the term ε_L one has to introduce and modelize, for instance centered with a variance σ^2.

5.2 Consequences on statistical calculations

With several levels $L_i = 10 \log (s+\varepsilon_i)/s_0$ i = 1…n, or $10 \log s/s_0 + \varepsilon_{Li}$ with related random and small independant terms ε_i and ε_{Li}, one has $Leq(L_i) = 10 \log\{1/n \sum 10^{Li/10}\} = 10 \log(1/n \sum s/s_0 + 1/n \sum \varepsilon_i/s_0) = 10 \log\{s/s_0 (1 + 1/ns \sum \varepsilon_i)\} = L + 10 \log (1 + 1/ns \sum \varepsilon_i)$ # $L + M/ns \sum \varepsilon_i$.

Then there is a new additive errored term e_{Leq} for L, and because all the relations $M\ \varepsilon_i$ # $s\ \varepsilon_{Li}$ this term is equal to $\varepsilon_{L.} = 1/n \sum \varepsilon_{Li}$, here centered with variance σ^2/n.

5.3 Consequences on the h-dispersions

a) about the h-dispersion inter,

in presence of the both levels sub-populations $L_i = L_{pop1} + \varepsilon_i$ and $G_j = L_{pop2} + \varepsilon_j$, one has $Leq(L_i) = L_{pop1} + \varepsilon_{L.pop1}$ with a random additive term of variance σ^2/n_1 , $G_j = L_{pop2} + \varepsilon_j$ and $Leq(G_j) = L_{pop2} + \varepsilon_{L.pop2}$ of variance σ^2/n_2 , $\varepsilon_{L.pop1}$ and $\varepsilon_{L.pop2}$ independant.

* with the δ, acoustical effect $L_{pop2} = L_{pop1} + \delta$, the arithmetical mean $1/n_+ \{n_1 Leq(L_i) + n_2 Leq(G_j)\}$ becomes $L_{pop1} + \delta\ n_2/n_+ + 1/n_+ (n_1\ \varepsilon_{L.pop1} + n_2\ \varepsilon_{L.pop2})$ with the additive random term $n_1/n_+\ \varepsilon_{L.pop1} + n_2/n_+\ \varepsilon_{L.pop2}$.

* for the h-mean of the two Leq we obtain $Leq\{Leq(L_i), Leq(G_j)\} = 10 \log\{1/n_+ (n_1\ 10^{Lpop1/10}\ 10^{\varepsilon L.pop1/10} + n_2\ 10^{Lpop2/10}\ 10^{\varepsilon L.pop2/10})\}$

$= 10 \log\{1/n_+ (n_1\ 10^{Lpop1/10}\ 10^{\varepsilon L.pop1/10} + n_2\ 10^{Lpop1/10}\ 10^{\delta/10} 10^{\varepsilon L.pop2/10})\}$

$= L_{pop1} + 10 \log\{1/n_+ (n_1\ 10^{\varepsilon L.pop1/10} + n_2\ 10^{\delta/10}\ 10^{\varepsilon L.pop2/10})\}$

$L_{pop1} + 10 \log\{1/n_+ [n_1 (1 + \varepsilon_{L.pop1}/M) + n_2\ 10^{\delta/10} (1 + \varepsilon_{L.pop2}/M)]\}$

$$= L_{pop1} + 10 \log\{1/n_+ [n_1 + n_2\ 10^{\delta/10} + \frac{n_1\ \varepsilon_{L.pop1} + n_2\ 10^{\delta/10}\ \varepsilon_{L.pop2}}{M}]\}$$

$$\# \ L_{pop1} + 10 \log(1/n_+ (n_1 + n_2\ 10^{\delta/10})) + \frac{n_1\ \varepsilon_{L.pop1} + n_2\ 10^{\delta/10}\ \varepsilon_{L.pop2}}{n_1 + n_2\ 10^{\delta/10}} .$$

* by mere difference the h-dispersion inter has a constant component $f_{n1,n2}(\delta)$ and a random component $n_1 (\frac{1}{n_1 + n_2\ 10^{\delta/10}} - \frac{1}{n_1 + n_2})\ \varepsilon_{L.pop1} + n_2 (\frac{10^{\delta/10}}{n_1 + n_2\ 10^{\delta/10}} - \frac{1}{n_1 + n_2})\ \varepsilon_{L.pop2}$ of variance $\sigma^2 \frac{n_1 n_2}{n_+} \frac{(10^{\delta/10}-1)^2}{(n_1 + n_2\ 10^{\delta/10})^2}$. This variance is positive, equal to 0 at the origin, growing on R^+ and alike to $\sigma^2\ n_1/n_+\ n_2$ when δ tends towards $+\infty$.

As a result the presence of random errored terms in noise levels provides a random behaviour for the h-dispersion inter, with a known variance. In these conditions, knowing the standard deviation $\text{stdh}(\delta) = \sigma\,(\frac{n_1 n_2}{n_+})^{1/2}\,\frac{10^{\delta/10}-1}{n_1 + n_2\,10^{\delta/10}}$ we may deduce some confidence intervals for δ such as δ ± 1,96 stdh(δ) for the level of confidence 0,95, or δ ± 1,65 stdh(δ) for the level 0,90.

b) unfortunately the same calculations provide an additive errored term zero when applied to the h-dispersion intra. In this case the approximations due to $\ln (z + \varepsilon)$ # $\ln z + \varepsilon/z$ are too crude and schematic, and cannot be convenient.

6 Towards two-way designs

The convenient common property related to an additive decomposition for the variance and the h-dispersion, § 3.2, cannot be extended to the designs with two different and independent factors, [2], because of the term of interaction between factors, [10]. However we have an interesting partial continuation.

Let L_{ijr} be the levels data with a first bimodal factor f_1 of index i = 1, 2, a second bimodal factor f_2 of index j= 1, 2, and the index r for the repetitions (new notations). In this two-way design we cannot obtain an additive decomposition of the total h-dispersion of the L_{ijr}, but of course we may consider two separate and resulting one-way designs. First the $L_{ir'}$ data with levels sorted only by the i index of factor 1 and the repetition index r' implying the factor 2 ; and second the $L_{jr''}$ data with levels sorted only by the j index of factor 2 and the repetition index r" implying the factor 1. For each of them we have the previous results of § 4 and § 5 in order to make the evaluation of the respective acoustical effect δ_1 and δ_2, and consequently we may observe which factor has the most important effect, and compare their respective magnitude, (possibly with confidence intervals, § 5).

7 Discussion and some conclusions

In acoustics the research for a compromise between the arithmetical usual calculations and the logic of levels needs some technical keenness in order to bring closer their more or less incompatible calculating rules. It is necessary for each pole to move a little, because there is a theoretical as well as a practical need for means, dispersions and statistical processings.

The h-dispersion in § 3 and § 4, and the random errored terms in § 5 are some possible convenient bridges. And more specifically the h-dispersion renders possible an additive decomposition in two parts for a tool similar to variance, as one usually does in a one-way analysis of variance. Later the random complements yield to associated confidence intervals for the evaluation of an acoustical effect between two sub populations of noise levels.

References

[1] Aczél J., 1987, A short course on functional equations, D. Reidel P. Company

[2] Brownlee K. A., 1965, Statistical theory and methodology in science and engineering, J. Wiley and sons

[3] Hardy G.H., Littlewood J.E., Pólya G., 1934, Inequalities, Cambridge Un. Press

[4] Johnson N.L., Kotz S., 1970, Continuous univariate distributions, volume 1, New-York, J. Wiley and sons

[5] Liénard P., 1978, La famille décibel, Masson

[6] Maurin M., 1998, Integrated versus instantaneous measurements continued, what about precision ?, Euronoise 98, Munich

[7] Maurin M., 1999, "The logic of levels", What about variances, covariances and linear regression ?, Inter-Noise 99, Fort Lauderdale

[8] Maurin M., 2000, "The logic of levels", What about multilinear regression ?, Inter-Noise 2000, Nice

[9] Maurin M., 2003, Logarithme, décibels et logique des niveaux, rapport INRETS-LTE n° 0304, version de juin 2006

[10] Maurin M., 2007, Un amusement sur la dispersion des données, la h-dispersion, Journées SFdF 2007, Angers