inter.noise 2000
The 29th International Congress and Exhibition on Noise Control Engineering
27-30 August 2000, Nice, FRANCE

I-INCE Classification: 0.0

THE 'LOGIC OF LEVEL' CONTINUED, WHAT ABOUT MULTI-VARIABLE DATA PROCESSING

M. Maurin

INRETS-LTE, 25 avenue François Mitterrand - Case 24, 69675, Bron, France

Email: maurin@inrets.fr

Keywords:

ABSTRACT
This note deals with statistical precautions, every time one intends to run statistical processings on multivariate data including noise levels (formally irrelevant for means and variances for instance ...).

1 - INTRODUCTION

Because levels are not usual numbers, but logarithms, they might not be handled as usual numbers in mathematical or statistical processing (and "blind" softwares also), for instance the addition is not relevant. They ought to be employed in accordance with ad hoc and rigorous rules that we proposed to name "the logic of levels".
The correctness is trite and classic for first order moment or average with the well-known equivalent noise levels L_{eq}; but for second order moments such as variances, covariances and regression coefficients we need to introduce new precautions inspired by the transformations of variables in Statistics. They appear as a *compromise* in order to keep useful tools in statistics while avoiding the non-additivity of levels [1].

2 - A MULTIVARIATE CONTEXT

These previous results are related to two variables regressions such as
i) $u = a + b\ L$ or $u = a + b\ y$, and i2) y or $L = a_2 + b\ _2\ u$
with a transformed y or $L = f(x)$, or again
ii) $t = h(u) = a' + b'\ y$ or $a' + b\ '\ L$, and ii2) y or $L = a'_2 + b'_2 h\,(u) = a'_2 + b'_2 t$
with a second transform $t = h(u)$, in the case of u and x are ordinary numerical variables for which arithmetical statistics m_u, m_x, σ^2_{u}, σ^2_{x} and cov(u,x) are relevant; here L for level when $f(x) = 10 \log x$ in acoustics.
But we know that questions are often multivariate, then we have to deal with the more global model

$$\mathrm{t = a + \Sigma_{j=1...J}\ b_j\ \ y_j}$$

including all the transformed variables $t = h(u)$, $y_j = f_j(x_j)$.
The classical matter is to minimize the sum of squares

$$\mathrm{Q = \Sigma_{i=1...n}\ \ (\ t_i - a - \Sigma_{j=1...J}\ b_j\ y_{ji})^2}$$

leading to usual normal Lagrange equations

$$\frac{\partial \mathrm{Q}}{\partial \mathrm{a}} = -2\Sigma_{\mathrm{i=1...n}}\ \ (\mathrm{t_i - a - \Sigma_j\ b_j\ y_{ji}}) = 0$$

$$\frac{\partial \mathrm{Q}}{\partial \mathrm{b_k}} = -2\ \Sigma_{\mathrm{i=1...n}}\ \mathrm{y_{ki}}\ \ (\ \mathrm{t_i - a - \Sigma_j\ b_j\ y_{ji}}) = 0,\ \ k = 1 \dots J$$

or

$$\frac{\partial \mathrm{Q}}{\partial \mathrm{a}} = -2\ \mathrm{n}\ \left(\mathrm{m_t}\ - \mathrm{a} - \Sigma_\mathrm{j}\ \mathrm{b_j}\ \mathrm{m_{yj}}\right)$$

$$\frac{\partial \mathrm{Q}}{\partial \mathrm{b_k}} = -2\ \mathrm{n}\ \left[\mathrm{cov}\left(\mathrm{t},\mathrm{y_k}\right)\ + \mathrm{m_t}\ \mathrm{m_{yk}} - \mathrm{a}\ \mathrm{m_{yk}} - \mathrm{b_k}\ \left(\ \sigma^2_{\mathrm{yk}} + \mathrm{m}^2_{\mathrm{yk}}\right)\right.$$
$$\left. -\Sigma_{\mathrm{j}\neq\mathrm{k}}\ \mathrm{b_j}\ \left(\mathrm{cov}\left(\ \mathrm{y_k},\mathrm{y_j}\right) + \mathrm{m_{yk}}\ \mathrm{m_{yj}}\right)\right]$$
$$= -2\ \mathrm{n}\ \left[\ \mathrm{cov}\left(\mathrm{t},\mathrm{y_k}\right) - \mathrm{b_k}\ \sigma^2_{\mathrm{yk}} - \Sigma_{\mathrm{j}\neq\mathrm{k}}\ \mathrm{b_j}\ \mathrm{cov}\left(\mathrm{y_k},\mathrm{y_j}\right) + \mathrm{m_{yk}}\ \left\{\mathrm{m_t} - \mathrm{a} - \Sigma_\mathrm{j}\ \mathrm{b_j}\ \mathrm{m_{yj}}\right\}\right],$$

that is to say

$$m_t - a - \Sigma_j\ b_j\ m_{yj} = 0$$
$$\mathrm{b_k}\ \sigma^2_{\mathrm{yk}} + \Sigma\ _{\mathrm{j}\neq\mathrm{k}}\ \mathrm{b_j}\ \mathrm{cov}\left(\mathrm{y_k},\mathrm{y_j}\right) = \mathrm{cov}\left(\mathrm{t},\mathrm{y_k}\right)$$

$k = 1...J$.
We observe that the b_k coefficients are merely solutions of a $J \times J$ linear system without a, and lastly

$$\mathrm{a} = \mathrm{m_t} - \Sigma_\mathrm{j}\ \mathrm{b_j}\ \mathrm{m_{yj}}$$

3 - THE MULTIVARIATE LOGIC OF LEVELS

This is a very convenient situation for the "logic of levels" precautions, because this renders easily a way to calculate first and second order moments for variables (called equivalent moments in accordance for equivalent level L_{eq}).
We remember that

$$\mathrm{m_{eqy}} = \mathrm{f}\left(\mathrm{m_x}\right)\ = \mathrm{f}\left(\mathrm{m_{f^{-1}(y)}}\ \right)$$

for first order, and

$$\sigma^2{}_{\mathrm{eqy}} = \sigma^2_\mathrm{x}\ \mathrm{f}'^2\ \left(\mathrm{m_x}\right) = \sigma^2_{\mathrm{f}^{-1}\ (\mathrm{y})}\ \mathrm{f}'^2\left(\mathrm{m_{f^{-1}(y)}}\right)$$

$$\mathrm{cov_{eq}}\left(\mathrm{t},\mathrm{y}\right)\ = \mathrm{cov}\left(\mathrm{u},\mathrm{x}\right)\ \mathrm{h}'\left(\ \mathrm{m_u}\right)\ \mathrm{f}'\left(\ \mathrm{m_x}\right) = \mathrm{cov}\left(\ \mathrm{h}^{-1}\left(\mathrm{t}\right),\mathrm{f}^{-1}\ \left(\mathrm{y}\right)\right)\ \mathrm{h}'\left(\mathrm{m_{h^{-1}(t)}}\ \right)\ \mathrm{f}'\left(\mathrm{m_{f^{-1}(y)}}\right)$$

for second order moments [1]; then we may logically define b_{eqk} equivalent coefficients as solutions of the renewed $J \times J$ system

$$\mathrm{b_{eqk}}\ \sigma^2_{\mathrm{eqyk}} + \Sigma_{\mathrm{j}\neq\mathrm{k}}\ \mathrm{b_{eqj}}\ \mathrm{cov_{eq}}\left(\mathrm{y_k},\mathrm{y_j}\right) = \mathrm{cov_{eq}}\left(\ \mathrm{t},\mathrm{y_k}\right),\ k = 1\ldots J$$

and the last a_{eq} coefficient as the final value

$$\mathrm{a_{eq}} = \mathrm{m_{eqt}} - \Sigma_\mathrm{j}\ \mathrm{b_{eqj}}\ \mathrm{m_{eqyj}}$$

once b_{eqk} are known.

4 - THE MULTIVARIATE ADJUSTMENT

1 – Of course, as for two variables these equivalent coefficients are a compromise between the least square approach in order to get the best regression coefficients, and the need of statistical relevancy for means, variances and covariances and discarding meaningless calculations [1].
Also we have to note that all of this is obtained in a very convenient and favorable situation because all researched coefficients are acquired in a mere two-steps numerical processing, once suitable statistics σ^2_{eqy}, $\mathrm{cov_{eq}}(t,\ y)$ and m_{eqy} are introduced.
2 – Moreover, a final remark renders here the logic of levels easier again; indeed this logic is the special case of logic of transform in statistics [2] and we observed that it comes from the use of the first terms of Taylor expansions in the vicinity of $m_x = E(X)$. Consequently if we pose new variables called "equivalent variables"

$${}_{\mathrm{eq}}y_k = f_k\left(m_{xk}\right) + \left(x_k - m_{xk}\right)\ f'_k\left(m_{xk}\right)$$
$${}_{\mathrm{eq}}\mathrm{t} = \mathrm{h}\left(\mathrm{m_u}\right) + \left(\mathrm{u} - \mathrm{m_u}\right)\ \mathrm{h}'\left(\mathrm{m_u}\right)$$

we note arithmetical statistics are relevant (because they are linear functions in x_k and u) and that we get $m_{eq y_k} = f_k \; (m_{xk})$, that is to say our previous equivalent mean m_{eqyk}, $\sigma^2_{eq y_k} = \sigma^2{}_{xk} \; f'^2_k \, (m_{xk})$, that is to say our equivalent variance σ^2_{eqyk}, and the same result again for ${}_{eq}t$ and covariances.
This means that the ordinary statistical processings applied to these equivalent variables provide the equivalent moments of first and second order which are needed by the logic of transform or the logic of levels.
And because all the usual statistical processing are included in common softwares, all of them provide immediately suitable results for levels under the condition that the initial variables y_k and t are replaced with the equivalent variables

$$ {}_{eq}y_k = f_k \left(m_{f_k^{-1}(y_k)} \right) + \left\{ f_k^{-1} \, (y_k) - m_{f_k^{-1}(y_k)} \right\} \; f'_k \left(m_{f_k^{-1}(y_k)} \right) $$

and

$$ {}_{eq}t = h \left(m_{h^{-1}(t)} \right) \; + \left\{ h^{-1} \, (t) \; - m_{h^{-1}(t)} \; \right\} \; h' \left(m_{h^{-1}(t)} \right) $$

3 − **In conclusion**, the numerical and statistical incorrectness when dealing with crude usual statistics for levels in processings are quite resolved when one replaces original (non arithmetic) data with above equivalent transformed data ${}_{eq}y$ and ${}_{eq}t$ whatever f_k and h transformations are, and softwares are allowed every time the substitution is made before the first run (by the mean of the implementation of a short previous adaptive program). This is a very convenient conclusion which may reconcile the data processings and the necessary numerical precautions implied by levels.
(please note here a ***call for multivariate data*** including noise levels and other variables in order to apply equivalent variables and compare with crude results).

REFERENCES

1. **Maurin M.**, The "logic of levels": What about variances, covariances and linear regression, In *Internoise 99*, pp. 1395-1398, 1999

2. **Kendall G., Stuart A.**, *The advanced theory of statistics*, Griffin, 1963