PLANNING OF EXPERIMENTAL WORK AND ANALYSING OF THE MEASURED DATA

by

Dr. W. Schmid, Grundlagen PKW Aufbauten, Daimler-Benz AG. Sindelfingen

A vast number of data and results have been obtained in recent years through intensive examinations of road accidents and their simulation on experimental plants. Surprising statements were frequently derived, the credibility of which gave many an expert a headache. Now as before it is the biomechanical principles which, because still little explored, aggravate a consequent improvement of vehicles in the sense of increasing the passive safety. With the relevant biomechanical experiments, it is often extremely difficult to uncover safe and generally valid connections, since a great number of influence parameters always differ from experiment to experiment. An important step with the treatment of such data quantities is the use of socalled multivariant systems increasingly realised in modern statistics. Four relevantly applicable multivariant models will be described here. Prerequisites to establish practically interpretable results take effect already when planning the experimental work. Particularly useful information can be gained from experiments only if the marginal conditions from the project definition to the final report are maintained.

Some projects showed that six important steps must be observed when conducting experimental investigations:

1. Defining the projects.
2. Planning the experiments.
3. Measuring.
4. Calculating.
5. Interpreting the results.
6. Reporting.

If this sequence is changed, first errors will creep in due to the fact that many experiments are pointless, often causing more confusion than contributing towards a solution of the set task.
Of course, it will often be necessary to jump from a later step back to an earlier one.

1. Defining the project.

The task for an experimental investigation can be set by principals, originate from one's own initiative or be the resulting project from an earlier investigation. Careful studies of literature and good contacts to other institutions and groups of researchers help to avoid double work. The investigation target must be clearly formulated and should be set down in writing to avoid later differences. The task must be carefully demarcated. The things that cannot be established or are to be established with the experimental investigation should be clear.

2. Planning of experiments.

As a rule, the scope and the number of the experiments are limited by the means available. Such means are experimental installations which exist or must be newly established, measuring instruments, the availability of suitably trained persons, the frequently set time and the availability of measuring objects - be it vehicles, volunteers or similar.
The choice of the experimental installation is determined by the objective of the investigation; measuring accuracy, systematic and coincidental measuring errors are important criteria for the selection of the measuring instruments.

The characters to be measured can be divided into independent and dependent characters; dependent characters are such quantities, the influencing through changes of the independent characters of which are to be uncovered by the experiments.

The independent characters can be divided into two classes; into set independent characters which cannot be changed by the experimentator, such as height, weight etc. and selectable, independent characters which can be determined by the experimentator from case to case, such as impact speed, acceleration etc.

The experiment plan has the form of an n-dimensional matrix (with n independent characters), in which the number of experiments with certain preset character configuration is entered as elements. The experimentator now has the possibility of combining the preset independent characters with the selectable ones in such a way that all characters initially considered as independent will finally be independent of each other as far as possible. This can be checked by dividing the n-dimensional experiment matrix into

(1) $$N = \frac{n * (n - 1)}{2}$$

2 dimensional confidence tables for each 2 independent characters and establishing the mutual independence with the chi-square-test or by means of the contingency coefficient.
If a genuine independence of all "independent" characters cannot be attained, the number of criteria will have to be reduced to

as complete as possible a set of actually independent characters. The factor analysis has proved to be a useful aid for this purpose.

3. Measuring.

Measurable are all characters the values of which can be read on a scale. Such cardinally scaled quantities (such as HIC, a, ...) are particularly suitable for a system analysis with the aid of models. However, it is also possible to include ordinally scaled quantities (such as AIS and similar). The interpretation of calculation results is then somewhat more difficult.

However, if characters are only nominally scaled (e.g. type of the experimental dummy, type of the restraint systems, ...), the data quantity must be divided into suitable classes in order to prove the influence of these criteria. With the aid of the chi-square-test or the contingency coefficient it can then be determined whether an influence on the dependent character is considerable or not. Prior to a measurement it is therefore necessary to always establish the most practical scale for each character. It must be maintained, since subsequent experiments can no longer be included in the total quantity of the measuring data. It is practical to record additional information regarding the flow of the experiment in a measuring protocol in addition to the numerical values (and the accuracy) - even if at first considered as subordinate marginal conditions - and to observe and note down special events. Such protocols are an extremely valuable aid when results obtained from a model calculation are to be interpreted.

4. Calculating.

After and frequently even during the performance of extensive measurements the measuring results will be graphically represented. As a rule, the scales of 2 characters, an independent one and one depending on the former, will be used as abscissa and ordinate axes and the relevant realisations of these two criteria be entered as measuring points. This provides a first optical impression as to the interdependence of these two criteria.

4.1 Correlations

With nominally scaled characters the chi-square-test or the contingency coefficient indicates whether a noticeable influence of a certain independent character on an observed dependent character can be proven. If all data are classified according to the nominally scaled criteria, an order of rank (with respect to the "effect") can often be established. In this way nominal scales can often be converted into a rank or ordinal scale. With certain limits, nominally scaled criteria can be included in the further calculation in this way. An initial impression regarding more or less intense relations between ordinally or cardinally scaled characters is given by the correlation coefficient, e.g. calculated according to BRAVAIS-PEARSON and SPAERMAN.

4.2 Regression

With a regression calculation it is assumed that the relation between two characters can be approximated by a linear function. The approximation, as a rule, takes place with the method of the smallest quadratic deviations or according to the method "Maximum Likelihood Quotient".

4.3 Models

With a greater number of independent characters, influences of individual independent characters are frequently concealed by the influences of other characters, especially when the mutual influence is roughly indentical. In such cases it is advisable to employ so-called multivariant systems. With these systems the type of the approximating function is always selected and attempts are made to find the function parameters so that for instance the squared differences between measured value and the approximation become smallest. We will introduce four different types of function, popularly called models.

The data of each experiment are classified in a line vector, the elements of which are the characters under consideration. The total number of the evaluable experiments determines the number of such line vectors which are now arranged in form of a matrix. We obtain the matrix of the measuring data, the number of columns of which is identical with the number of the characters under consideration and the number of lines of the latter is identical with the number of the evaluable experiments. Consequently a column vector consists of all realisations of a certain character.

4.3.1 The linear model.

The linear model in the literature is sometimes called multiple regression. However, this designation is also applied to methods with which the usual regression calculation is successively applied at first to the character with the greatest influence, then to those with the second-greatest influence etc.
With the linear model the column vector of the dependent character under consideration ist represented as linear function of the column vectors of selected independent characters.

If $\vec{y}$ is the column vector of the dependent character,
$\underline{\underline{x}}$ is the matrix of the independent characters under consideration, a sub-matrix of the measuring data matrix,

$\vec{b}$ is the vector of the linear coefficients, which is determined by suitable calculation methods,
$\vec{\varepsilon}$ the vector of the "approximation deviations",

then

(2) $$\vec{y} = \bar{\bar{x}} * \vec{b} + \vec{\varepsilon}$$

THE LINEAR MODEL. Seen geometrically, the relation (3)

(3) $$\vec{y}^{o} = \bar{\bar{x}} * \vec{b}$$

represents a multi-dimensional (corresponding to the dimension of $\vec{b}$ and $\vec{\varepsilon}$) plane which approximates the measured data cloud as good as possible, e.g. in the sense of GAUSS with smallest quadratic deviation, equation (4):

(4) $$\vec{\varepsilon} = \vec{y} - \vec{y}^{o} = \vec{y} - \bar{\bar{x}}\vec{b} \overset{!}{=} \text{Min}$$

The elements b_j of $\vec{b}$ are the respective gradients of the approximating plane in the direction of the j'th character axis. In order to be able to compare these, the characters x_j were rendered dimensionless through the transformation (5).

(5) $$x_{i,j} \rightarrow \frac{x_{i,j} - \bar{x}_j}{\bar{x}_j}$$

$\bar{x}_j$ being the respective arithmetic mean value of the value supply of the j'th criterion. For these new criteria, the gradients b_j^* can also be interpreted as influence factors.

The so-called moments can be calculated also for measuring values distributed not normally, but at random. The moment of the 1st order was already encountered as arithmetic mean in the transformation (5). Geometrically it can be used to calculate the "gravitational centre" of the n-dimensional data "cloud". It can be shown that the approximating plane always passes through this point. The moment of 2nd order is better known as standard deviation and can be used as a good standard for confidence limits. The moment of 3rd order is called obliquity and it indicates whether the shallows on the right or left of the mean value are longer or shorter.

4.3.2 The product model.

If the numerical values of all characters are logarithmated, the linear model

$$\log y_i^{o} = \log b_o + \sum_{j=1}^{p} b_j * \log x_{ij}$$

is equivalent to the PRODUCT MODEL (6).

$$(6) \qquad y_i = b_o * x_{i1}^{b_1} * x_{i2}^{b_2} * \dots x_{ip}^{b_p} + \varepsilon_i = \varepsilon_i + b_o \prod_{j=1}^{p} x_{ij}^{b_j}$$

Consequently one can use the same algorithm as in 4.3.1.

4.3.3 The model with adapted exponents.

The MODEL WITH ADAPTED EXPONENTS works with the formulation (7)

$$(7) \qquad y_1 = b_o + \sum_{j=1}^{p} b_j x_{ij}^{\varkappa_j}$$

With this formulation, first $\varkappa_j$ and than b_j are determined with the methods from 4.3.2 and 4.3.1. Extensive investigations with gradient methods have shown that at least a local minimum of the quadratic deviations are found.

4.3.4 The hyperbola-parabola model.

The HYPERBOLA-PARABOLA MODEL (8) requires the most calculation work of the models introduced here.

$$(8) \qquad y_i = b_o + \sum_{j=1}^{p} b_j x_{ij} + \sum_{k=1}^{p} \sum_{j=1}^{p} c_{jk} x_{ij} x_{ik}$$

At first the linear proportion is determined as with the linear model. With these raw values it is then possible to find the complete set of the parameters b_j and c_{jk} for instance with the aid of the method of FLETCHER AND REEVES (conjugated gradients).

4.4 Model selection.

Each of the introduced models has its own characteristic "curve shape", which can be bent and shifted through the choice of the coefficient to a greater or lesser degree. Consequently the decision as to what model should be applied depends on the problem and the type of the measured data. The fact that the model which renders the remaining quadratic deviation between measurement and approximating model smallest as compared with other models should be applied each time, can be used as a decision rule.
The quantity LQ of the relation (9) can be considered as a type of likelihood quotient.

$$(9) \qquad LQ = \frac{se}{s}$$

The standard deviation s applies to the model, s_e to the measuring data of the dependent character.

5. Interpretation, conclusions.

The practical use of an experimental task depends greatly on the extent to which one succeeds in interpreting the obtained results and in converting them into logical statements. However, a significant determination or at least a statement regarding the magnitude of the error probability belongs to each derived statement. Such information can be established from the confidence ranges of the individual coefficients and the residual deviations of the models.

When formulating statements, the protocols and records mentioned in the 3rd chapter proved very helpful. If so required, an established, uncertain statement must be backed up or refuted by jumping back to step 2.

6. Reporting.

Unfortunately too little significance is often attached to reporting. It is a fact that even an experimental work can only be evaluated after the final report. The expenditure for this necessary work is often highly underestimated.
It is desirable that at least the reporter is informed about all stages of the experimental investigations to be lectured on, and knows the applied measuring and calculation methods at least to the extent that rough errors by wrong application of these today still quite complex aids are avoided.