Pergamon

Ann. occup. Hyg., Vol. 39, No. 6, pp 873–877, 1995
Elsevier Science Ltd
British Occupational Hygiene Society
Printed in Great Britain
0003–4878/95 $9.50 + 0.00

COMPUTER PROGRAM REVIEW

HYGINIST: A Computer Program for the Lognormal Evaluation of Air Exposure Data, Vers. 2.12. By Theo M. L. Scheffers, Scheffers IHPC, Cramer van Brienenstraat 1f, P.O. Box 5552, 6202 XB Maastricht, The Netherlands. (ISBN 90 801900 1 2.) Price £160.00 (+£10 handling), Dfl 395 (+40 Dfl handling).

INTRODUCTION

The first thing about HYGINIST is that the presentation is first class—the 140-odd page manual is well printed on high quality A4 paper in a four-ring binder. It won't fit on the usual shelf reserved for most other computer and program manuals, but then it is a bit different. HYGINIST is a stand-alone .EXE PC-compatible program of about 150 k compiled from BASIC and used for calculating lognormal statistics from air exposure data. There are 36 data files which are employed to illustrate the uses to which the program can be put.

THE MANUAL

If you buy HYGINIST you will need to study the manual carefully. This is nothing to do with the sometimes awkward English translation, or the misprints, or even the 'difficulty' of the subject. It is simply that to use the program effectively you cannot afford not to understand the manual.

You will find that the program does all it claims to do. The cross-references are numerous and helpful, and each chapter is devoted to one particular aspect of the analysis of occupational hygiene data.

Of course, you do not have to absorb the manual at a single sitting, or even understand it fully, before you can get started. You could load HYGINIST and run it, and you will find that it is constructed of pages or screens, called 'windows' which are numbered on-screen from 1 to 34, with subdivisions in some cases. They are accessible one from another in logical sequence, performing different tasks, such as entering data from keyboard or file, calculating summary statistics, plotting and so on. These pages are all individually detailed in the manual, as are the routes to reach them.

THE MANUAL IN MORE DETAIL

The contents list includes all the windows, tables, examples and figures. In addition there is a section, with references, which is a most thorough lexicon of the statistics needed for handling occupational hygiene data. A table of symbols and abbreviations is included which you will need to consult frequently, since some symbols may not signify what you expect (μ for the mean of a normal distribution and EXP(μ) for the geometric mean of the lognormal distribution. I would have

preferred to have followed, for example, the symbols in Leidel *et al.* (1977), using μ for the AM and μ_l for the mean of the logged data and hence EXP(μ_l) for the GM, or μ, μ_ℓ and EXP(μ_ℓ) as in Tuggle (1982).

There is a useful glossary and index, and an Annex which contains background material on exposure assessment, the lognormal distribution and the statistics and algorithms behind HYGINIST, sections on installing HYGINIST on a computer, the hardware and software needed and an outline of how to run the program.

Chapter 1 introduces HYGINIST. Subsequent chapters are prefaced by a flow chart, indicating which areas (windows) are active in the chapter.

Chapter 2 deals with data management, keyboard/file input, data editing and initial inputs needed, such as the units you are using and gets as far as the summary statistics means and standard deviations. These take you on to the next screens in the sequence.

Chapter 3 brings us to plotting, regression analysis of fitted lines, and tests for goodness of fit using the Shapiro–Wilks W-statistic. This is done for the lognormal distribution, but offers six other data transformations to try. Although there are some powerful statistics here the on-screen plot is disappointing. The range of the plot is limited to that defined by the sample extremes and the corresponding lowest and highest rankits. There is no scale or graticule on the plot which makes it difficult to understand even though the current scale transformation is indicated. Getting a hard copy of the screen plot requires an additional program (usually bundled with the operating system) but it is still a simple screen dump. And be warned that the alternative transformations offered do just that to the data—the normal probability (x-axis) scale of the plot is never transformed to another, for example, the (Log)Extreme Value Distribution.

Here too it is possible to set limits to censor upper (for example, above an overload or saturation limit) and lower (below a limit of detection) data values on the plot. This is *not* the same as removing outliers.

In Chapter 4 the lognormal descriptive statistics are dealt with. It is here that the hygiene limit is introduced and inferences drawn, but only on the assumed lognormal distribution.

Chapter 5 brings us to compliance testing using a number of approaches (including NIOSH, Wilks and Tuggle). It would probably be prudent to know what you are dealing with before you begin quoting results from this area of the program. Just because the statistics are there, the program can calculate them and you can read the results does not mean that the number you quote is the one you want.

Chapter 6 provides for the comparison of samples, given the previously calculated summary statistics. The F-test is carried out first to validate the application of the unmatched t-test. Only lognormally distributed data can be used. The t-test is used to determine whether the GM is different from the GM of another sample, including the case when the GSDs are different. Comparisons of sample GSD and GM with the (known) population values can also be done, as can comparisons when the variances are significantly different.

Chapter 7 deals with the topic of "How many samples do I need to take?" to which the answer, by whatever route, is always "More than the Company can afford." The Chapter covers the minimization of sample size for compliance testing, given the GM and GSD and some other parameters.

PHILOSOPHY

The statistics used

It seems to be a fact of life that statisticians tend to prefer the statistics they have come to love and cherish, and may believe in them utterly whereas an occupational hygienist may need to say that experience and common sense sometimes override what the statistics seem to be telling him.

For example, while the insistence on, and remarking on, maintaining the full working precision of the BASIC used (as in the computation of Student's t to $\delta t/t < 10^{-6}$ on p. 63) is commendable, it is not really a selling point. We all know the real accuracy and precision of occupational hygiene data, although some of the algorithms used may require high precision. I certainly agree with the author when he says that the credibility of the data, the method of their collection and the occupational hygiene behind them is more important than any facility to run a stats package.

I also thought that the consideration of the effect of statistical (or other) outliers (they do occur) was somewhat dismissive and limited to the proposition that their presence hardly influences the mean. While this is true the presence of one or two outliers, especially if the sample size is small, can markedly influence the choice of transformation to obtain the best-fitting straight line. It is not made clear whether outliers are to be included, but only as numerically undefined data with rankit values or whether they should be excluded from further consideration. This was always my understanding of the purpose of testing for, and dealing with, outliers.

The lognormal distribution

I found it disappointing that the author follows so closely the dictum that all occupational hygiene data are lognormally distributed, or nearly so, supporting the assertion with a number of quotations. Much of what follows in HYGINIST (especially in Chapters 4–6) is immutable as a consequence. The F-test and t-test cannot be applied to normally distributed data, for example. My own feeling is that occupational hygiene data are lognormally distributed (if they are) because of a lack of control in the processes which give rise to the data. If *real* control can be achieved then the data may tend to become normally distributed.

In Dewell (1994b) it was noted that not all data sets were lognormally distributed. Also with the passing of time and improvements to control systems the GSD of some corresponding lognormally distributed data sets decreased, and the occurrence of some normally distributed data sets increased in frequency. It is conceded that the program (see its title) is intended for use on *airborne* concentrations, for which the insistence of lognormality may be more likely, but other occupational hygiene data do exist.

The problem of lack of control (both in occupational hygiene and statistical terms) seems to have serious implications for the application of control (Shewhart) chart theory to occupational hygiene data. There is, in the manual, some confusion (too common) between control charts on the one hand and time series or trend analysis on the other. Neither are handled by this version of HYGINIST although a couple of time trend plots (called 'control charts') are included. It is hoped that such confusion can be resolved in preparing the next version.

I just feel that it is a pity that (other than allowing the plotting of alternative transformations) the opportunity is denied for the simpler normal distribution to have an equal place in the scheme of things, as in OH_STATS. Whether there is a need for other distributions to be considered (as opposed to other data transformations) is perhaps a moot point, but at least the normal distribution should have a place alongside the lognormal. Unless, of course, it has been decided otherwise by Brussels.

This aspect, and the undistinguished graphics noted earlier, were what disturbed me most.

CONCLUSIONS

I hope that this review will stimulate hygienists to re-examine their data in a new light using tools which already exist. I cannot, in all conscience, recommend HYGINIST rather than OH_STATS (or vice versa). They both do similar things in a similar way. They each do things the other does not. They both benefit from additional programs to support their graphics.

HYGINIST will be useful, with closer acquaintance and frequent use, to hygienists who have an understanding of how their samples arise and a smattering of statistical understanding of occupational hygiene data. The same can be said of OH_STATS, of course (Dewell, 1994a). They can both take hygienists far along the path of a better understanding of their occupational hygiene data.

Perhaps the real question is "Can a statistics package be used to guide us in improving working conditions or confirming that they are now of adequate quality?" If the answer is "Yes" then the package is acceptable. To this extent both HYGINIST and OH_STATS perform well, but somewhat differently. The statistics used in HYGINIST may be more powerful than those in OH_STATS. The rigid format of, and insistence on lognormality in, HYGINIST must be compared with the potentially more flexible format of OH_STATS, which also handles normal distribution statistics and includes a few more applications. The rather better graphics achievable in OH_STATS may appeal to those who think this important. No doubt both packages have their place. It is only a pity perhaps that they are in competition when they might have been combined. As my father used to say "Two heads are better than one, even if they are both sheep's heads".

I am indebted to Tom Gillanders (Hon. Sec. of the Institute of Occupational Hygienists) for inviting me to review HYGINIST and to Theo Scheffers for making a copy available to him to pass on to me.

P. Dewell
17 Brotherton Avenue,
Redditch,
Worcestershire B97 5SA, U.K.

REFERENCES

Dewell, P. (1994a) *OH_STATS—Some Occupational Hygiene Statistics Programs in GW-BASIC for PCs.* H & H Scientific Consultants Ltd, Leeds, U.K.

Dewell, P. (1994b) *Concentrations of Cadmium in Air and Urine in an Alkaline Battery Works: A Case Study*. BOHS Technical Handbook Series No. 1, Supplement A. H & H Scientific Consultants Ltd, Leeds, U.K.

Leidel, N. A., Busch, K. A. and Lynch, J. R. (1977) *Occupational Exposure Sampling Strategy Manual*. DHEW Publication No. 77-173. U.S. Department of Health, Education and Welfare, NIOSH, Cincinatti, Ohio.

Tuggle, R. M. (1982) Assessment of occupational exposure using one-sided tolerance limits. *Am. ind. Hyg. Ass. J.* **43**, 338–346.