RATS - THE SAS SOFTWARE CONNECTION

Arthur C. Singer, Wyeth Laboratories, Inc.
Caroline M. Bobik, Wyeth Laboratories, Inc.

Abstract

RATS (Rodent Activity and Thermogenesis System) is a completely automated computer-controlled data collection system for drug studies. Oxygen consumption, carbon dioxide production, and motor activity in rodents are recorded. Drug-related changes in motor activity which affect oxygen consumption are of primary importance and present a complex problem in statistical model building.

Data from 10- or 20-day experiments were transmitted by modem via an HP3000 to an IBM 3033 mainframe. Approximately 30K successive unidentified numbers daily were stored in a fixed-pattern data file; identification of these numbers required modulo arithmetic. The data were formatted into observations on a time basis conditioning them for SAS procedures.

An upgrade of RATS, using an HP98695A Coax Interface with appropriate software, now allows the HP9836C computer to emulate an IBM 3278 terminal with direct data transmission to the mainframe at 9600 baud. Data formatting into SAS* observations now occurs before transmission.

The model building involves determining the relationships among 23 variables in order to assess the degree of multicollinearity and to select those variables of greatest influence upon the dependent variable (oxygen consumption). SAS procedures, including regression and multivariate analyses, are being applied to determine an appropriate model.

Introduction

RATS (Rodent Activity and Thermogenesis System) is a completely automated computer-controlled modular system used for in vivo studies in order to assess the thermogenic potential of test compounds. Oxygen consumption, which may be directly related to thermogenic activity, carbon dioxide production, and motor activity are measured and recorded as unbiased objective data. The open-flow system uses a custom-designed cage contained within an activity monitor, in line with mass flow transducers, diaphragm pumps, an oxygen analyzer, and a CO2 analyzer to provide the information required for drug screening. All system control and data acquisition is performed by an HP3497A Data Acquisition/Control Unit and an HP9836C Computer.

Originally RATS had two experimental cages; recently a third cage was added to the system. For any experiment requiring more than three treatments, a balanced incomplete block design (BIBD) was used, providing the maximum amount of information from the three-cage prototype system. Experiments using three treatments or less were run as balanced designs with appropriate randomization. Drug screening studies require three control groups and at least three test compound groups per study. The control groups consist of a vehicle control, a thermogenic control, and a motor activity control.

At the conclusion of each day (block), a data stream of approximately 30K unidentified numbers was transmitted by modem to an HP3000 through the HP9836C using HP Series 200 terminal emulator software. When transmission had been completed, the data were sent to the IBM 3033 mainframe through MultiLeaving Remote Job Entry (MRJE), a subsystem of the HP3000 which provides for the submission of multiple batch job streams to remotely-located host computers. Each set of numbers generated by RATS was stored in a repetitive pattern, creating a sequential data file.

A program was written to handle the output from a system of 16 experimental cages; currently a three-cage prototype is in use. The first 104 numbers in the stream identify the rat code numbers, body weights, gender, treatment groups, dosages, and routes of treatment administration for a 16-cage system. For each cage and time of observation, a series of variables exists; these include time, status (baseline, treated), 21 motor activity variables, floor and ceiling temperatures, one air flow readout, actual O2 and CO2 percentages in exit air, cage number identification, calculated O2 consumption (STPD), and barometric pressure.

The variables were identified by the remainders resulting from the application of modulo arithmetic (function MOD) to the automatic variable counted by SAS* software using the total number of variables at a given time of observation as the modulo base. After identification, each variable could be used for statistical analysis.

An upgrade of RATS, using an HP98695A 3270 Coax Interface with appropriate software, now allows the HP9836C computer to emulate an IBM 3278 terminal. The data are transferred to an ASCII file. All variables recorded at a given time are packed within a string of 255 ASCII characters to form a SAS* observation. This file is uploaded directly to the IBM 3033 mainframe using HP3270 terminal emulator software. Such a simple step transfer is time-saving, more reliable and facilitates data editing.

Statistical Model Building

Drug screening for active compounds requires an understanding of the relationships among the response variables used as indicators of significant drug activity.
A preliminary biological hypothesis points to a relationship between motor activity and oxygen consumption. Intuition indicates that the horizontal, vertical, and stereotypical components of activity may be correlated; all variables may not be necessary for the model.

After data transfer to the IBM system in order to apply SAS® software, it was discovered that PROC GLM was not able to handle the large volume of data in the model for these experiments. It was necessary to write an algorithm involving the repetitive use of PROC MEANS for sums of squares and degrees of freedom needed for the required sources of variation in the model. The algorithm was verified with a small data set which PROC GLM was able to analyze without memory constraints. Data sets created by PROC MEANS were later merged, and an analysis of variance table was derived from the merged set.

Amphetamine (5 mg/ml, p.o.) has a pronounced effect on the central nervous system, which results in increased motor activity, providing a control situation for assessing the effect of such activity on oxygen consumption. A two-cage experiment using rats dosed with water or amphetamine yielded data suitable for initial model-building.

Preliminary statistical analyses and graphic displays (PROC PLOT) of the data revealed a great amount of variability in response among rats within treatment groups. Methods for reducing this heterogeneity of variances were attempted, as well as the assessment of multicollinearity, using several SAS® procedures. The most appropriate procedures appeared to be PROC STEPWISE, PROC FACTOR, PROC SCORE, PROC REG, and PROC RSQUARE. Some of these procedures were suggested by the SUGI 10 (1985) tutorial, entitled "Multicollinearity and Variable Selection", presented by Rudolf J. Freund.

Factor analysis via PROC FACTOR is a statistical technique developed for reducing a large number of correlated variables (observed) to a smaller number of uncorrelated variables (hypothetical). PROC STEPWISE provides equations that are useful for prediction with R² close to 1 and/or use only a few independent variables. PROC SCORE combines values from two SAS® data sets, one of which usually contains the original data. PROC REG is a general purpose procedure for regression with each independent variable having one degree of freedom associated with it. PROC RSQUARE finds subsets of independent variables that best predict a dependent variable by linear regression in the given data set.

Because of the variability in response, the genetic background of the rats under test was investigated. They were identified as a maximally outbred Sprague Dawley strain; all current work uses the inbred Fischer 344 strain, from a specific-pathogen free colony.

The large volume of data for each outbred rat indicated that the analyses be done on individual rats. PROC STEPWISE was used in an attempt to choose those activity variables having a significant effect on oxygen consumption. This procedure showed a different pattern of variables for each rat. Initial consultation with Dr. Freund led to the use of the following sequence: PROC FACTOR, PROC SCORE, PROC REG, and PROC RSQUARE.

Since there are three kinds of movement observed for the activity variables, three factors were requested by PROC FACTOR to see if these types of movement were separated out. After varimax rotation, in most instances the movements were grouped by type, e.g., the horizontal movements would have high loadings within one factor, the vertical movements and stereotypy within other factors. Use of either more or less factors made no appreciable difference.

PROC SCORE combines the results of PROC FACTOR with the data from the original data set. A representative variable from each factor, empirically selected for optimum influence on oxygen consumption, was used in the model statements of PROC REG, to determine if multicollinearity existed, and PROC RSQUARE, to test the effects of the influential activity variables on oxygen consumption.

An exploratory three-cage experiment was designed using one water control rat and two amphetamine control rats per block. The data obtained from this experiment are being used in model building. The results of factor analysis showed that many of the activity variables loaded at 0.8 or above on three factors corresponding generally to the horizontal, vertical, and stereotypical components of motor activity.

Dr. Freund suggested the use of composite variables derived from the summation of the standardized values of the motor activity variables which loaded on each factor. The standardized values were obtained using PROC STANDARD defining the mean as 1 and the standard deviation as 1.

Future Considerations

Time trends of the response to treatment for the activity variables and oxygen consumption are an important aspect that have just started to be investigated in this project. Of greatest interest are the linear and quadratic components. The linear trend would indicate if there was a significant change from start to finish over the duration of the treatment period. It is expected that the significant maximum quadratic effect would indicate both the onset time of treatment effect and maximal mean differences among treatments. Preliminary attempts at such analyses indicated that coding of the time values to reduce coefficient magnitudes may be necessary.
Acknowledgements

The authors wish to express their appreciation to Rudolf J. Freund, Ph.D., June Deitz, the Duplicating Department in Great Valley, and the Audio-Visual Department in Radnor for their valuable assistance. They also wish to thank their spouses for their forbearance during the experimentation and analytical periods. All correspondence should be directed to:

Arthur C. Singer
Wyeth Laboratories
P.O. Box 8299
Philadelphia, PA 19101

References


*SAS is the registered trademark of SAS Institute, Inc., Cary, NC, USA.