

Teaching statistics at E.U.T. with the use of a computer

Citation for published version (APA):

Dijkstra, J. B., & Doornbos, R. (1985). *Teaching statistics at E.U.T. with the use of a computer*. (Computing centre note; Vol. 25). Technische Hogeschool Eindhoven.

Document status and date:

Published: 01/01/1985

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Eindhoven University of Technology
Computing Centre Note 25

Teaching Statistics at E.U.T. with
the use of a Computer

R. Doornbos and J.B. Dijkstra

Prepared for the Seminar on Mathematics in Engineering Education;
the Impact of Computers.

March 27-30, 1985,
Copenhagen, Denmark.

Teaching Statistics at E.U.T. with
the use of a Computer

R. Doornbos and J.B. Dijkstra

Eindhoven University of Technology
The Netherlands

Keywords: Applied Statistics, Regression Analysis, Generalised Linear Models, Multivariate Statistical Methods, Teaching Statistics, GLIM, CONSTAT, STATLIB.

Abstract

In three courses in Applied Statistics at the Eindhoven University of Technology the computer plays an important role. These courses are:
(1) Regression Analysis, (2) Generalised Linear Models and
(3) Multivariate Statistical Methods. Students in their final studies also use the computer for validating statistical tests, as well as for some other problems.

1. Introduction

The Department of Mathematics consists of several groups, one of which is dealing with applied statistics. In this group are working five persons and their tasks are teaching statistics, research and consultation. The teaching is not restricted to the Department of Mathematics; in other departments the students also have to learn some statistical techniques.

In the Computing Centre the field of Statistical Applications is covered by four persons. They develop statistical software and assist the people who want to use it. They also buy programs and implement them on the local computers. Courses in several software packages are given.

Between these groups at the Department of Mathematics and the Computing Centre exists a close collaboration and this results (among other things) in the increasing role of the computer in teaching statistics at the Eindhoven University of Technology.

2. Regression Analysis

The computer is not used in every course; at this moment only in the courses that are mentioned in the abstract. The first example is Regression Analysis and it is intended for students in their second year. The course is based on the book by Montgomery and Peck [1] and the statistical package GLIM [2]. A typical problem is given in table 1. The students are asked to produce a model that describes the gasoline mileage for automobiles. Eleven possible predictors are given and it is easy in GLIM to apply transformations to them. The model has to contain only variables that are relevant and the students have to examine the residuals very carefully.

Automobile	y	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉	x ₁₀	x ₁₁
Apollo	18.90	350	165	260	8.0:1	2.56:1	4	3	200.3	69.9	3910	A
Omega	17.00	350	170	275	8.5:1	2.56:1	4	3	199.6	72.9	3860	A
Nova	20.00	250	105	185	8.25:1	2.73:1	1	3	197.7	72.2	3510	A
Monarch	18.25	351	143	255	8.0:1	3.00:1	2	3	199.9	74.0	3890	A
Duster	20.07	225	95	170	8.4:1	2.76:1	1	3	194.1	71.8	3365	M
Jenson Conv.	11.2	440	215	330	8.2:1	2.88:1	4	3	184.5	69	4215	A
Skyhawk	22.12	231	110	175	8.0:1	2.56:1	2	3	179.3	65.4	3020	A
Monza	21.47	262	110	200	8.5:1	2.56:1	2	3	179.3	65.4	3180	A
Scirocco	34.70	89.7	70	81	8.2:1	3.90:1	2	4	155.7	64	1905	M
Corolla SR-5	30.40	96.9	75	83	9.0:1	4.30:1	2	5	165.2	65	2320	M
Camaro	16.50	350	155	250	8.5:1	3.08:1	4	3	195.4	74.4	3885	A
Datsun B210	36.50	85.3	80	83	8.5:1	3.89:1	2	4	160.6	62.2	2009	M
Capri II	21.50	171	109	146	8.2:1	3.22:1	2	4	170.4	66.9	2655	M
Pacer	19.70	258	110	195	8.0:1	3.08:1	1	3	171.5	77	3375	A
Bobcat	20.30	140	83	109	8.4:1	3.40:1	2	4	168.8	69.4	2700	M
Granada	17.80	302	129	220	8.0:1	3.0:1	2	3	199.9	74	3890	A
Eldorado	14.39	500	190	360	8.5:1	2.73:1	4	3	224.1	79.8	5290	A
Imperial	14.89	440	215	330	8.2:1	2.71:1	4	3	231.0	79.7	5185	A
Nova LN	17.80	350	155	250	8.5:1	3.08:1	4	3	196.7	72.2	3910	A
Valiant	16.41	318	145	255	8.5:1	2.45:1	2	3	197.6	71	3660	A
Starfire	23.54	231	110	175	8.0:1	2.56:1	2	3	179.3	65.4	3050	A
Cordoba	21.47	360	180	290	8.4:1	2.45:1	2	3	214.2	76.3	4250	A
Trans Am	16.59	400	185	NA	7.6:1	3.08:1	4	3	196	73	3850	A
Corolla E-5	31.90	96.9	75	83	9.0:1	4.30:1	2	5	165.2	61.8	2275	M
Astre	29.40	140	86	NA	8.0:1	2.92:1	2	4	176.4	65.4	2150	M
Mark IV	13.27	460	223	366	8.0:1	3.00:1	4	3	228	79.8	5430	A
Celica GT	23.90	133.6	96	120	8.4:1	3.91:1	2	5	171.5	63.4	2535	M
Charger SE	19.73	318	140	255	8.5:1	2.71:1	2	3	215.3	76.3	4370	A
Cougar	13.90	351	148	243	8.0:1	3.25:1	2	3	215.5	78.5	4540	A
Elite	13.27	351	148	243	8.0:1	3.26:1	2	3	216.1	78.5	4715	A
Matador	13.77	360	195	295	8.25:1	3.15:1	4	3	209.3	77.4	4215	A
Corvette	16.50	350	165	295	8.5:1	2.73:1	4	3	185.2	69	3660	A

y : Miles/gallon

x₁ : Displacement (cubic in.)

x₂ : Horsepower (ft-lb)

x₃ : Torque (ft-lb)

x₄ : Compression ratio

x₅ : Rear axle ratio

x₆ : Carburator (barrels)

x₇ : No. of transmission speeds

x₈ : Overall length (in.)

x₉ : Width (lbs)

x₁₀ : Weight (lbs)

x₁₁ : Type of transmission (A-automatic,

M-manual)

Source: Motor Trend, 1975

Table 1

Evidently it is almost impossible to solve such problems without a computer. It is only a few years ago that the students were only confronted with artificial and very small problems that could be evaluated by hand. Now it is possible for them to apply statistical techniques to life situations. This seems an important improvement in Engineering Education.

3. Generalised Linear Models

This is a course for students in their third year. It is based on the book by Dobson [3] and here we also use GLIM. The name of this statistical package comes from Generalised Linear Interactive Modelling and so it seems an appropriate choice. The students have to solve practical problems in the following fields: (1) Multiple Regression, (2) Analysis of Variance and Covariance, (3) Binary Variables and Logistic Regression and (4) Contingency Tables and Log-linear Models. An example of such a problem is given in table 2. The students have to construct a hierarchical log-linear model with as few parameters as possible that describes these data.

		H			
		A	1	2	3
male	20 - 34	20	20	16	16
	35 - 64	40	30	48	64
	65 -	6	6	20	10
female	20 - 34	56	45	42	22
	35 - 64	52	55	57	66
	65 -	9	9	27	13

Table 2

The frequencies in this table are classified by three factors. The subjects are people living in a certain neighbourhood where they can choose between four family doctors (factor H). The other factors are A (age) and S (sex). The resulting model contains only one interaction and that is A*H, which seems to be reasonable. S cannot be omitted but it is no part of any relevant interaction.

Such models can only be found with a trial and error technique that gives the user of an interactive statistical program some feeling for the data. Data analysis as a topic is closely related to statistical inference. And it is only since the students can use a terminal that they really get to know some of its techniques.

4. Multivariate Statistical Methods

This is a course for students in their fourth year. We use a text by Bosch and Doornbos [4] and a collection of almost 300 statistical routines in Algol and Fortran called STATLIB [5] that was developed at the local Computing Centre. This collection will also be available in Pascal next year. As an alternative students can use CONSTAT [6] which is a conversational statistical program that uses the routines of STATLIB.

The topics in this course are: (1) Principal Components, (2) Factor Analysis, (3) Canonical Correlation, (4) Discriminant Analysis and (5) Cluster Analysis.

A typical problem is given in table 3. The subjects are 17 statistical packages that are suitable for microcomputers. For every program the price and the number of pages of the manual is given. For nine topics in statistics the number of possible tests are given, and the students have to perform a Factor Analysis on these data where the number of factors has to be chosen carefully. The resulting factor matrix can be rotated to improve the case of interpretation.

name	price in \$	pages documentation	T ₁	T ₂	T ₃	T ₄	T ₅	T ₆	T ₇	T ₈	T ₉
ABSTAT	395	100	11	8	10	5	3	5	2	8	2
AIDA	235	90	9	12	3	4	2	1	2	6	1
ANOVA II	150	52	5	2	2	0	0	1	0	0	7
A-STAT	200	140	9	8	6	4	2	1	2	7	3
BASIC SS	100	100	2	0	0	3	0	0	13	0	0
BIOSTAT	95	35	6	0	2	0	2	1	2	6	2
COMMODORE	60	300	2	1	8	3	6	2	5	6	2
COMPSTAT	1500	1000	4	4	9	5	5	7	10	7	6
DESCRIPTIVE	60	58	3	1	5	1	0	0	0	0	0
ED-SCI	100	75	12	1	3	3	3	0	2	4	2
EDA	165	354	1	3	5	2	0	0	0	0	0
HAL	95	80	8	6	5	3	1	0	0	3	1
KEYSTAT	130	88	3	0	7	3	2	0	5	1	5
MDA	349	208	7	4	4	2	0	0	0	6	0
MICROSTAT	375	110	8	8	7	4	6	8	9	10	3
MSUSTAT	500	123	4	3	6	5	4	7	5	7	6
NUM CRUNCH	200	70	5	7	6	4	3	8	7	9	7

Table 3

The students also have to compute the factor scores and discuss their results. They know what the topics T_1 to T_9 are, but we will not mention them here.

This kind of problem, especially if one is trying several rotations, cannot be solved without a computer. An interesting observation is that the students seem to like dealing with such problems. They are searching for structures in the data and are trying several approaches in order to find them. And during this process they learn a lot about the statistical techniques involved.

5. Students in their Final Studies

In the last three sections the students applied statistical programs to data. They could study the methods that lie behind these programs but for them these methods were merely tools that could be used but not modified or replaced by products of their own skill. If a student chooses Applied Statistics as a topic for his final studies, we expect him to go a bit further. Such a student has to deal with statistical methods as his subject. He has to be able to develop new methods or to adapt existing ones for non-standard situations.

A typical problem is to construct an approximate test for a situation in which no exact solution exists. Such a test will then be added to STATLIB so that other users can benefit from the study. If the subject seems to be of general interest it sometime happens that a publication will follow.

The validation of an approximate test usually involves a simulation study in which the test is applied thousands of times to pseudorandom data from some chosen distributions. It goes without saying that this is only possible with a computer. In this area the computer is not only important for teaching, but also for the development of new statistical methods.

Some recent topics were:

- (1) The Behrens-Fisher problem for more than two groups
- (2) Adaptive nonparametric analysis of variance
- (3) Outlier-resistant analysis of variance
- (4) Comparison of two location parameters from multinormal distributions with unequal covariance matrices
- (5) The lack of fit of a Gram-Charlier model for data from various distributions of the Pearson family
- (6) Adaptive two-sample tests.

These topics were chosen as a consequence of the statistical consultation in the Computing Centre. And so the students are motivated by the knowledge that their results will be used in practice.

6. References

1. Douglas C. Montgomery and Elizabeth A. Peck
Introduction to linear Regression Analysis
Wiley, New York, 1982
2. R.J. Baker and J.A. Nelden
The GLIM system, release 3
NAG Central Office, Oxford, 1978
3. Annette J. Dobson
Introduction to Statistical Modelling
Chapman and Hall, London, 1983
4. A.J. Bosch and R. Doornbos
Multivariate Analysis
Unpublished lecture notes of E.U.T.
5. Jan B. Dijkstra e.a.
STATLIB, A library of statistical routines in Algol and Fortran
Almost 300 routines, divided into 21 chapters that can be obtained
separately
6. P.C.F. de Witte
CONSTAT, A Conversational Statistical Program
Report of E.U.T. Computing Centre