STATISTICAL METHODS FOR DEVELOPING AND DISTINGUISHING MULTINOMINAL RESPONSE MODELS IN THE TRAUMATOLOGICAL ANALYSIS OF SIMULATED AUTOMOBILE IMPACTS

W. Härdle
Rheinische Friedrich-Wilhelms-Universität Bonn
D-5300 Bonn, Federal Republic of Germany

D. Kallieris
Ruprecht-Karls-Universität Heidelberg
D-6900 Heidelberg, Federal Republik of Germany

R. Mattern
Johannes Gutenberg-Universität Mainz
D-6500 Mainz, Federal Republic of Germany

ABSTRACT

Simulated car-to-car side impacts, designed for the analysis of traumatological aspects, involve two sets of variables. Predictors include exogenous biomechanical factors as well as anthropometric variables, such as age. The response is measured a scale of injuy scores and is thus multinominal.

It is the aim of a statistical analysis of such data to devise a multinominal response model that explains possible patterns of injury as a function of a suitable set of predictor variables. Several approaches for modelling such a multinominal response relationship have been proposed in the literature, among them the Logistic and the Weibull regression models. Two major questions in applying such models are as follows: What model is appropriate and how should different models be compared. Another concern is how the quality of a given model should be presented for varying sets of predictors.

In this paper we discuss the first question by constructing a goodness-of-fit test based on bootstrapping flexible, non-parametric alternatives to a given parametric candidate model. Secondly, we present several graphical techniques that allow relatively simple comparisons of different models.

1. Modelling the influence of anthropometric and mechanical parameters on trauma indices:

The aim of the statistical analysis of simulated car impacts is to develop models that allow one to understand how the severity of impacts depend on observable input variables. Typically such input variables can be divided into two types. The first set of variables is describing the test subject's physical characteristics, such as height or age. A second set is concerned with the actual experimental setting, and contains such parameters as velocity of the impact and acceleration measured at various places. These input variables determine jointly the response variable. The observed response variable is a trauma index usually scaled according to some injury scale, e.g. AIS (1980). The AIS trauma index, for example, is a discrete variable in $\{0,1,2,3,4,5,6\}$, with the lightest (or non) injury indexed by "0" and the severest injury indexed by "6". The input variables are mostly of continuous nature, i.e. they can possibly take each value in a certain interval.

Phrased in terms of statistical methodology we are given a discrete regression problem, i.e. discrete response variables (trauma index) are regressed on various kinds of predictor variables (possibly continuous or also discrete).(See Bickel and Doksum (1977), Neter and Wasserman (1974, Chapter 9)). The aim of this statistical problem is to construct suitable models for explaining the probability of a certain level of trauma index as a function of the given covariables. In this paper we denote by (X_i, Y_i), $i = 1,\ldots, n$, the data points from such an experiment; X standing for the vector of predictor variables (input) and Y denoting the discrete response variable (output vector). Since the response variable is multinominal (i.e. takes values in a discrete ordered set) it is reasonable to define the regression function as the probability that Y is bigger than some value c. Hence,we are dealing with a set of regression functions

$$p_c(x) = P(Y \geq c \mid X=x).$$

where c runs through the discrete set of possible response values (trauma indices). In determing such functions p_c one would like to have some basis requirements fulfilled that are direct consequences of the experimental setup. These are

(1.1) Monotonicity, i.e. if the input variables are ordered in some natural way then increasing the strength of impact or increasing age, the probability of having a trauma index greater than or equal to c should also increase.

(1.2) Consistency, i.e. $p_{c_1} \geq p_{c_2}$ for $c_1 \leq c_2$

Consistency means that the curves p_c should be so that the probability of having trauma index greater than c increases if c decreases.

In the next section we discuss several multinominal response models. In section 3 we show how nonparametric smoothing techniques help in selecting a suitable response model. In section 4 we discuss some graphical methods for enhancing the summary statistics of a given fit when the set of predictor variables is varied. In section 5 the application of these methods to the Heidelberg side impact data is presented. Section 6 is devoted to conclusions.

2. Multinominal Response Models

There are two different approaches to model the dependence of the conditional probability $p_c(x) = P(Y \geq c | X=x)$ as a function of the covariables x. The first approach is to assume that this function p_c is a member of a specific class of parameterized functions. The second approach is called non-parametric since the form of p_c is not restricted by any requirement except those of (1.1) and (1.2) above. The parametric approach has the advantage of easier interpretation of coefficients and also of numerical computations, whereas the non-parametric approach has the advantage of not being bound to any functional form. Both should serve each other as an alternative and should not be seen as mutually exclusive models. Well-known parametric models include the Logistic and the Probit regression models. The basic structural assumption for both approaches is the same; both are models based on linear combinations (projections) of the predictor variable x, i.e. the function p_c is modelled as

$$p_c(x) = G_c(\beta^T x).$$

with a link function G_c and parameter β. The parametric approach consists of fixing the function $G_c(.) = G_c(\alpha_c + .)$ to a certain shape whereas the non-parametric approach does not prescribe the form of G_c. In the following we just write G to describe the general form of G_c.

In a Logit analysis one assumes that G is of the form of a logistic distribution function, i.e.

$$G(z) = \exp(z)/(1+\exp(z)).$$

The functions P_c are determined by the maximum likelihood method, i.e. one maximizes for each c

$$\Pi_{i=1}^{n} P(Y_i \geq c \mid X_i = x_i)$$

$$= \Pi_{i=1}^{n} G(\alpha_c + \beta^T x_i)^{Y_i^c}(1 - G(\alpha_c + \beta^T x_i))^{(1 - Y_i^c)},$$

$$Y_i^c = I(Y_i \geq c).$$

subject to the consistency condition. In the same way other models like the Probit model with G equal to the standard normal distribution function can be adapted. Yet another shape function is the Weibull distribution function.

The non-parametric approach does not fix the shape function G, but rather lets it be any smooth function following the requirements (1.1) and (1.2). Given the parameter vector ß the link function G is determined by a non-parametric smoothing technique, such as spline or kernel, see Härdle (1988). The kernel smoother $\hat{G}_h(z)$ at the point

$$z = \beta^T x \text{ for data } (Z_i = \beta^T X_i,\ Y_i)$$

is defined by

$$\hat{G}_h(z) = n^{-1}\Sigma_{i=1}^{n} K_h(z-Z_i)Y_i / n^{-1}\Sigma_{i=1}^{n} K_h(z-Z_i)$$

where $K_h(u) = h^{-1}K(u/h)$ is a delta function sequence with bandwidth h and kernel K, where K is a continuous probability density. The kernel smoother is a consistent estimate of G if $h \longrightarrow 0$ as the sample size n tends to infinity. The parameter ß can be determined in various numerical ways, since the function G is not determined up to scale. One of the possibilities is to determine G and ß jointly by minimizing the Residual Sum of Squares (RSS) or other measures of accuracy. This amounts to finding G and ß such that

$$n^{-1}\Sigma_{i=1}^{n}(Y_i - G(\beta^T X_i))^2$$

is minimal. This minimization is done iteratively by searching over all possible directions ß, that is why this method is called Projection Pursuit Regression (PPR), see Friedman and Stuetzle (1981). Another method is called Average Derivative Estimation (ADE). In ADE estimates of ß are obtained in a direct way without involving the link function G. This estimate of ß is defined as

$$\hat{\beta} = n^{-1}\Sigma_{i=1}^{n} Y_i \hat{f}'(X_i)/\hat{f}(X_i)$$

where $\hat{f}$ denotes an estimate of the partial derivatives of f, the density of X. For details see Härdle and Stoker (1988).

3. Selecting a suitable model

The task finding a suitable model among the many possible parametric and non-parametric alternatives involves the statistical precision of the model as well as the numerical applicability. It is widely known that the Logistic regression model can be quite easily fitted numerically, SAS Supplementary User's Guide (1985). Other link functions G, for example the Probit curve have a similar shape (see Berkson, 1951) but require more computational effort. Also the non-parametric smoothing method requires a lot more on computations but has the advantage of not being restricted in its functional form. In particular the symmetry of the link function that is inherent to the Logit model is no restriction for the non-parametric approach. Indeed the response of the side impact experiments is somewhat asymmetric, as was pointed out by several people who tried a skewed Weibull distribution as a link function G. The price one has to pay though for this additional feature is that the number of parameters, and thus the numerical cost and precision of the algorithm, increase.

Since the non-parametric alternative allows fitting in a much wider class of functions it seems reasonable that it can be used in a formal test of goodness of fit of low dimensional parametric models. To simplify matters let us consider only a binominal response model of one dimensional X variables, i.e. Y takes the values 0 or 1. the proposed test is based on smoothing the response variables of a given parametric fit $p(x;\hat{\beta})$. One defines the kernel smoother on data (X_i, Y_i) as

$$\hat{p}(X_j) = n^{-1}\Sigma_{i=1}^{n} K_h(X_j - X_i)Y_i / n^{-1}\Sigma_{i=1}^{n} K_h(X_j - X_i).$$

The smoothing parameter h can be determined by crossvalidation, see Härdle (1988). The test is described formally as follows.

1. Fit a candidate parametric model $(p\,(x;\hat{\beta})$

2. Simulate new observations (X^*_i, Y^*_i) from this model by using a pseudo random number generator based on $p(x;\hat{\beta})$ (bootstrapping).

3. Determine for each X_i that has been observed the empirical 5 % quantiles of a kernel smoother of the simulated data.

4. Center these 5 % bands around the assumed parametric candidate model.

5. Check whether the kernel smoother based on the original data lies in between these bands.

Figure 3.1

Another method is based on comparing the likelihood for different models with a bias correction for different number of parameters. This is related to ideas of Akaike (1977) and works as follows. One compares the Log-Likelihoods under both models, i.e.

$$n\, L_1(\dot{\beta}_1) - n\, L_2(\beta_2) - (\dim(\text{model}_1) - \dim(\text{model}_2)).$$

Based on the limiting chi-square distribution of twice the likelihood ratio statistic one cannont distinguish the two models if the magnitude of the above difference is less than 0.5.

4. Comparing similar models

If the above models are run for several types and sets of input variables it is important to compare the output of the different fits. In the study of the Heidelberg data we found the following, mostly graphically oriented tools very convenient.

Concomitant pairs

Concomitant pairs are defined through all pairs of observations with different response values. Now count all pairs of observations where the current model fit predicted a higher probability for the higher Y-value. Then compute the share of these pairs among all pairs with different Y-values. Certainly if this share of concomitant pairs is close to 1 the model fits quite well. The procedure LOGIST of the SAS system computes this number on request.

Prediction Table

The prediction table is simply a frequency table of the observed trauma indices versus the predicted trauma indices. The number of correctly predicted response variables is the classification rate. This number lies between 0 and 1. Certainly a number close to one is desirable. It is quite intuitive that the empirically determined classification rates are over optimistic since the data is used to determine the model as well as to judge it. An unbiased estimate of the classification rate can be obtained by, for example, cross validation. In this method the whole analysis is performed n times on n subsamples each of size n-1 (leave one out method). The left out observation is predicted by the model constructed from the rest of the observations. This leads to an unbiased estimate of the prediction error, as was shown by Stone (1974).

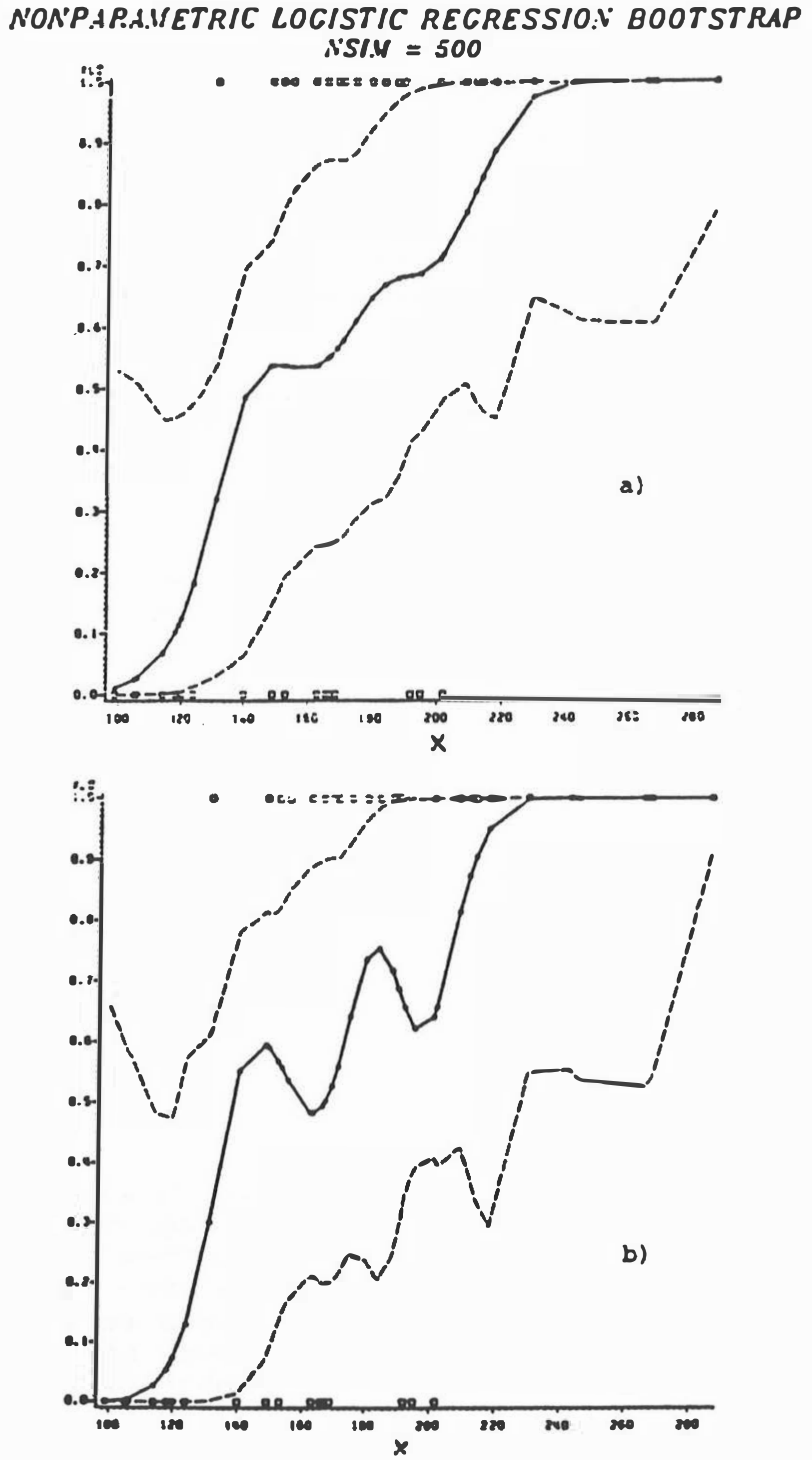

Fig. 3.1 Nonparametric logistic distribution of the injury severity (y = 1 for AIS > 3 and Y = 0 for AIS ≤ 3) over the TTI with 5 % confidencebands for 500 simulations according to the bootstrap method.
a) bandwidth h = 13
b) bandwidth h = 9

The enhanced histogram of prediction errors

This is a histogram of the observed differences between the observed trauma index and the predicted index where large indices are marked in a special way. The procedure is as follows. 1. Compute all the differences predicted response - observed response. 2. Index all large trauma values (for the AIS values (predicted or observed) greater than 4.
3. Draw a histogram of these differences where the big trauma indices get marked by using special symbol.

In figure 4.1 we show an enhanced histogram for the TTI (Eppinger et al., 1984) as a predictor variable for the TOAIS (thorax AIS).

Figure 4.1

This Thoracic Trauma Index is defined through

$$TTI = 1.4\ AGE + 0.5\ FORCE.$$

One sees from this enhanced histogram of prediction erros that the TTI leans toward over estimating the true responses. Indeed, the histogram is skewed to the right. There are 11 observations involving the thorax AIS value of 4. Two of these eleven observations have prediction error zero.One observation has been predicted to have AIS value 4, but really had value 3 (prediction error 2 to the right in the histogram), and eight observations had AIS value 4 but were wrongly classified as 3. One should therefore search for a model that more faithfully predicts the high AIS values.

A distortion measure

As a measure of distortion of current fit we would like to propose two subintegrals of the above histogram. This pair of numbers tells first whether the fit is skew, i.e. has a bias towards over- or underestimating the true response value. Secondly the size of the subintegrals relative to the sample size immediately gives a goodness of fit criterion. The first subintegral just counts the number of positive exceedances (to the right of the column zero in figure 4.1). The second subintegral counts the number of negative exceedances, in this case -8. This together gives the distortion mesure (-8, 35) which describes in a very condensed form the skewness of the prediction and how much the true values are missed by the above model.

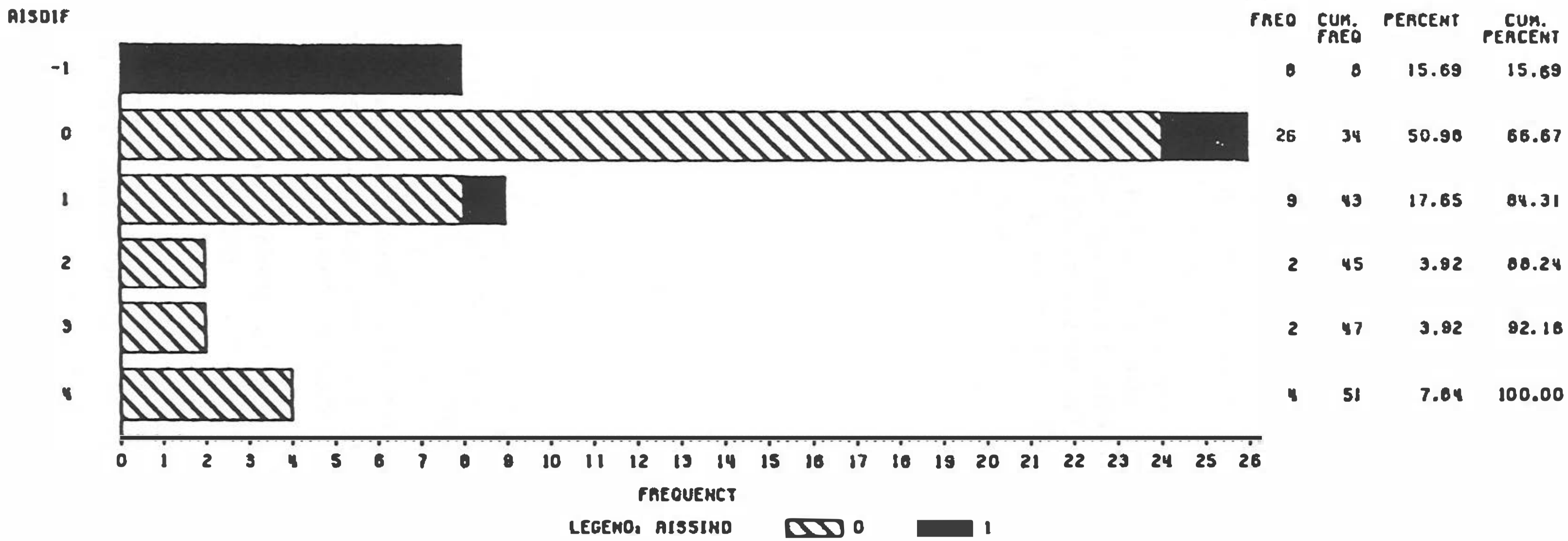

Fig. 4.1 Logistic prediction of the body-AIS (TAAIS) from TTI. Enhanced AIS-difference-histogram between the observed predicted TAAIS.
Marked black: body-AIS-predictions, in which one TAAIS 5 is involved.

The Isoquants

The plot of isoquants is designed for two dimensional predictor variables and shows in a graphical way what trauma indices are to be expected given all possible combinations of covariables. In figure 4.2 we show the predicted thorax AIS classes as a function of AGE and FORCE, as defined in Kallieris, Mattern and Härdle (1986).

Figure 4.2

The region indicated by the letter A would be the region of (AGE, FORCE) combinations where AIS = 0 would be predicted. The region with AIS = 3 is shown by D and the highest AIS value of 4 is marked by an E. Overlaid in this plane are the original data values (0,1,2,3,4). This plot allows simple comparison of different fits by simply studying the regions that determine the AIS values. Given for instance the age of 30 one can easily determine by raising the values of FORCE at what points of FORCE the prediction to higher AIS classes would happen. (FORCE level 140 jumg to predicted AIS 3, FORCE level 250 jump to predicted AIS 4).

5. Application to the Heidelberg data

Only a few research onsets are suited to determine the connection beween mechanical influence and injury severity when measured in AIS degrees. There are real accident analyses on one hand and crash tests with post mortem human subjects (PMHS) on the other hand. Both research onsets are not ideal. The advantage of crash tests with PMHS is, e.g., that by defined conditions of the accident severity, loads acting on the body can be measured in physical magnitudes like acceleration at ribs, sternum, vertebral bodies and head. This is not possible in the real accident analyses. Differences of the injury limits against the living human beings are criticized as a disadvantage of the crash tests with PMHS. The load values measured on the bodies of the PMHS however, are indispensable basis data for the construction of dummies, if these dummies should be qualified for the injury prediction in crash tests.

At the Institute for Legal Medicine of the University of Heidelberg crash tests were conducted with PMHS and dummies for many years to investigate this research concept. As follows, the investigation of lateral collisions should represent which connections exist between loading parameters at the body of the PMHS, anthropometric data and injury severity and how these connections can be used for injury prediction by utilization of the statistical methods described above. Basis of the connection analyses are 58 90-degree lateral collisions. In these collisions PMHS have been loaded in near side position in the impacted/standing vehicle.

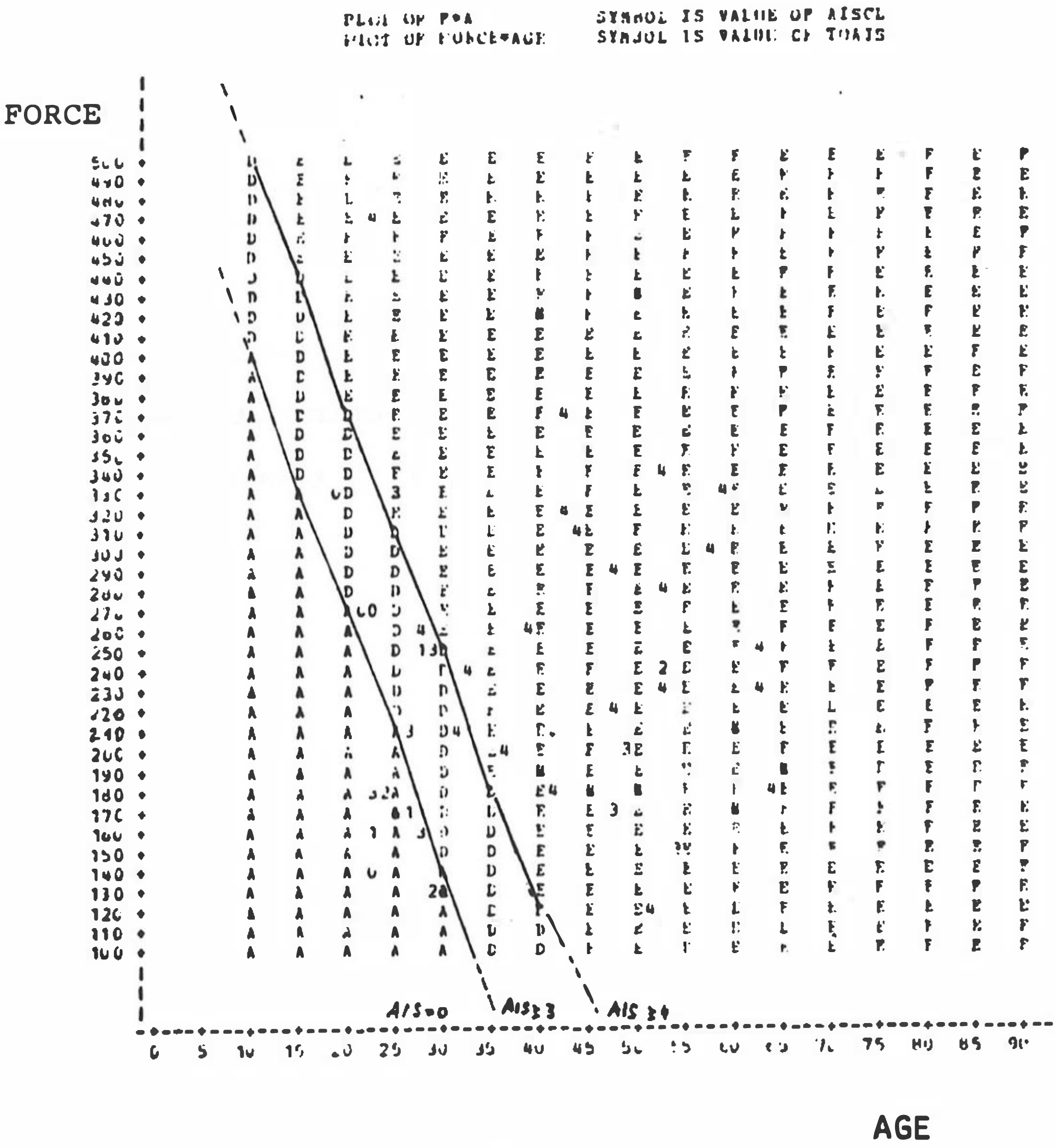

Fig. 4.2 Isoquantplot for the illustration of the prediction results of the logistic regression from AGE and FORCE.
Zone A: prediction of TOAIS = 0
Zone B: prediction of TOAIS = 3
Zone E: prediction of TOAIS = 4
Numbers in the zones: observed thorax-injury degrees
FORCE = 1/2 (accel.max. 4th rib impacted side + max. result. accel. Th 12) x bodymass / 75

The crash tests have been conducted at impact velocities of 40, 45, 50 and 60 km/h (Kallieris et al., 1987). In the PMHS 22 acceleration values at head, thorax, spinal column and pelvis have been recorded for each test. The injuries of the PMHS have been scaled according to AIS 80. It was seen in the statistical analyses that the injury levels could be most effectively predicted by the method of logistic regression. In the 90 degree lateral collisions the body injury severity (TAAIS) was generally leading and determined the maximum injury severity (MAIS). Therefore, the prediction of the body injury severity for right side lateral collisions is presented here as an example. Among the 22 as maximum and 3 ms values recorded accelerations the following proved to be the best predictors:

1. Acceleration (3 ms value) in x-direction at lower sternum (BUX3) (g);
2. acceleration (3 ms value) at the 12th thoracic vertebra in y-direction (T12Y3) (g);

The further improvement of the injury prediction has been reached in considering the Body Mass (BMASS) (kg) as covariable. With these covariable combination, the logistic model estimated the following parameters for the injury index Z:

$$Z = 0.15 \text{ BMASS} + 0.08 \text{ T12Y3} + 0.06 \text{ BUX3}.$$

The probability curves for TAAIS rankings 0,4 and 5 are shown in figure 5.1, for impacts from the right. The three tests with TAAIS 2 and 3 in the test series were not considered.

Figure 5.1

Below a Z value of 18.3, the envelope of the AIS probability curves indicates a high probability to be uninjured (the highest probability is below Z = 18). Between Z = 18.3 and Z = 20, a TAAIS of 4 is largely to be expected and above Z = 20 the probability for TAAIS 5 of about 45 % increases continuously to 100 % (at Z = 25). The enhanced TAAIS difference histogram (see section 4) in figure 5.2 shows that the above mentioned covariable combination as correctly predicts 59 % of the cases. The model predicts the TAAIS in 19 % too high and in 15 % a level too low; each one time, the model underestimated the observed injury for two and 4 AIS degrees.

Figure 5.2

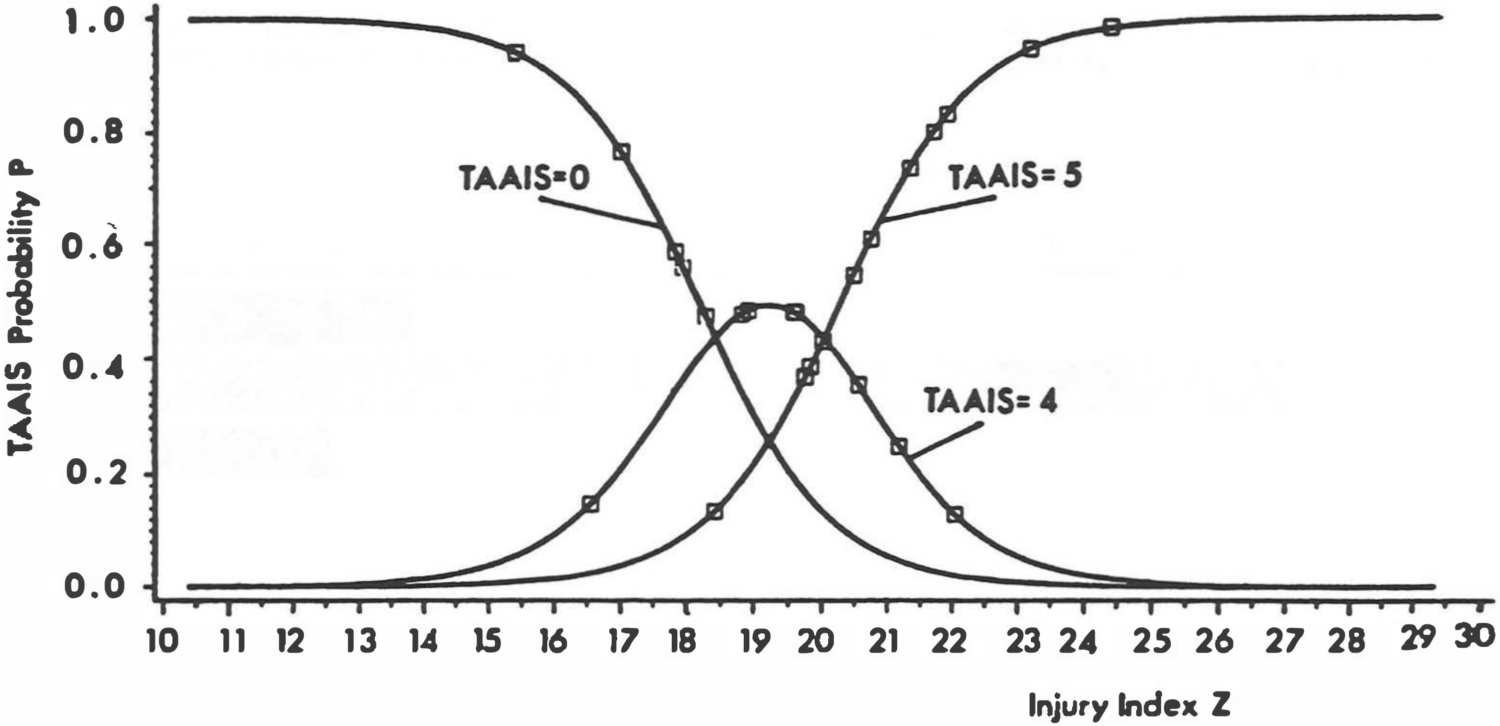

Fig. 5.1 Predicted probability of torso injury (TAAIS) for impacts from the right

□ Observed degree of injury

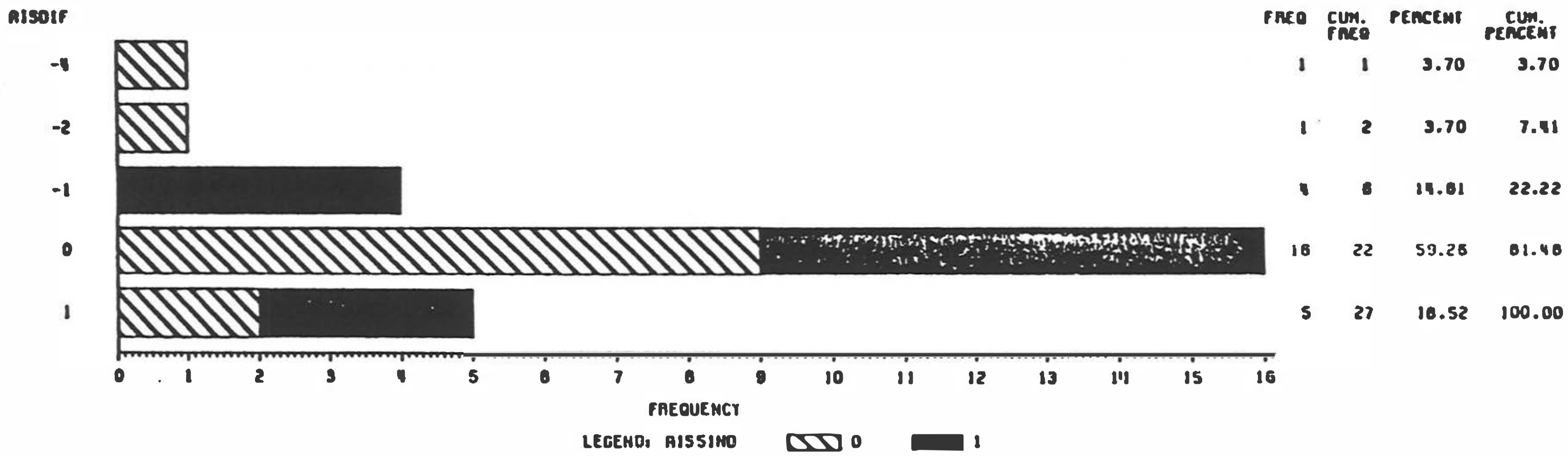

Fig. 5.2 Enhanced AIS-difference-histogram for the logistic prediction of the body-AIS (TAAIS) grom Z = 0,15 AGE + 0,08 accel. Y-direct. Th 12 + 0,06 3 ms accel. X-direct. upper sternum

6. Conclusions

We have presented several multinominal response models of parametric and non-parametric nature. A way of comparing these models and deciding which one is more appropriate than others is given by considering non-parametric alternatives in the construction of a simulation band. This simulation band technique (section 3) lead for the Heidelberg data to the conclusion tht the Logistic response model is appropriate for the analysis of car-to-car side impacts. Comparing the Likelihoods of the Logistic and the Weibull link functions we found no better fit for the Weibull model, see Kallieris, Mattern and Härdle (1986). We furthermore presented a variety of graphical techniques which are of great assistance when looking for suitable predictor variables X, see section 4. Using these techniques we found for example that the Logistic model using the trauma index

$$Z = 0.15 \text{ BMASS} + 0.08 \text{ T12Y3} + 0.06 \text{ BUX3}$$

had good prediction properties for the TAAIS, see section 5.

REFERENCES

AIS (1980) States JD, Huelke DF, Baker SP, Bryant RW et. al. The Abbreviated Injury Scale, 1980 Revision.

Akaike H (1977) On Entropy Maximization Principle. In Applications of Statistics, Ed.P.R. Krishaniah. Amsterdam, North Holland

Berkson J (1951) Why I prefer Logits to Probits. Biometrics 7: 327-339

Bickel P, Doksum K (1977) Mathematical Statistics. Holden-Day Inc., San Fransisco

Eppinger RH, Marcus JH, Morgan RM (1984) Development of Dummy and Injury Index for NHTSA's thoracic side impact protection research program, SAE technical paper series 840885, Government/Industry Meeting and Exposition Washington D.C.

Friedman J, Stuetzle W (1981) Projection Pursuit Regression. J.Amer.Statist.Assoc. 76: 817-823

Härdle W, Stoker T (1988) Investigating smooth multiple regression models by the method of Average Derivatives. J. Amer.Statist.Assoc., to appear

Härdle W (1988) Applied Nonparametric Regression. Book to appear

Kallieris D, Mattern R, Härdle W (1986) Belastbarkeitsgrenzen und Verletzungsmechanik des angegurteten Pkw-Insassen beim Seitaufprall. Phase II: Ansätze für Verletzungsprädiktionen. Schriftenreihe der Forschungsvereinigung Automobiltechnik e.V. (FAT) Nr. 60, Frankfurt/Main 17

Kallieris D, Schmidt Gg, Mattern R (1987) Vertebral Column Injuries in 90 degree Collisions - A study with Post Mortem Human Subjects. Proc. of Intern. IRCOBI Conf. on the Biomechanics of Impacts, Birmingham, 189-192

Neter J, Wasserman W (1974) Applied Linear Statistical Models. Richad D. Irwin, Inc. Homewood Illinois

SAS - Statistical Analysis System, Cary North Carolina

SAS - Supplementary User's Guide, Cary North Carolina

Stone M (1974) Crossvalidatory choice and assessment of statistical predictions (with discussion). Journal of the Royal Statistical Society, Series B, 36, 111-147

This paper is kindly dedicated to Prof. Dr. med. Georg Schmidt, Heidelberg, on the occasion of his 65th birthday.