10176 / 338

UCLA Statistics Series #20

November, 1988

The relation between in and output variables in various groups of individuals

with applications to school effectiveness research

Jan de Leeuw
Ita Kreft

UCLA Classification: Applications. Social Sciences-Education

Abstract

In this paper we review some recent developments in the field of *multilevel analysis*, or more generally in the field of relating sets of variables in different groups of individuals. We mainly concentrate on example from school effectiveness research (do schools make a difference ?). Our treatment emphasizes discrete nonparametric probability models and de-emphasizes the linear model. This is because nonparametric models are more simple and at the same time more realistic in most educational research situations.

Version of the second author

Introduction

One of the basic problems in schoolcareer research, and of status attainment research in general, is to establish the relationships between different sets of variables describing individuals. The first set usually consists of various variables describing *background*, such as socioeconomic status of the family, number of siblings, religion, race, and sex. The second set of variables describes the *career* of the individual, such as his schoolresults, his choice of profession, or his income at various stages of his life. To some extent the classification of variables into these sets is arbitrary, and depends on the ~~purposes~~ of the investigation. In some investigations, for example, the intelligence ~~test~~ score at age 12 is taken as a background variable, in others it is considered to be a step in the career. This depends, obviously, on the particular point in time where one thinks the career starts. It also depends, a bit less obviously, on the preconceived notions the investigator has about intelligence. If intelligence is genetically determined ability, then it is best considered to be background, if it is a result of primary schooling and learning at home, then one can easily think of it as a part of the career. In the same way some studies, in which for instance the main purpose of the investigation is to predict later academic achievement, consider the whole primary school career of the individual as background. This means that often the distinction between background and career is operationally equivalent to the distinction between *independent* and *dependent* variables, or between *exogeneous* and *endogeneous* variables. Because we want a terminology that is quite general, and not too esoteric, we prefer to speak of *input* and *output* variables. It is the choice of the researcher to define why some variables are considered input while others are output

appear to achievement and as [illegible]

An important point in this context is that very often educational sociologists and psychologists go into much more theoretical detail about the relationships between the variables in their study. They do not simply divide variables into two sets, but they use more sets and also talk about relationships within sets. This leads to the notion of a path model, which is a considerable, although straightforward, generalization of the distinction between dependent and independent variables. Although we think that the use of path models is, in itself, perfectly legitimate and potentially very interesting, we also think that there are not many useful applications of complicated path models in the social science literature. Compare Freedman (1987) for a forceful criticism of path models, and also compare the discussion ~~following his pa~~per. Similar criticisms have been voiced by Cliff (1983), De Leeuw (1985), and Dessens and Jansen (1986). Simple multiple regression models and perhaps simple causal chains can sometimes be applied, but generally the data are not strong enough, and the prior theoretical knowledge is too weak, to justify use of complicated models. Consequently we do not think it to be a great loss of generality if we restrict our attention to a very simple class of descriptive path models, in which there are just two sets of variables.

in Kreft (1987)

A third introductory point is that most educational researchers assume, already at the very first steps of their analysis, that their data can be considered to be a random sample from a multivariate normal distribution. If we think of the usual situations in schoolcareer research this assumption is simply ludicrous. Most of the variables that are collected are discrete, with only a few possible values, and even these few values are often not unambiguously ordered. This indicates that weak nonparametric statistical models are necessary, and of course the use of nonparametric models means that samples have to be large, and that only a few variables can be studied at the same time.

The typical modern study in school career research or in status attainment research is based on samples between 1000 and 100.000 individuals, but the number of variables modeled simultaneously is often 20 or more. Even with an average of 3 categories this means 3^{20} cells, and even if we have a million individuals we still have, on the average, only one individual in every 3500 cells of the table. This means that researchers are caught between two extremes: either they do rather pedestrian studies with only a few variables and huge samples, or they are forced to make many silly assumptions in order to be able to proceed. It seems that sociologists and educationalists almost exclusively choose the second option, and fit very complicated models which seem to suggest a lot of detailed prior knowledge, and which seem to be able to make many very precise statements about career attainment. We suggest that most of this precision is spurious, just capitalizing on chance, combined with forcing many prejudices (Kalman, 1983) on the data analysis. As a consequence we shall concentrate in this paper on simple models for discrete variables. And we shall try to stay as far away as possible from the linear model. The problems we study here have usually been discussed in the context of the linear model. This is confusing, because problems specific to the linear model, such as multicollinearity, become confounded with the basic probabilistic issues that deal with the sampling design. The mathematical level of the paper is consequently quite low, we merely require a basic knowledge of the rules for computing conditional and marginal probabilities.

Ultimately our discussion is based on the particular interpretation we have of the practice of fitting statistical models. We shall not discuss this interpretation in detail, because that would take us to far astray, but the idea is briefly that models are used to *smooth* data structures. Such smoothing is necessary for purposes of *communication*, because we cannot publish data matrices which are too large, for purposes of *interpretation*, because relations between

It is also necessary

variables may not be obvious, ~~and~~ for reasons of *prediction*, because smoothing can lead to a more stable basis for extrapolation. Various aspects of this point of view are discussed in more detail in De Leeuw (1984, 1986, 1987, 1988).

when changing the raw data

The role of the sampling design

There are several statistical complications specific to schoolcareer research which we mention first. They have to do with sampling. In the simplest case the data consist of a single sample of individuals from a well-defined population. In more complicated cases this sample is stratified: we decide to include a fixed number of individuals from black families and a fixed number of individuals from white families. Now being white or black is not a random variable any more, but we actually deal with two different populations from which we have sampled. It can very well be the case that the relationship between input and output is different in the two populations. A related situation is the following. We have a single random sample, but we decide to look at the boys and girls separately. This corresponds with the statistical operation of conditioning, because we look at the relation between input and output, given that the person is a boy or a girl. Clearly sex cannot be among the input variables any more in this setup, because it has by definition no variance in each of the two populations.

The situation becomes even more complicated in the case of two-stage sampling. Suppose we draw a sample of high schools, and then for each school we draw a sample of students from that school. Suppose also that we have some variables describing schools, such as the resources they have, the teaching style they use, and the type of discipline they enforce. Obviously we want to incorporate both the variables describing individuals and the variables describing schools in our analysis. There are two simple ways to do this. The first one is to *aggregate* all individual variables to school level by computing within school averages. The analysis is then done at school level, because we now have variables at school level only. The second procedure is to *disaggregate* the school level variables, by assigning to each individual the teaching style or discipline type of his school. The analysis is then done at the individual level. Both procedures are two-step procedures: in the first step they aggregate or disaggregate to bring all variables on the same level, then in the second step they analyze on that common level. Neither of them is wrong, but they both have their dangers. If we aggregate, then the sampling characteristics of the aggregated individual variables will generally be different from those of the school variables. The variance of an aggregated individual variable will, for instance, be inversely proportional to the size of the school. This must be taken into account in the school level analysis. If we disaggregate school level variables, then we must take into account that individuals in the same school will necessarily get the same value on the disaggregated school variable, and thus scores on disaggregated variables cannot be independent between individuals of the same school. This also must be taken into account in the final analysis.

If we briefly summarize these considerations, we see that we must take the sampling characteristics of the data into account, and if we apply a two-step procedure we have to take the operations in the first step into account when analyzing our data in the second step. This seems elementary advice, but until recently it was ignored, or it could not be implemented very well because the technical apparatus was not available. Some valiant, but incomplete,

Inference from school level to individual level not valid (Kreft, 1987, Kreft & De Leeuw, 1988) attempts are Burstein, Linn, and Capell (1976), Burstein, Miller, and Linn (1980), and Van den Eeden and Saris (1984). In order to explain these approaches, we first translate them to our framework, starting with a simple case.

Notation and examples

But let us introduce some notation to start with. The input variables are collected in a vector x, and the output variables in a vector y. We are interested in predicting y from x, or in the dependence of y on x, which means in probabilistic terms in the conditional distribution p(y|x). Because we assume throughout the paper that the variables in x and those in y are discrete, the joint distribution p(x,y) of x and y can be collected in a single bivariate contingency table or *cross table*, with the values of x defining the rows and the values of y defining the columns. This does not mean that we assume that x and y are univariate. If x is defined by SEX, with 2 possible values, father's profession SES, with 6 possible values, and (grouped) IQ, with 7 possible values, then these three variables can be combined to one single variable with $2 \times 6 \times 7 = 84$ values, and the cross table has 84 rows. This trick shows that in the context of categorical data analysis the dimensionality of the input and output does not really matter. The data can be collected in a single cross table. Of course if we have many variables in x and/or y we rapidly run into practical problems, because the number of rows and/or columns will soon be enormously large, and the table will be very empty indeed. But for our formulas and for general discussion these practical problems do not matter, and we shall ignore them for the moment. Observe that the conditional distributions p(y|x) are found by dividing each entry in the table by the corresponding row-total. Thus we compute percentages row-wise.

We shall illustrate our calculations by using two datasets from the Netherlands. The first is the SMVO data set (CBS, 1982). This is a cohort of 37000 students in the Netherlands, who left primary school in 1977. The Central Bureau of Statistics collected background information on these students, and followed their careers in secondary education. We shall use a selection from the cohort made by Meester and De Leeuw (1983). They eliminated all students for which there were no intelligence test scores, and all students with non-working parents. This leaves 16236 students. We also use variables and codings used by Meester and De Leeuw. The first variable is SEX, with two values M and F. The second one is father's profession SES, with the six values WOR (blue collar), FAR (farmers), SHO (shopkeepers), LOW (lower white collar), MID (middle white collar), HIG (professionals, higher white collar). The third variable is the intelligence test score TIB, categorized in seven categories. And, finally, the fourth variable we use is ADV, the sixth grade teacher's advice about the most appropriate form of secondary education for the pupil. This has four levels LBO (lower blue collar), MAV (lower white collar), HAV (higher white collar), VWO (pre-university).

The SMVO data are very interesting and powerful, because of the large and carefully constructed sample, but they have no information on the schools that the pupils go to. In order to be able to study school effects we use the GALO data set (Peschar, 1973). These data were collected in the city of Groningen. We use the 1959 cohort, which consists of 1270 students in 37 schools. For each of the students we have information about SEX, IQ, SES, and ADV, as in SMVO. The categorizations are a bit different, however. IQ in GALO is a continuous variable, with a population mean of 100 and a population standard deviation of 16. SES is categorized in WO1 (unskilled labour), WO2 (skilled labour), while LOW, SHO, MID, HIG are the same as in SMVO. Variable ADV has seven levels, NFE (no further education), VLO (extended primary education), LBO (as in SMVO: lower technical and domestic), LLT (lower agricultural), ULO (corresponds with MAV in SMVO), MMS (higher general education for girls), VHMO (corresponds with VWO). Again there are more details in Meester and De Leeuw (1983).

Ignorability

The first problem we discuss is a relatively simple one, but it is very important in practice. It is sometimes called the *selection problem*. It occurs if we have incomplete information about the relationship between input and output, or if we only have information for a not necessarily representative subset of the population we are interested in. In educational research a common example of the situation is that we know the relationship between variables in a particular school or district, and we want to make a statement about the relation in the general population. Missing data problems and restriction of range problems are both special cases of the selection problem. As a matter of fact we shall discuss the somewhat more general situation in which we have complete information about the distribution of our variables in a population $\mathcal{P}_0$ and we have incomplete information about the distribution another populations $\mathcal{P}_1$. We want to use the information we have about $\mathcal{P}_0$ to complete our information about $\mathcal{P}_1$. Combine this with the notation introduced earlier. We know the joint distribution p(x,y) of x and y in a *selected population* $\mathcal{P}_0$, but we are really interested in the distribution in the *target population* $\mathcal{P}_1$. Let us write these joint distributions as p(x,y|z), with z either 0 (selected) or 1 (target). We do not know the joint distribution p(x,y|1), and we want to estimate it in some way or another. This goes beyond the data, and thus the only way to do the estimation is to substitute assumptions for data.

We illustrate this with one of our examples right away. $\mathcal{P}_0$ is the population defined by our selection from the SMVO data (of about 16000 pupils), and $\mathcal{P}_1$ is the population of all SMVO students (of about 37000). Variable y is ADV, and x is the background variable SES. Let us suppose that the conditional distribution of y given x is the same in $\mathcal{P}_0$ and $\mathcal{P}_1$, i.e. that p(y|x,0) = p(y|x,1). This means, roughly, that although our populations may differ in background, the relationship between input and output is the same in both populations. Also suppose that we do know the marginal p(x|1). Then it follows, by the definition of conditional probability, that p(x,y|1) = p(y|x,1)p(x|1) = p(y|x,0)p(x|1) = p(x,y|0)w(x), where w(x) = p(x|1)/p(x|0). This means that, in order to find p(x,y|1), we take each row of p(x,y|0) and multiply it with w(x). This is, of course, the weighting procedure that has been used in survey research for a long time. We see that it is based on the assumption that the conditional distribution of y given x is the same in both populations.

Table 1 has the necessary numbers, taken from Meester and De Leeuw (1983, pages 189, and 197-199). The 6 x 4 table on the left is p(x,y|0), the bivariate distribution in the selected sample of about 16000, at the bottom of the table we see the weights w(x) which must be applied to the rows of the table, and the 6 x 4 table on the right is the adjusted table $p(x,y|1)$. The adjustments are minor, but it is still obvious from the weights that the lower classes are under-represented in our subsample and the higher classes are over-represented. In the case that it is really the marginal distribution of y that we are interested in we can use $p(y|1) = \Sigma$ $p(x,y|0)w(x)$. This is compared with $p(y|0)$ in Table 2.

Let us now consider a slightly more complicated example. Suppose TIB and ADV are both in y, the selected population $\mathcal{P}_0$ is the SMVO cohort, and the target population $\mathcal{P}_1$ is the population of all primary school leavers in the Netherlands. Thus we want to find $p(y|1)$, the cross table of TIB and ADV in the Netherlands, having observed only the SMVO sample. The first job we have to do is to find variables x such that it is reasonable to assume that $p(y|x,1) = p(y|x,0)$. We could decide, for instance, that using SEX and SES in x is sufficient to make this assumption reasonable. If it is not, we have to add more variables (such as the regio the student comes from, or some characteristics of her school), and we may get into empty cell problems rapidly. We then measure $p(y|x,0)$ in our selected population of students, and use the formula to estimate $p(x,y|1)$ and $p(y|1)$. Again for this it is necessary that we have $p(x|1)$, the SEX * SES distribution in the Netherlands, but this will often be available from census or other official data.

We can now easily generalize the notation $p(y|x,z)$ with z either 0 or 1 to the case in which we have more populations, i.e. more values of z. The key assumption remains the same. In the terminology of Rubin (1977) we say that if the conditional distribution $p(y|x,z)$ is the same for all z, i.e. for all populations $\mathcal{P}_j$, then membership of subpopulations is *ignorable*. We can also write the condition as $p(y|x,z) = p(y|x)$, where $p(y|x)$ is the conditional distribution in the union of all populations. Another, more symmetrical, way of writing the condition is also interesting. Simple manipulations show us that $p(y,z|x) = p(y|x)p(z|x)$, or given x the variables y and z are independent. If we make a cross table of y and z for each value of x, then the chi square of all these cross tables must be zero.

GALO

Let us look at a school effect situation where this theory can be applied. Suppose indivuals are assigned to schools solely on the basis of the measurements of the background variables in x. This means that there is a new random variable z describing school membership, and that we have $z = f(x)$, for some function f. The values of f are the different schools in our study. Each school corresponds to a population $\mathcal{P}_j$. It is not necessary that we actually know f, it suffices to assume that such a function exists. Because $p(x,y) = \Sigma\, p(x,y,z) = \Sigma\, p(y|x,z)p(z|x)p(x)$ we find that $p(y|x) = \Sigma\, p(y|x,z)p(z|x)$. But z is a function of x, which means that $p(z|x)$ is either one or zero. It is equal to one if and only if $z = f(x)$. We can formalize this by defining $\delta(x,z)$, which is equal to 1 if $z = f(x)$ and equal to 0 otherwise. Thus $p(y|x) = \Sigma\, p(y|x,z)\delta(x,z) = p(y|x,f(x))$. Thus in each school f(x) the conditional distribution of y given x is the same, and assignment to schools is ignorable. In the aggregation oriented literature (compare, for instance, Langbein, 1979) this condition is also known as *grouping on the independent variables.*

It will be clear that the plausibility of the ignorability assumption depends strongly on the input variables one includes in the conditioning operation. If there are many input variables that one controls for, then the assumption becomes more plausible. But at the same time less useful, because having many variables means that we run into the empty cell problem again, and we have problems collecting sufficiently many data points. Clearly we can maintain that ignorability implies that schools do not make a difference, because differences in the distribution of y are only due to differences in individual variables x. If we control for x, then there are no differences. And conversely we can say, that schools do make a difference if ignorability does not obtain, even for an infinite number of individual background variables. It is important to remember that in order to say that schools make a difference we not only have to say in which respect (i.e. for which y), but also we have to list the background variables in x we control for.

This is illustrated for the GALO data in Table 3. For y we use ADV, for x there are three different choices. The first choice is to use SEX * SES, with 12 values, the second one is to use SES only, with 6 values, and the final analysis only uses SEX, with two ~~variables~~ values. We can now test ignorability, i.e. we can find out if schools make a difference. Thus we want to test if $p(y|x,z) = p(y|x)$, where z takes on the values corresponding with the 37 schools. For our first choice of x, for instance, we make 12 tables of dimensions 37 x 7, and we add the chi squares of these 12 tables. Each table adds 36 x 6 = 216 degrees of freedom to the total chi square. In Table 3 we see that chi square becomes higher if we control for more variables, but if we take the degrees of freedom into account it becomes lower. Thus ignorability becomes more true if we control for more background variables, as expected. On the other hand column 3 in the table shows the percentage of zero cells in the table, which is already a disconcerting 50% if we only control for SEX, and which becomes ~~almost~~ 80% if we control for SEX and SES. Observe that we have not even tried to control for IQ. This means that the statistical theory for chi square, which is based on tables which are relatively well-filled, clearly does not apply any more. Thus we really cannot test the chi squares for significance, using the usual tables.

Selection theorems with linearity

linear

Now let us suppose the ignorability assumption is true, and let us assume at the same time that the regression of y on x in the target population $\mathcal{P}_0$ is linear. We do not assume linearity in the selected population $\mathcal{P}_1$. We now introduce some convenient notation. Let us write m_x^0 and m_x^1 for the averages of x in the two populations. In the same way define m_y^0 and m_y^1. We use the same type of notation for variances and covariances. Thus V_{xy}^0 is the covariance matrix of x and y in the target population, and V_{yy}^1 is the covariance matrix of y in the selected population. We know the means and dispersions in the selected population, but at least some information about the target population is missing.

Linear regression of y on x in the target population means that $\mathbf{E}(y|x,0) = Ax + b$. Ignorability implies that $\mathbf{E}(y|x,0) = \mathbf{E}(y|x,1)$. Thus A and b can be estimated from the selected population. It follows that $\mathbf{E}(y|0) = A\mathbf{E}(x|0) + b$. If $\mathbf{E}(x|0)$ is known, we can actually estimate $\mathbf{E}(y|0)$, without sampling the target population. More interesting results become available if we also assume homoscedasticity in the target population. Thus, with V the marginal

for dispersion, we assume that $V(y|x,0)$ is independent of x. It follows that $V(y|0) = AV(x|0)A' + V(y|x,1)$. Clearly we can compute $V(y|x,1)$, we already know A, and thus if we also know $V(x|0)$ we can compute $V(y|0)$.

Let us give a concrete example again. Suppose we want to know the distribution of a battery of intelligence tests in the population of the Netherlands, given a the distribution in a sample from the population of first year students at the University of Groningen. The trick is to find selection variables x first such that it is plausible that $p(y|x,0) = p(y|x,1)$. If x is SES, for instance, then we assume that p(intelligence|SES) is the same for the students as for the general population. We now compute the regression of intelligence on SES and the conditional dispersion of intelligence given SES for the students. We use the mean and variance of our SES indicators for the general population, which we presumably know from other sources, to compute mean and variance of the test scores in the general population.

The same basic reasoning can be used with many other types of groups. Suppose for example that we give students a test x at the end of primary school. Those whose score lower than x_0 go to school type 1, those who score between x_0 and x_1 go to type 2, and those who score higher than x_1 go to type 3. Suppose schools do not make a difference. Thus p(success|test) is the same for all three types. Nevertheless the distribution of success y will be very different in the three types, because of the selection. We find, for instance, $p(y|1) = \Sigma \{p(x)p(y|x) \mid x \leq x_0\} / \Sigma \{p(x) \mid x \leq x_0\}$. Thus, if the regression is linear, $\mathbf{E}(y|1) = A\mathbf{E}(x \mid x \leq x_0) + b$. If both y and x are univariate, and x is standard normal, then we find $\mathbf{E}(y|1) = -a\phi(x_0) + b$ and $\mathbf{E}(y|3) = a\phi(x_1) + b$, with ϕ the ordinate of the standard normal density. If y_1 and y_2 are two measures of school success, then we find for their covariance in type 1 schools $C(y_1,y_2|1) = C(y_1,y_2|x) + a_1a_2V(x \mid x \leq x_0)$, and so on. It could be, for instance, that y_1 is success in the first grade of secondary education, while y_2 is success in the sixth grade. It is clear that even if schools do not make a difference, the regression of sixth grade success on first grade success can be very different in different school types due to selection on the test x. Observe that in general this regression will even be nonlinear (again due to selection).

$p(z|x)$

$\left\{ \begin{array}{l} p(x, z \ldots \\ (x, y \mid 0) \end{array} \right.$ [illegible]

$p(\ldots)$

In the situations discussed so far, applied to school effectiveness studies, can be described as follows. Because p(y|x) is the same in all schools, we can clearly say that schools do not make a difference. But because p(x) may be different in all schools it follows that p(y) can be different, and thus schools may seem to make a difference. If one thinks about this a little bit more the problem becomes somewhat of a pseudo-problem. If we control for an infinite number of relevant variables, then ignorability becomes almost necessarily true, although the actual measurements can no longer be carried out. And of course ignorability always remains an assumption which cannot really be tested in practice, because we do not have the required measurements in the target population. If we had them, there was no need to assume anything in the first place.

Nonignorable membership

If membership is non-ignorable, i.e. if $p(y|x,0) \neq p(y|x,1)$, then we are in an essentially more complicated situation. The problem is that ignorability is defined clearly, and it is a strong enough assumption to have interesting consequences. But processes can be non-ignorable in many different ways, and we have to be more specific before we can actually use the concept. This means that we actually have to model the form of non-ignorability, and in order to do this in a responsible way we need informations which is unfortunately rather scarce in school effectiveness studies.

Suppose, for example, that we *group on the dependent variable*. Thus $z = f(y)$. It follows that $p(x,y,z) = 0$ if $z \neq f(y)$ and $p(x,y,z) = p(x,y)$ if $z = f(y)$. Suppose $f^{-1}(z)$ is the set of y for which $f(y) = z$. Then $p(x,z) = \Sigma \{p(x,y) \mid y \in f^{-1}(z)$. Thus $p(y|x,z) = 0$ if $z \neq f(y)$, and $p(y|x,z) = p(x,y)/p(x,z) = p(x,y)/\Sigma \{p(x,y) \mid y \in f^{-1}(z)\}$ otherwise. $E(y \notin f^{-1}(z))$

group on ~~latent~~ latent variable.

LV

$p(x,y,z,\chi) = p(x,y,z|\chi)p(\chi) = p(x|\chi)\, p(y|\chi)\, p(z|\chi)\, p(\chi)$

$p(y|x,z) = \dfrac{\sum_\chi p(x|\chi)\, p(y|\chi)\, p(z|\chi)\, p(\chi)}{\sum_\chi p(y|\chi) \ldots p(\chi)}$

Σ

$\int y\, p(y|x,z)\, dy$

$= \dfrac{\int_{y \in f^{-1}(z)} p(\ldots)\, dy}{\int p(\ldots|y)\, dy}$

	LBO	MAV	HAV	VWO	LBO	MAV	HAV	VWO
WOR	.1561	.1456	.0460	.0237	.1488	.1388	.0438	.0226
FAR	.0167	.0285	.0101	.0063	.0157	.0268	.0095	.0060
SHO	.0320	.0428	.0166	.0109	.0313	.0419	.0162	.0106
LOW	.0305	.0565	.0283	.0179	.0321	.0594	.0297	.0188
MID	.0361	.0777	.0527	.0420	.0380	.0816	.0554	.0441
HIG	.0101	.0415	.0361	.0354	.0105	.0435	.0379	.0371

weights .9530 .9410 .9782 1.0511 1.0505 1.0486

Table 1 : SES * ADV observed and adjusted, with weights

$p(y \mid 0)$	.2815	.3926	.1897	.1362
$\Sigma\ p(x, y \mid 0)w(x)$	.2764	.3919	.1925	.1392

Table 2: Marginals for ADV, observed and adjusted