Primary Education in Madagascar: Admissions, Enrolment, Delay, Drop-outs

J.K. Lindsey, R. Rakotondrazaka, and C. Tibi,
with the help of
Patrick Lindsey, ChangLin Mei, and Alberto Ferreira Poblete

Coordinated by
the UNESCO
and
the Ministry of National Education, Madagascar

Preface

This study was supported and financed in connection with the programme MAGPLANED, Programme d'Amélioration de la Planification de l'Education à Madagascar, and coordinated by the UNESCO and the Ministry of National Education of Madagascar. The study was possible thanks to the SAT1 funds of the UNDP, along with other resources from MAG/91/001 and 700-MAG-10, that permitted its financing.

Part of the analysis work was done by ChangLin Mei and Alberto Ferreira Poblete during the months of July, August, and September, 1995 at the Limburgs Universitair Centrum, Diepenbeek, Belgium, and another part by Patrick Lindsey in the summer of 1996.

Diepenbeek and Liège J.K.L.
March, 1996

Contents

1
La Grande Isle: Madagascar

Madagascar has the reputation of being a natural paradise with many rare and unusual species. However, when one flies over the country, from the African continent to Tananarive, one has the impression of a stark and inhospitable terrain, of bare hills with little life, human or other. Madagascar is an immense country, full of contrasts.

1.1 Geography

Madagascar is the fourth largest island in the world, after Greenland, New Guinea, and Borneo. It is situated in the Indian Ocean to the south-east of the African continent and is as large as France and the Benelux countries put together. It is 1600 km long and 600 km at the widest point and is surrounded by many small islands and coral atolls.

The country is divided into six provinces (*faritany*): Antananarivo, Antsiranana, Fianarantsoa, Mahajanga, Toamasina, and Toliara, and 111 districts (*fivondronana* or CISCO). The smallest administrative division (*fokontany*) is the neighbourhood in town and the village in the countryside; there are 13,476 of these.

The capital, Antananarivo (Tananarive or, most often, Tana), is located in the centre on the high plateau.

1.1.1 REGIONS

Highlands The central highlands, one-third of the country, consist of high plains and hills. They range in altitude from one thousand to 1600 metres. North of Antananarivo, the plateau extends for several hundred kilometres almost without trees. To the south, however, the country is covered with rice paddies in the valleys and terraced fields on the hills. One also finds thick pine forests and icy falls.

East The east side of the island, covered (originally) with tropical rain forest, is a zone of broken terrain 25 to 100 km wide, separated from the highlands by abrupt escarpments. Small isolated plains, alternating with low hills and separated by numerous estuaries, stretch parallel to the coast for about 20 km in width. In the middle, a 430 km long canal unites these

various estuaries from the north to the south. The often impenetrable forest starts at the foot of the chain of mountains.

West The western plains and plateaus of savanna, and baobabs (*renala*), have a softer form, as does the extreme south. Thus, the highlands drop slowly to the coast in a series of steps. Vast stretches are almost desert, scattered with high grass, palms, or baobabs. The coast is cut up and full of small coves, with many small islands offshore.

South In the southern zone, wooded slopes and steppes alternate, where only thorny bushes, baobabs, and palms grow. This area receives very little rain. The southwest coast boasts a series of splendid beaches.

North The north is a complex meeting place of volcanic forms, basins, and deltas.

1.1.2 CLIMATE

The climate varies from humid tropical in the east through highland tropical to dry tropical in the west. The far north is subject to violent cyclones. On the east coast, the hot, humid southeast wind brings rain throughout the year. The south has a ten month period of dry period each year, too often accompanied by drought (*kéré*), the most recent being in 1992. In most parts of the island, the hot rainy season lasts from November to April. It rarely happens that the sky is overcast the whole day, much less several days.

1.1.3 ECOLOGY

Thus, the island has an extraordinary diversity of natural environments, from coral reefs to rain forests, and from arid deserts to highlands, all unfortunately often gravely threatened. It has flora and fauna that are unique in the world: orchids, palm trees, cactuses, baobabs, lemurs, insectivores, and chameleons.

The population places enormous pressure on the ecosystem of the land. The forest must meet the needs of the population in wood for cooking, heating, and construction. Each year, several hundreds of thousands of hectares of savanna burn and about 0.2 million hectares of tropical forest disappear through the use of cut and burn techniques (*tavy*). In all, it is estimated that 30% of the surface area of the country is (re)burned each year; 85% of the natural forest cover has disappeared. Erosion sends masses of red earth to the sea.

The tropical forest, that covered three quarters of the country a few centuries ago, now only occupies one fifth of the area. The practice of cut and burn in the forests has clearly increased in the last fifteen years,

closely connected with the increasing poverty of the rural population. It threatens the forest cover in a definitive fashion, especially on the east coast. Numerous regions will soon be desert.

1.1.4 POPULATION

The Malagasy population is estimated to be about 12 million inhabitants, but the density is only, on average, 21 people per km^2. It is young, with one-half less than 20 years old, and essentially rural (80%), mainly living in villages. The inhabitants of Madagascar are spread among 18 tribes, themselves split into clans.

1.1.5 TRANSPORT

The road system is 25,500 km long, with 5500 km asphalted, one-half being dirt tracks. Bush taxi is the fundamental means of transportation. Road traffic handles 80% of passenger transportation and 40% of goods.

There are 800 km of railway, six international and 12 coastal ports, as well as 57 airports open to the public, of which three are international.

1.2 Some history

1.2.1 KINGDOMS

The history of Madagascar is fairly well documented for at least eight hundred years. But, much earlier, immigrants from Africa and, especially, Asia arrived in dug-out out-rider canoes. The Malagasy language has a clear Malaysian origin. On the south-west coast, we find funeral sculptures directly influenced by Indonesians sources.

A highly developed social stratification existed in the Highlands, but also elsewhere in the monarchical societies, for example in the southeast. The society was hierarchical and decentralized. All the kingdoms had administrative divisions into districts and sub-districts, classified by hierarchical order according to the social status of the inhabitants, complemented by an age classification.

The population was characterized by inequalities based on differences of status. The principal inequality was the distinction between free people and slaves. The latter often represented two-thirds of the population, even in the nineteenth century. Thus, from the tenth century, slavery formed the background for all of the political, economic, and social life of all Malagasy communities. The members of the family of someone in debt could be reduced to slavery for the debts. Slavery, along with the international slave trade, especially since the sixteenth century, had profound and durable effects on the demography and mentality of the Malagasy people. It was abolished at the end of the nineteenth century.

Royal power and the popular assemblies, in the form of *kabary*, are well described by oral tradition from the seventeenth century, but can be traced

back to the twelfth. Across the centuries, the cult of ancestors, a popular practice, became the cult of the royal ancestors. The army, that had been developed when circumstances required it, took an importance of its own from the second half of the eighteenth century and became a permanent institution in the nineteenth.

The king was the absolute owner of all land. He held an eminent right on all of the land. The family and clan only held the right of use, the land being collectively owned. The sale of land to foreigners was forbidden because it belonged to the ancestors, to the clan, to the family. Individual property was only recognized in 1881.

The consolidation and modernization of the royal administration occurred especially on the Highlands. Of particular note are the major works on the plain of Antananarivo from the seventeenth century and the construction of dikes (*fefiloha*) and canals using the royal corvée that delocalized the traditional clans. These royal corvées (*fanompoana*), with the markets (*tsena*), represented the major part of the state budget.

The markets allowed the population to make their commercial transactions in all security. The courts in these markets forced the people progressively to abandon barter and to use only money for their exchanges. The markets regularized the slave trade by eliminated the theft of persons. The growing role of money in the markets contributed greatly to the growth of agricultural and artisanal production.

1.2.2 EUROPEANS

Europeans explorers, the Portuguese, arrived in Madagascar at the end of the fifteenth century. In this period, the Merina of the highlands had established their domination over almost all of the island. At the beginning of the nineteenth century, English and French missionaries arrived in mass. After fifty years of military attacks, with the English helping the Merina, the French declared Madagascar their colony in 1896. The 'pacification' lasted many years; the repression of the nationalist uprising of 1947 resulted in over 100,000 deaths.

The Malagasy Republic was born in 1958, becoming independent two years later. At the beginning of the 70s, trouble broke out, especially in the south where the cattle, the main wealth of the region, were destroyed by an epidemic. A general strike turned into revolution. Madagascar left the zone of the French franc (CFA), followed by a series of devaluations. In 1975, the Democratic Republic of Madagascar was created with a open Marxist-Leninist policy.

Early 1991 saw months of demonstrations involving hundreds of thousands. From May 1991 to January 1992, government, economy, and transportation ground to a halt due to general strikes, the 'transition period'. This led to the third republic, the Republic of Madagascar, but with an economy in ruins and ready for a series of World Bank and IMF structural

adjustments. The devaluation of the CFA early in 1994 did not immediately affect the country, but it was forced to let the Malagasy franc float shortly thereafter.

In November, 1995, the Queen's Palace, the country's much visited history museum, named a world heritage by the UNESCO, burned, destroying many of the major objects of Madagascar's cultural patrimony and creating a national shock.

1.3 Society

1.3.1 ORGANISATION

The extended family, predominantly patrilineal, is the basic unit of historical Malagasy social organization. From the outside, the family seems to be organized according to this this patriarchal model; the men represent the family to the exterior. But, inside the family, the woman has practically the same rights as the man.

Traditional marriage was a convention passed between two families in order to establish or to strengthen lasting relationships. The second function of marriage was to reinforce the demographic strength of family relationships by multiplying their members. In this way, the foundation of the village community (*fokonolona*) is formed, based on parental relationships (*fihavanana*).

Malagasy people respect age, the village elders (*rayamandreny*), their parents, but also their ancestors (*razana*). The funeral ceremonies have great importance because the dead can intervene in daily life. The dead are morned but great rejoicing also follows a death. Sometimes the tombs are larger than the houses of the living.

Traditional religion has neither temples nor churches. Instead, the sacred place is the tomb. About 45% of the population calls itself Christian. The place occupied by churches (sometimes two) in the villages is striking, but their role is primarily social: schools and clinics. Protestantism dominates on the highlands and Catholicism in the coastal regions.

In a country with a strong oral tradition, eloquent speech has a fundamental importance. The term, *kabary*, has been extended to groups of people gathering for the pleasure of speaking, at festivals and ceremonies. These last several hours and follow precise rules.

Interdicts or taboos (*fady*) vary from one locality to another; they are many and varied, and scrupulously respected. The local healer (*ombiasy*), with his ointments, infusions, and other concoctions (*fanafody*), is the rival, and the partner, of the doctor. His services, paid in kind, can be five times more expensive.

1.3.2 LANGUAGES

Since the nineteenth century, Protestant and Catholique missionaries had developed the school system. After 1905, the government created a parallel public system that, in 1940, had 120,000 students. From 1909, certain Malagasy, who spoke French and fulfilled certain other conditions, could become French citizens.

During the colonial period, until 1965, French had the status of mother tongue at school, Malagasy being considered to be a second foreign language after English which was studied as the first foreign language! Only in 1965 did Malagasy become a compulsory subject. French is essentially the language of the elite, through which social selection operates.

From 1971, Malagasy became the national language. But certain Malagasy consider it to be an ethnic language, that of the Merina who dominated the others before the colonization. From 1973, French officially became a foreign language. During the transition period, from 1991 to 1993, an educational reform proposed to make French the official teaching language throughout the entire duration of education, from the lowest classes to the highest level, while still keeping Malagasy as the national language.

1.3.3 VILLAGES

Each region has a specific type of house, but a commonality is the form, which is always rectangular. If the coastal huts must be constructed for maximum protection from the heat, those on the highlands must meet the cold of the dry season.

The coastal huts are mainly built of plant materials: wood or raffia palm. In the east, all houses have a wooden basis, never earth or daub. In the highland villages, the houses are narrow and high, with few openings. The walls of bricks, either sun-dried or baked and covered with baked earth, are thick and the roof is thatched, although corrugated steel is replacing this. In the north, the traditional house is made of plaited bamboo.

In the countryside, drinking water is an eternal problem. In the south, 90% of the rural population must, on average, make a round trip of ten to fifteen km in search of water. This tedious corvée is the work of the women, one pail on the head and two others hung from the ends of a yoke.

1.4 Economy

The successive oil crises struck severe blows to the country. In 1980, it already became impossible to continue paying the debt. Between 1971 and 1991, per capita income fell by 40%, the number of people living in poverty growing dramatically.

1.4.1 AGRICULTURE

Agriculture plays a fundamental role in the Malagasy economy, representing 40% of the GDP and 80% of exports and occupying 85% of the population.

Rice Rice (*vary*) occupies the first place in agriculture, both by the area involved and by its contribution to feeding the population. It is the basic food, often the only food, of the population: 500g per day per person. Until 1970, Madagascar had been a major exporter of rice, but at that time, it had to start importing.

Of the three millions hectares cultivated, rice occupies 1.3 millions, with the same number of workers. In each village, the rice has its building: a storehouse in the form of a hut similar to a human habitation that it dominates by its place on stilts, always in the centre. The life of the local village people turns around it.

Rain-dependent cultivation of rice occurs in the east of the country and is characterized by the small size of the plots, often depending on cut and burn techniques (*tavy*). It covers one-half the cultivated area (0.65 million hectares). But, with its small yield (0.8 t/ha), it only provides 20% of production. On the other hand, the irrigated rice fields yield up to 4 t/ha. The most beautiful rice paddies, terraces cut into the sides of the hills as in south-east Asia, are found in the highlands. Some regions manage to obtain two crops a year.

The rice is harvested by hand with a sickle. After threshing it, again by had, striking the sheaves to knock out the grains, the unshelled rice (paddy) is put out to dry on mats. Then, it must be pounded to remove the cuticle from the grain, work shared by mothers and daughters.

Other subsistence crops The other main subsistence plants cultivated are manioc, the second crop of the country, sweet potato, and corn, which altogether only represent a 10% contribution to the basic ration of the consumers. Barley and wheat are being developed, but remain at insignificant at the national level.

The oil-producing crops are peanuts, the oil palm, and the coconut palm.

Export crops

Coffee Arabica, grown on the highlands, is used for internal consumption. Robusta, grown on the east coast (0.2 millions hectares) is exported. About 90% of production comes from small mixed farms where coffee is associated with subsistence crops, especially rice and manioc. All of the large farms disappeared by 1975.

Vanilla Madagascar produces 70% of the vanilla consumed in the world. The crop was introduced to the island in 1870 and is localized on the north-east coast, a region very susceptible to climatic variation (cyclones). This represents from 15 to 20% of the total revenue from exports.

Pepper The extensive cultivation of the liana is located on the east coast, with two crops per year. The production of pepper represents 5% of the world market and 13% of export revenue of the country.

Cloves Cloves are a labour-intensive 'picking' crop, with 80,000 small growers in the region of Toamasina (east coast). Production suffers from strong competition from Zanzibar and Indonesia.

Industrial crops

Cotton Cotton has only recently (1960) been introduced. Small growers occupy 70% of the area cultivated and contribute 50% of the crop. Shelling, spinning, weaving, and garment-making are all done locally.

Sugar Sugar cane, the third crop of the country, is grown in the north and on the west coast. Several agro-industrial complexes produce sugar and alcohol (rum).

Stock farming Madagascar has the reputation of a country for herds of zebu, cattle with a hump. It has as many zebus as people. The size of the herd always contributes to the measure of fame and social position of a Malagasy person. Pastures cover 60% of the land.

Brush fires are, in fact, savanna fires. They are very spectacular by their size at the end of the dry season. Some talk of an ecological disaster by erosion, but stock raising, in the western half of the country, depends on this practice in order to maintain the grassy savanna. As well, the regrowth of graminaceous species is accelerated and certain external parasites are reduced.

Drastic drops in the number of zebus due to drought have fomented the activities of cattle rustlers (*dahalo*), especially in the south, increasing insecurity.

The Malagasy eat little zebu meat, simply to accompany their rice. Their favourite piece is the hump that contains a lot of fat and that is used to make the national dish, the *romazafa*, a stew containing many kinds of vegetables.

1.4.2 FISHING

Both fresh and salt water fishing are important. The irrigation system for the rice can also serve for pisciculture. The catch may be sold fresh,

smoked, salted, or dried. Fishing produce (shrimp, crabs, tuna, crayfish) occupies the third position in the export receipts of the country.

1.4.3 FORESTRY

Humid tropical forests on the east coast, dry tropical forests in the west and southwest, bush in the south, and mangroves along the northwest coast make up a total of 21% of the area of the country. Commercial species include lumber, precious and ornamental woods, and medicinal plants. More than 50% of the existing forests are protected (natural reserves and national parks).

1.4.4 INDUSTRY AND COMMERCE

The Malagasy industrial sector, mainly consisting of small and medium sized industries, was originally conceived for the local market under a policy of import substitution. A process of privatisation of public enterprises in difficulty was begun in 1988.

Food The food industry, with 40% of production, is the most important in this sector. Sugar occupies by far the largest part. The main products are food oils, preserves, meat products, milk, tapioca, and flour.

Textiles The textile industry represents 30% of industrial production. It covers much of local demand.

Mines Close to 10% of export receipts come from minerals. We find industrial minerals (chrome, quartz, coal, graphite, bauxite), ornamental stones (marble, cristals), and rare jewels (rubies, emeralds, saphires).

Markets Each village of sufficient size has its weekly market. A trip to the nearest market village with an ox cart to sell one's produce takes three days.

Every Friday, the Tananarive market (*zoma*) takes place on Independence Avenue. It is one of the largest open air markets in the world. The sellers arrive the day before, spending the night there to be ready in the morning. Each type of merchant is grouped together according to the type of product sold.

2
Introduction to the survey

2.1 General context

Primary school in Madagascar includes five grades (classes 11 to 7, according to the French system), while secondary school has a total of seven years. In the 70s, universal primary education was almost achieved but, in the 80s, total enrollment stagnated and then began to decline.

For the last few years, enrollment in primary education in Madagascar has tended to stabilize, with significant transfers from the public to the private schools. We see an apparent descolarisation at all levels, in terms of rates, one of the causes probably being the economic crisis that has affected the country. Officially, the enrollment rate is 60%.

At the same time, we see a growing preference for private schools, probably because they have an image of better quality associated with more discipline. The events of 1991 certainly played an important role in the acceleration of this phenomenon because of the strikes in the public schools. We must also notice the poor state of the public schools and their low level of quality.

The goals and objectives of primary education specify that it must be, above all, useful, preparing all children for the real pursuit of an occupation.

2.1.1 SCHOOLS

After having increased regularly since 1975/76, the number of primary schools decreased from 1987/88, although the number actually operating had started to decrease in 1982/83. Altogether, the drop was about 12% by 1992/93, corresponding to 1500 schools disappearing. The main cause of closing seems to be that the school buildings were destroyed or in bad condition. The teachers leaving, whether appointed elsewhere and not replaced or resigning due to poor working conditions, also play an important role.

2.1.2 SCHOOL PROGRAMS

Primary school programs have often been overhauled. For example, the changes in 1990 and 1992 consisted essentially in

- eliminating redundancies ;

- removing any ideological or political content;
- introducing new themes related to the protection of the environment, hygiene, nutrition, civics, and morals;

If, before, the program was expressed uniquely in terms of contents, consisting of a list of themes to be studied by each class and each level, from 1985, it took on a new look with the introduction of programs by objective. The instructions became more detailed and precise, with both general and specific objectives. The contents was cut up into themes spread over the 30 weeks of the school year for the nine subjects on the program:

(1) Malagasy,
(2) French,
(3) calculation,
(4) socio-economics,
(5) moral and civic education,
(6) common knowledge,
(7) artistic education,
(8) productive activities,
(9) physical education.

In 1991, an educational reform proposed to make French the language of teaching at all levels of the system.

2.1.3 PEDAGOGICAL ORGANIZATION

Classes normally last 28 1/2 hours a week, including 1 2/3 hours of recreation. But for various reasons (numbers, distance from school, lack of teachers, lack of nutrition), most schools, especially in rural areas, can only operate half time. Then, each child only goes to class half a day. Thus, the average weekly teaching period is 22 hours.

Three subjects take up one half of the timetable: Malagasy, French, and calculation.

The size of class normally can vary from 25 to 50 children. But the average student/teacher ratio varies from 20 in Toliara to 69 in Antsiranana. The average size of school varies from 126 children in Toliara to 238 in Antananarivo. This small size (with five classes) may be explained by various factors, such as the dispersion of the population.

Multigrade classes involve teachers how, in principle, look after two or three classes, but cases exist of up to five classes. In 1992, they represented 17% of the total. Classes operating half time are usually in schools with a single teacher. They represented 21% of the total in 1992.

Most teachers have received no initial training (ex-volunteers from National Service who stayed on to teach) or have received only basics (three months training in centres for those holding a lower secondary school 'college' diploma, BEPC). Legally, teaches must teach 27 1/2 hours per week.

Those in schools operating half-time can work 15 1/2 hours. Directors of schools with at least ten classes do not teach.

Each month, many teachers must leave for anywhere from three to ten days to go into town by bush taxi to receive their salary.

2.1.4 STUDENTS

Malagasy primary education is characterized by low promotion rates and high rates of failure, repeating, and dropping out, especially in areas outside the large urban centres. Both repeating and dropping out are very common from the first year on.

It seems that many children, especially in rural areas and in the least well-off families, drop out soon after the beginning of the school year for a variety of reasons:

- the need to help with the work in the fields;
- the beginning of the rainy season and the difficulty to get to school;
- the lack of food during the three or four months before the harvest.

These children are often enrolled in school again the following year by their parents, through the pressure of the local authorities or the director of the school or because the parents have not yet abandoned hope of educating their children.

Because of a combination of dropouts and a low rate of success at the final examination, only about 11% of the children enrolled in primary school eventually complete the five grades and obtain a leaving certificate (CEPE).

2.2 Objectives of the study

The study to be described in this text was carried out for the Ministry of National Education of Madagascar and the UNESCO in the autumn of 1994. It involved 4012 children between the ages of six and 16 living in 300 different villages of the six provinces of Madagascar. The 2514 in primary school in 1993 represent about one child in 500 attending school in rural areas. A very large number of items were investigated, including information about the children themselves, the characteristics of the families, the villages, and the provinces sampled.

The principal objective of this study is to identify the factors influencing the admission to, attendance at, and dropping out of school, taking into account explanatory variables related to characteristics of the child, the family, the school, and the village.

The study relies on the hypothesis that enrollment and attendance at school are determined by five large groups of factors:

(1) economic factors such as direct costs of attendance and opportunity costs, considered from the point of view of possibilities for work and

revenue;
(2) socio-cultural factors, values carried by the school, religion, taboos, and so on;
(3) characteristics offered by the educational system: distance to travel to neighbouring schools and their physical accessibility, quality of services, language for teaching, etc.
(4) results of the children at school;
(5) certain local conditions such as banditism that may lead to schools closing.

The objectives of this study are to analyze different aspects and characteristics of primary education in Madagascar. The study is mainly focused on analyzing factors which influence children's attendance at school. It includes the aspects:

(1) factors affecting admission to primary school;
(2) determinants of enrollment in 1993;
(3) reasons why children delay starting to school;
(4) factors influencing dropping out.

The aim of the study is not only to identify the important explanatory factors which influence children's education, but also to provide information which could be helpful in decisions to improve Madagascar's education system in the future.

Information was also collected on factors related to closure of schools. This aspect will not be covered in the present text.

2.3 Description of the investigation

In this section, the method for the investigation is briefly described.

2.3.1 SAMPLE SELECTION

Because the accessibility to primary education differs between rural and urban areas, different investigations would need to have been carried out to account for the characteristics of the two areas; instead, it was decided to take a sample in the rural areas where about 80% of the population lives.

The sampling unit was the household, although the unit of analysis is the child. 1500 families were sampled and a total number of 4012 children whose ages were between six and 16 years old were included. Because an up-to-date census of all the people or households was not available at a national level, it was necessary to work through clusters. The smallest unit available was the village, so this was taken as the first sampling unit.

To ensure a similar distribution of the sample throughout the island, the sample was stratified by district (CISCO). Because the existence of a school in the village is an important determinant of school attendance

within each CISCO, the villages were stratified by the presence or not of a functioning school.

Of 111 CISCOs in the country, 104 are in rural areas. Every cluster (a village) contained five families, for a total of 300 villages. In the 46 most densely populated CISCOs (i.e. with more than 109,000 inhabitants), four villages were randomly sampled for each CISCO, two with and two without school. In the 58 less populated CISCOs, two villages were randomly sampled for each, one with and another without a school. In this way, the sample of 300 villages was obtained.

Because the census is not compiled above the local level, the investigators had to do a local census in the villages and choose five families randomly. A total of 1500 families was obtained, stratified by the CISCO and by the existence of a school.

2.3.2 QUESTIONNAIRES

Four types of questionnaires were used; these concerned the village, the school, the family, and the child. They were prepared by the central teams and were tested in several villages before being accepted and translated from French into Malagasy.

General information was registered in the investigation of the villages. A second general questionnaire recorded detailed information about each school in the village, whether currently functional or closed. Each family chosen in the cluster had to fill in a questionnaire, where aspects like their financial situation, level of education, attitude towards the school, etc. were obtained. The family also had to fill in a second form, one for each child between the age of six and 16 years in the family. Here, the specific activities of the child in the family and his/her academic records were compiled.

2.3.3 PERIOD OF INVESTIGATION

Several constraints were present for the choice of surveying dates. To start with, in some parts of the island, the rainy season starts in mid-November, which makes some villages unreachable after that date.

Furthermore, the academic year starts at the beginning of October while the registrations are processed by the end of that month. Hence, the first possibility was to carry out the investigation from the end of October. However, in that case there were just fifteen days to make the investigation, given the proximity of the rainy season. There was also the problem that the people working at the Ministry of National Education would be very busy at that time, due to the start of the academic year.

Finally, the months of October and November are also the period of sowing crops for many families who are hence busy in their fields. This would leave little time for them to fill in the questionnaires.

The second possibility was to carry out the investigation during the month of September, before the start of the academic year. In that case, the investigation would have to be made about the attendance at school in the previous academic year, 1993–94, and about the intentions for the following year. This second choice was retained.

2.3.4 INVESTIGATORS

Training The training of the field investigators, carried out by members of a central team, was done in the six provincial capitals during September 1994. Of the 300 candidates examined, 150 were kept as investigators, along with 18 supervisors (chosen amongst the best of the candidates). Most often, they came from the CISCO concerned and had a good knowledge of their region.

After in-class training of two days, the candidates were sent during two further days to villages. Thus, the candidates arrived on Monday morning, after which we distributed copies of the questionnaires and the instructions. During the morning, they were to study these, in order to show their ability to be responsible and to work individually. At the same time, we trained the local team. The classroom training of the candidates took place on Monday afternoon and Tuesday. Then, on Wednesday and Thursday, they went into villages near the provincial capitals in order to learn, in the field, how to administer the questionnaires and to show their ability to form good contacts with the local people.

The selection of the investigators was done according to four criteria:

- individual work;
- participation in the class training;
- interaction in the village;
- presentation of the completed questionnaires.

Finally, Friday afternoon, the chosen candidates received detailed instructions before leaving for the villages to which they were assigned.

Specific instructions

Questionnaires Each field investigator had to visit two villages, except in the cases when the distance to one of the villages was too long; all interviewing had to be finished in a three week period. Each member of the team received three questionnaire forms for villages, three for open schools, two for closed schools, 12 for families and 60 for children.

When they received the questionnaires, the investigator each had to number all the pages with the codes of the CISCO, the villages, and the families, making sure that no pages were missing. Everything was to be written in pencil because the ink could be removed by the rain.

Villages The villages sampled were placed on regional maps which afterwards helped to produce a national map of the villages sampled. If a village could not be located on the map, the provincial team were to contact the central team in Tananarive to choose another village. No other changes in the choice of village were allowed. Fortunately, this possibility did not arise.

Choice of households When the investigators arrived in the villages, they had to obtain the local census from the head of the village or from the local security committee president. They were to check and to complete it themselves if necessary. The heads of the families were numbered consecutively and the total divided by five. The number so obtained was kept and five families were picked from the census at the positions corresponding to the first five multiples of that number; in this way, the investigators should get a total of five families randomly sampled. If the family chosen had no children between the ages of 6 and 16, the investigators had to sample the preceding family in the census list.

Controls The investigators received one half of their daily expenses before leaving for the villages and the other half upon returning, once the questionnaires had been verified.

The investigator had to obtain the signature of the head of the village at the arrival and at the departure. All filled in questionnaire forms had to be signed. The forms were controlled in the provincial capital on the return of the investigators.

At the same time as the investigation was being carried out, the 18 supervisors passed through three randomly chosen villages each (a total of 54 villages), so that one village in every six was controlled. The list of villages to be controlled in each province was chosen in Tananarive and put in closed envelopes that would be open only when the investigators had left for the field.

2.4 Methodology

2.4.1 DATA MANAGEMENT

Although there appear to be over one thousand variables in the data set, this is largely illusory. Under the harmful influence of the French statistical school, the local team split most of the multi-category variables into sets of binary variables either at the questionnaire or at the coding stage (at the last moment, without consulting the authors). This unnecessarily multiplied the size of the data set and introduced many uncontrollable errors. (For example, 25 children are classified as being in villages where the *main* activity is agriculture, herding, *and* fishing.)

From the total number of 4012 children, those observations correspond-

ing to the six-year-old children were excluded, because half of them had not started school at the time of the investigation (their starting date was the following academic year). When fitting the models, the observations with missing values in the explanatory variables included in the model were eliminated. This assumes that these values are missing at random, which, as we shall see, was often clearly not the case. Thus, in some analyses, we have had to eliminate completely the variables for which values seemed to be missing in a non-random fashion.

From the many variables in the data set, about fifty have been selected as important as influencing children's education as explanatory variables (see Appendix A and Appendix B for the explanation of each variable and the corresponding code used in the analyses). If the variables which were used in creating the responses are included, nearly 100 variables in the data set were used.

The explanatory variables were divided into five groups:

(1) variables about the children themselves (age, sex, activities at home, etc.);
(2) the characteristics of the family (size, education of the parents, size of the house, etc.);
(3) variables for the characteristics of the villages sampled (main activity of the village, existence of market, etc.);
(4) the variables for the characteristics of the school existing in the village (type of school and existence of a closed school);
(5) the different provinces sampled in the country (Antananarivo, Fianarantsoa, etc., taken as one factor variable).

When fitting the models, the explanatory variables were fitted group by group, taking into consideration the relative importance of each group of variables.

2.4.2 RESPONSE VARIABLES

Enrollment Enrollment is defined as registration of a child in a given school year in a given school. It does not imply attendance for any period of time. For a given cohort, we can define the probability of enrollment at a given age in a given school year. By aggregating over years of study, we obtain the probability of enrollment for a given age. By aggregating over ages, we obtain the probability of enrollment in a given school year. By addition over both years and age, we obtain the global probability of enrollment.

On the other hand, attendance can be defined by the number of days in the school year that the child actually was present at school, in relation to the total number of school days.

Admission Admission is defined as the first enrollment in primary school. The probability of admission can be defined in several different ways:

- the probability of admission relative to the cohort:
  - the probability of admission at a given age;
  - the probability of admission before some given age;
  - the global probability of admission of a cohort.
- the probability of admission of all children, irrespective of cohort;
- the conditional probability of admission, given that the child had never previously been enrolled.

Delay The delay in admission is defined with respect to the 'normal' age of entry to primary school. We take this to be six years old although certain children begin earlier. Thus, for example, a child enrolled in primary school for the first time at eight years old has two years delay.

Repeating We only consider repeating for two consecutive years, although some children spend more than two years in the same class.

Dropout We take dropout to mean that a child has completely abandoned school. In practice, a difficulty arises because some children leave for a year or two, but subsequently begin again. We take a child who enrolled in school to have dropped out if he or she is not enrolled again for three consecutive years.

2.4.3 MODELS

Two kinds of generalized linear regression models were fitted in the analyses. For the models referring to children entering school, to their enrollment in school in 1993, and to their dropout from school, logistic models were fitted because the response variables are binary (children entered school or not in 1993; they were enrolled or not in 1993; they dropped out of school or not after 1990)

When the response variable was a count (i.e. how many years a child delayed starting to school), the model used was of the log linear type with Poisson distribution, which means that the log average delay is a linear function of the explanatory variables.

Although the study was designed to have 50% of the villages without a functioning school, according to the information we had in the sampling frame, as explained in Section 2.3.1 above, this turned out not to be the case. Only 18.4% of the children in the study come from villages without an open school (see Section A.4). Thus, there are apparently far fewer closed schools than the government statistics show. Because we have no idea of the correct proportion of villages with closed schools, we provide no correction to the data as it stands. Theoretically, the villages without schools should

be over-represented. This would imply that our results *under-estimate* the success of primary education for the children in Madagascar (see the tables in Section B.5).

Because the study is mainly about the children themselves, we have taken the variables for the children's characteristics as the most important factors influencing their education, going down in importance to the variables not relating directly to the children (the characteristics of the family, the village, etc.). When fitting the models, all the significant variables for the children themselves (the first group of the variables) were first put into the model. (Sex is kept end until the final elimination, even if it is not significant.) Then, the second group of variables (about the families) was added into the model. This process continued until all four groups of variables had been checked, after which the ones that had become non-significant in the meantime were deleted by backward procedure. At each stage, interactions of all variables with sex and age are also considered.

We considered a variable to be significant if the deviance was reduced by two when it was entered into the model. (This is the Akaike information criterion, AIC, for model selection.) In order to avoid too strict cuts, borderline variables were left temporarily in the model at each stage. If they were still almost significant until the fourth group of the variables was added, a final decision was made. At this point, we obtained our final model.

Because we are least interested in the difference among the six provinces for children's education, we added the fifth group last.

Naturally, different ways of introducing the variables into a model could give different results, with different variables being included in the final model. However, our method of introducing variables seems logical given that we were more interested in the factors that affect children themselves than in any other factors. Furthermore, other methods of introducing variables were tested, leading in most cases to very similar results (both in terms of the variables included in the final model and in terms of the estimates of the parameters).

For the analysis of the data, the statistical package GLIM was used.

2.4.4 PRESENTATION OF THE RESULTS

Our data contain a vast amount of information about the trajectories of the children through school. This has not yet been used in the analyses. We do, however, present tables at the beginning of each chapter showing how the various response variables have evolved across cohorts and ages within our study.

Modelling results are presented as a series of tables for each response variable, in a separate chapter. The four main groups of variables are presented first, in sequence, before eliminating non-significant variables, and finally adding the provinces.

In each case, we try to find and to measure the influence of variables that act on the response variable in question. For the dichotomous variables, the coefficients obtained have values that represent the influence of the variables, and are comparable. However, for the quantitative variables, we shall look mainly at the signs of the coefficients because it is not possible to compare the values; they depend on the units of measurement of the variables, which are not comparable.

In the tables, N indicates the number of observations used and M the number lost due to missing values. Variables omitted due to non-random missing values are indicated by a star (*) if non-significant and by a plus (+) if significant.

3
School enrollment

In this chapter, we are interested in analyzing the reasons why children are enrolled in school or not. This will provide us with an overview before we look at admissions, delay, and dropouts. The enrollment that we are interested in is defined as whether the children were registered or not in school, leaving aside whether the children actually attended school and for how much time.

From the table on enrollment by age in Section B.2, we see that more than 75% of children between seven and 13 years old were enrolled in school in 1993. Thus, we can see that the great majority of children go to school, at one age or another. Those who do not probably belong to families who either reject schooling or are extremely poor or live in remote villages with difficult communications not covered by the school system or with a closed school.

3.1 Cohort analysis

We first look at how enrollment has evolved over time, as illustrated in Table 3.1. If we look at the left-hand columns, we see that children have been entering school at progressively younger ages. Consequently, the more recent is the cohort, the higher is enrollment for a given age.

Table 3.1. School enrollment (%) by cohort and age.

	Age										
Cohort	7	8	9	10	11	12	13	14	15	16	N
1977			21.6	28.8	40.2	62.5	62.9	63.3	57.2	42.8	264
1978		8.9	24.9	38.6	61.8	64.5	68.9	63.1	49.5		293
1979	7.3	13.8	25.7	53.5	65.7	73.0	67.8	63.8			370
1980	9.2	21.6	47.6	61.2	74.3	80.2	76.6				338
1981	8.5	26.5	43.7	62.7	69.1	71.9					437
1982	20.1	39.3	64.1	75.1	82.2						354
1983	23.5	45.6	66.0	76.5							430
1984	32.3	56.6	76.8								371
1985	45.6	72.9									432
1986	73.5										404

If we read across the lines of Table 3.1, we see that, for each cohort, the rate of enrollment increases with age. For the oldest cohorts, the maximum occurs at 12 years old. Afterwards, some of the children quit school and others go on to secondary school. In the more recent cohorts, the maximum occurs in the year of the survey, meaning that the rate of enrollment may still increase further.

3.2 Enrollment, 1993

In our modelling in this chapter, we shall only take into account children registered in school at the beginning of the academic year in 1993. A more adequate model would take into account the career of the children throughout the years when he or she should be at school. Of the 3693 children aged seven to sixteen, 135 went to secondary school during the period that concerns us here.

Because the variable 'whether the child is registered or not in school' is binary, we fit a logistic model, i.e. the log odds (the log ratio of the probability for children to go to school to the probability for children not to go to school in 1993) is fitted as a linear regression model in the explanatory variables, and the distribution is binomial. The formula of the model is as follows:

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K \tag{3.1}$$

where π is the probability to be enrolled in school (then $1-\pi$ is the probability not to be enrolled in school), the x_k, $k = 1, \ldots, K$, represent the explanatory variables included in the model, and the β_k are the parameters of the model.

The logistic model used for the study of enrollment leads to the use of a considerable number of explanatory variables. Remember that many combinations of variables have been tried before arriving at the groups of variables to be presented here as best explaining the observed situation. For enrollment in 1993, we have explanatory variables from all five groups chosen *a priori*: those linked with the children, the family, the village, the school, and the province.

3.2.1 CHILDREN

We start from 3693 observations (all the children who were more than 6 years old) and a null model with deviance 4513.7. According to our rule for excluding non-significant variables stated in the first chapter, seven out of the 12 variables in this group have been retained. (For the definition of all the variables, see Appendix A.) The deviance of the model is 3229.7 with 3068 degrees of freedom from 3078 observations. At this stage, 615 observations have been weighted out due to missing values in some of the explanatory variables. The result is shown in Table 3.2.

Table 3.2. Enrollment at school with variables for the children. ($N = 3078, M = 615$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	1.705	0.244	WATER$^+$
AGE	−0.140	0.017	BROTHER
SEX	−0.024	0.115	AFOOD
BIOLOGIC	0.349	0.157	ANIMALS
RICE	1.008	0.150	MEALS
CLOTHES	0.598	0.126	
SHOPPING	0.735	0.090	
FIELD	−0.072	0.139	
SEX.RICE	−0.530	0.202	
SEX.FIELD	−0.591	0.208	

One of the variables, fetching water, has missing values that apparently are not random. Although very significant when added to the model (all positively related to school enrollment), this variable has not been included for this reason.

From these results, we can see that the age of the children affects negatively their probability of going to school. This is shown by a negative estimate of the parameter for the variable concerning the age of the children. It means that the older the child is, the lower the probability of being in school will be. The reason for this result may be that older children must stay at home to help their family.

For the sex of the children, it seems that girls have lower probability to go to school than boys, but this depends on their involvement in preparing the rice and working in the fields.

If a child is a biological member of the family (if not, the child could belong, for example, to some relatives or friends), the probability of being enrolled in school is higher than that of a child who is not. This is surprising as one would expect a child to be sent to some relatives' family in order to go to school.

When we look at the tasks that the children must do in the household, those that are found to be significant are pounding the rice, washing up the clothes of the family, and going on small shopping errands. Children doing them have a higher probability of going to school, although for rice the influence is much stronger for boys than girls. It seems that these are lighter, more suitable tasks for school children to do. In contrast, working in the fields is negatively related to school enrollment, especially for girls. On the other hand, looking for food for the animals of the family and taking care of the animals are not included. Participation in domestic

Table 3.3. Enrollment at school with variables for the children and the family. ($N = 2400, M = 1293$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	4.328	0.830	HWALL
AGE	−0.309	0.063	SPMEAT
SEX	−0.054	0.151	ACTIV
BIOLOGIC	0.222	0.246	FREAD
RICE	0.887	0.184	DISTMIN
CLOTHES	0.408	0.152	DISTWAT
SHOPPING	0.695	0.112	
FIELD	−0.089	0.167	
SEX.RICE	−0.418	0.248	
SEX.FIELD	−0.521	0.248	
HAREA	0.008	0.003	
FAMSIZE	−0.101	0.044	
NUMBCH	−0.108	0.108	
FATHED	0.159	0.050	
MREAD	0.480	0.122	
RELIGION2	−0.734	0.651	
RELIGION3	−2.756	0.618	
SOCACT	0.362	0.110	
REASON2	0.386	0.147	
HOLPRF2	−0.321	0.137	
LANGSC2	−0.336	0.141	
AGE.NUMBCH	0.015	0.008	
AGE.RELIGION2	0.037	0.053	
AGE.RELIGION3	0.148	0.051	

tasks is perhaps occasional, as opposed to those of other children who cannot attend school because they are employed for more continuous tasks for their parents, particularly agricultural work.

3.2.2 FAMILY

We now keep these significant variables for the children in the model and add the 16 variables referring to the family's characteristics. After excluding the non-significant variables one by one, ten of them were significant. The new deviance for our model is now 2140.9 with 2376 degrees of freedom from 2400 observations. 678 observations were lost because of the missing values of these variables, for a total of 1293. The results are shown in Table 3.3.

Before analyzing the results for the new variables, it is important to

note that all the variables that were in the model before remain significant with similar estimated values except for whether the child is a biological member of the family.

The area of the house where the family lives can be taken as one measure of the wealth of the family. It may represent the synthesis of several factors: besides the wealth of the family, more space available for school work, perhaps the activity of the father. Thus, probability of being enrolled in school increases with the size of house. On the other hand, given that variable, it does not depend on two other closely related measures, the type of construction material of the walls of the house or the amount of money spent on meat by the family. However, the fact that a family eats more meat is at least partly linked to greater wealth but it is not associated with enrollment. Could it be that the work of the children permits certain families to eat more meat? The links between enrollment and the 'wealth' of the family are not obvious.

The probability of a child being enrolled in school increases with the number of children (under 16) in the family, this effect increasing with age, but, at the same time, decreases with the size of the family (these two variables can be expected to be closely related: the correlation coefficient is 0.82). Apparently, families with a large number of adults have less probability of sending the young children to school.

As might be expected, the probability of being enrolled in school increases with the level of education of the father and when the mother can read. Given the father's level of education, the indicator as to whether or not he can read is not necessary.

The religion of the head of the family (taken to be the father) also seems to have a significant explanatory effect on the probability of children going to school. We find that Protestants have the highest probability of their children going to school, with Catholics somewhat lower; however, not being either of the two was highly significant with respect to these two religions. The differences diminish as the age of the child increases. The negative estimate of the coefficient for this variable means that when the father of the family belongs to another religion than Catholic or Protestant, the children have a much lower probability of going to school. Other religions include Muslims, Anglicans or Traditional; some of them are based on more traditional roles of the family members, reducing the possibilities of education for their children.

School enrollment also depends on whether the father participates in various social and religious activities in the village.

When we look at the main reason why parents decide to put their children in school, we find that those who give the relationship among the director of the school, the teachers, and the parents have a higher probability of being enrolled in school. Apparently, a good tri-partite relationship gives parents confidence about the authority and the personality of the di-

rectors and the teachers, inducing them to send their children to school. Other reasons, like safety on the way to the school, distance to the school or timetable of the lessons are not significant enough for parents to decide to take their children to school.

For some reason, school enrollment is lower in families where the preference is to teach the children at school both French and Malagasy rather than one or the other. It is also lower in families who would prefer the main holidays for the school to be during harvest period rather than the rainy season or unchanged.

Other variables that were checked but found to be non-significant are whether the father is a farmer or not, the distance from the house to the closest source of drinkable water, and the time (in minutes) spent by children on the way to school.

3.2.3 VILLAGE

After adding all of the variables for the village group, and deleting those that were non-significant, only two of the nine variables remain significant, according to the rules we use. At this point, two further variables, whether the child is involved in washing up the clothes and the number of children in the family, become non-significant. The rest of the variables that were already in the model remain significant, with the same sign, so that again all the conclusions made for these variables still hold.

The new model, which includes variables for the child's and family's characteristics, as well as those for the differences across villages, has a deviance of 2002.8 with 2203 degrees of freedom from 2230 observations. At this stage, 180 further observations were weighted out, for a total of 1477. The results are shown in Table 3.4.

Only the size of the village and whether the economic activity is predominantly herding enter the model here. Children having more chance of being enrolled in school if they live in large villages, although the effect decreases with age, and where the main activity is not herding.

The variables corresponding to the characteristics of the village that were found to be non-significant, included the distance to the faritany (district capital); the existence of a merchant in the village; whether there is a market in the village or not; the main economic activity of the village (the two variables for main activities: agriculture or fishing); and whether the village was mainly Catholic or Protestant.

3.2.4 SCHOOLS

At this point we add several variables referring to the schools existing in the villages into the model. The types of open and closed schools in the village, as well as the number of schools in the village, are significant enough to be kept in the model.

Table 3.4. Enrollment at school with variables for the children, the family, and the village. ($N = 2230, M = 1463$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	3.610	0.883	DISTF
AGE	−0.262	0.068	SHOP
SEX	−0.016	0.156	MARKET
BIOLOGIC	0.199	0.252	AGRV
RICE	0.856	0.190	FISHV
CLOTHES	0.295	0.160	CATHV
SHOPPING	0.706	0.116	PROTV
FIELD	−0.023	0.173	
SEX.RICE	−0.381	0.257	
SEX.FIELD	−0.511	0.256	
HAREA	0.009	0.003	
FAMSIZE	−0.082	0.046	
NUMBCH	−0.116	0.111	
FATHED	0.142	0.052	
MREAD	0.460	0.126	
RELIGION2	−0.639	0.660	
RELIGION3	−2.698	0.632	
SOCACT	0.362	0.114	
REASON2	0.396	0.150	
HOLPRF2	−0.260	0.140	
LANGSC2	−0.374	0.147	
AGE.NUMBCH	0.014	0.008	
AGE.RELIGION2	0.027	0.053	
AGE.RELIGION3	0.133	0.053	
HERDV	−0.148	0.121	
VSIZE	0.004	0.002	
AGE.VSIZE	−0.000	0.000	

The deviance of our new model is 1817.1 with 2165 degrees of freedom from 2202 observations. The variable, type of school, has 28 extra missing values, for a total of 1491. The results are presented in Table 3.5. At this point, the parents' preference for the language in the school and whether the main activity of a village is herding becomes non-significant. Whereas if the child is involved in washing up the clothes changes from non-significant in the last model to quite relevant in this one.

As might be expected, probability of enrollment increases with the number of schools in the village, but this effect decreases with age. It is also considerably higher if there is an open private school (type 3) in the village,

Table 3.5. Enrollment at school with variables for the children, the family, the village and the school. ($N = 2202, M = 1491$)

Significant variables	Estimate	Standard error
Constant	1.827	1.008
AGE	−0.235	0.077
SEX	−0.050	0.300
BIOLOGIC	0.176	0.265
RICE	0.803	0.198
CLOTHES	0.466	0.169
SHOPPING	0.606	0.122
FIELD	0.047	0.181
SEX.RICE	−0.279	0.272
SEX.FIELD	−0.565	0.270
HAREA	0.010	0.003
FAMSIZE	−0.057	0.048
NUMBCH	−0.111	0.116
FATHED	0.149	0.056
MREAD	0.499	0.136
RELIGION2	−1.024	0.687
RELIGION3	−2.701	0.673
SOCACT	0.282	0.122
REASON2	0.358	0.160
HOLPRF2	−0.406	0.151
LANGSC2	−0.226	0.157
AGE.NUMBCH	0.011	0.009
AGE.RELIGION2	0.066	0.055
AGE.RELIGION3	0.154	0.056
HERDV	−0.022	0.131
VSIZE	0.004	0.002
AGE.VSIZE	−0.000	0.000
NUMBSCH	1.487	0.561
TYPESCH2	0.804	0.343
TYPESCH3	0.639	0.411
TYPESCH4	1.063	0.653
TYPECSCH2	−0.170	0.176
TYPECSCH3	−1.167	0.305
SEX.TYPESCH2	−0.166	0.307
SEX.TYPESCH3	1.165	0.496
SEX.TYPESCH4	−1.505	0.568
AGE.NUMBSCH	−0.067	0.040

especially for girls, but lower if there is a closed private school. Enrollment is stronger where the offering of educational facilities is greater.

3.2.5 VARIABLE REMOVAL

At this stage, we remove the variables that have become non-significant, starting with those that entered the model most recently. Thus, we shall take out whether or not it is a herding village, language choice, number of children in the family, and whether the child is a biological member of the family, as well as some interactions.

This procedure reduces the number of missing values to 1292, leaving 2401 observations. The deviance is now 2018.8 with 2370 degrees of freedom. The results are shown in Table 3.6.

3.2.6 REGIONAL DIFFERENCES

When introducing the variable that takes into account the different characteristics of the regions in the country, we find that living in certain regions can have a significantly different effect on whether children are enrolled in school or not. The deviance of our new model is 1951.4 with 2365 degrees of freedom again from 2401 observations. The results are shown in Table 3.7.

The analysis of the influence of geographical variables can be interesting. It has sometimes been noticed that the location of a family in one province rather than another can have an incidence on enrollment. This variable takes into account the specific characteristics of each province not explicitly handled by the other variables in the model. For example, the powers that be in a province may have an unfavourable attitude to schooling of children.

In two of the provinces, Fianarantsoa and Antsiranana, we find that in comparison with the region of Antananarivo (taken as the reference level), the probability of children going to school is significantly lower. In contrast, for Mahajanga and Toliara, it is higher. For the other province, Toamasina, there is no difference in the probability of children going to school with respect to the region of Antananarivo.

These results contrast with the gross data in Section B.6. There, enrollment in Fianarantsoa, Antisranana, and Toamasina is considerably less than in the other three provinces. Thus, the variables in the model can explain the low enrollment in Toamasina but not in the other two provinces.

3.3 Conclusions

From our final model, we can see that the main sub-groups of variables affecting the enrollment at school refer to the characteristics of the children themselves and to their family; several variables for the work that the children do at home are significant. Children from richer more educated families also have a better chance of going to school. The main variables

Table 3.6. Enrollment at school with variables for the children, the family, the village and the school, after removal of unnecessary variables. ($N = 2401, M = 1292$)

Significant variables	Estimate	Standard error
Constant	0.666	0.726
AGE	−0.142	0.058
SEX	−0.070	0.270
RICE	0.577	0.130
CLOTHES	0.524	0.158
SHOPPING	0.576	0.115
FIELD	−0.039	0.169
SEX.FIELD	−0.517	0.247
HAREA	0.010	0.003
FAMSIZE	−0.031	0.021
FATHED	0.178	0.052
MREAD	0.520	0.126
RELIGION2	−0.616	0.649
RELIGION3	−2.344	0.635
SOCACT	0.265	0.115
REASON2	0.324	0.150
HOLPRF2	−0.384	0.143
AGE.RELIGION2	0.029	0.052
AGE.RELIGION3	0.124	0.053
VSIZE	0.005	0.002
AGE.VSIZE	−0.000	0.000
NUMBSCH	1.438	0.533
TYPESCH2	0.799	0.322
TYPESCH3	0.610	0.389
TYPESCH4	0.928	0.614
TYPECSCH2	−0.231	0.163
TYPECSCH3	−1.081	0.301
SEX.TYPESCH2	−0.211	0.291
SEX.TYPESCH3	0.928	0.457
SEX.TYPESCH4	−1.355	0.545
AGE.NUMBSCH	−0.065	0.039

decreasing the probability to go to school are religion and holiday preference, as well as the variables for the age and the sex of the children, which show that older children and girls have generally a lower probability to enroll in school than younger children and boys, respectively.

No variables referring to the villages where the children live are signifi-

Table 3.7. Enrollment at school with variables for the children, the family, the village, the school, and the provinces. ($N = 2401, M = 1292$)

Significant variables	Estimate	Standard error
Constant	0.928	0.758
AGE	−0.135	0.059
SEX	−0.117	0.276
RICE	0.556	0.135
CLOTHES	0.559	0.163
SHOPPING	0.532	0.119
FIELD	−0.045	0.173
SEX.FIELD	−0.715	0.253
HAREA	0.013	0.003
FAMSIZE	−0.047	0.022
FATHED	0.195	0.053
MREAD	0.419	0.131
RELIGION2	−0.698	0.659
RELIGION3	−2.026	0.648
SOCACT	0.184	0.119
REASON2	0.365	0.160
HOLPRF2	−0.457	0.148
AGE.RELIGION2	0.035	0.053
AGE.RELIGION3	0.112	0.054
VSIZE	0.004	0.002
AGE.VSIZE	−0.000	0.000
NUMBSCH	1.475	0.545
TYPESCH2	0.937	0.330
TYPESCH3	0.647	0.400
TYPESCH4	1.216	0.631
TYPECSCH2	−0.008	0.172
TYPECSCH3	−1.168	0.322
SEX.TYPESCH2	−0.158	0.298
SEX.TYPESCH3	1.053	0.463
SEX.TYPESCH4	−1.365	0.559
AGE.NUMBSCH	−0.070	0.039
Fianarantsoa	−0.699	0.186
Mahajanga	0.327	0.229
Antisranana	−0.938	0.225
Toamasina	−0.073	0.223
Toliara	0.807	0.315

cant in the final model. On the other hand, the probability for the children to go to school increases considerably with the availability of schools in the village.

As a final remark, we may note that whether we include the children who went to secondary school or not does not change the factors that influence significantly the probability of children being enrolled in school. This is reasonable since only few children went to secondary school in the period concerned. However, it would be interesting to see what factors actually influence the children going to secondary school. But if we weight out the observations with missing values in the explanatory variables, we have too few observations for the children going to secondary school.

4
School admissions

In this chapter, we are interested in analyzing the factors influencing the admission of children to school for the first time.

Global admissions (public and private schools), that had increased from the beginning of the 1980s, fell 10% in 1989/90, not really taking off again until 1993. The evolution between 1987 and 1992 might be explained by the results of the economic crisis on the demand for education, notably the direct and indirect costs of schooling. It could also result from the political events of 1991 that may have discouraged certain parents from sending their children to school during the strikes, only going back to school later.

Since 1987, new admissions to the public schools have regularly decreased, but this trend has been counterbalanced by an increase in admissions to private schools. The rate of participation of the private sector in admissions has thus passed from 17% in 1987 to 25% in 1991. This situation probably results first from the low quality of teaching in public schools, but also from the 1991 events that pushed parents to transfer their children to private schools because of the strikes and lack of safety.

4.1 Cohort analysis

Two indices of admission can normally be calculated from the global statistics derived from the files of the annual census of schools and from estimates of the population (extrapolated from the data in the population census of 1978). These are the gross enrollment ratio (GER: new admissions divided by the number of six year olds) and the net enrollment ratio (NER: admissions of six year olds divided by the total number of six year olds). While the GER gives an idea of the amplitude of annual admissions in relation to the number of children of the age to be admitted and of the capacity of the system to accept them, on the other hand, it gives no information on what is happening to a cohort with respect to admission. The NER only indicates the proportion of children of a cohort who enter school at the normal admission age.

In Madagascar, the GER tended to decrease for a number of years, before stabilizing at a level above 100%. The estimates of the GER from our study and from the Ministry of National Education are shown in Table 4.1. However, the latter concerns the whole country, including the urban

Table 4.1. Official GERs, along with school admissions, six year old children, and GERs estimated from the sample, along with their 95% confidence interval (CI).

	1990	1991	1992	1993
GER (Ministry)	104.7	103.5	100.5	100.5
Admissions	346	441	449	520
Children age 6	430	371	432	404
GER (Study)	80.5	118.9	103.9	128.7
CI	(69.6, 93.0)	(103.2, 136.9)	(90.8, 118.9)	(112.7, 147.0)

areas. The two sets of result thus agree fairly closely, given that those from the study have sampling error in both the numerator and denominator.

From the data collected from the survey, it is possible, in contrast, to estimate indices of admission that are much more precise and relevant:

- the rate of admission by age for a given cohort (admission of the children in the cohort at each age divided by the number in the cohort);
- the cumulative rate of admission at the latest at some given age, by cohort;
- the rate of admission of each cohort (obtained by adding the rates by age for a cohort for all possible ages of admission).

If precise and valid data are not collected on the distribution of ages of those newly enrolled during the annual census of schools, this survey provides the only instrument allowing one to follow the admission history each generation of children.

The rates of admission by age for each cohort in the study are given in Table 4.2 and those for the cumulative rate of admission at the latest at some given age, by cohort, in Table 4.3. The estimates in these tables show that, during the last few years, children from successive cohorts have tended to enroll in primary school at a younger age. This trend is striking by its size and regularity. It clarifies the reasons for the evolution of the GER recorded over recent years: to a large extent, this evolution results from a reduction in late admissions. The lack of admission of older children is disappearing. In other words, the lowering of the GER does not result from fewer children going to school, but this must be checked by following future admissions of the cohorts that have just begun school.

We see, for example, that the 1982 cohort has a global admission rate of 85% at 11 years old, in spite of the fact that it had only 18% at seven years. In contrast, the 1986 cohort already had a rate of 65% at seven years old. Thus, we can predict that at least 90% of the children in these recent cohorts have been or will be enrolled in school, at one moment or another.

Table 4.2. School admissions (%) by cohort and age.

	Age										
Cohort	6	7	8	9	10	11	12	13	14	15	16
1977					7.6	11.0	9.8	5.3	3.0	2.3	0.4
1978				10.2	10.6	14.7	6.5	5.1	0.3	0.3	
1979			6.5	10.0	21.1	13.5	6.8	2.7	2.7		
1980		5.9	12.4	22.8	14.8	14.2	7.4	1.5			
1981	2.5	4.3	16.9	17.8	19.2	9.8	7.1				
1982	4.2	14.1	20.1	25.1	13.0	8.2					
1983	7.7	14.9	22.3	22.1	14.4						
1984	10.2	20.5	26.1	20.2							
1985	11.6	29.2	28.9								
1986	20.0	44.8									
1987	30.1										

Table 4.3. Cumulative school admissions (%) by cohort and age.

	Age										
Cohort	6	7	8	9	10	11	12	13	14	15	16
1977					7.6	18.6	28.4	33.7	36.7	39.0	39.4
1978				10.2	20.8	35.5	42.0	47.1	47.4	47.7	
1979			6.5	16.5	37.6	51.1	57.9	60.6	63.3		
1980		5.9	18.3	41.1	55.9	70.1	77.5	79.0			
1981	2.5	6.8	23.7	41.5	60.7	70.5	77.6				
1982	4.2	18.3	38.4	63.5	76.5	84.7					
1983	7.7	22.6	44.9	67.0	81.4						
1984	10.2	30.7	56.8	77.0							
1985	11.6	40.8	69.7								
1986	20.0	64.8									
1987	30.1										

In other words, the country is close to global admission in rural areas, and probably also throughout the country.

The level of GER, greater than 100%, can thus be explained by catching up on admissions at the older ages, given that the rate of admission by cohort is over 90%. For the analyses of the data of this survey, one and only one admission was recorded per child (that of the school year when the parents first declared that the child was enrolled at school). Thus, it is clearly impossible that the admission data include reinscriptions after dropping out. In contrast, this situation could very well happen with the census statistics, given the way in which they are collected. It is also not possible to claim a bias resulting from the use of estimated demographic data, as for the calculation of the census GER and NER. The enrollment

Table 4.4. School admissions (%) for those not already in school by cohort and age.

	Age									
Cohort	7	8	9	10	11	12	13	14	15	16
1977				10.0	15.8	21.3	15.1	9.0	6.3	0.9
1978			12.1	14.8	28.7	17.9	15.3	1.2	1.0	
1979		7.1	11.9	31.5	29.9	21.0	10.1	8.8		
1980	6.1	13.7	30.4	28.9	36.9	31.3	7.9			
1981	1.1	18.9	24.6	34.9	27.2	23.8				
1982	21.4	25.4	41.6	37.4	33.7					
1983	16.4	29.4	41.3	43.1						
1984	23.3	39.1	49.3							
1985	35.6	53.6								
1986	62.0									

and population statistics come from the same information base, the households. The only risk of bias in this survey could concern an over-declaration of enrollment of their children by the parents (or an under-declaration of non-enrolled children).

The rates of admission by age in Table 4.2 will be used in Chapter 5 to study the delay in starting to school.

4.2 Admissions, 1993

In order to calculate usable probabilities of first admission, we need to consider, at any point in time, only those children who have never yet been enrolled. The data by cohort are given in Table 4.4. Here, we shall only look at admissions in 1993; this corresponds to the bottom diagonal of the table.

Because the variable 'whether the child enters school or not' is binary, we fit the logistic model of Equation (3.1) where π is here the probability of first admission.

4.2.1 CHILDREN

We start from 1412 observations (all the children who were more than 6 years old and who had not gone to school before) and a null model with deviance 1858.3. According to our rule for excluding non-significant variables stated in the first chapter, six out of the 12 variables in this group have been retained. The deviance of the model is 1037.5 with 1001 degrees of freedom from 1009 observations. At this stage, 403 observations have been weighted out due to missing values in some of the explanatory variables. The result is shown in Table 4.5.

Probability of starting school decreases with age, as one might expect. It is greater for those fetching water and doing the shopping but strangely

Table 4.5. Admissions to school with variables for the children. ($N = 1009, M = 403$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	3.352	0.374	BIOLOGIC
AGE	−0.438	0.038	RICE
SEX	−0.062	0.157	BROTHER
WATER	1.264	0.171	CLOTHES
MEALS	−2.049	0.917	ANIMALS
SHOPPING	0.641	0.155	AFOOD
FIELD	−0.415	0.212	
AGE.MEALS	0.141	0.085	

less for children helping prepare the meals (see the table in Section B.2). It is also lower for children working in the fields. It is interesting to note that there is no significant difference for sex, although this variable is kept in the model provisionally. Apparently, there is no discrimination against girls.

The fact that a child fetches water is the factor with positive effect on the probability of admission that is largest and most significant. Perhaps, parents take advantage of the daily trips of the school children to have them bring back water. This could be logical for the children not yet at school may be too young to do it, while those older than eight or nine are all either in school or involved in productive activities. This variable is perhaps more important than preparing meals or shopping because bringing water is a task that children of both sexes can be asked to do. In contrast to fetching water and shopping, the effect of preparing meals is negative but decreases with age.

4.2.2 FAMILY

We now keep these significant variables for the children in the model and add the 16 variables referring to the family's characteristics. After excluding the non-significant variables one by one, four of them were significant. The new deviance for our model is now 829.0 with 859 degrees of freedom from 872 observations. 137 observations were lost because of the missing values of these variables, for a total of 540. The results are shown in Table 4.6.

The probability of admission is higher if the father participates in the social activities of the village and if the mother can read. Surprisingly, it also increases with the distance from the supply of water (see the table in Section B.3), especially for girls. It is smaller for religions other than Catholic and Protestant.

Table 4.6. Admissions to school with variables for the children and the family. ($N = 872, M = 540$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	3.525	0.445	HAREA
AGE	−0.469	0.044	HWALL
SEX	−0.167	0.187	SPMEAT
WATER	1.104	0.193	ACTIV
MEALS	−1.267	1.040	FAMSIZE
SHOPPING	0.469	0.176	NUMBCH
FIELD	−0.424	0.238	FATHED
AGE.MEALS	0.066	0.098	FREAD
SOCACT	0.382	0.174	DISTMIN
MREAD	0.876	0.181	REASON
DISTWAT	0.153	0.101	LANGSC
RELOTHER	−0.826	0.183	HOLPRF
SEX.DISTWAT	0.298	0.144	

The ability of the mother to read, then, has a positive influence on the probability of admission. In contrast, the number of years of education is not a significant explanatory variable. Real abilities, and not just formal schooling, is what really counts. This tends to confirm other studies indicating that several years of school does not necessarily translate into the acquisition of basic knowledge (or that the latter tends to disappear with time). It could also indicate a large variability in learning among adults having taken the same number of years of schooling.

4.2.3 VILLAGE

After adding all of the variables for the village group, and deleting those that were non-significant, only one of the nine variables remained significant, according to the rules we use, the size of the village. The new model, which includes variables for the child's and family's characteristics, as well as that for the differences across villages, has a deviance of 784.8 with 819 degrees of freedom from 834 observations. At this stage, 38 further observations were weighted out, for a total of 578. The results are in Table 4.7.

The probability of starting school increases with the size of the village, especially for the younger children.

4.2.4 SCHOOLS

At this point we add variables referring to the schools existing in the villages into the model. The types of schools in the village are not significant enough

Table 4.7. Admissions to school with variables for the children, the family, and the village. ($N = 834, M = 578$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	2.799	0.634	AGRV
AGE	−0.411	0.064	HERDV
SEX	−0.196	0.192	FISHV
WATER	1.008	0.199	MARKET
MEALS	−1.538	1.102	SHOP
SHOPPING	0.445	0.182	DISTF
FIELD	−0.312	0.244	CATHV
AGE.MEALS	0.086	0.104	PROTV
SOCACT	0.449	0.181	
MREAD	0.786	0.186	
DISTWAT	0.158	0.101	
RELOTHER	−0.909	0.188	
SEX.DISTWAT	0.312	0.146	
VSIZE	0.006	0.003	
AGE.VSIZE	−0.000	0.000	

to be kept in the model. The deviance of our new model is 738.3 with 817 degrees of freedom from 834 observations. The two variables, open school and number of schools, have no extra missing values, for an unchanged total of 578. The results are presented in Table 4.8.

The probability of admission increases with the number of schools in the village and if there is an open school.

As one might expect, educational availability has is a significant factor in determining the probability of admission, have a strong positive effect. This relationship is especially clear when there is a functioning school in the village, as compared to when there is not (or a closed school). But, the number of schools also plays a role: it is not sufficient that one school be available in a village for all educational demand to be met.

4.2.5 VARIABLE REMOVAL

At this stage, we remove the variables that have become non-significant, starting with those that entered the model most recently. Thus, we shall take out whether or not the child does the shopping and works in the fields and the interaction between sex and preparing the meals.

This procedure reduces the number of missing values to 556, leaving 856 observations. The deviance is now 751.7 with 842 degrees of freedom. The results are shown in Table 4.9.

Table 4.8. Admissions to school with variables for the children, the family, the village and the school. ($N = 834, M = 578$)

Significant variables	Estimate	Standard error
Constant	1.363	0.690
AGE	−0.408	0.066
SEX	−0.269	0.199
WATER	1.131	0.206
MEALS	−0.296	0.254
SHOPPING	0.258	0.190
FIELD	−0.296	0.254
AGE.MEALS	0.053	0.109
SOCACT	0.376	0.187
MREAD	0.837	0.194
DISTWAT	0.124	0.101
RELOTHER	−0.653	0.199
SEX.DISTWAT	0.318	0.147
VSIZE	0.007	0.003
AGE.VSIZE	−0.001	0.000
NUMBSCH	0.473	0.278
OPSCH	1.206	0.399

Table 4.9. Admissions to school with variables for the children, the family, the village and the school, after removal of unnecessary variables. ($N = 856, M = 556$)

Significant variables	Estimate	Standard error
Constant	1.377	0.655
AGE	−0.413	0.062
SEX	−0.249	0.195
WATER	1.166	0.199
MEALS	−0.470	0.283
SOCACT	0.391	0.184
MREAD	0.824	0.191
DISTWAT	0.141	0.098
RELOTHER	−0.658	0.196
SEX.DISTWAT	0.357	0.141
VSIZE	0.007	0.003
AGE.VSIZE	−0.001	0.000
NUMBSCH	0.412	0.271
OPSCH	1.380	0.389

Table 4.10. Admissions to school with variables for the children, the family, the village, the school, and the provinces. ($N = 856, M = 556$)

Significant variables	Estimate	Standard error
Constant	1.483	0.710
AGE	−0.406	0.063
SEX	−0.295	0.199
WATER	1.303	0.215
MEALS	−0.524	0.289
SOCACT	0.379	0.192
MREAD	0.732	0.200
DISTWAT	0.091	0.101
RELOTHER	−0.443	0.216
SEX.DISTWAT	0.368	0.142
VSIZE	0.007	0.003
AGE.VSIZE	−0.001	0.000
NUMBSCH	0.509	0.294
OPSCH	1.402	0.403
Fianarantsoa	−0.606	0.308
Mahajanga	0.419	0.394
Antisranana	−0.914	0.368
Toamasina	0.111	0.353
Toliara	0.580	0.558

4.2.6 REGIONAL DIFFERENCES

When introducing the variable that takes into account the different characteristics of the regions in the country, we find that living in certain regions can have a significantly different effect on whether children was admitted to school or not. The deviance of our new model is 727.1 with 837 degrees of freedom from 856 observations. The results are shown in Table 3.7.

In two of the provinces, Fianarantsoa and Antsiranana, we find that in comparison with the region of Antananarivo (taken as the reference level), the probability of children starting to school is significantly lower. In contrast, for Mahajanga and Toliara, it is higher. For the other province, Toamasina, there is no difference in the probability of children going to school with respect to the region of Antananarivo.

As for enrollment, these results contrast with the gross data in Section B.6. There, the admission rates in Fianarantsoa, Antisranana, and Toamasina are about one-half those in the other three provinces. Thus, the variables in the model can explain the low admission rate in Toamasina but not in the other two provinces. The model also indicates that the province of Antananarivo has a higher admission rate than it should, given

the variables in the model.

4.3 Conclusions

From our final model, we can see that the main sub-groups of variables affecting the admission to school refer to the characteristics of the children themselves and to their family; several variables for the work that the children do at home are significant. Children from more active and more educated families also have a better chance of going to school. Indeed, they are the main variables changing the probability to start school (apart from the age of the children, which shows that older children have generally a lower probability to be admitted school).

No variables referring to the villages where the children live are significant in the final model. On the other hand, the probability for the children to go to school increases considerably with the availability of schools in the village.

5
Delay in starting school

In this chapter, we shall be interested in analyzing the possible explanations for children starting school late. In the Madagascar educational system, children should start school when they are six years old. However, this is not always the case; in the period from 1989 to 1993 some children delayed more than five years to start school and the average delay was 2.54 years. We only consider the years since 1989 because the earlier years of the older children are missing so that the mean delay is under-estimated.

5.1 Cohort analysis

Table 4.2 gave the delays in starting to school for the various cohorts (simply subtract six from the ages). That table can be rearranged by year, instead of cohort, as in Table 5.1. This shows the distribution of ages of children, each year, starting to school for the first time. However, the percentages for the shorter delays for the earlier years are overestimated because of the missing upper right corner of the table.

The reasons for children starting school late may have changed over time, due to the development of Madagascar during the last few years. Thus, we decided first to build a more general model where all the children starting school between 1989 and 1993 were taken into account, based on

Table 5.1. Delay in starting to school (% and number) by year.

	Delay										
Year	0	1	2	3	4	5	6	7	8	9	10
1989	12.3	16.2	24.0	25.3	25.3	14.0	8.1				
	38	50	74	78	78	43	25				
1990	12.8	15.8	17.8	20.0	12.5	13.0	4.8	3.5			
	51	63	71	79	50	52	19	14			
1991	14.6	14.8	18.7	17.1	16.1	9.3	4.9	2.9	1.6		
	75	76	96	88	83	48	25	15	8		
1992	17.9	23.2	17.9	17.3	8.3	7.9	4.6	1.7	0.2	0.9	
	97	126	97	94	45	43	25	9	1	5	
1993	16.0	30.0	20.6	11.7	9.7	4.8	4.3	0.8	1.7	0.2	0.2
	97	182	125	71	59	29	26	5	10	1	1

the data in this table; secondly, we look at the special case of children starting school in 1993 (the last line of the table). The average delay has decreased in recent years, with an average delay of 2.17 years to start school in 1993.

The delay for enrollment at school can go from zero years (if the child starts when he/she is six years old) up to ten years of delay. We shall fit a generalized linear regression model with a Poisson distribution. In this model, the log average delay has a linear relationship on the explanatory variables:

$$\log(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K$$

where μ is the average delay to start going to school, β_k $k = 1, \ldots, K$, are the regression coefficients for the different explanatory variables, and x_k are the explanatory variables.

The explanatory variables used and the method of introducing them are the same as in the previous chapters. The method to select the significant variables is unchanged. But in this case, the children who were six years old and had started going to school are taken into account.

5.2 Delay, 1989–1993

Here we shall discuss the delay observed for all the children who started school in the period 1989 to 1993. In the data set there were children who seemed to start school before they were six years old. This could be due to a coding error or to parents taking their children to school earlier. There are 65 children who started when they were five; these were included in our model as if they had no delay. However, those who started school when they were two, three or four years old according to the data set (overall, only 20 children) are excluded; if they are not coding errors, they are the children of the teachers.

To take into account the gradual change in delay over the years, we include a linear trend in time in the model.

5.2.1 TIME TREND FOR AVERAGE DELAY

Our first interest in this section is to check how the delay changes during the period observed. This can be done by creating a new variable (YEAR) which is given values from one to five to those children starting in 1989 up to 1993, respectively.

If we look at Table 5.2 for the average delay during the years observed, we find a smooth decrease from 1989 up to 1993. (Recall, from Table 5.1, that the earlier years are under-estimated.) For that reason, we decided to include in the model the variable YEAR to take into account the linear trend.

Table 5.2. Average delay on each year of the sampling period (1989–1993)

Year	1993	1992	1991	1990	1989
Average delay	2.17	2.33	2.77	2.79	2.87

Table 5.3. Delay to enroll at school with only the linear time trend as explanatory variable. ($N = 2447$)

Fitted variable	Estimate	Standard error
Constant	1.096	0.023
YEAR	−0.075	0.009

When we fit only this variable as explanatory variable for the delay to enroll at school (see Table 5.3), we find the linear term to be highly significant. The deviance for this model is 3913.1 with 2445 degrees of freedom from 2447 observations as compared to a null deviance of 3982.4 with 2446 degrees of freedom.

From this point, the usual procedure of adding the children, family, village, school, and region variables will be followed.

5.2.2 CHILDREN

Most of the variables chosen for the characteristics of the children are significant at this point. From the fifteen variables fitted in the model, ten were significant. Unfortunately, there is a major problem of non-random missing values so that we are only able to retain a three variables in the model. The deviance of the model is 3457.5 with 2226 degrees of freedom from 2232 observations (having started with 2447 observations). The results are shown in Table 5.4.

At first sight, it seems that girls delay slightly less in entering school than boys, keeping other variables fixed. The point estimate of the ratio of the average delay for girls to the average delay for boys is

$$\begin{aligned}\hat{\mu} &= \frac{\hat{\mu}_g}{\hat{\mu}_b} \\ &= \exp(-0.133) \\ &= 0.88\end{aligned}$$

where μ_g and μ_b are the average delay to enroll at school for girls and boys, respectively, μ is the ratio of the two averages, and the 'hat' means that they are estimated values (according to our model) for those quantities. This means that the girls' delay is, on average, 0.88 times that of boys. This

Table 5.4. Delay to start school with variables for the children. ($N = 2232, M = 215$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	1.156	0.031	BIOLOGIC
YEAR	−0.088	0.011	RICE$^+$
SEX	−0.162	0.032	SHOPPING$^+$
CLOTHES	−0.093	0.070	WATER$^+$
SEX.CLOTHES	0.161	0.068	MEALS$^+$
YEAR.CLOTHES	0.116	0.021	BROTHER*
			FIELD$^+$
			AFOOD*
			ANIMALS$^+$

type of calculation can be made in the same way for any of the variables in our model.

There is no significant difference in the delay between the children who are the biological members of the families sampled and those who are not.

Amongst the tasks found to have a significant relation to the delay, we find getting the water, pounding the rice, washing the clothes of the family, doing the small shopping, helping to prepare the meals for the family, taking care of the animals raised by the family, and working in the fields. All of them except shopping have a positive relationship to the delay to enroll at school, which means that children started school later if they had to do any of those tasks. Unfortunately, all of the variables except washing the clothes appear to have non-random missing values, so are not included in the model. For the domestic tasks, the missing values tend to be for children with short delays, while, for the agricultural tasks, it is for those with long delays. Two other variables were found to have non-significant influence on the delay to enroll at school: taking care of their brothers and sisters and getting the food for the animals of the family.

Washing the clothes increases the delay for the girls and has had an increasing influence over the years.

5.2.3 FAMILY

The introduction of the family group of variables into the model does not change any of the conclusions already made in the previous section. All the variables for children in the model before remain significant and do not change sign. For the new variables, only five out of the 18 in this group were found to be significant at this point. The deviance decreases to 2568.4 with 1717 degrees of freedom from 1731 observations. Because these five variables were added, 501 observations were weighted out at this stage.

Table 5.5. Delay to start school with variables for the children and the family. ($N = 1731, M = 716$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	1.469	0.147	HWALL
YEAR	−0.180	0.052	SPMEAT*
SEX	0.020	0.082	ACTIV
CLOTHES	−0.084	0.082	FAMSIZE
SEX.CLOTHES	0.175	0.081	SOCACT
YEAR.CLOTHES	0.095	0.025	MREAD
HAREA	−0.001	0.001	FREAD
NUMBCH	0.043	0.009	DISTWAT
FATHED	−0.093	0.013	REASON
DISTMIN	0.000	0.001	HOLPRF$^+$
LANGSC3	−0.393	0.136	RELIGION*
SEX.NUMBCH	−0.034	0.013	
YEAR.DISTMIN	0.001	0.000	
YEAR.LANGSC3	0.077	0.052	

The results are in Table 5.5.

The area of the house shows a negative relationship with the delay to enroll at school; this means that the delay for a child to start school decreases with house area where his/her family lives increasing. This is a logical conclusion because generally, larger house area means a wealthier family and a wealthier family generally takes more care with their children's education. On the other hand, in less wealthy families, paying for the school might produce financial problems, so that their children would be delayed in starting to school.

Families with a large number of children show a longer delay for their children to start school, especially the boys. This is in accordance with what we said before, because families with lots of children usually have more financial problems, leading to problems to keep their children in the school. Instead, they may let their children do some work for the family (e.g., getting water, taking care of the animals, etc.).

More education for the father means a shorter delay for his children to go to school. This can be seen from the negative estimate of the coefficient of that variable. Such a father may care more about the education of his children because he understands better the importance of having an educational background.

A important result obtained from our model refers to the increased delay to start school if the distance from the family to the school (measured in minutes) is longer, this effect increasing over the years. Obviously, parents

care about the safety of the children and know that the younger ones should not walk long distances every day to enroll at school. Thus, when the school is far away, perhaps parents would rather defer the entrance to the school several years than let their children start school at six years old.

For some reason, delay to start school is lower in families where the preference is to teach the children at school in both French and Malagasy rather than one or the other, but this decreases over the years.

Many variables in this group were found to have non-significant effect on the delay to enroll at school (see Table 5.5). Some of those variables could have certain associations with the variables already in the model (e.g. the size of the family and the number of children in the family; whether the father and the mother can read, and the level of education of the father, etc.); so when one of them is significant, the rest of variables might have a weaker influence on the delay to go to school.

5.2.4 VILLAGE

At this stage, only four of the ten village variables checked in our model are significant. All the variables that were in the model before remain significant and with no changes in sign, so that similar conclusions as before hold for this section. The new deviance for our model is 1928.2 with 1286 degrees of freedom from 1305 observations. By adding these variables, 426 observations were weighted out at this stage. Almost one half of the observations are now missing. The results are given in Table 5.6.

For the villages where the main activity is breeding animals, where the majority of people are Catholic or which are further form the district's capital, their children show more delay in beginning school than children in other villages. On the other hand, for the villages with a shop, the average delay for boys to go to school is smaller.

Other variables found to be non-significant on the average delay are the size of the village; whether the village is based on agriculture or fishing; and whether there is a market in the village or not.

5.2.5 SCHOOLS

No variables in this group are found to be significant at this point. The model therefore stays the same as previously. All the previous ones included are still needed.

5.2.6 REGIONAL DIFFERENCES

To see how the delay for children to start school changes across the different provinces, we add the variable for the provinces sampled into the model (province Antananarivo is taken as the reference level). The results including this new variable, removing the interaction between distance to water and year, are given in Table5.7. The deviance of this model is 1916.1 with 1282 degrees of freedom from 1305 observations.

Table 5.6. Delay to start school with variables for the children, the family and the village. ($N = 1305, M = 1142$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	1.260	0.201	AGRV$^+$
YEAR	−0.068	0.071	FISHV
SEX	−0.107	0.105	MARKET
CLOTHES	0.141	0.096	VSIZE
SEX.CLOTHES	0.213	0.096	PROTV*
YEAR.CLOTHES	0.088	0.030	
HAREA	−0.001	0.001	
NUMBCH	0.041	0.010	
FATHED	−0.087	0.016	
DISTMIN	0.000	0.001	
LANGSC3	−0.170	0.187	
SEX.NUMBCH	−0.034	0.015	
YEAR.DISTMIN	0.001	0.000	
YEAR.LANGSC3	0.033	0.071	
HERDV	0.083	0.039	
SHOP	−0.162	0.051	
DISTF	0.000	0.000	
CATHV	0.098	0.037	
SEX.SHOP	0.166	0.075	

From the table we see that the children living in Antananarivo show a significantly longer delay to start school in comparison with the children living in other provinces, once the other variables in the model have been taken into account. The children living in Antisranana and Toliara have significantly less delay than the four other regions, which are all similar.

5.3 Delay, 1993

In this section we just take the children who actually started school in 1993, as a special case of the model above; children starting earlier are not used. As for the previous sections, the observations with missing values for the explanatory variables were excluded from the model. The starting number of observations is 606 children with a deviance of 1008.2.

5.3.1 CHILDREN

Firstly we introduce the variables for the different characteristics of the children. Almost all the significant variables refer to the tasks that the children must do at home. The deviance of our model is 766.0 with 514

Table 5.7. Delay in 1989 to start school with all five groups of the variables. ($N = 1305, M = 1142$)

Significant variables	Estimate	Standard error
Constant	1.361	0.126
YEAR	−0.103	0.019
SEX	−0.117	0.106
CLOTHES	−0.113	0.097
SEX.CLOTHES	0.194	0.096
YEAR.CLOTHES	0.088	0.030
HAREA	−0.001	0.001
NUMBCH	0.038	0.010
FATHED	−0.091	0.016
DISTMIN	0.000	0.001
LANGSC3	−0.212	0.094
SEX.NUMBCH	−0.032	0.015
YEAR.DISTMIN	0.001	0.000
HERDV	0.062	0.041
SHOP	−0.162	0.053
DISTF	0.000	0.000
CATHV	0.097	0.039
SEX.SHOP	0.155	0.075
Fianarantsoa	−0.054	0.056
Mahajanga	−0.047	0.060
Antisranana	−0.151	0.071
Toamasina	−0.060	0.069
Toliara	−0.243	0.079

degrees of freedom from 528 observations. The results are given in Table 5.8.

If we compare this model with the one in the previous section which included all the children starting school between 1989 and 1993, we see that almost all of the variables are required in this model, with the exception of the variable for the time trend. A major difference is that, here, there is no longer an indication of the missing values being non-random.

The tasks that the children must do such as housework (carrying the water, pounding the rice, washing the clothes, helping to prepare the meals, going on short shopping errands) and agriculture (taking care of the animals, working in the fields, finding food for the animals) are all, except three, factors which influence positively the delay: they are related to an increased delay to start school.

Doing the shopping decreases the delay for the girls but not the boys.

Table 5.8. Delay to start school in 1993 with variables for the children. ($N = 528, M = 78$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	0.339	0.134	BROTHER*
SEX	0.054	0.101	
BIOLOGIC	0.159	0.112	
WATER	0.185	0.070	
RICE	0.165	0.068	
CLOTHES	−0.295	0.175	
MEALS	0.282	0.099	
SHOPPING	0.043	0.086	
ANIMALS	0.179	0.078	
FIELD	0.329	0.079	
AFOOD	−0.190	0.170	
SEX.CLOTHES	0.624	0.193	
SEX.SHOPPING	−0.254	0.121	
SEX.AFOOD	−0.618	0.400	

Washing the clothes increases the delay for the girls but decreases it for the boys. Finding animal food decreases the delay much more for the girls than for the boys.

Among the non-work variables, there appears to be a slightly longer delay for boys to enter school than for girls, but it is non-significant. If the child is a biological member of the family, there is also more chance for the child to start school late.

Although most of the variables related to work could not be included in the more general model for all the children, not all were significant in any case, whereas here they all are. This might imply an increasing need of the children help at home over the years, for example due to a lowering standard of living.

In this section only one variable is not significant, this is whether the child must take care of his brothers or sisters. This variable has non-random missing values but is not significant once included in the model; therefore this task does not seem to influence the delay in starting school.

5.3.2 FAMILY

From this family sub-group, four variables are kept in the model, in contrast to only two included in the model for all the children starting school in the full period (the area of the house and the number of children); they are both significant and the signs are the same. The two other variables included at this point are whether the mother can read and the main reason for putting

Table 5.9. Delay to start school in 1993 with variables for the children and the family. ($N = 492, M = 114$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	0.073	0.195	HWALL*
SEX	0.119	0.205	SPMEAT*
BIOLOGIC	0.163	0.132	ACTIV
WATER	0.196	0.074	FAMSIZE
RICE	0.154	0.070	RELIGION*
CLOTHES	−0.125	0.186	SOCACT
MEALS	0.208	0.105	FATHED$^+$
SHOPPING	0.014	0.091	FREAD$^+$
ANIMALS	0.202	0.081	DISTWAT
FIELD	0.291	0.085	DISTMIN$^+$
AFOOD	−0.219	0.176	LANGSC
SEX.CLOTHES	0.434	0.209	HOLPRF$^+$
SEX.SHOPPING	−0.257	0.128	
SEX.AFOOD	−0.624	0.406	
HAREA	−0.007	0.002	
NUMBCH	0.084	0.018	
MREAD	−0.204	0.093	
REASON2	0.203	0.069	
SEX.HAREA	0.005	0.003	
SEX.NUMBCH	−0.063	0.025	
SEX.MREAD	0.233	0.132	

children in school (school not far away, free books and the relationships between the director and the teachers). The deviance of the new model is 681.8 with 471 degrees of freedom from 492 observations. The results are in Table 5.9.

As we said in the previous section, the area of the house has a negative relation to the delay to enroll at school (so children living in the family with larger houses delay less), but here the effect is mainly for boys. We can also see that the larger the number of children in the family, the longer the delay to start school, once again mainly for the boys. The results for the variable referring to whether the mother is able to read has a negative influence on delays as expected because a higher level of education for the mother implies less delay in starting school, but once more only for the boys.

Several of the main reasons for the parents to put the children in school are significant; if the school is not far away (REASON1, the number corresponds to the numbered reason in Section B.3), whether the books are free

(REASON3), and if the relationships between the director and the teachers are good (REASON9) are all similar. All of them are compared to the rest of the reasons given to put the children in the school (these are used as the reference level). Those parents giving the first ones show a higher average delay to enter school than the rest of reasons (so that if parents care more about those reasons they will be more likely to keep the children at home until they get older).

Finally, four other variables were significant but have non-random missing values and could not be included in the model. The first two are related to the level of education the father has. The other two variables are whether the parents would like to change the main school holidays for the children and the time in minutes to go from the house to the school.

5.3.3 VILLAGE

Four more variables enter the model at this stage; only two were included in the model for all the children starting school in the full period (the distance to the district capital and whether the majority of the people in the village are Catholic or not); they are both significant but the sign of the first has changed. The other two variables that have entered the model are whether the village has a market or not and whether the majority of the people of the village are Protestant. The deviance of the new model is 552.3 with 391 degrees of freedom from 416 observations. The results are in Table 5.10.

As could be expected the presence of a market in the village influences negatively the delay in starting school; this result is the same for the distance to the district's capital as the delay for starting school decreases as the distance shortens.

For villages where the majority of people are Protestant, their female children show a shorter delay in enrollment at school than in other villages.

5.3.4 SCHOOLS

Two of the school variables are significant and enter the model. These are the presence of both a private and a public school in the village, and if any are closed or not. The deviance of the new model is 539.2 with 381 degrees of freedom from 408 observations. The results are in Table 5.11.

The result for type of school is surprising because the presence of more schools implies more important delays for starting school (see the table in Section B.5). On the other hand, greater delay when there is a closed public or private school is to be expected.

5.3.5 VARIABLE REMOVAL

At this stage, we remove the variables that have become non-significant, starting with those that entered the model most recently. Thus, we shall

Table 5.10. Delay to start school in 1993 with variables for the children, the family, and the village. ($N = 416, M = 190$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	0.176	0.234	AGRV*
SEX	0.339	0.231	HERDV*
BIOLOGIC	0.166	0.160	FISHV*
WATER	0.190	0.083	SHOP
RICE	0.200	0.079	VSIZE
CLOTHES	−0.044	0.207	
MEALS	0.247	0.121	
SHOPPING	0.103	0.101	
ANIMALS	0.239	0.091	
FIELD	0.301	0.095	
AFOOD	−0.144	0.207	
SEX.CLOTHES	0.293	0.230	
SEX.SHOPPING	−0.377	0.139	
SEX.AFOOD	−0.377	0.503	
HAREA	−0.007	0.003	
NUMBCH	0.087	0.021	
MREAD	−0.282	0.102	
REASON2	0.239	0.077	
SEX.HAREA	0.005	0.003	
SEX.NUMBCH	−0.077	0.029	
SEX.MREAD	0.271	0.143	
MARKET	−0.302	0.121	
DISTF	−0.001	0.000	
PROTV	−0.030	0.126	
SEX.PROTV	−0.350	0.182	

take out whether the child must prepare the meals, find the food for the family animals, or feed the animals, and if the child is a biological member of the family.

This procedure leaves the number of missing values at 198, with 408 observations remaining. The deviance is now 541.5 with 384 degrees of freedom. The results are shown in Table 5.12.

5.3.6 REGIONAL DIFFERENCES

When we include in the model the variable to take into account the variability across the provinces, we find the results in Table 5.13. For the two provinces, Antisranana and Toliara, the average delay in starting school is smaller than the others, as in the more general model taking into account

Table 5.11. Delay to start school in 1993 with variables for the children, the family, the village, and the school. ($N = 408, M = 198$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	0.187	0.236	NUMBSCH
SEX	0.331	0.235	
BIOLOGIC	0.140	0.161	
WATER	0.213	0.085	
RICE	0.190	0.080	
CLOTHES	−0.072	0.207	
MEALS	0.271	0.122	
SHOPPING	0.122	0.102	
ANIMALS	0.238	0.092	
FIELD	0.300	0.097	
AFOOD	−0.113	0.207	
SEX.CLOTHES	0.362	0.232	
SEX.SHOPPING	−0.397	0.142	
SEX.AFOOD	−0.359	0.503	
HAREA	−0.007	0.003	
NUMBCH	0.088	0.021	
MREAD	−0.295	0.104	
REASON2	0.214	0.079	
SEX.HAREA	0.005	0.003	
SEX.NUMBCH	−0.086	0.029	
SEX.MREAD	0.301	0.145	
MARKET	−0.351	0.123	
DISTF	−0.001	0.000	
PROTV	−0.080	0.132	
SEX.PROTV	−0.299	0.187	
TYPESCH2	0.320	0.134	
TYPECSCH2	0.511	0.265	

all the years observed. The deviance of this model is 524.5 with 380 degrees of freedom from 408 observations.

5.4 Changes in delay, 1989–1993

If we compare the model including all the children who started school over the period studied with the model for starting school in 1993, we find that the reasons explaining the average delay change to some degree. The main group of variables affecting both models refers to the activities that the children must do at home, but these have a lot of non-random missing values in the first case. These variables have a positive and significant

Table 5.12. Delay to start school in 1993 with variables for the children, the family, the village, and the school, after removal of unnecessary variables. ($N = 408, M = 199$)

Significant variables	Estimate	Standard error
Constant	0.324	0.191
SEX	0.329	0.234
WATER	0.211	0.085
RICE	0.199	0.080
CLOTHES	−0.069	0.207
MEALS	0.280	0.122
FIELD	0.293	0.096
SEX.CLOTHES	0.356	0.232
SEX.SHOPPING	−0.402	0.142
HAREA	−0.008	0.003
NUMBCH	0.088	0.021
MREAD	−0.287	0.103
REASON2	0.216	0.079
SEX.HAREA	0.006	0.003
SEX.NUMBCH	−0.087	0.029
SEX.MREAD	0.294	0.145
MARKET	−0.351	0.123
DISTF	−0.001	0.000
PROTV	−0.084	0.132
SEX.PROTV	−0.313	0.187
TYPESCH2	0.327	0.134
TYPECSCH2	0.530	0.264

influence on the average delay (i.e. they increase the delay).

In 1993, the family and village variables are slightly different from the general model. Hence, we could say that the delay to enroll at school is mainly affected by the children's variables (or more precisely, by decisions made by their parents). As more housework tasks are present in 1993, that increase the delay, it might be assumed that the parents' point of view has changed in recent years. They may now require more help from their children perhaps due to the increasing living costs.

We can conclude that the recent drop in new admissions at the first year of primary school seems largely explained by a normalization of the age of admission, with children tending more and more to enter school around six or seven years old.

Table 5.13. Delay in 1993 to start school with all five groups of the variables. ($N = 408, M = 198$)

Significant variables	Estimate	Standard error
Constant	0.339	0.216
SEX	0.255	0.232
WATER	0.263	0.087
RICE	0.248	0.082
CLOTHES	0.206	0.108
MEALS	0.266	0.123
FIELD	0.255	0.097
SEX.SHOPPING	−0.326	0.143
HAREA	−0.008	0.003
NUMBCH	0.085	0.021
MREAD	−0.308	0.106
REASON2	0.275	0.080
SEX.HAREA	0.005	0.003
SEX.NUMBCH	−0.082	0.029
SEX.MREAD	0.337	0.145
MARKET	−0.374	0.125
DISTF	−0.001	0.000
PROTV	−0.067	0.139
SEX.PROTV	−0.290	0.187
TYPESCH2	0.221	0.137
TYPECSCH2	0.429	0.265
Fianarantsoa	0.125	0.118
Mahajanga	−0.023	0.123
Antisranana	−0.385	0.152
Toamasina	0.116	0.148
Toliara	−0.284	0.174

6
Repeaters

In this chapter, we shall look at the factors that influence children repeating their year. This is not the same as the pass rate because it does not take into account the children dropping out after a failure. Madagascar follows the French system of high failure rates in primary school, not the virtually automatic passing found in many English-speaking countries. This operates from the very first year of primary education.

6.1 Cohort analysis

The rates of repeating for each cohort by age are given in Table 6.1. Given the sampling variation, these values are fairly constant over the cohorts. They are around 30% until about nine or ten years old, dropping somewhat for the older ages. This reduction may be partly explained by children who have failed several times dropping out. Of course, some of the older children are in secondary school where the failure rate appears to be lower. Thus, we can conclude that the rate of repeating has not really changed over the years.

Table 6.1. School repeaters (%) by cohort and age when they failed.

	Age								
Cohort	7	8	9	10	11	12	13	14	15
1977			33.3	24.0	24.0	17.7	18.2	19.8	24.0
1978		26.7	31.1	20.2	17.2	27.0	19.3	13.6	
1979	11.1	29.4	29.5	26.5	21.8	21.5	24.7		
1980	36.7	27.8	22.4	20.8	24.0	28.3			
1981	27.0	30.4	26.8	21.7	25.6				
1982	27.1	29.7	20.7	32.5					
1983	30.6	28.6	31.5						
1984	30.3	31.8							
1985	34.7								

Table 6.2. Repeaters with variables for the children. ($N = 1853, M = 262$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	0.829	0.401	BIOLOGIC
AGE	−0.225	0.037	MEALS
SEX	−1.462	0.539	BROTHER
WATER	0.306	0.128	CLOTHES
RICE	0.278	0.112	ANIMALS
SHOPPING	0.169	0.106	AFOOD
CLASS	0.197	0.045	FIELD
SEX.AGE	0.110	0.045	

6.2 Repeaters, 1992–1993

We shall study the children who were in the same year at school in both 1992 and 1993. This corresponds to the bottom diagonal of Table 6.1. Of the 2115 children for whom we have data in these two years, some (57) were already in the same class the previous year.

6.2.1 CHILDREN

We start with 2115 children: those who were at least seven in 1992 and who were in school in both 1992 and 1993 and for whom we have information about their year of study. The null deviance is 2524.6. Six variables are significant, giving a deviance of 2158.2 with 1845 degrees of freedom for 1853 observations. Missing values account for the remaining 262 observations. The results are given in Table 6.2.

The probability of repeating is lower for girls and diminishes with age, but less rapidly for the girls. On the other hand, given the other variables in the model, it increases with the year in school.

A child has more chance of repeating if he or she has to fetch water, pound the rice, or do the shopping. This contrasts with the effect of these variables for the other response variables studied, where they favourably affect the education of the child. It may be in the poorest families, that are however motivated to send their children to school, that the children must perform these domestic tasks. Notice that none of the agriculture tasks enter the model.

6.2.2 FAMILY

After checking the family variables, we find that a large number (7), are significant. (Although the standard errors for choice of language are large, removing it increases the deviance by 5.4.) However, 338 observations are

Table 6.3. Repeaters with variables for the children and the family. ($N = 1515, M = 600$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	1.168	0.626	RELIGION
AGE	−0.289	0.042	REASON
SEX	−1.465	0.595	DISTMIN
WATER	0.334	0.143	DISTWAT
RICE	0.278	0.126	ACTSOC
SHOPPING	0.090	0.117	SPMEAT
CLASS	0.305	0.056	NUMBCH
SEX.AGE	0.102	0.050	HOLPRF
HAREA	−0.005	0.003	HWALL
ACTIV	−0.298	0.154	
FAMSIZE	0.073	0.024	
FATHED	−0.239	0.063	
MREAD	−0.295	0.138	
FREAD	0.451	0.192	
LANGSC2	0.004	0.391	
LANGSC3	0.369	0.357	

lost due to missing values. The new deviance is 1757.6 with 1499 degrees of freedom for 1515 observations. The results appear in Table 6.3. Among the previous variables, doing the shopping becomes non-significant.

The probability of repeating decreases with the wealth of the family, as measured by the size of the house, and when the father is active in the village. On the other hand, it increases with the size of the family.

Not surprisingly, all of the education variables enter the model. The probability of repeating is lower in families where the father has more education and where the mother can read. But, given the other variables, it is higher if the father can read. It is also higher if the family prefer the child to be taught in both Malagasy and French.

6.2.3 VILLAGE

Only two of the village variables enter the model. The deviance is 1656.5 with 1429 degrees of freedom from 1448 observations. Here, 66 more observations are lost due to missing values. The results are in Table 6.4.

The most striking result is that repeating occurs much less in villages where the main activity is herding (see the table in Section B.4), especially for the younger children. On the other hand, it is higher in Protestant villages than the others.

Table 6.4. Repeaters with variables for the children, the family, and the village. ($N = 1448, M = 667$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	1.722	0.680	AGRV
AGE	−0.326	0.047	FISHV
SEX	−1.548	0.615	MARKET
WATER	0.302	0.148	SHOP
RICE	0.245	0.130	VSIZE
SHOPPING	0.060	0.122	DISTF
CLASS	0.294	0.058	CATHV
SEX.AGE	0.111	0.052	
HAREA	−0.004	0.003	
ACTIV	−0.214	0.159	
FAMSIZE	0.065	0.025	
FATHED	−0.290	0.067	
MREAD	−0.238	0.144	
FREAD	0.463	0.201	
LANGSC2	0.166	0.398	
LANGSC3	0.419	0.360	
HERDV	−2.033	0.670	
PROTV	0.268	0.133	
AGE.HERDV	0.129	0.056	

None of the other variables, especially those concerning the size or the wealth of the village, enter the model.

6.2.4 SCHOOLS

Only the type of school is significant at this stage. The deviance is 1597.6 with 1392 degrees of freedom from 1414 observations (34 additional missing values). The results are given in Table 6.5.

The important result here is that private schools have a considerably higher rate of repeating than do public schools (see the table in Section B.5).

As one might expect, neither the number of schools nor the fact that there is a closed school in the village influences the rate of repeating.

6.2.5 VARIABLE REMOVAL

Three variables can now be eliminated, as non-significant, with a gain of 40 observations. These are doing the shopping, if the father is active in the village, and the choice of language at school. The deviance is now 1647.6

Table 6.5. Repeaters with variables for the children, the family, the village, and the schools. ($N = 1414, M = 701$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	1.545	0.714	NUMBSCH
AGE	−0.337	0.048	TYPECSCH
SEX	−1.837	0.630	
WATER	0.327	0.151	
RICE	0.224	0.133	
SHOPPING	0.114	0.125	
CLASS	0.301	0.060	
SEX.AGE	0.134	0.053	
HAREA	−0.004	0.003	
ACTIV	−0.203	0.164	
FAMSIZE	0.068	0.025	
FATHED	−0.283	0.068	
MREAD	−0.289	0.148	
FREAD	0.474	0.207	
LANGSC2	0.123	0.399	
LANGSC3	0.362	0.361	
HERDV	−2.190	0.684	
PROTV	0.232	0.139	
AGE.HERDV	0.143	0.057	
TYPESCH2	0.213	0.219	
TYPESCH3	0.577	0.276	
TYPESCH4	0.465	0.282	

with 1436 degrees of freedom from 1454 observations. The results are in Table 6.6.

None of the other relationships already discussed change greatly.

6.2.6 REGIONAL DIFFERENCES

After introducing the differences among the provinces, we can eliminate two variables: the area of the house and whether the child pounds the rice (with a gain of 52 observations). Apparently, these two variables vary greatly among the provinces. The new deviance is 1650.2 with 1485 degrees of freedom from 1506 observations.

The rate of repeating is much higher in Mahajanga and Antisranana and lower in Fianarantsoa than in the three other provinces (see the table in Section B.6). The socio-economic variables in the model are not able to account for these differences.

Table 6.6. Repeaters with variables for the children, the family, the village, and the schools, after removing non-significant variables. ($N = 1454, M = 661$)

Significant variables	Estimate	Standard error
Constant	1.969	0.594
AGE	−0.354	0.048
SEX	−1.935	0.621
WATER	0.346	0.149
RICE	0.182	0.129
CLASS	0.322	0.058
SEX.AGE	0.141	0.052
HAREA	−0.004	0.003
FAMSIZE	0.061	0.025
FATHED	−0.267	0.065
MREAD	−0.280	0.145
FREAD	0.491	0.202
HERDV	−2.489	0.673
PROTV	0.267	0.135
AGE.HERDV	0.165	0.056
TYPESCH2	0.176	0.217
TYPESCH3	0.557	0.269
TYPESCH4	0.453	0.280

6.3 Conclusions

As one would expect, the level of education of the parents plays a central role in whether the children repeat or not. However, the positive effect of the father reading, increasing the probability of repeating, is somewhat puzzling.

The other important results are the role of domestic tasks as associated with more repeating, in contrast to the results in other chapters, and the good showing of the herding villages.

Obviously, considerable work has to be done to equalize chances among the provinces.

Table 6.7. Repeaters with variables for the children, the family, the village, the schools, and the provinces. ($N = 1506, M = 609$)

Significant variables	Estimate	Standard error
Constant	0.880	0.624
AGE	−0.313	0.047
SEX	−1.642	0.619
WATER	0.330	0.147
CLASS	0.319	0.058
SEX.AGE	0.122	0.052
FAMSIZE	0.069	0.025
FATHED	−0.289	0.066
MREAD	−0.242	0.149
FREAD	0.486	0.208
HERDV	−2.056	0.672
PROTV	0.276	0.143
AGE.HERDV	0.132	0.056
TYPESCH2	0.396	0.224
TYPESCH3	1.111	0.285
TYPESCH4	0.826	0.297
Fianarantsoa	−0.356	0.211
Mahajanga	0.956	0.195
Antisranana	0.864	0.236
Toamasina	−0.083	0.246
Toliara	0.081	0.234

7
Dropouts from school

It is interesting and important to understand the reasons why children do not continue on to finish their education (mainly primary education) once they start school. In this chapter, we shall fit logistic models to analyse the reasons why children dropped out of school.

A dropout here is defined as follows: if a child was in school in 1990, but he/she stopped going to school for the next three years, this child is considered to be a dropout. We chose the three year period during which the child is not in school so as not to include children who drop out temporarily for a year or two. We code our response variable as one for a child who dropped out of school ; otherwise, we code the response variable as zero.

7.1 Cohort analysis

The dropout rate was calculated by age and by cohort, taking the number of dropouts observed divided by the total number of enrolled children (of a given age and cohort), given that the child did not return to school within three years. In other words, when a child was no longer found in the educational system for three years, we considered her or him to have dropped out.

Taking into account the three year delay, we can only observe a restricted number of cohorts: we cannot go further than the 1983 cohort who were seven years old in 1990 and for whom we must wait until 1993 to determine if they really dropped out at seven. In each cell of Table 7.1, we divided the observed number of dropouts by the number enrolled that year.

The rate of dropouts is never very high. It seems to be a bit larger for the older cohorts. A child who drops out before 12 seems to do it most often early on, that is about seven. But the number of observations is too small to be able to draw very strong conclusions.

7.2 Dropouts after 1990

From the total number of 4012 children, we excluded the children who were less than or equal to six-year-old in 1990. From the children whose

Table 7.1. Dropouts from school (%) by cohort and age.

	Age						
Cohort	7	8	9	10	11	12	13
1977			8.8	5.3	3.8	9.7	9.0
1978		6.5	2.7	4.4	8.8	2.1	
1979	11.1	0.0	4.2	4.5	2.1		
1980	6.5	1.4	3.1	1.9			
1981	8.1	3.4	3.1				
1982	5.6	1.4					
1983	4.0						

Table 7.2. Dropouts with variables for the children. ($N = 1041, M = 195$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	−9.815	9.252	RICE
AGE	0.115	0.143	BROTHER
SEX	−4.772	2.953	ANIMALS
BIOLOGIC	5.737	9.118	FIELD
WATER	−0.678	0.411	SHOPPING
CLOTHES	−1.068	0.607	AFOOD
MEALS	−1.046	0.672	
SEX.AGE	0.470	0.261	

ages were more than six years, we also excluded these children who were not enrolled in school in 1990, because these children do not provide any information about the dropouts also. After excluding these children, we have total 1236 observations for 1990, among which there are 40 children who dropped out of school the next year (for at least three years) and 1196 children who were enrolled in school for at least a part of the period in question.

Starting from these remarks, we fit a logistic model to analyze the important reasons why children drop out of school. We shall use the model of Equation (3.1) where π is here the probability of dropout. The method to introduce the explanatory variables and to select the significant ones is the same as before.

7.2.1 CHILDREN

For the total 16 variables in the group for the children, six of them were found to be significant. The result is shown in the Table 7.2. The deviance of the model is 217.8 with 1033 degrees of freedom from 1041 observations.

At this stage, 195 observations were weighted out because of missing values.

For the three variables personally about the child (age, sex and being a biological member of the family or not), we know from the result that the age of a child positively affects the dropout. i.e., the older a child, especially a girl, the larger probability to drop out of school. Biological members of the family have higher probability of dropping out. This may be because non-biological members of the household are sent to the family specially to go to school. Sex is not significant but is retained until final variable removal.

The other three variables in the model concern the domestic tasks that children have to do for their family (getting the water, washing the clothes, and helping to prepare the meals). For these variables, we see that the children who have to do these tasks have a lower probability to drop out of school than other children. The other variables of this kind such as taking care of the animals raised by their families, taking care of brothers or sisters, working in the fields, etc. have no significant influence on dropping out.

7.2.2 FAMILY

After adding all the variables in the family group and excluding the non-significant ones, only two are kept in the model. They refer to the activity of the father in the village and whether the religion of the head of the family is Protestant. The result is shown in the Table 7.3. The deviance of the model is 185.8 with 916 degrees of freedom from 927 observations. 114 observations were weighted out because of adding these three variables about the families.

If the father is a farmer, the probability of dropping out increases with the age of the child.

The religion of the head of the family seems to have a significant explanatory effect on the probability of children dropping out of school. We find that Protestants are the only ones to have a significant higher average of dropouts. The positive estimate of the coefficient for this variable means that when the father of the family belongs to another religion than Protestant, the children have a much lower probability of dropping out of school.

It is worth noting that the education level of the parents (FATHED, FREAD, and MREAD) does not have a significant influence on their children's dropout.

At this stage, all the significant variables coming from the former group are still significant. Most of the coefficients are unchanged. The conclusions we drew in the previous section still hold after adding the variables for the families.

Table 7.3. Dropouts with the variables for children and family. ($N = 927, M = 309$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	19.890	41.64	HAREA
AGE	−4.319	4.667	HWALL
SEX	−4.457	2.973	SPMEAT
BIOLOGIC	7.690	25.52	FAMSIZE
WATER	−0.576	0.431	NUMBCH
CLOTHES	−1.105	0.622	SOCACT
MEALS	−0.991	0.704	FATHED
SEX.AGE	0.455	0.263	MREAD
ACTIV	−33.030	32.950	FREAD
RELPROT	1.002	0.429	DISTWAT
AGE.ACTIV	4.534	4.671	DISTMIN
			REASON
			LANGSC
			HOLPRF

7.2.3 VILLAGES

When adding the variables for the village, only one of the ten variables in this group was significant and hence was kept in the model. This is the variables concerning whether there is a shop in the village or not. At this point, the model had the deviance of 157.5 with 817 degrees of freedom from 830 observations. 97 observations were weighted out at this stage. The results are given in Table 7.4.

There being a shop in the village seems to have strong negative link to the dropouts of the children in the village,especially as their age increases. i.e., if there are shops in the village, the (older) children in this village drop out of school less. This is reasonable because there being shops in the village generally means that the village is more developed and richer.

The other variables such as the size of the village, whether the village being mainly agricultural, fishing, or breeding animals or not, and the majority religion of the people in the village, do not have the significant influence on the children's dropouts.

7.2.4 SCHOOLS

When we introduce the variables for the schools into the model, we find that none are significant. This may not be surprising because we are here concerned with retaining children already in school.

Table 7.4. Dropouts with the variables for children, family, and village. ($N = 830, M = 406$)

Significant variables	Estimate	Standard error	Omitted variables
Constant	19.420	45.33	AGRV
AGE	−4.278	4.932	HERDV
SEX	−5.618	3.311	FISHV
BIOLOGIC	7.601	29.01	MARKET
WATER	−0.748	0.469	DISTF
CLOTHES	−1.029	0.680	CATHV
MEALS	−0.513	0.751	PROTV
SEX.AGE	0.551	0.292	VSIZE
ACTIV	−33.960	34.79	
RELPROT	0.986	0.467	
AGE.ACTIV	4.682	4.932	
SHOP	3.790	3.324	
AGE.SHOP	−0.436	0.294	

7.2.5 VARIABLE REMOVAL

With our complete model for the four groups of variables, we can now consider eliminating any variables that have become non-significant. The only ones that can be removed refer to the children. We remove being a biological member of the family and if the child helps prepare the meals. The new deviance is 164.2 with 830 degrees of freedom from 841 observations. The results are given in Table 7.5.

None of the relationships change very much from the previous models.

7.2.6 REGIONAL DIFFERENCES

In order to see if there are differences in dropout among the six provinces, we add this variable to the model. The result are shown in Table 7.6. The deviance of the model decreases to 149.7 with 825 degrees of freedom by adding this variable; there are still 841 observations.

With Antananarivo as the reference level, we see that children who live in Fianarantsoa, and perhaps Antsiranana and Toamasina, drop out of school more than those who live in Antananarivo. Toliara has no drop outs in the sample.

To confirm our conclusions, we also give the table of the frequencies and the percentages of the dropouts (see Table 7.7). There are two parts in this table; one is for all the observations (i.e., the children who were more than six-year-old in 1994 and went to school that year) and another is with only the observations used in the final model (i.e., deleting the observations which have missing values in the explanatory variables of the model)

Table 7.5. Dropouts with the variables for children, family, village, and school, after removing non-significant variables. ($N = 841, M = 395$)

Significant variables	Estimate	Standard error
Constant	27.520	34.56
AGE	−4.344	4.895
SEX	−5.619	3.185
WATER	−0.789	0.458
CLOTHES	−1.337	0.649
SEX.AGE	0.554	0.283
ACTIV	−34.410	34.53
RELPROT	1.090	0.458
AGE.ACTIV	4.741	4.894
SHOP	3.755	3.204
AGE.SHOP	−0.447	0.284

From this table, we also see that Fianarantsoa and Toamasina have larger percentages of the dropouts. The percentages for Antananarivo and Antsiranana are almost the same, as are those for Mahajanga and Toliara. However, these values do not take into account differences among the regions in the other variables in the model.

It is worth to note that by deleting the missing values in the explanatory variables, the percentage of the dropouts for Toamasina decrease while that for Fianarantsoa increases. This is an indication that missing values are not randomly missing. For example, for Toamasina, there are proportionally more missing values in the explanatory variables for the children who dropped out of school than those who did not drop out of school.

7.3 Conclusions

There are several factors which have the significant influence on the dropouts:

(1) the child's age;
(2) the tasks that children do for their families;
(3) the family's position in the village;
(4) the prosperity of the village, as indicated by it having a shop.

As in all of the chapters, it is worth emphasizing that these conclusions should be carefully used in practice. There are a large number of missing values in the data set, and they are generally not randomly missing.

In contrast to what is often thought, the dropout rate in Madagascar is fairly low. This result agrees with the high rate of enrollment up to 13 years old.

Table 7.6. Dropouts for all five groups of variables. ($N = 841, M = 395$)

Significant variables	Estimate	Standard error
Constant	26.590	33.20
AGE	−4.313	4.692
SEX	−6.844	3.483
WATER	−0.659	0.475
CLOTHES	−1.252	0.660
SEX.AGE	0.677	0.309
ACTIV	−35.550	33.16
RELPROT	1.434	0.500
AGE.ACTIV	4.791	4.693
SHOP	5.328	3.493
AGE.SHOP	−0.567	0.308
Fianarantsoa	1.618	0.611
Mahajanga	−0.503	1.129
Antsiranana	1.044	0.911
Toamasina	0.844	0.823
Toliara	−6.974	16.07

Table 7.7. Comparison of dropouts within the provinces.

	Dropout (with all observations at the beginning)			Dropout (with observations in the final model)		
Provinces	0	1	Percentage	0	1	Percentage
Antananarivo	349	8	2.2%	260	6	2.3%
Fianarantsoa	242	15	4.1%	152	11	7.2%
Mahajanga	210	2	0.9%	145	1	0.7%
Antsiranana	130	4	3.0%	62	2	3.2%
Toamasina	145	11	7.1%	87	3	3.4%
Toliara	120	0	0.0%	112	0	0.0%

8
Discussion

As a conclusion to this report, we shall give some discussion about the data set, the variables and the models we have fitted and the results obtained. We shall also point out the problems we found at the different stages of the analyses.

8.1 Data set

As we said in Chapter 2, this data set was obtained by questioning each family about their children. So, as with any other data sets obtained in this way, it is very common to obtain many missing values, especially for large scale investigations like this.

In our data set, most of the explanatory variables included missing values. Some of them even included more than 500 missing values. The combination of the missing values in the different variables made the number of observations available in the final model decrease sometimes to close to half the size of the set we started a model with. Furthermore, some of the variables with many missing values were very important both for the enrollment and for the delay to start school.

A very important problem we found with the missing values was that they were not randomly missing. Generally, there were more missing values for the children who did not go to school than for the children who did go to school.

The fact that the missing values were not randomly missing had perverse effects on the response variable; in the case of the logistic models, the combination of missing values across explanatory variables in the model can make too many observations disappear from one of the categories of the response variable (either too few zeros or too few ones). For example, for the enrollment at school in 1993, before fitting any variables we had 1109 children not enrolled at school and 2584 children enrolled; but in the final model, we only had 564 children not enrolled at school but 1826 children enrolled. This means that 49% of the observations for children not enrolled at school were deleted, but only 29% of the observations for children enrolled at school were deleted. This occurred in spite of the fact that some variables with a lot of missing values were not used because we believed that these values were not missing at random. There were also

similar problems for the response variables in other chapters.

The perverse effects also appeared in the explanatory variables, changing sometimes the real direction of the relationship. For example, for the distance to the source of drinkable water, if we only exclude the missing values for that variable, the average distance walked for children who were in school in 1993 is 246.8 metres and for children who were not in school in 1993 is 304.9 metres. When we take into account the missing values for the explanatory variables in the model for enrollment, the average distance walked by the children who were in school in 1993 is reduced to 236.5 metres, while for the children who were not in the school, the distance is 222.7 metres. In other words, the combination of missing values changes the average size of the explanatory variable differently for each level of the response variables. This variable was not used in the final models (for this reason).

Instead of weighting out the missing observations, we could use other techniques to deal with the missing values, such as fitting a regression model to predict them. However, given the size of the data set and the large number of missing values for many variables, this would require considerable extra work.

We also found several problems due to the coding of the variables. Either the official coding was not followed or the variables were not coded appropriate to measure what it should have. Some variables had to be eliminated from our models and many others were left unused because of their misleading definitions.

8.2 Variables

The number of variables in this data set was over 1000; this is a very large number for the number of observations we had, so a selection of variables was needed if we wanted to keep a good proportion between variables and observations.

We chose approximate 50 variables that in our opinion could have had an important effect on the response variables. Of course, there could be other variables which we did not take into account but which could have a significant explanatory effect on the models we fitted.

8.3 Models

A usual way to check whether the models are well fitted or not, is to use Q-Q plots (i.e. plot of the ordered standardized residuals against the Normal quantiles). If the model is well fitted, the ordered residuals should represent a 45 degree line. However, we must admit that it is impossible for the standardized residuals to fit that line exactly (residuals are never perfect!).

We drew a Q-Q plot for most of the logistic models we have fitted. A

regression line slightly deviated from the actual 45 degree line, but for well fitted models, the regression line becomes almost equal to that 45 degree line.

The Q-Q plot for the log linear model of the average delay to start school with a Poisson distribution is quite close to the 45 degree line, except for the smaller residuals. There is a slight curvature on the plot, which could be due to a trend in the residual plot (perhaps because there are important variables not included in the model). However, considering the large number of observations and variables in our model, we think these plots are quite good; this would mean that the models used for the response variables are reasonable.

8.4 Results

8.4.1 ENROLLMENT

The probability of enrollment at school appears to have its most important adversaries in the activities that the children must do at home (working on the fields, getting food for the animals of the family, etc.) but unfortunately these all seem to have non-random missing values. Because of this, the variables for the characteristics of the family seem to have taken on an inordinately significant effect while the local differences across villages do not even appear in the model. On the other hand, the presence and type of school have a major impact (although must be nuanced due to the problems with the variables for the children). We may presume that larger, more developed villages give more chances to children to enrolling at school.

Light domestic work is linked to a greater probability of attending school, whereas agriculture tasks go in the opposite direction. Children who go to school look after the light work whereas those who stay at home do the heavier work that requires full time involvement and/or a timetable incompatible with school hours. We do not know if this division occurs within or between families.

The educational level and the wealth (size of the house) of the parents, as well as the father's participation in social activities, are favourable to attending school.

The existence of a school, public or private, influences attendance. It is striking that girls have a higher chance of attending school in villages having only a private school. The fact that a school is available does not necessarily exhaust the demand. In so far as the number of schools is not confounded with the size of village, it seems that there may be a place for both public and private schools from the point of view of the behaviour of the families. In the same way, the closure of a school, public or private, tends to decrease the probability of attendance.

8.4.2 ADMISSION

The probability of admission only depends on one variable each of the domestic and agriculture tasks, being favourably linked with the first and negatively with the second.

Several parental variables also have an effect, but none of the village ones. The mother's education level (evaluated by her ability to read, not the level of study reached) and being a Catholic or Protestant increase the chances of admission.

The size of the village could cover the positive impact of the global level of wealth, openness to the outside, and consciousness of the effects of education.

However, the single most important factor is if there is an open school in the village. It is the variable that has the strongest effect on attending school.

8.4.3 DELAY

The delay to start school has decreased during the years sampled. Here we looked at 1989 to 1993. Again, because many of the children's variables have non-random missing values and cannot be included in the model, the family variables take extra prominence (this is confirmed by the results for 1993 alone). Several of the village variables are also significant.

The delay to enroll at school in 1993 was positively affected by several of the activities that the children must do at home (carrying water, washing up the clothes of the family, etc.). Here, for once, they have the same relationship as the agricultural tasks, both being adverse to schooling. Several parent and village variables also have important relationships. Different aspects of the family had a significant effect on the delay to start school. Most of them could be related to the educational and economic differences across families: children in families with fewer children, larger houses, and a mother who could read, showed a shorter delay to start school. Children living in villages with a Protestant majority showed a shorter average delay to start school.

The higher is the level of education of the parents, the more often the child tends to start early. For all years together, it is the father's education, while for 1993 it is the mother's. The larger is the house, the earlier the child starts; the more children in the family, the greater is the delay.

For the series of years, neither the presence of a school nor its distance affected the delay. However, in 1993, just the existence of a private school tended to be linked with increased delay. This result may seem paradoxical if we remember the strong growth of the private schools since 1991, but this movement has occurred mainly in the towns. An explication may be the lowering standard of living in the past few years. It may be affecting rural families to such a point that they now prefer the public schools for

their children.

8.4.4 REPEATING

As would be expected, the parents' education has a prime rôle in whether the children repeat their year at school. Given that helping with domestic tasks increases the risk of repeating, in contrast to the effect of these factors on the other response variables, it appears that children in the families closest to the margin of survival have the most chance of repeating.

8.4.5 DROPOUT

For the dropouts, the children's personal variables (age, sex, whether the children are the biological member of the family) play a greater role in the early stages of the model building. In the end, only a few variables stay in the model. None refer to the schools. We see, especially, that the probability of dropping out increases rapidly with the age of the child in agricultural families.

8.5 Conclusions

On the basis of the data in this study, it appears clear that educational policy concerning admitting and retaining children in primary school should concentrate on four complementary objectives:

(1) reduce the number of repeaters;
(2) concentrate action in favour of better retention of children in primary education;
(3) encourage the movement towards admission at the normal age of six;
(4) act to encourage the enrollment of the 5 to 7% of children who do not now enter school.

However, there is great variability among regions. The provinces of Fianarantsoa, Antisranana, and Toamasina are generally far behind the others. Only some of this can be explained by socio-economic factors, especially for Toamasina.

Many analyses remain to be done based on the data collected: analysis of the length of enrollment according to the age of admission; study of the wealth and spending of the families; analysis of the costs of education for the family in relation to total spending and the number of children; choice of a public or private school; relationships between the opinions expressed by the parents and their actual behaviour; and so on.

For a realistic and effective analysis of the available data, a team of five to ten statisticians would have had to work for at least a year.

This study clearly brings out the weight and complexity of the interaction among presence of a school, family organization, and the educational level and wealth of the parents in making decisions concerning their children enrolling in, attending, and dropping out of school in rural areas. In

this set of factors, the relative weight of the variables for the child and the family is very high: the Ministry of National Education obviously has no possibility to intervene here. Its only area of action concerns the existence and closeness of schools. The potential effect of the quality of teaching and other school variables has not been taken into account in this study. These results may lead the Ministry to reevaluate its room for action with regard to the objective of universal enrollment.

Appendix A
List of explanatory variables

A.1 Children

SEX of the child.

Boys	Girls
2051	1961

AGE of the children.

6	7	8	9	10	11	12	13	14	15	16
319	404	432	371	430	354	437	338	370	293	264

BIOLOGIC: Whether the child was born within the family or not (for example, he or she may be living with relatives).

No	Yes
313	3699

BROTHER: Whether the child takes care of brothers and sisters or not.

No	Yes	MV
2755	756	501

CLOTHES: Whether the child washes the clothes or not.

No	Yes	MV
2706	820	486

WATER: Whether the child gets the water or not.

No	Yes	MV
1270	2162	580

RICE: Whether the child pounds the rice or not.

No	Yes	MV
2095	1423	494

MEALS: Whether the child helps prepare the meals or not.

No	Yes	MV
2811	702	499

SHOPPING: Whether the child goes shopping or not.

No	Yes	MV
1756	1813	443

FIELD: Whether the child works in the fields or not.

No	Yes	MV
2606	929	477

ANIMALS: Whether the child takes care of the animals or not.

No	Yes	MV
2749	789	474

AFOOD: Whether the child gets food for the animals or not.

No	Yes	MV
3226	282	504

CLASS93: Year of school of those children enrolled in 1993.

1	2	3	4	5	6	7	8	9	10	11
844	640	486	292	252	59	54	20	9	5	3

A.2 Family

NUMBCH: Number of children in the family from six to 16 years old (mean = 5.56).

1	2	3	4	5	6	7	8	9	10	11	12
112	243	409	610	652	619	518	294	200	102	107	19
13	14	MV									
14	5	108									

FAMSIZE: Size of the family (mean = 8.08).

ACTIV: Activity of father (whether he is a farmer or has another activity).

Other	Farmer	MV
718	3102	192

HAREA: Area of the house (mean = 30.79 m^2).

HWALL: Material of construction of the walls of the house.

Other	Hard	MV
3671	309	32

SPMEAT: Amount of money (mean = 109,111 Malagasy Francs) spent on meat by the family during the year.

DISTWAT: Distance from the house to the closest source of drinkable water (mean = 0.392 km).

DISTMIN: Time in minutes to go from the house to the school (mean = 23.67).

RELIGION: Religion of the head of the family.

Protestant	Catholic	Other	MV
1316	1307	1123	266

RELOTHER: Variable constructed from the previous one: whether the religion of the head of the family is other than either Catholic or Protestant.

FATHED: Level of education of the father of the family (mean = 1.52).

0	1	2	3	4	5	6	MV
1120	755	1229	516	264	19	5	104

FREAD: Whether the father reads or not.

No	Yes	MV
964	2890	158

MREAD: Whether the mother reads or not.

No	Yes	MV
1532	2436	44

SOCACT: Whether the father participates in various social and religious activities or not.

No	Yes
1911	2101

REASON (1 to 10): Most important reason why the parents decided to put their child in the school (for an explanation of the ten levels of this variable, see Appendix B).

1	2	3	4	5	6	7	8	9	10	11	MV
1001	280	217	604	142	26	24	32	131	556	754	245

HOLPREF: When would the parents prefer the main school holidays for the children.

Rainy season	Harvest	No change	MV
1389	683	1716	224

LANGSCH: What language do parents want the children to learn at the school.

Malagasy	French	Both	MV
156	540	3055	261

A.3 Village

DISTF: Distance to the faritany (district's capital, mean = 234.0 km).
VSIZE: Size of the village (mean = 177.8 houses).
AGRV: Whether the village is mainly agricultural or not.

No	Yes	MV
333	3650	29

HERDV: Whether the village's main activity is breeding animals or not.

No	Yes	MV
2556	1359	97

FISHV: Whether the village's main activity is fishing or not.

No	Yes	MV
3278	137	597

SHOP: Whether there is a shop in the village or not.

No	Yes	MV
1099	2371	542

MARKET: Whether the village has a market or not.

No	Yes	MV
3408	472	132

CATHV: Whether the majority of the people in the village are Catholic or not.

No	Yes	MV
2520	1327	165

PROTV: Whether the majority of the people in the village are Protestant or not.

No	Yes	MV
2857	955	200

A.4 Schools

NUMBSCH: Number of schools in the village (mean = 0.953).

0	1	2	3
806	2670	455	81

OPSCH: Variable constructed from the previous one: whether there is an open school in the village or not.

TYPESCH: Type of open school in the village.

None	Public	Private	Both	MV
739	2543	381	282	67

TYPECSCH: Type of closed school in the village.

None	Public	Private
3063	848	101

A.5 Regions

PROV: Province where the child lives.

Antananarivo	Fianarantsoa	Mahajanga
862	826	608
Antsiranana	Toamasina	Toliara
557	817	342

A.6 Time trend

YEAR: Year of admission to school.

Appendix B
Some descriptive statistics

In this appendix, we provide three tables for each of the explanatory variables in our models. For the continuous variables, we give a table of the average value of that variable for the enrollment, admission, and drop out response variables. For the categorical variables, we give a two-way table of the marginal frequencies and the percentages for whether the children are in school or not in 1993, for admission in 1993, and for drop outs in 1990 as well as the average age to start the school in 1993 (subtract six to obtain the delay). In some cases, e.g. for those variables with distinct categories but used in our models as continuous variables (e.g. age), we give a table with all the marginal frequencies for each category as well as a table with the average value of the explanatory variable.

The only observations that were not included in the tables correspond to those children with age 6, because they were not included in our model except in the model for delay.

N indicates the number responding, M the number of missing values, and MV the missing values. For binary variables, the first category is coded zero and the second is one.

B.1 Response variables

ENROLLMENT: Children attending school in 1993 or not. $N = 3693$

	Average	Enrollment		
	starting age	0	1	%
Total	8.172	1109	2584	70.0
Total (not incl. those in secondary)		1098	2460	69.1

ADMISSION: Children beginning school in 1993. $N = 1412$

	Admissions		
	0	1	%
Total	892	520	36.8

AGE89: Age of children starting to school from 1989 to 1993. $N = 2447$

AGE89	6	7	8	9	10	11	12	13	14	15	16
	358	497	463	410	315	215	120	43	19	6	1

AGE93: Age of children starting to school in 1993. $N = 606$

AGE93	6	7	8	9	10	11	12	13	14	15	16
	97	182	125	71	59	29	26	5	10	1	1

REPEAT: Children repeating their year in 1993. $N = 2115$

	Repeaters		
	0	1	%
Total	1514	601	28.4

DROPOUT: Children attending school in 1990, with those dropping out for the three following years. $N = 1236$

	Drop outs		
	0	1	%
Total	1196	40	3.2

B.2 Children

SEX of the children observed.

$N = 3693, M = 0$

	Average	Enrollment		
SEX	starting age	0	1	%
Boys	8.242	567	1334	70.2
Girls	8.105	542	1250	69.8

$N = 1412, M = 0$

	Admissions		
SEX	0	1	%
Boys	469	262	35.8
Girls	423	258	37.9

$N = 2115, M = 0$

	Repeaters		
SEX	0	1	%
Boys	761	314	29.2
Girls	753	287	27.6

$N = 1236, M = 0$

	Drop outs		
SEX	0	1	%
Boys	586	21	3.5
Girls	610	19	3.0

Average AGE of the children.

$N = 3693, M = 0$		
	Enrollment	
	0	1
Average AGE	11.88	10.87

Frequency of the AGE of the children sampled.

$N = 3693, M = 0$			
	Enrollment		
AGE	0	1	%
7	107	297	73.5
8	117	315	72.9
9	86	285	76.8
10	101	329	76.5
11	63	291	82.2
12	123	314	71.9
13	79	259	76.6
14	134	236	63.8
15	148	145	49.5
16	151	113	42.8

$N = 1412, M = 0$			
	Admissions		
AGE	0	1	%
7	101	181	64.2
8	108	125	53.6
9	77	75	49.3
10	82	62	43.1
11	57	29	33.7
12	99	31	23.8
13	58	5	7.9
14	103	10	8.8
15	102	1	1.0
16	105	1	0.9

$N = 2115, M = 0$

	Repeaters		
AGE	0	1	%
7	128	68	34.7
8	144	67	31.6
9	191	88	31.5
10	179	86	32.5
11	224	77	25.4
12	195	77	28.3
13	189	62	24.7
14	147	39	21.0
15	117	37	24.0

$N = 1236, M = 0$

	Drop outs		
AGE	0	1	%
7	97	4	4.0
8	137	2	1.4
9	185	6	3.1
10	203	4	1.9
11	238	5	2.1
12	185	4	2.1
13	151	15	9.0

BIOLOGIC: Whether the child is the biological member of the family or not.

$N = 3693, M = 0$

	Average	Enrollment		
BIOLOGIC	starting age	0	1	%
No	8.113	102	179	63.7
Yes	8.178	1007	2405	70.5

$N = 1412, M = 0$

	Admissions		
BIOLOGIC	0	1	%
No	82	51	38.3
Yes	810	469	36.7

$N = 2115, M = 0$

	Repeaters		
BIOLOGIC	0	1	%
No	100	40	28.6
Yes	1414	561	28.4

$N = 1236, M = 0$

	Drop outs		
BIOLOGIC	0	1	%
No	76	3	3.8
Yes	1120	37	3.2

BROTHER: Whether the child takes care of brothers and sisters or not.

$N = 3255, M = 438$

	Average	Enrollment		
BROTHER	starting age	0	1	%
No	8.177	720	1827	71.7
Yes	8.277	141	567	80.1
MV	7.756	248	190	43.4

$N = 1175, M = 237$

	Admissions		
BROTHER	0	1	%
No	588	361	38.0
Yes	102	124	54.9
MV	202	35	14.8

$N = 1934, M = 181$

	Repeaters		
BROTHER	0	1	%
No	1065	417	28.1
Yes	315	137	30.4
MV	134	47	26.0

$N = 1109, M = 127$

	Drop outs		
BROTHER	0	1	%
No	853	23	2.6
Yes	229	4	1.7
MV	114	13	10.2

CLOTHES: Whether the child must wash the clothes of family or not.

$N = 3252, M = 441$

	Average	Enrollment		
CLOTHES	starting age	0	1	%
No	8.059	712	1751	71.1
Yes	9.239	158	641	80.2
MV	7.830	239	192	44.5

$N = 1178, M = 234$

	Admissions		
CLOTHES	0	1	%
No	595	413	41.0
Yes	106	64	37.6
MV	191	43	18.4

$N = 1941, M = 174$

	Repeaters		
CLOTHES	0	1	%
No	947	390	29.2
Yes	440	164	27.1
MV	127	47	27.0

$N = 1118, M = 118$

	Drop outs		
CLOTHES	0	1	%
No	677	25	3.6
Yes	412	4	1.0
MV	107	11	9.3

WATER: Whether the child goes to get the water or not.

$N = 3184, M = 509$

	Average	Enrollment		
WATER	starting age	0	1	%
No	7.774	413	696	62.8
Yes	8.466	354	1721	82.9
MV	7.978	342	167	32.8

$N = 1085, M = 327$

	Admissions		
WATER	0	1	%
No	355	181	33.8
Yes	249	300	54.6
MV	288	39	11.9

$N = 1965, M = 150$

	Repeaters		
WATER	0	1	%
No	397	124	24.7
Yes	1005	439	30.2
MV	112	38	25.3

$N = 1103, M = 133$

	Drop outs		
WATER	0	1	%
No	272	13	4.6
Yes	804	14	1.7
MV	120	13	9.8

RICE: Whether the child pounds the rice at home or not.

$N = 3258, M = 435$

	Average	Enrollment		
RICE	starting age	0	1	%
No	7.914	601	1283	68.1
Yes	8.735	256	1118	81.4
MV	7.935	252	183	42.1

$N = 1154, M = 258$

	Admissions		
RICE	0	1	%
No	512	312	37.9
Yes	164	166	50.3
MV	216	42	16.3

$N = 1959, M = 156$

	Repeaters		
RICE	0	1	%
No	722	247	25.5
Yes	681	309	31.2
MV	111	45	28.8

$N = 1117, M = 119$

	Drop outs		
RICE	0	1	%
No	532	20	3.6
Yes	556	9	1.6
MV	108	11	9.2

MEALS: Whether the child help to prepare the meals or not.

$N = 3249, M = 444$

	Average	Enrollment		
MEALS	starting age	0	1	%
No	8.067	716	1850	72.1
Yes	9.373	145	538	78.8
MV	8.000	248	196	44.1

$N = 1170, M = 242$

	Admissions		
MEALS	0	1	%
No	596	425	41.6
Yes	98	51	34.2
MV	198	44	18.2

$N = 1938, M = 177$

	Repeaters		
MEALS	0	1	%
No	1010	413	29.0
Yes	369	146	28.3
MV	135	42	23.7

$N = 1111, M = 125$

	Drop outs		
MEALS	0	1	%
No	722	25	3.3
Yes	361	3	0.8
MV	113	12	9.6

SHOPPING: Whether the child must go to do small shopping or not.

$N = 3304, M = 389$

	Average	Enrollment		
SHOPPING	starting age	0	1	%
No	8.376	576	1077	65.2
Yes	8.060	304	1347	81.6
MV	7.937	229	160	41.1

$N = 1197, M = 215$

	Admissions		
SHOPPING	0	1	%
No	472	197	29.4
Yes	232	296	56.1
MV	188	27	12.6

$N = 1959, M = 156$

	Repeaters		
SHOPPING	0	1	%
No	678	241	26.2
Yes	723	317	30.5
MV	113	43	27.6

$N = 1127, M = 109$

	Drop outs		
SHOPPING	0	1	%
No	571	21	3.5
Yes	526	9	1.7
MV	99	10	9.2

FIELD: Whether the child works in the fields or not.

$N = 3271, M = 422$

	Average	Enrollment		
FIELD	starting age	0	1	%
No	8.004	631	1726	73.2
Yes	9.146	250	664	72.6
MV	8.163	228	194	46.0

$N = 1184, M = 228$

	Admissions		
FIELD	0	1	%
No	542	400	42.5
Yes	167	75	31.0
MV	183	45	19.7

$N = 1942, M = 173$

	Repeaters		
FIELD	0	1	%
No	923	369	28.6
Yes	465	185	28.5
MV	126	47	27.2

$N = 1127, M = 109$

	Drop outs		
FIELD	0	1	%
No	660	16	2.4
Yes	439	12	2.7
MV	97	12	11.0

ANIMALS: Whether the child must take care of animals of the family or not.

$N = 3271, M = 422$

	Average	Enrollment		
ANIMALS	starting age	0	1	%
No	8.069	687	1827	72.7
Yes	8.522	191	568	74.8
MV	8.317	231	189	45.0

$N = 1190, M = 222$

	Admissions		
ANIMALS	0	1	%
No	560	371	39.8
Yes	147	112	43.2
MV	185	37	16.7

$N = 1936, M = 179$

	Repeaters		
ANIMALS	0	1	%
No	1044	418	28.6
Yes	345	129	27.2
MV	125	54	30.2

$N = 1116, M = 120$

	Drop outs		
ANIMALS	0	1	%
No	811	20	2.4
Yes	275	10	3.5
MV	110	10	8.3

AFOOD: Whether the child must go to get food for the animals of the family.

$N = 3246, M = 447$

	Average	Enrollment		
AFOOD	starting age	0	1	%
No	8.184	814	2157	72.6
Yes	7.967	70	205	74.5
MV	8.163	225	222	49.7

$N = 1188, M = 224$

	Admissions		
AFOOD	0	1	%
No	659	447	40.4
Yes	53	29	35.4
MV	180	44	19.6

$N = 1914, M = 201$

	Repeaters		
AFOOD	0	1	%
No	1234	495	28.6
Yes	138	47	25.4
MV	142	59	29.4

$N = 1104, M = 132$

	Drop outs		
AFOOD	0	1	%
No	947	25	2.6
Yes	127	5	3.8
MV	122	10	7.6

CLASS: Child's class at school.

$N = 2115, M = 0$

	Repeaters		
CLASS	0	1	%
1	444	171	33.2
2	408	154	27.4
3	290	122	29.6
4	199	53	21.0
5	92	84	47.7
6	48	9	15.8
7	18	6	25.0
8	10	2	16.7
9	4	0	0.0
10	1	0	0.0

$N = 1229, M = 7$

	Drop outs		
CLASS	0	1	%
1	453	10	2.2
2	347	10	2.8
3	209	8	3.7
4	107	7	3.3
5	49	4	7.5
6	17	0	0.0
7	6	0	0.0
8	0	1	100.0
9	0	0	—
10	1	0	0.0

B.3 Family

NUMBCH: Number of children in the family.

$N = 3588, M = 105$

	Enrollment	
	0	1
Average NUMBCH	5.715	5.511

$N = 1373, M = 39$

	Admissions	
	0	1
Average NUMBCH	5.673	5.426

$N = 2053, M = 62$

	Repeaters	
	0	1
Average NUMBCH	5.605	5.663

$N = 1190, M = 46$

	Drop outs	
	0	1
Average NUMBCH	5.723	5.590

FAMSIZE: Size of the sampled family.

$N = 3693, M = 0$

	Enrollment	
	0	1
Average FAMSIZE	8.329	7.975

$N = 1412, M = 0$

	Admissions	
	0	1
Average FAMSIZE	8.265	7.977

$N = 2115, M = 0$

	Repeaters	
	0	1
Average FAMSIZE	7.980	8.213

$N = 1236, M = 0$

	Drop outs	
	0	1
Average FAMSIZE	8.131	8.075

ACTIV: Activity of father (whether he is a farmer or he is in any other activity area).

$N = 3511, M = 182$

	Average	Enrollment		
ACTIV	starting age	0	1	%
Other	8.293	125	533	81.0
Farmer	8.188	971	1882	66.0
MV	7.444	13	169	92.9

$N = 1374, M = 38$

	Admissions		
ACTIV	0	1	%
Other	100	83	45.4
Farmer	779	412	34.6
MV	13	25	65.8

$N = 1982, M = 133$

	Repeaters		
ACTIV	0	1	%
Other	310	126	28.9
Farmer	1099	447	28.9
MV	105	28	21.1

$N = 1153, M = 83$

	Drop outs		
ACTIV	0	1	%
Other	278	2	0.7
Farmer	836	37	4.2
MV	82	1	1.1

HAREA: Area of the house (in squared metres).

$N = 3653, M = 40$

	Enrollment	
	0	1
Average HAREA	26.79	32.60

$N = 1392, M = 20$

	Admissions	
	0	1
Average HAREA	26.98	28.73

$N = 2097, M = 18$

	Repeaters	
	0	1
Average HAREA	33.66	31.60

$N = 1221, M = 15$

	Drop outs	
	0	1
Average HAREA	34.76	27.05

HWALL: Material of construction of the walls of the house.

$N = 3665, M = 28$

	Average	Enrollment		
HWALL	starting age	0	1	%
Other	8.210	1072	2303	68.2
Hard	7.795	35	255	87.9
MV	7.333	2	26	92.9

$N = 1403, M = 9$

	Admissions		
HWALL	0	1	%
Other	1045	566	35.1
Hard	35	44	55.7
MV	3	6	66.7

$N = 2093, M = 22$

	Repeaters		
HWALL	0	1	%
Other	1337	543	28.9
Hard	159	54	25.4
MV	18	4	18.2

$N = 1221, M = 15$

	Drop outs		
HWALL	0	1	%
Other	1046	36	3.3
Hard	136	3	2.3
MV	14	1	6.7

SPMEAT: Amount of money spent in meat by the family (in Malagasy francs).

$N = 3685, M = 8$

	Enrollment	
	0	1
Average SPMEAT	114,215	106,682

$N = 1412, M = 0$

	Admissions	
	0	1
Average SPMEAT	114,206	97,674

$N = 2107, M = 8$

	Repeaters	
	0	1
Average SPMEAT	106,392	113,064

$N = 1231, M = 5$		
	Drop outs	
	0	1
Average SPMEAT	119,751	77,543

DISTWAT: Distance from the house to the closest source of drinkable water (km).

$N = 3610, M = 83$		
	Enrollment	
	0	1
Average DISTWAT	0.304	0.246
$N = 1375, M = 37$		
	Admissions	
	0	1
Average DISTWAT	0.325	0.436
$N = 2066, M = 49$		
	Repeaters	
	0	1
Average DISTWAT	0.191	0.214
$N = 1199, M = 37$		
	Drop outs	
	0	1
Average DISTWAT	0.210	0.191

DISTMIN: Time in minutes from the house to the school.

$N = 3009, M = 684$		
	Enrollment	
	0	1
Average DISTMIN	25.40	23.13
$N = 1042, M = 370$		
	Admissions	
	0	1
Average DISTMIN	26.65	23.11
$N = 1822, M = 293$		
	Repeaters	
	0	1
Average DISTMIN	22.63	23.64
$N = 1055, M = 181$		
	Drop outs	
	0	1
Average DISTMIN	23.60	21.78

RELIGION: Religion of the head of the family.

$N = 3352, M = 341$

RELIGION	Average starting age	Enrollment 0	1	%
Protestant	7.881	240	988	80.5
Catholic	8.183	264	941	78.1
Other	8.346	476	543	53.3
MV	9.030	129	112	46.5

$N = 1262, M = 150$

RELIGION	Admissions 0	1	%
Protestant	186	179	49.0
Catholic	155	175	53.0
Other	428	139	24.5
MV	123	27	18.0

$N = 2032, M = 83$

RELIGION	Repeaters 0	1	%
Protestant	589	215	26.7
Catholic	593	223	27.3
Other	280	132	32.0
MV	52	31	37.3

$N = 1190, M = 46$

RELIGION	Drop outs 0	1	%
Protestant	468	23	4.7
Catholic	472	8	1.7
Other	210	9	4.1
MV	46	0	0.0

Average level of education of the father of the family.

$N = 3595, M = 98$

	Enrollment 0	1
Average FATHED	1.062	1.721

FATHED: Level of education of the father of the family.

$N = 3595, M = 98$

FATHED	Average starting age	Enrollment 0	Enrollment 1	Enrollment %
Never in school	8.463	487	536	52.4
Primary				
First cycle	8.137	198	489	71.2
Second cycle	8.255	258	900	77.7
Secondary				
First cycle	7.577	101	359	78.0
Second cycle	7.645	30	215	87.8
Superior				
First cycle	—	1	16	94.1
Second cycle	—	0	5	—
MV	8.250	34	64	65.3

$N = 1374, M = 38$

FATHED	Admissions 0	Admissions 1	Admissions %
Never in school	435	139	24.2
Primary			
First cycle	146	112	43.4
Second cycle	190	172	47.5
Secondary			
First cycle	79	57	41.9
Second cycle	15	28	65.1
Superior			
First cycle	1	0	0.0
Second cycle	0	0	—
MV	26	12	31.6

$N = 2063, M = 52$

FATHED	Repeaters 0	Repeaters 1	Repeaters %
Never in school	280	134	32.4
Primary			
First cycle	264	130	33.0
Second cycle	543	202	27.1
Secondary			
First cycle	217	85	28.1
Second cycle	153	36	19.0
Superior			
First cycle	11	4	26.7
Second cycle	4	0	0.0
MV	42	10	19.2

$N = 1206, M = 3$

	Drop outs		
FATHED	0	1	%
Never in school	227	8	3.4
Primary			
First cycle	205	11	5.1
Second cycle	418	12	2.8
Secondary			
First cycle	194	6	3.0
Second cycle	113	0	0.0
Superior			
First cycle	11	0	0.0
Second cycle	1	0	0.0
MV	27	3	10.0

FREAD: Whether the father of the child can read or not.

$N = 3545, M = 148$

	Average	Enrollment		
FREAD	starting age	0	1	%
No	8.660	430	449	51.1
Yes	8.014	634	2032	76.2
MV	8.200	45	103	69.6

$N = 1355, M = 57$

	Admissions		
FREAD	0	1	%
No	380	127	25.0
Yes	474	374	44.1
MV	38	19	33.3

$N = 2032, M = 83$

	Repeaters		
FREAD	0	1	%
No	250	96	27.7
Yes	1200	486	28.8
MV	64	19	22.9

$N = 1192, M = 44$

	Drop outs		
FREAD	0	1	%
No	197	11	5.3
Yes	955	29	2.9
MV	44	0	0.0

MREAD: Whether the mother of the child can read or not.

$N = 3651, M = 42$

	Average	Enrollment		
MREAD	starting age	0	1	%
No	8.302	636	765	54.6
Yes	8.111	461	1789	79.5
MV	7.667	12	30	71.4

$N = 1399, M = 13$

	Admissions		
MREAD	0	1	%
No	550	171	23.7
Yes	332	346	51.0
MV	10	3	23.1

$N = 2088, M = 27$

MREAD	Repeaters		
	0	1	%
No	440	193	30.5
Yes	1054	401	27.6
MV	20	7	25.9

$N = 1218, M = 18$

MREAD	Drop outs		
	0	1	%
No	354	15	4.1
Yes	824	25	2.9
MV	18	0	0.0

REASON: Most important reason why the parents decided to put the child in the school.

$N = 3491, M = 202$

REASON	Average starting age	Enrollment 0	1	%
School not far away	7.840	318	589	64.9
Safe way to the school	8.255	58	200	77.5
Free books in school	8.045	64	136	68.0
Free material in school	8.229	133	428	76.3
School's restaurant	8.870	53	77	59.2
School's schedule	9.000	3	23	88.5
School's opening time	7.000	8	14	63.6
(Unknown)	8.273	8	22	73.3
Director–teachers relationships	8.167	36	88	71.0
Director–teachers–parents relationships	8.182	108	403	78.9
Competence of teachers	8.426	209	493	70.2
MV	7.765	111	111	50.0

$N = 1315, M = 97$

REASON	Admissions 0	1	%
School not far away	249	128	34.0
Safe way to the school	49	43	46.7
Free books in school	50	21	29.6
Free material in school	109	82	42.9
School's restaurant	45	20	30.8
School's schedule	3	5	62.5
School's opening time	7	1	12.5
(Unknown)	5	10	66.7
Director–teachers relationships	27	14	34.1
Director–teachers–parents relationships	86	74	46.3
Competence of teachers	178	109	38.0
MV	84	13	13.4

$N = 1996, M = 119$

	Repeaters		
REASON	0	1	%
School not far away	361	135	27.2
Safe way to the school	100	50	33.3
Free books in school	78	44	36.1
Free material in school	251	94	27.2
School's restaurant	48	12	20.0
School's schedule	12	6	33.3
School's opening time	11	2	15.4
(Unknown)	10	4	28.6
Director–teachers relationships	53	20	27.4
Director–teachers–parents relationships	228	91	28.5
Competence of teachers	268	118	30.6
MV	94	25	21.0

$N = 1171, M = 65$

	Drop outs		
REASON	0	1	%
School not far away	272	16	5.6
Safe way to the school	91	3	3.2
Free books in school	61	1	1.6
Free material in school	207	6	2.8
School's restaurant	38	2	5.0
School's schedule	5	1	16.7
School's opening time	7	0	0.0
(Unknown)	5	0	0.0
Director–teachers relationships	44	2	4.3
Director–teachers–parents relationships	169	3	1.7
Competence of teachers	233	5	2.1
MV	64	1	1.5

SOCACT: Whether the father of the family participates in various social and religious activities or not.

$N = 3693, M = 0$

	Average	Enrollment		
SOCACT	starting age	0	1	%
No	8.235	590	1163	66.3
Yes	8.114	519	1421	73.2

$N = 1412, M = 0$

	Admissions		
SOCACT	0	1	%
No	484	251	34.1
Yes	408	269	39.7

$N = 2115, M = 0$

	Repeaters		
SOCACT	0	1	%
No	686	258	27.3
Yes	828	343	29.3

$N = 1236, M = 0$

	Drop outs		
SOCACT	0	1	%
No	524	13	2.4
Yes	672	27	3.9

HOLPREF: When would the parents prefer the main school holidays for the children.

$N = 3489, M = 204$

	Average	Enrollment		
HOLPREF	starting age	0	1	%
During rainy season	8.264	364	913	71.5
During harvest period	8.300	197	428	68.5
No changes	8.074	411	1176	74.1
MV	7.937	137	67	32.8

$N = 1268, M = 144$

	Admissions		
HOLPREF	0	1	%
During rainy season	280	188	40.2
During harvest period	169	84	33.2
No changes	312	235	43.0
MV	131	13	9.0

$N = 2062, M = 53$

	Repeaters		
HOLPREF	0	1	%
During rainy season	556	191	25.6
During harvest period	241	101	29.5
No changes	677	296	30.4
MV	40	13	24.5

$N = 1197, M = 39$

	Drop outs		
HOLPREF	0	1	%
During rainy season	430	9	2.1
During harvest period	181	7	3.7
No changes	551	19	3.3
MV	34	5	12.8

LANGSCH: What language do parents want the children to learn at the school.

$N = 3446, M = 247$

	Average	Enrollment		
LANGSCH	starting age	0	1	%
Malagasy	8.500	53	92	63.4
French	8.033	105	392	78.9
Both	8.155	841	1963	70.0
MV	8.486	110	137	55.5

$N = 1281, M = 131$

	Admissions		
LANGSCH	0	1	%
Malagasy	44	24	35.3
French	81	76	48.4
Both	669	387	36.6
MV	98	33	25.2

$N = 2007, M = 108$

	Repeaters		
LANGSCH	0	1	%
Malagasy	55	18	24.7
French	235	67	22.2
Both	1140	492	30.1
MV	84	24	22.2

$N = 1169, M = 67$

	Drop outs		
LANGSCH	0	1	%
Malagasy	37	2	5.1
French	180	3	1.6
Both	915	32	3.4
MV	64	3	4.5

B.4 Village

DISTF: Distance to the faritany (district's capital).

$N = 3317, M = 376$		
	Enrollment	
	0	1
Average DISTF	270.4	216.2

$N = 1255, M = 157$		
	Admissions	
	0	1
Average DISTF	265.7	214.6

$N = 1911, M = 204$		
	Repeaters	
	0	1
Average DISTF	219.5	224.8

$N = 1097, M = 139$		
	Drop outs	
	0	1
Average DISTF	208.9	207.0

VSIZE: Size of the village (number of houses).

$N = 3532, M = 161$		
	Enrollment	
	0	1
Average VSIZE	156.8	188.7

$N = 1364, M = 48$		
	Admissions	
	0	1
Average VSIZE	161.0	170.1

$N = 2013, M = 102$		
	Repeaters	
	0	1
Average VSIZE	188.7	191.0

$N = 1168, M = 68$		
	Drop outs	
	0	1
Average VSIZE	189.6	246.5

AGRV: Whether the village is mainly agricultural or not.

$N = 3666, M = 27$

	Average	Enrollment		
AGRV	Starting age	0	1	%
No	7.818	44	266	85.8
Yes	8.189	1063	2293	68.3
MV	8.667	2	25	92.6

$N = 1409, M = 3$

	Admissions		
AGRV	0	1	%
No	40	31	43.7
Yes	851	487	36.4
MV	1	2	66.7

$N = 2092, M = 23$

	Repeaters		
AGRV	0	1	%
No	153	61	28.5
Yes	1345	533	28.4
MV	16	7	30.4

$N = 1223, M = 12$

	Drop outs		
AGRV	0	1	%
No	145	1	0.7
Yes	1038	39	3.6
MV	13	0	0.0

HERDV: Whether the village's main activity is breeding animals or not.

$N = 3603, M = 90$

	Average	Enrollment		
HERDV	starting age	0	1	%
No	8.207	704	1674	70.4
Yes	8.039	390	835	68.2
MV	8.684	15	75	83.3

$N = 1387, M = 25$

	Admissions		
HERDV	0	1	%
No	580	350	37.6
Yes	302	155	33.9
MV	10	15	60.0

$N = 2055, M = 60$

	Repeaters		
HERDV	0	1	%
No	913	427	31.9
Yes	559	156	21.8
MV	42	18	30.0

$N = 1199, M = 37$

	Drop outs		
HERDV	0	1	%
No	774	28	3.5
Yes	385	12	3.0
MV	37	0	0.0

FISHV: Whether the village's main activity is fishing or not.

$N = 3250, M = 443$

	Average	Enrollment		
FISHV	starting age	0	1	%
No	8.201	982	2160	68.7
Yes	7.800	22	86	79.6
MV	8.091	137	414	75.1

$N = 1227, M = 185$

	Admissions		
FISHV	0	1	%
No	746	430	36.3
Yes	34	17	33.3
MV	112	73	39.5

$N = 1778, M = 337$

	Repeaters		
FISHV	0	1	%
No	1208	499	29.2
Yes	57	14	19.2
MV	249	88	26.1

$N = 1927, M = 219$

	Drop outs		
FISHV	0	1	%
No	941	32	3.3
Yes	44	0	—
MV	211	8	3.7

SHOP: Whether there is a shop in the village or not.

$N = 3197, M = 496$

	Average	Enrollment		
SHOP	starting age	0	1	%
No	8.386	383	621	61.9
Yes	8.053	505	1688	77.0
MV	8.261	221.0 275	55.4	

$N = 1133, M = 279$

	Admissions		
SHOP	0	1	%
No	300	142	32.1
Yes	392	299	43.3
MV	200	79	28.3

$N = 1918, M = 197$

	Repeaters		
SHOP	0	1	%
No	383	140	26.8
Yes	995	400	28.7
MV	136	61	31.0

$N = 1103, M = 133$

	Drop outs		
SHOP	0	1	%
No	249	18	6.7
Yes	819	17	2.0
MV	128	5	3.8

MARKET: Whether the village has a market or not.

$N = 3575, M = 118$

	Average	Enrollment		
MARKET	starting age	0	1	%
No	8.198	990	2151	68.5
Yes	7.841	91	343	79.0
MV	8.538	28	90	76.3

$N = 1369, M = 43$

	Admissions		
MARKET	0	1	%
No	798	444	35.7
Yes	74	53	41.7
MV	20	23	53.5

$N = 2046, M = 69$

	Repeaters		
MARKET	0	1	%
No	1267	494	28.0
Yes	198	87	30.5
MV	49	20	29.0

$N = 1199, M = 37$

	Drop outs		
MARKET	0	1	%
No	1003	34	3.3
Yes	158	4	2.5
MV	35	2	5.4

CATHV: Whether the majority of the people in the village were Catholic or not.

$N = 3543, M = 150$

	Average	Enrollment		
CATHV	starting age	0	1	%
No	8.082	735	1596	68.5
Yes	8.339	292	920	75.9
MV	7.786	82	68 45.3	

$N = 1322, M = 90$

	Admissions		
CATHV	0	1	%
No	622	319	33.9
Yes	192	189	49.6
MV	78	12	13.3

$N = 2060, M = 55$

	Repeaters		
CATHV	0	1	%
No	898	387	30.1
Yes	570	205	26.5
MV	46	9	16.4

$N = 1207, M = 29$

	Drop outs		
CATHV	0	1	%
No	746	26	3.4
Yes	422	13	3.0
MV	28	1	3.4

PROTV: Whether the majority of the people in the village were Protestant or not.

$N = 3507, M = 186$

	Average	Enrollment		
PROTV	starting age	0	1	%
No	8.220	832	1786	68.2
Yes	8.021	184	705	79.3
MV	8.174	93	93	50.0

$N = 1306, M = 106$

	Admissions		
PROTV	0	1	%
No	669	372	35.7
Yes	138	127	47.9
MV	85	21	19.8

$N = 2046, M = 69$

	Repeaters		
PROTV	0	1	%
No	1057	399	27.4
Yes	399	191	32.4
MV	58	11	15.9

$N = 1201, M = 35$

	Drop outs		
PROTV	0	1	%
No	805	30	3.6
Yes	356	10	2.7
MV	35	0	0.0

B.5 Schools

NUMBSCH: Number of schools in the village.

$N = 3693, M = 0$

	Enrollment	
	0	1
Average NUMBSCH	0.669	1.106

$N = 1412, M = 0$

	Admissions	
	0	1
Average NUMBSCH	0.602	1.038

$N = 2115, M = 0$

	Repeaters	
	0	1
Average NUMBSCH	1.072	1.106

$N = 1236, M = 0$

	Drop outs	
	0	1
Average NUMBSCH	1.061	0.775

TYPESCH: Type of school.

$N = 3628, M = 65$

	Average	Enrollment		
TYPESCH	starting age	0	1	%
No school	8.135	452	218	32.5
Public school	8.096	530	1812	79.5
Private school	8.354	77	278	69.4
Both schools	8.667	36	225	95.9
MV	8.900	14	51	78.5

$N = 1392, M = 20$

	Admissions		
TYPESCH	0	1	%
No school	400	47	10.5
Public school	387	376	49.3
Private school	69	54	43.9
Both schools	26	33	55.9
MV	10	10	50.0

$N = 2076, M = 39$

	Repeaters		
TYPESCH	0	1	%
No school	165	43	20.7
Public school	1058	404	27.6
Private school	140	73	34.3
Both schools	131	62	32.1
MV	20	19	48.7

$N = 1215, M = 21$

	Drop outs		
TYPESCH	0	1	%
No school	149	14	8.6
Public school	796	20	2.5
Private school	110	2	1.8
Both schools	120	4	3.2
MV	21	0	0.0

TYPECSCH: Type of closed school in the village.

$N = 3693, M = 0$

	Average	Enrollment		
TYPECSCH	starting age	0	1	%
No school	8.179	680	2151	76.0
Public school	8.139	398	368	48.0
Private school	8.176	31	65	67.7

$N = 1412, M = 0$

	Admissions		
TYPECSCH	0	1	%
No school	530	411	43.7
Public school	340	94	21.7
Private school	22	15	40.5

$N = 2115, M = 0$

	Repeaters		
TYPECSCH	0	1	%
No school	1231	527	30.0
Public school	238	66	21.7
Private school	45	8	15.1

$N = 1236, M = 0$

	Drop outs		
TYPECSCH	0	1	%
No school	978	29	2.9
Public school	200	11	5.2
Private school	18	0	0.0

B.6 Provinces

PROV: Province where the child lives.

$N = 3693, M = 0$

	Average	Enrollment		
PROV	starting age	0	1	%
Antananarivo	8.132	118	683	85.3
Fianarantsoa	8.248	274	484	63.9
Mahajanga	8.450	108	466	81.2
Antsiranana	8.042	223	273	55.0
Toamasina	8.136	351	393	52.8
Toliara	7.789	35	285	89.1

$N = 1412, M = 0$

	Admissions		
PROV	0	1	%
Antananarivo	76	117	60.6
Fianarantsoa	207	102	33.0
Mahajanga	73	95	56.5
Antsiranana	207	52	20.1
Toamasina	305	109	26.3
Toliara	24	45	65.2

$N = 2115, M = 0$

	Repeaters		
PROV	0	1	%
Antananarivo	412	161	28.1
Fianarantsoa	339	77	18.5
Mahajanga	218	165	43.1
Antsiranana	136	77	36.2
Toamasina	234	62	20.9
Toliara	175	59	25.2

$N = 1236, M = 0$

	Drop outs		
PROV	0	1	%
Antananarivo	349	8	2.2
Fianarantsoa	242	15	4.1
Mahajanga	210	2	0.9
Antsiranana	130	4	3.0
Toamasina	145	11	7.1
Toliara	120	0	0.0

Bibliography

1. Boch, M. (1968) Tombs and conservatism among the Merina of Madagascar. *Man* **3**, 94–104.
2. Boch, M. (1971) Decision-making in councils among the Merina of Madagascar. in Richards, A.I. and Kuper, A. (ed.) *Councils in Action.* Cambridge: Cambridge University Press, pp. 29–62.
3. Boch, M. (1975) Property and the end of affinity. in Bloch, M. (ed.) *Marxist Analyses and Social Anthropology.* London: Malaby, pp. 203–228.
4. Boch, M. (1977) The disconnection between power and rank as a process. An outline of the development of kingdoms in central Madagascar. *Archives européennes de Sociologie* **18**, 107–148.
5. Boch, M. (1978) Marriage amongst equals: an analysis of the marriage ceremony of the Merina of Madagascar. *Man* **13**, 21–33.
6. Bourniquel, V. and Vidal, J.P. (1990) *Bonjour Madagascar.* Montpelier: Pélican.
7. Brieu, S. (1995) La Grande Ile sous l'influence des églises. *Le monde diplomatique*, octobre.
8. Bui, R. (1989) Madagascar ne s'agenouille pas. *Le monde diplomatique*, décembre.
9. Coussy, J. and Hugon, P. (1986) Trois pays, trois types de contraintes. *Le monde diplomatique*, avril.
10. Därr, W. and Wodtcke, A. (1994) Madagascar Comores. Paris: Arthaud.
11. Heneveld, W. (1995) Madagascar. Towards a School-based Strategy for Improving Primary and Secondary Education. World Bank Report No. 13450-MAG.
12. Langellier, J.P. (1996) Survivre à Madagascar. *Le Monde*, 8 avril.
13. Leymarie, P. (1989) Le président Ratsiraka, champion de l'entreprise privée. *Le monde diplomatique*, janvier.
14. Leymarie, P. (1989) Opération 'sécurité intégrée' contre les voleurs de zébus. *Le monde diplomatique*, janvier.
15. Leymarie, P. (1991) Révolution ou restauration à Madagascar? *Le monde diplomatique*, octobre.
16. Leymarie, P. (1995) Longue patience à Madagascar. *Le monde diplomatique*, octobre.
17. MAGPLANED (1995) *Diagnostic et Scenarios de Développement des Enseignements primaire et secondaire.* Ministère de l'Education nationale et Commission nationale malgache pour l'UNESCO. 117 pages.

18. Oberlé, P. (1987) *Antananarivo et l'Imerina.* Antananarivo: Société malgache d'Edition.
19. Raison-Jourde, F. (1984) Les élites malgaches, ou, comment changer pour mieux conserver? Travail Capital et Société 17, 174–195.
20. Raison-Jourde, F. (1995) La mémoire en cendres de Madagascar. *Le monde diplomatique*, décembre.
21. Rakoto, I., Ramiandrasoa, F., and Razoharinoro-Randriamboavonjy (1995) Des institutions juridiques, politiques et économiques à Madagascar depuis les origines jusqu'à la proclamation de l'indépendence en 1960. *Revue d'Information de la Commission nationale malgache* **4**, 107–110.
22. Rambeloson-Rapiera, J. (1994) La langue française à Madagascar. *Revue d'Information de la Commission nationale malgache* **3**, 95–97.
23. Ravololomanga, B. and Schlemmer, B. (1994) De l'enfant richesse à l'enfant fardeau: l'enfant au travail à Madagascar et sa place dans l'imaginaire social. Travail Capital et Société **27**, 216–232.
24. Swaney, D. and Willox, R. (1994) *Madagascar and Comoros.* Hawthorn: Lonely Planet.
25. Vérin, P. (1994) Madagascar. Paris: Karthala.