

Background :

In 1999, the French national institute for economic and statistical studies (INSEE) conducted a survey called HID on impairment, disability and handicap in ordinary household. But as the target population of this survey is quite rare, a simple random sampling would have given us poor estimates (in term of variance of the the survey variables).

Following, Onu stat recommandations¹, it was decided to conduct a two phase survey :

- a short self-administered questionnaire (the VQS questionnaire) was given to 410 000 people at the time of the population census in march 1999.
- The 359 000 respondants to VQS where classified in 10 stratums according to their age and the severity of their handicap (6 groups of various estimated severity where constructed) which was estimated using VQS data. The 21 700 sample of HID was created by a stratified sampling on those stratums.

The quality of the HID survey (in term of variance of the the survey variables) depends a lot on the quality of the stratums created with the VQS data. The quality of the stratum depends on 2 factors:

1: the VQS questions set

2: the algorithm used to create the stratums

the questions studied in this poster are:

1: what is the efficiency of the VQS questions set?

2: what is the efficiency of the algorithm used to create the stratums?

¹ United Nations, Department for Economic and Social Information and Policy Analysis, Statistics Division : « Manual for the Development of Statistical Information for Disability Programmes and Policies », Statistics on Special Population Groups, Series Y, n° 8, New York, 1996

A: presentation of the VQS questionnaire

The VQS questionnaire contains around 20 questions for each member of the household.

set on questions on incapacity: not asked if the person is less than 16 years old

- Variable CLOSE_SIGHT :Has the person difficulty seeing the ordinary characters of a newspaper (with spectacles or contact lenses if she has)? YES/NO
- Variable FAR_SIGHT :has the person difficulty to recognise the face of someone the other side of the room ou the other site of the road (with spectacles or contact lenses if she has)? YES/NO
- Variable FORM :Usually has the person difficulty to fill a form on her own? YES/NO
- Variable COMMUNICATION : Has the person difficulty to talk and make yerself understood? YES/NO
- Variable HEARING : Has the person difficulty to hear a conversation between several persons? YES/NO
- Variable PICK : Has the person difficulty to bend and pick an object on the floor? YES/NO
- Variable DRESS : Has the person difficulty to dress and undress? YES/NO

Set of questions on help:

- Variable HELP : Due to an health condition, does the person need the help of someone in the daily life? YES/NO
- Variable ARRANGEMENT : Did you do or do you plan to do arrangements of your housing? YES/NO
- Variable TECH_AIDE: Due to an health condition, does the person uses regularly a prosthesis or a technical aide (stick, crutch, wheelchair,...) to the exclusion of spectacles and dental prosthesis. YES/NO

Other questions: one on activites limitations, one on auto attribtution of a handicap, and three on recognition of a handicap

- Variable LIMITATION : Is the person limited in the quantity or sort of activities that she can do? YES/NO
- Variable HANDICAP : Does the person considers that she has a handicap? YES/NO
- Variable APPLICATION: Has the person or has someone made for the person an application to get an official recognition of a handicap? YES/NO
- Variable RECOGNITION : If yes, was this application accepted? YES/NO/PENDING
- Variable SPE_CLASS: If the person is a child, is he in a specialised class or institution, due to an health condition or to learning problems? YES/NO

In fact there were 2 more questions but they where not taken into account by the algorithm used to create the stratums.

The goal of this questionnaire was to detect all handicaped people through at least one question. So the questions are often very broad. The conceptors of this questionnaire decided that a few questions on incapacity wouldn't suffice so the 3 questions on help on the other question set were added.

To analyse the VQS data we can do a classification on the data but there are several **issues**:

Issue 1: we have to separate the children less than 16 years old from the other VQS respondents for the incapacities questions where not asked to them.

Issue 2: we have to separate the people age less than 60 from the people over 60 years old because those 2 populations don't answer the questions the same way. For exemple:

- the official recognition of a handicap in France are not the same at all when you are under or over 60.
- A study made on the VQS data² shows that when you are more than 60 you tend less to consider that you have a handicap than where you are less than 60 (the data is controled on the VQS questions on incapacity, help and activity limitation).

Issue 3: As the variable are qualitatives, we have to do a multiple correspondence analysis (proc CORRESP in SAS) and then a classification on the coordinates of the individuals on the different axis.

Issue 4: As the data set is quite large, a ascendant hierarchical classification takes too much time of computation, so a mixed classification is used : we do agregation around the mobile center (proc FASTCLUS in SAS) and we then do an ascendant hierarchical classification of the clusters obtained previously (proc CLUSTER in SAS).

Issue 5: the multiple correspondence analysis calculates distance between individuals using the chi2 distance. A modality of answer to a question can have a high weight in the distance calculus between individuals if this modality is rare.

Issue 6: We made a separate group with the persons who haven't answer YES to a single question. We don't have to put them in the classification for we know that they will form a homogeneous group.

Issue 7: There are some missing values but the classification should be made on complete data. So we did a bit of manual imputation and we put in a separate group the people with incomplete data that could't be completed easily.

The classification on the people aged 60 or more years old :

1: we create manually the group of the persons who haven't answer YES to a single question.

2: we did a bit of manual imputation in quite obvious cases and we then put in a separate group the people with incomplete data.

3: The variables APPLICATION and RECOGNITION give the same information except when the official recognition is pending or refused which is rare. So to adress the issue 5, we excluded the variable RECOGNITION from the analysis and make 2 distincts groups with the persons whose recognition is pending or refused.

4: We put the variables COMMUNICATION and ARRANGEMENT as supplementary variables. We separated the data between those who have made an official recognition of a handicap and the others and we made a classification on each group.

Some of the VQS respondents have answered the HID survey, so for each of our groups we can add information from HID. In HID we have the number of impairment for each HID respondent and we can construct an incapacity indicator using all HID questions on incapacity and by putting a weight on each answer. The KATZ indicator can also be calculated using the HID questions.

The results are summarised in the following table :

² INED review population 2002 RAVAUD, LETOURMY and VILLE

For each VQS variable I give the percentage of person who answered YES to the question

	People who haven't answer YES to a single question	A group from classification on people who made an application	A group from classification on people who made no application	people with incomplete data	A group from classification on people who made no application	application to get an official recognition rejected	A group from classification on people who made an application	A group from classification on people who made no application	application to get an official recognition pending	A group from classification on people who made an application
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
Numbers of VQS respondant per group	31 449	16510	1067	6715	5010	652	3201	3690	396	1083
CLOSE_SIGHT	0%	28%	21%	41%	20%	36%	25%	71%	39%	85%
FAR_SIGHT	0%	1%	8%	17%	0%	20%	0%	100%	23%	100%
FORM	0%	23%	19%	26%	27%	34%	29%	59%	41%	79%
COMMUNICATION	0%	4%	5%	8%	4%	9%	11%	13%	15%	28%
HEARING	0%	35%	26%	40%	45%	38%	31%	49%	41%	56%
PICK	0%	39%	30%	49%	55%	63%	69%	56%	66%	79%
DRESS	0%	9%	9%	19%	24%	30%	39%	27%	45%	56%
HELP	0%	10%	9%	24%	32%	38%	47%	39%	52%	72%
ARRANGEMENT	0%	3%	4%	8%	9%	12%	17%	10%	20%	26%
TECH_AIDE	0%	0%	20%	29%	100%	38%	50%	34%	45%	60%
LIMITATION	0%	40%	0%	62%	67%	79%	100%	65%	82%	100%
HANDICAP	0%	18%	63%	44%	52%	81%	92%	47%	80%	96%
APPLICATION	0%	0%	100%	27%	0%	91%	100%	0%	78%	100%
Numbers of HID respondant	1663	2394	313	1071	954	151	1023	640	95	287
Average number of impairment for HID respondant	1,2	2,2	2,4	2,7	2,8	3	3	3	3,1	3,5
Average incapacity indicator for HID respondant	0,26	0,66	0,69	1,12	1,22	1,25	1,52	1,60	1,63	2,75
HID respondants classified B with Katz indicator	2%	5%	7%	8%	10%	9%	12%	11%	12%	17%

Analysis of the table:

- there is a lot of heterogeneity among people who have made an application to get an official recognition of a handicap and got an official recognition (for the refused and pending recognition have been put apart).
- The questions on incapacity and the others questions are **complementary** : the groups 7 and 8 seems to be constituted of people with quite high handicap according to the average number of impairment for HID respondent and the average incapacity indicator for HID respondent but people from group 8 were detected principally by the incapacity questions and people from group 7 were detected principally by the other questions.
- People who have a pending official recognition don't have the same characteristics that those who have a rejected application for an official recognition.

The classification on the children less than 16 years old:

The method is the same than for the people aged 60 or more.

1: we create manually the group of the persons who haven't answer YES to a single question.

2: we did a bit of manual imputation in quite obvious cases and we then put in a separate group the people with incomplete data.

3: The variables APPLICATION and RECOGNITION give the same information except when the official recognition is pending or refused which is rare. So to address the issue 5, we excluded the variable RECOGNITION from the analysis and make 2 distinct groups with the persons whose recognition is pending or refused.

4: We separated the data between those who are less than 5 years old and the others and we made a classification on each group.

The reason is that variable SPE_CLASS as no sense for the younger (and the question is often unanswered) and most parents answered YES to the question HELP even if the children had no health condition or handicap.

The results are summarised in the following table :

	People who haven't answer YES to a single question	A group from classification on children less than 5	A group from classification on children 5 or more	A group from classification on children 5 or more	A group from classification on children 5 or more	children with incomplete data	application to get an official recognition rejected	application to get an official recognition pending	A group from classification on children less than 5	A group from classification on children 5 or more
classes	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8	Group 9	Group 10
Numbers of VQS respondent per group	70 898	800	490	304	498	431	73	74	115	439
HELP	0%	70%	78%	9%	3%	26%	25%	54%	71%	64%
ARRANGEMENT	0%	18%	36%	0%	0%	21%	8%	38%	23%	21%
TECH_AIDE	0%	12%	10%	0%	29%	26%	26%	28%	44%	34%
LIMITATION	0%	24%	23%	11%	77%	45%	51%	65%	73%	75%
HANDICAP	0%	10%	7%	12%	37%	25%	66%	61%	89%	91%
APPLICATION	0%	0%	0%	0%	0%	17%	74%	53%	97%	99%
SPE_CLASS	0%	5%	1%	100%	1%	20%	23%	27%	23%	68%
Numbers of HID respondent	635	149	93	145	110	99	29	24	70	220
Average number of impairment for HID respondent	0,35	0,6	0,6	1	1	1,2	1,6	1,9	1,6	1,9
Serious mental impairment for HID respondent								12,5%	14%	17%
Light mental impairment or learning problems for HID respondent				16%		9%	10%	17%		

Analysis of the table:

- the variables SPE_CLASS and HELP don't give valuable information for the children less than five years old. The percentage of children needing help is nearly the same for the groups 2 and 9 but children from group 9 are highly handicapped contrary to children from group 2.
- We find in group 4 quite a high proportion of children with Light mental impairment or learning problems. Those children are detected in this group by the variable SPE_CLASS only so this variable is very useful.

The study of the VQS data has not been conducted for the adults.

We saw with those 2 classifications that it was possible to construct various groups (10 for each classification) with very different patterns and that there was huge difference of the proportion of people answering YES to any question among the groups. This tends to prove that the VQS questions are useful.

But do we have a complete picture of the various types of handicap with those questions?

To answer that question we used the 6 groups of estimated severity of handicap created by the conceptors of the survey and the HID data to get estimates of the numbers of blind people, heavily physically handicapped people (tetraplegic, paraplegic or hemiplegic),...

the 6 groups of estimated severity of handicap created by the conceptors of the survey :

group 1 : People who haven't answer YES to a single question of VQS

group 2 : people who have a single difficulty

group 3 : people who declare a handicap or have applied for an official recognition or declare activity limitation or need help or suffer from several difficulties

group 4 : people who declare a handicap or have applied for an official recognition and declare activity limitation and have several difficulties

group 5 : people who declare a handicap or have applied for an official recognition and have a lot of difficulties

group 6 : people who have an official recognition of handicap

We get the following results :

VQS Groups by conceptors of the survey	Pourcentage of blinds less than 60 in each group	Pourcentage of blinds 60 or more in each group	Pourcentage of heavily physically handicaped people less than 60 in each group	Pourcentage of heavily physically handicaped people 60 or more in each group
1	1%	1%		2%
2	12%	1%		10%
3	10%	7%	4%	8%
4	3%	14%	1%	6%
5	10%	34%	18%	26%
6	64%	42%	77%	45%

VQS Groups by conceptors of the survey	Pourcentage of people less than 60 who have seen a doctor for mental of psychical illness in the past 3 months	Pourcentage of people 60 or more who have seen a doctor for mental of psychical illness in the past 3 months	Pourcentage of people less than 60 who have consistency behaviour problem ³	Pourcentage of people 60 or more who have consistency behaviour problem
1	60%	10%	67%	17%
2	4%	12%	5%	9%
3	4,5%	12%	4%	6%
4	3%	16%	2%	12%
5	7%	25%	5%	26,5%
6	20%	25%	17%	30%

The VQS questions screen well people with sensorial impairment, physical impairment but not psychical impairment.

³ communication problem not related to hearing problem, agressive behaviour, strange behaviour

Let's move on the second question : **what is the efficiency of the algorithm used to create the stratums?**

the conceptors groups seem quite good as we can see in the following table (for the persons 60 or more):

	Numbers of HID respondant	Average number of impairment for HID respondant	Average incapacity indicator for HID respondant
Group 1	1652	1,2	0,3
Group 2	771	1,8	0,4
Group 3	987	2,2	0,6
Group 4	1255	2,6	1
Group 5	2107	2,9	1,4
Group 6	1819	3	1,6

the groups obtained by the classification seem quite good as we can see in the following table (for the persons 60 or more): some of the classification groups have been grouped

	Numbers of HID respondant	Average number of impairment for HID respondant	Average incapacity indicator for HID respondant
Group 1	1663	1,2	0,3
Group 2+group 3	2707	2,2	0,6
Group 4+ Group 5+group 6	2176	2,7	1,2
Group 7+ Group 8+group 9	1757	3	1,5
Group 10	287	3,5	2,7

How can we judge the quality of those groups?

Who can calculate the correlation between VQS groups (the conceptors groups or groups issued from the classifications) and an interest quantitative variable (z) : the number of impairment for HID respondant and the incapacity indicator for HID respondant using the formula:

$$\rho^2 = \frac{\sum_{group_j} nj(\text{avg}(z_j) - \text{avg}(z))^2}{N \cdot \text{var}(z)}$$

The ρ^2 will be between 0 and 1

In fact calculating a ρ^2 is the same thing that calculation an inter-group variance (at a coefficient near).

We know that that for any variable the inter-group variance plus the intra-group variance is constant. VQS groups will be good if the intra-group variance is low and inter-group variance.

Results for the people aged 60 or more:

	quantitative variable :number of impairment for HID respondant	quantitative variable : incapacity indicator for HID respondant
VQS Groups by conceptors of the survey	$\rho^2 = 0,055$	$\rho^2 = 0,10$
VQS Groups obtained by the classification	$\rho^2 = 0,057$	$\rho^2 = 0,17$

The groups made by the classification don't seem better than the groups made the conceptors of the survey.

In fact the usefulness of a classification for creating groups is limited in our case because we had to put some variables as supplementary and because a classification give the same weight to each variable.

Perhaps a better solution would consist to do a discriminant analysis (proc DISCRIM with SAS) on the VQS data. We would make the groups using a global indicator like the number of impairment for HID respondant or incapacity indicator for HID respondant (if a HID respondant has an incapacity indicator > level he is group number X,...). The results of this discriminant analysis could be used for the next VQS/HID surveys.