

SAS INPUT CONVENTIONS

November 7, 2003

The Center for Disease Control/National Center for Health Statistics, is releasing SAS programs to analyze over twenty years of data. Data from the Nation Health Examination Surveys (1959-1970), National Health and Nutrition Examination Surveys (1971-1980), and Hispanic Health and Nutrition Examination Survey (1982-1984) can now be analyzed with the assistance of these SAS programs.

The data was previously available in a simple text format, but required researchers to write input statements in order to analyze the data. Researchers had to write code based on the tape position of specific items and labels from hardcopy codebooks or data tape documentation. The CDC/NCHS has now written SAS code for these files, and expects them to significantly improve the ability to analyze this data.

These files are provided as a starting point to perform advanced statistical queries on the data. They are to be used in conjunction with legacy codebooks or data tape documentation to analyze the data. Over 55 files from six surveys are now available.

SAS File Naming Conventions

Each SAS file has the same name of the tape data file, but ends with the extension .SAS. All tape data files start with DU and are followed by their tape number. The Psychological Distress Questionnaire of the NHANES survey data file has a tape number of 4067. The data file is thus named DU4067.txt. The SAS input program would be named DU4067.SAS.

Many of the questionnaire sections have formats attached to the questions. For sections that have formatting available, we have provided two versions of the file – one with formatting and one without. The file naming conventions are the same, except those files with formatting have a _F appended to the end of the file name.

Variable Naming Conventions

Variable names start with the two letter abbreviation of the survey described above. It is then followed by a two letter abbreviation of the questionnaire section, followed by four characters representing the tape position of the variable in the data file. The “Age Head of Household” question of the Physical Measurements questionnaire in the NHES 1 survey has a tape position of 17. The variable name for this data item then is H1PM0017. In order to determine a variables name:

- 1 Find the question in the data tape documentation.
- 2 Determine its starting tape position (often located to the left of the question).

- 3 Add this number to survey and section abbreviations and pad it with zeros to the left so that it is 4 characters long. The total number of characters for the variable will be eight.

Examples:

1. NHES 1 Body Measurements Questionnaire – Question “Elbow Rest Height” at tape position 73 would have variable name H1BM0073.
2. NHES 2 – Question “Age when first walked” has tape position 167 would have variable name H2ID0167.
3. NHES 3 – Question “Age of father” has tape position 180 would have variable name H3ED0180.
4. NHANES 1 – Medical History Childhood Questionnaire – Question “How tall is he?” has tape position 306, would have variable name N1CH0306.

Exceptions:

1. The sequence number of each questionnaire has is name SEQN.
2. Hispanic Hanes variables are named slightly differently. Variables up to tape position 400 are demographic variables and have a section abbreviation of DM instead of the questionnaire abbreviation.

Label Naming Conventions

The labels provided in the SAS statements are meant as a guide to assist researchers. The labels have been truncated at forty characters instead of being adjusted to fit within the 40 character SAS limit. For the full label, please check the data tape documentation.

Additional Notes:

NHANES 2 data tape file number 5252 (Chest X-Ray Examination) has SEQN starting at position 328 instead of the first 5 tape positions. The data tape file and the data tape documentation are different.

For several questions in NHES 2 questionnaire 2IDT (Integrated Data Tape), the formatted responses are different from those listed in data tape documentation. ‘&’ should be ‘V’ and ‘-’ should be ‘X’. See question 565 for an example.

File Naming List

A. National Health Examination Survey I (NHES 1)

- a. Years: 1959-1962
- b. SAS File Prefix: H1
- c. SAS File Count: 8

Tape Number	Name	Abbreviation
1001	Demographic Characteristics	DM
1002	Psychological Distress	MH
1003	Physical Measurements	BM
1004	Cardiovascular Data	CD
1005	Osteoarthritis and Rheumatoid Arthritis	AR
1006	Dental Findings	OH
1007	Diabetes Findings	DI
1008	Vision	VI

B. National Health Examination Survey II (NHES 2)

- a. Years: 1963-1965
- b. SAS File Prefix: H2
- c. SAS File Count: 1

Tape Number	Name	Abbreviation
2IDT	Integrated Data Tape	ID

C. National Health Examination Survey III (NHES 3)

- a. Years: 1966-1970
- b. SAS File Prefix: H3
- c. SAS File Count: 1

Tape Number	Name	Abbreviation
3EDT	Extended Data Tapes	ED

D. National Health and Nutrition Examination Survey I (NHANES 1)

- a. Years: 1971 – 1975
- b. SAS File Prefix: N1
- c. SAS File Count: 20

Tape Number	Name	Abbreviation
4067	Medical History Questionnaire, Ages 1-11 years	CH
4081	Medical History Questionnaire, Ages 12-74 years	AH
4091	Health Care Needs, General Medical History Supplement, and Respiratory and	GM

	Cardiovascular Supplements	
4111	Anthropometry Goniometry, Skeletal Age, Bone Density, and Cortical Thickness	BM
4121	Arthritis Data	AR
4140	Computer Measurements and Interpretations of Electrocardiograms	EK
4151	Dermatology	DE
4161	Ophthalmology	OP
4163	Near and Distant Vision	VI
4171	General Well-Being and the CES-D Depression Scale	MH
4233	Medical Examination	ME
4235	Dental	OH
4241	Audiometric	AU
4250	Spirometry-Best Trials Only	SP
4251	Chest X-Ray, Pulmonary Diffusion, and Tuberculin Test Results	XR
4701	Dietary Frequency and Adequacy	FF
4702	Model Gram	MG
4703	Nutrient Composition	NC
4704	24-Hour Food Consumption Intake	DR
4800	Biochemistry, Serology, Hematology, Blood Slides, Urine Dipst.	LB

E. National Health and Nutrition Examination Survey II (NHANES 2)

- a. Years: 1976-1980
- b. SAS File Prefix: N2
- c. SAS File Count: 14

Tape Number	Name	Abbreviation
5010	Medical History Questionnaire, Ages 6 months-11 years	CH
5020	Medical History Questionnaire, Ages 12-74 years	AH
5252	Chest X-Ray Examination	XR
5301	Anthropometry	BM
5302	Physician's Examination	PE
5305	Health History Supplement	SH
5306	Audiometric Air Conduction Test	AU
5309	Allergy Skin Testing	AL
5317	Behavior Questionnaire	BQ
5411	Hematology and Biochemistry	LB

5701	Total Nutrient Intake, Food Frequency, and Other Related Dietary Data	DR
5702	Model Gram	MG
5703	Nutrient Composition	NC
5704	24-Hour Recall, Specific Food Item	DR

F. Hispanic Health and Nutrition Examination Survey (HHANES)

- a. Years: 1982-1984
- b. SAS File Prefix: HH
- c. SAS File Count: 16

Tape Number	Name	Abbreviation
6501	Body Measurements	BM
6502	Hearing	AU
6504	Gallbladder Ultrasound Data	GB
6505	Dental Health	OH
6506	Diabetes and OGTT Data	DI
6507	Vision	VI
6509	Physician's Examination	PE
6511	Blood and Urine Assessments	LB
6521	Adolescent and Adult History Questionnaire	AH
6522	Child History Questionnaire	CH
6523	Measures of Depression	MH
6525	Dietary Practices, Food Frequency and Total Nutrient Intake	DB
6526	24-Hour Recall	DR
6533	Alcohol Consumption Data	AL
6540	Measurement and Interpretation of Electrocardiograms	EK
6543	Drug Abuse	DU