From: Trond Petersen
To: Course Participants
RE: Codebook for data for Homework Assignments

The data files for computer exercises

Below follows the description of a data file made available for the course.

The name of the data file is HIRE771.DTA.

It is given as STATA file.

There are 3,990 records of data, corresponding to 3,990 individuals. For more information, see below.

Some background information on the data:

These data are taken from information on every individual hired into a large U.S. organization in the year 1977. It gives data on every hired person into the organization in that year, a total of 3990 persons.

The organization is hierarchically organized into so-called salary grade levels, from 1 (the lowest) to 20 (the highest). Salary grade levels above 20 pertain to Vice Presidents and Officers of the corporation. The organizational hierarchy is quite typical of large U.S. organizations, in retail trade, banking, insurance, the federal bureaucracy, and also many manufacturing organizations.

Missing codes:

The codes -9, -99, etc. usually mean that the information is missing. This is hence not a valid a code. In those cases it should not be included in the analysis. In some cases the code -9 may mean that the variable is not relevant for the individual. One example of such a variable is BAMAJOR. If the EDUC variable equals 6 (= Bachelor's Degree) or higher, 7, 8, or 9, then the person has taken college or more education and should also have a college major. However, BAMAJOR may be equal to -9, in which case the information on the BA Major is missing. If, however the person has not taken college, then BAMAJOR should also be equal to -9. In that case, the information on BAMAJOR is not missing. The person does just not have a BAMAJOR because the person did not obtain a college degree.

There will also be some codes in the data that are not listed below. Those should be treated as either missing or put in separate category, called OTHER. When you compute means, medians, percentiles, and the like, cases with such values should not be included. When you compute frequencies and other measures relevant for nominal variables the cases may be included.

Variables not yet documented:

The following variables have yet to be documented:

```
EMPLTYPE       (this has been 2/3 documented)
MAMAJOR        (virtually no one has a Master)
MAYEAR
PHDMAJOR       (virtually no one has a PhD)
PHDYEAR
```

Those will be documented at a later point in time.

Format for data:

The relevant format for the ASCII data is this.

There are 24 variables on the file. Each variable is separated by several columns. Each variable is right justified. Each record of data (one record per person) has record length 200.

The Variables on the Data File:

There are 24 variables in the file. They are listed below along with short explanations. Thereafter follows an explanation of codes for some of the variables, that is, those that need additional explanations.

Variable Name: Explanation:

ID   Identification number for individual
SEX   Sex (=0 if male) (=1 if female)
ETHNIC   Ethnicity
BIRTHDAT   Birthdate: year (two digits), month, and day
EDUC   Educational level when started in company
FSTEMPDAT   First employment data: year, month, and day
PRES77   Present in 1977 (=1), not present in 1977 (=0)
PRES78   Present in 1978 (=1), not present in 1978 (=0)
EMPLTYPE   Employee type
HOMEOFF   Home office
MARITAL   Marital status in 1977
SALARY   Weekly salary in 1977: in cents, insert two decimals
SALCODE   Salary code
WAGECODE   Wage hour code
JOBFOC   Job focus
SALGRADE   Salary Grade Level
QUITDATE   Date Company was left: year, month, and date
QUITREAS   Reason Company was left
EDUC77   Educational level in 1977
BAMAJOR   Bachelor's Major
MAMAJOR   Master's Major
MAYEAR   Master's Year
PHDMAJOR   PhD Major
PHDYEAR   PhD Year

Explanation of Codes on Variables:

SEX   Sex

                0= Male
                1= Female

ETHNIC   Ethnicity

                0= White
                1= African-American
                2= Asian
                3= Native American (merge with "0" in analysis)
                4= Hispanic

BIRTHDAT   Birthdate: year (two digits), month, and day

                year (two digits)
                   If birth year is in the range 80-99,
                   it means that the person was born in the
                   period 1880-1899.
                   If birth year is in the range 00-78,
                   it means that the person was born in the
                   period 1900-1978 (noone was born after about 1960)
                month (two digits)
                day (two digits)

EDUC   Education level

Valid codes: 0-9; missing: -9

This is the code for EDUC and EDUC77 (Education level):

0   Less than four years of high school
1   High school graduate (4 years)
2   High school graduate (4 years) plus secretarial
           or business school
3   College courses or certificates, less than 60 credits
4   College courses, 60 or more credits but degree not
           received
5   Junior or community college degree
6   Bachelor's degree
7   Graduate school courses, advanced degree not received
8   Master's degree
9   Doctorate

```
EMPLTYPE                Employee type

Valid codes:  1-11;  missing:  -9

1         Full-time employee, working regular hours
2         Permanent employee, shift worker
3         Temporary employee, with regular hours
4         Working regular hours, less than  full time
5         Junior-senior High School students
6-11      Other employee types with nonregular employment, not to be used


HOMEOFF                 Home office

Valid codes:  1-2;   missing:  -9

1         Home office
2         Field Office (elsewhere than the home office)


MARITAL                 Marital status in 1977

Valid codes:  0-4;  missing:  -9

0         Single
1         Married
2         Divorced
3         Widower
4         Separated


SALCODE                 Salary code

Valid codes:  0-5;  missing:  -9

0         Hourly
1         Annual
2         Weekly
3         Monthly
4         Semi monthly (e.g., on the 1. and 15. of each month)
5         Bi weekly (paid every two weeks, e.g., every second Friday)


SALGRADE                Salary Grade Level

Valid codes:  1-20, 21-22, 50-51;  missing:  -9

1-20      The standard grade level hierarchy
21-22     Officers of the organization (Vice Presidents, etc.)
50-51     Some special grades
```

```
QUITREAS                Reason Company was left

Valid codes:  1-90;   missing:  -9

          Controllable reasons:

01        Higher earnings
03        Better working conditions
04        Greater opportunity
05        Nearer home or better transportation
07        More interesting or suitable work
50        Enter agent's contract

          Uncontrollable reasons:

20        Pregnancy
24        Change of residences
28        Return to school
32        Household duties
36        Health-personal
40        Illness in family
44        Military service leave
48        Other

          Involuntary reasons:

65        Dismissal
70        Position abolished
80        Retirement
90        Death

66      Disappeared from tape. There should be no cases with
        this code on the tape, they should be recoded to -9.


WAGECODE                Wage hour code

Valid codes:  0-3;  missing:  -9

0       Nonexempt (i.e., eligible for overtime pay)
1       Administrative
2       Professional
3       Executive
```

```
JOBFOC                    Job focus

This is the code for variable JOBFOC (Job Focus), describing the primary skill
or experience used in performing a particular job:

Valid codes:  1-49, 50, 51;  missing:  -9

01        Accounting/bookkeeping clerk
02        Coder
03        Correspondence clerk
04        EAM/EDP operator
05        Figure clerk
06        File clerk
07        Input/output clerk
08        Keypunch operator
09        Messenger
10        Other general clerks
11        Other machine operators
12        Preparation assembly clerk
13        Secretary/stenographer
14        Transcriber
15        Typist
16        Statistical typist
17-21     To be used for future clerical additions
22        Accounting
23        Auditing
24        MPA cost analysis
25        Claims analysis/approval
26        Contract/analysis and development
27        Mathematical analysis
28        EDP programming
29        EDP machine operations
30        Systems design
31        Creative writing/editing
32        Employee development and instruction:
          Career development programs
33        Employee development and instruction:
          Job related instruction
34        Sales support
35        Staff administration/technical
36        Corporate relations
37        Employee relations
38        Group patron relations
39        Inspection
40        Policy change
41        Underwriting
42        Investment functions
43        Other
44        Incentive compensation personell
45-49     To be used for future administrative additions
50-51     Applies to jobs where job focus us not applicable,
          i.e. jobs in grades 13 and above; jobs in grades 6
          through 12 not having an administrative wage hour

Codes 1-21 are clerical jobs, while codes 22-49 are adminstrative and
professional jobs.
```

```
BAMAJOR                 Bachelor's Major

This is the code for Major for Bachelor's degree:

-9      missing or no college

1       Accounting
2       Actuarial science
3       Advertising
4       Commercial art
5       Economics
6       Education
7       Engineering
8       English
9       Business administration, management, marketing
10      History
11      Industrial relations/personell
12      Insurance
13      Journalism
14      Law
15      Foreign language
16      Library science
17      Mathematics
18      Medicine/nursing
19      Music
20      Data processing/programming
21      Political science
22      Psychology/sociology
23      Real estate
24      Sciences
25      Secretarial studies
26      Statistics
27      Other
```

[FILE: BAS_CODEBOOK_HIRE771.DOC (in directory \BAS\DAT\]