Data Sets

1. DebTrivedi data

The dataset “DebTrivedi” were obtained from the US National Medical Expenditure Survey (NMES) for 1987/88, including 4406 individuals, aged 66 and over, who are covered by Medicare, a public insurance program. These data were analyzed by Deb and Trivedi (1997) with the original objective to model the demand for medical care by the elderly—as captured by the number of physician/non-physician office and hospital outpatient visits— based on the independent variables (covariates) available for the individuals.

In our lecture notes for Ch3, we have adopted the number of physician office visits **ofp** as the dependent variable and used some of the other variables as covariates to illustrate GLMs for count data. There is a total of 19 variables in the data set, including the health status variables hosp (number of hospital stays), health (self-perceived health status), numchron (number of chronic conditions), as well as the socioeconomic variables gender, school (number of years of education), and privins (private insurance indicator) etc.
In your project, you may consider other variables, except **ofp**, as dependent variables based on your study interest.

The full dataset can be found in R package MixAll by
Library(MixAll)
data(DebTrivedi)

Reference:
Deb P, Trivedi PK (1997). Demand for Medical Care by the Elderly: A Finite Mixture Approach. Journal of Applied Econometrics, 12, 313–336.

2. Bank Credit data

When a bank receives a loan application, based on the applicant’s profile the bank has to make a decision regarding whether to go ahead with the loan approval or not. Two types of risks are associated with the bank’s decision:

- If the applicant is a good credit risk, i.e. is likely to repay the loan, then not approving the loan to the person results in a loss of business to the bank
- If the applicant is a bad credit risk, i.e. is not likely to repay the loan, then approving the loan to the person results in a financial loss to the bank

The second risk may be more serious from the bank perspective as the bank (or any other institution lending the money to an untrustworthy party) had a higher chance of not being paid back the borrowed amount. So it is important for the bank or other lending authority to evaluate the risks associated with lending money to a customer.

Suppose that you are given this data set by the bank and required to address the risk problems by using the applicant's demographic and socio-economic profiles of customers.

A bank credit dataset "**germancredit.csv**" contains information about 1000 loan applicants who defaulted or did not default on their loans. The dataset includes total 21 variables, such as their account balance, credit amount, age, occupation, loan records, etc.

The dataset contains 21 variables

Variable	**Description**	**Codes/Values**
default	default on loan	1=default on loan, 0=not default
checkingstatus	Status of existing checking account	A11 : ... < 0 DM A12 : 0 <= ... < 200 DM A13 : ... >= 200 DM / salary assignments for at least 1 year A14 : no checking account
duration	Duration in month	
history	Credit history	A30 : no credits taken/ all credits paid back duly A31 : all credits at this bank paid back duly A32 : existing credits paid back till now A33 : delay in paying off in the past A34 : critical account/ other credits existing (not at this bank)
purpose	Purpose of loan	A40 : car (new) A41 : car (used) A42 : furniture/equipment A43 : radio/television A44 : domestic appliances A45 : repairs A46 : education A47 : (vacation-does not exist?) A48 : retraining A49 : business A410 : others
amount	Credit amount	Numerical
savings	Savings account/bonds	A61 : ... < 100 DM A62 : 100 <= ... < 500 DM A63 : 500 <= ... < 1000 DM A64 : .. >= 1000 DM A65 : unknown/no savings acc
employ	Present employment since	A71 : unemployed A72 : ... < 1 year A73 : 1 <= ... < 4 years A74 : 4 <= ... < 7 years A75 : .. >= 7 years
installment	Installment rate in percentage of disposable income	Numerical

status	Personal status and sex	A91 : male : divorced/separated A92 : female : divorced/separated/married A93 : male : single A94 : male : married/widowed A95 : female : single
others	Other debtors / guarantors	A101 : none A102 : co-applicant A103 : guarantor
residence	Present residence since	Numerical
property	Property	A121 : real estate A122 : if not A121 : building society savings agreement/ life insurance A123 : if not A121/A122 : car or other, not in attributee 6 A124 : unknown / no property
age	Age in years	Numerical
otherplans	Other installment plans	A141 : bank A142 : stores A143 : none
housing	Housing	A151 : rent A152 : own A153 : for free
cards	Number of existing credits at this bank	Numerical
job	Job	A171 : unemployed/unskilled-non-resident A172 : unskilled - resident A173 : skilled employee / official A174 : management/ self-employed/ highly qualified employee/ officer
liable	Number of people being liable to provide maintenance for	numerical
tele	Telephone	A191 : none A192 : yes, registered under the customers name
foreign	Foreign worker	A201 : yes A202 : no