

HGB DataSet

Attribute	Definition
ID	Unique identifier for each subject
group	Water Consumption Group 1 = Tap Water Only 2 = Bottled/Filtered Water Only 3 = Combination of Tap and Bottled/Filtered
hgb9	Week 9 Hemoglobin (g/dL)
prenatal	Adequate Prenatal Care 1=Yes 0=No
psmoke	Pre-Pregnancy Smoker 1=Yes 0=No
age	Age at initial visit
wtgain	Weight gain throughout pregnancy
parity	Number of previous births 0 = None 1 = One 2 = Two 3= Three or more
ed	Educational Attainment 0 = Less than HS 1 = HS/GED 2 = College or more
change	Change in hemoglobin (hgb9-hgb36)
hgb36	Week 36 Hemoglobin (g/dL)
income	Annual Household Income
water	Amount of Tap Water Consumed (only applicable for those in Groups 1 and 3)

Nets.csv DataSet

Attribute	Definition
rural	Rural Household 1 = Rural 0 = Urban
age	Age of head of household
famsize	Size of the family (including head of household)
hcdist	Distance from household to healthcare facility 1 = Within 15 miles 2 = Between 15 and 50 miles 3 = Greater than 50 miles
child	Household has children under the age of 5 years 1 = Yes 0 = No
wealth	Wealth index of household 1 = Lowest 25% 2 = Next lowest 3 = Next highest 4 = Highest 25%
net	Used an ITN the previous night 1 = Yes 0 = No

Bank.csv DataSet

age: Age of the client. Numeric.

job: Client's job. Categorical. One of "admin.", "unknown", "unemployed", "management", "housemaid", "entrepreneur", "student", "blue-collar", "self-employed", "retired", "technician", "services"

marital: Marital status. Categorical. One of "married", "divorced", "single"

education: Education level. Categorical. One of "unknown", "secondary", "primary", "tertiary"

default: Whether the client defaulted on credit. Binary. "yes", "no"

balance: Average yearly balance. Numeric

housing: Whether the client has a mortgage. Binary. "yes", "no"

loan: Whether the client has a personal loan. Binary. "yes", "no"

contact: Type of last contact communication. Categorical. One of "unknown", "telephone", "cellular"

duration: Duration of the last contact, in seconds. Numeric.

campaign: Number of contacts performed during the campaign. Numeric.

y: Whether the client has subscribed to a term deposit. Binary. "yes", "no"

Housing DataSet

Crime: Per capita crime rate (numerical)

Zone: Proportion of residential land zone (numerical)

Ind: Proportion of industrial land zone (numerical)

Rooms: Average number of rooms per house (numerical)

<1940: Proportion of houses built prior to 1940 (numerical)

Distance: Distance to the employment centers (numerical)

Highway: Distance to highways (numerical)

Tax: Property tax rate per \$10,000 (numerical)

Teacher: Student-teacher ratio (numerical)

Median: Median value of the home in this suburb (nominal: VH, H, L, VL)

Contraceptives DataSet

W_age: Wife's age (numerical)

W_ed: Wife's education (categorical) 1=low, 2, 3, 4=high

H_ed: Husband's education (categorical) 1=low, 2, 3, 4=high

Children: Number of children ever born (numerical)

W_work: Wife's now working? (binary) 0=Yes, 1=No

SOL: Standard-of-living index (categorical) 1=low, 2, 3, 4=high

Media: Media exposure (binary) 0=Good, 1=Not good

Contra: Contraceptive method used (class attribute) 1=No-use, 2=Long-term, 3=Short-term

Workers DataSet

Education: Number of years of education.

South: Indicator variable for Southern region (1=Person lives in South, 0=Person lives elsewhere).

Sex: Indicator variable for gender (1=Female, 0=Male).

Experience: Number of years of work experience.

Wage: Wage (dollars per hour).

Age: Age (years).

Race: Race (1=Other, 2=Hispanic, 3=White).

Occupation: Occupational category (1=Management, 2=Sales, 3=Clerical, 4=Service, 5=Professional, 6=Other).

Sector: Sector (0=Other, 1=Manufacturing, 2=Construction).

Marr: Marital Status (0=Unmarried, 1=Married)

Union: Indicator variable for union membership (1=Union member, 0=Not union member).