**Housing Values in Suburbs of Boston**

**Description:** The Boston data frame has 14 columns, and refers to the information about different districts (or suburbs) of Boston, USA using data collected in 1970.

**Format:** This data frame contains the following columns:

crim per capita crime rate by suburb.

zn proportion of residential land zoned for lots over 25,000 sq.ft.

indus proportion of non-retail business acres per suburb.

chas Charles River dummy variable (= 1 if suburb bounds river; 0 otherwise).

nox nitrogen oxides concentration (parts per 10 million).

rm average number of rooms per dwelling.

age proportion of owner-occupied units in the suburb built prior to 1940.

dis weighted mean of distances to five Boston employment centres.

rad index of accessibility to radial highways.

tax full-value property-tax rate per $10,000.

ptratio pupil-teacher ratio by suburb.

black uses formula *1000(Bk - 0.63)^2* where *Bk* is the proportion of blacks by suburb.

lstat lower status of the population (percent).

medv median value of owner-occupied homes in $1000s.

**Source**

Harrison, D. and Rubinfeld, D.L. (1978) Hedonic prices and the demand for clean air. *J. Environ. Economics and Management* **5**, 81–102.

Belsley D.A., Kuh, E. and Welsch, R.E. (1980) *Regression Diagnostics. Identifying Influential Data and Sources of Collinearity.* New York: Wiley.

TASK: Predict **medv** using the other variables as predictors

**Credit Card Balance Data**

Description**:** A simulated data set containing information on bank customers.

Format

A data frame with the following 4 variables.

ID Identification

Income Income in $10,000's

Limit Credit limit

Rating Credit rating

Cards Number of credit cards

Age Age in years

Education number of years of education

Gender factor with levels Male and Female

Student factor with levels No and Yes indicating whether individual was a student

Married factor with levels No and Yes indicating whether individual was married

Ethnicity factor with levels African American, Asian, and Caucasian

Balance Average credit card balance in $.

Source

Simulated data.

References

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013) An Introduction to Statistical Learning with applications in R, [www.StatLearning.com](http://127.0.0.1:27858/help/library/ISLR/html/www.StatLearning.com), Springer-Verlag, New York

TASK: Predict **Balance** using the other variables as predictors