

Data sets description for Homework #1

AirPollution.csv: Air pollution data in 41 US cities.

- SO2: sulphur dioxide content of air in micrograms per cubic metre. This is a measure of the air pollution of the city.
- Temp: average annual temperature in Fahrenheit.
- Manuf: number of manufacturing enterprises employing 20 or more workers.
- Pop: population size (1970 census); in thousands.
- Wind: average annual wind speed in miles per hour.
- Precip: average annual precipitation in inches (1 inch = 2.54 cm).
- Days: average number of days with precipitation per year.

```
> data4=read.csv("data4.csv",header=TRUE,sep=",")
> head(data4)
```

TrackRecords.txt: National track records for men in 55 countries. Variables are, in order, the country and different distances from shortest to longest:

- Country
- 100 m (in seconds)
- 200 m (in seconds)
- 400 m (in seconds)
- 800 m (in minutes)
- 1500 m (in minutes)
- 5000 m (in minutes)
- 10000 m (in minutes)
- Maratón (in minutes)

```
> data8=read.table("data8.DAT")
> head(data8)
> colnames(Grupo8)=c("country", "100m(s)", "200m(s)", "400m(s)", "800m(m)",
+ "1500m(m)", "5000m(m)", "10000(m)", "Marat(m)")
```

Business.txt The dataset consists of 100 observations on a number of variables and is an example of a segmentation study for a business to business situation. It was obtained from a large company which is an industrial supplier. Each of these variables was measured on a graphic rating scale, where a ten-centimeter line was drawn between the endpoints, labeled *Poor* = 0 and *Excellent* = 10.

- Delivery speed: amount of time it takes to deliver the product once an order has been confirmed.
- Price level: perceived level of price charged by product suppliers.
- Price flexibility: perceived willingness of the company representatives to negotiate price on all types of purchases.
- Manufacturer's image: overall image of the manufacturer/supplier.
- Service: overall level of service necessary for maintaining a satisfactory relationship between supplier and purchaser.
- Salesforce's image: overall image of the manufacturer's sales force
- Product quality: perceived level of quality of a particular product (e.g., performance or yield).
- Size of firm: size of the firm relative to others in this market. This variable has two categories: 1 =large, and 0 =small.
- Usage level: how much of the firm's total product is purchased from this company, measured on a 100-point percentage scale, ranging from 0 to 100 percent.
- Satisfaction level: how satisfied the purchaser is with past purchases from the firm.
- Structure of procurement: method of procuring/purchasing products within a particular company. This variable has two categories: 1 =centralized procurement, and 0 =decentralized procurement.
- Type of industry: industry classification in which a product purchaser belongs. This variable has two categories: 1 =*industry A classification*, and 0 =*other industries*.
- Type of buying situation: type of situation facing the purchaser. This variable has three categories: 1 =*new task*, 2 =*modified rebuy* (other products), and 3 =*straight rebuy* (the same products).

```
> business=read.csv("data7.csv", header=TRUE, sep=",")
```

Cars.csv Information on different variables on a set of different car models.

- mpg: miles per gallon, represents fuel efficiency.
- cylinders: number of cylinders in engine.

- engine displacement: volume swept by all pistons inside the cylinders. It is an estimate of the maximum power that the engine can produce.
- horsepower: unit of measurement of power.
- weight: car weight.
- acceleration: time to reach 100 *Km/h* (60 mph), in seconds.
- model year: year the model was released.
- origin: origin of the car, 1 = *US*, 2 = *EU*, 3 = *Japan*.
- car name: unique identifier consisting of the manufacturer model name of the car.

rec.rds Residential energy consumption: survey conducted every 2-3 years. Many variables, the number reduced to 21 here. Some of them are:

- division: US census division.
- typehuq: Type of housing unit
- cd65: number of cooling degree days (outside air temperature was lower than a specific "base temperature")
- hd65: number of heating degree days (outside air temperature was higher than a specific "base temperature")
- cost, coolcost, heatcost: energy spend (total, cooling, heating)
- moneypy: gross household income per year
- recbath: hot tub used
- othrooms: number of rooms other than bedrooms and bathrooms
- ncombath: number of full bathrooms
- wheatsiz: main water heater size (if storage tank)
- numfrig: number of refrigerators used
- totsqft, tothsqft: Total and total heated sqft

```
>rec=readRDS("rec.rds")
```