

Number	Variable Name	Description
1	Identification number	1–43
2	Number of orders	Number of heating equipment orders during month
3	Interest rate	Prime rate in effect during month
4	New homes	Number of new homes completed and for sale in sales region during month
5	Discount	Percent discount (0–5) offered to distributors during month; value is usually 0, indicating no discount
6	Inventories	Distributor inventories in warehouses during month
7	Sell through	Number of units sold by distributor to contractors in previous month
8	Temperature deviation	Difference between average temperature for month and 30-year average for that month
9	Year	1999, 2000, 2001, or 2002
10	Month	Coded 1–12

1	2	3	4	5	6	7	8	9	10
1	121	0.0750	64	0	3536	615	2.22	1999	1
2	227	0.0750	64	0	3042	813	0.28	1999	2
3	446	0.0750	65	0	2456	704	0.79	1999	3
...	...	...	...	...	...	...	...	...	...
41	754	0.0475	64	0	1417	927	0.81	2002	6
42	1098	0.0475	65	0	1244	877	0.28	2002	7
43	1158	0.0475	65	0	1465	809	0.50	2002	8

## Data Set C.9 Ischemic Heart Disease

A health insurance company collected information on 788 of its subscribers who had made claims resulting from ischemic (coronary) heart disease. Data were obtained on total costs of services provided for these 788 subscribers and the nature of the various services for the period of January 1, 1998 through December 31, 1999. Each line in the data set has an identification number and provides information on 9 other variables for each subscriber. The 10 variables are:

Variable Number	Variable Name	Description
1	Identification number	1–788
2	Total cost	Total cost of claims by subscriber (dollars)
3	Age	Age of subscriber (years)
4	Gender	Gender of subscriber: 1 if male; 0 otherwise
5	Interventions	Total number of interventions or procedures carried out
6	Drugs	Number of tracked drugs prescribed
7	Emergency room visits	Number of emergency room visits
8	Complications	Number of other complications that arose during heart disease treatment
9	Comorbidities	Number of other diseases that the subscriber had during period
10	Duration	Number of days of duration of treatment condition

1	179.1	63	0	2	1
2	319.0	59	0	2	0
3	9310.7	62	0	17	0
...	...	...	...	...	...
786	2677.7	68	0	3	2
787	1282.2	58	0	7	2
788	586.0	56	0	4	4

## Data Set C.10 Disease Outbreak

This data set provides information from a study based on a random sample within two sectors in a city. Each line of the data set has an identification number and provides information on 5 other variables for a single subscriber.

Variable Number	Variable Name	Description
1	Identification number	1–196
2	Age	Age of person in years
3	Socioeconomic status	1 = upper middle class; 2 = lower middle class
4	Sector	Sector where person lives
5	Disease status	1 = with disease; 0 = without disease
6	Savings account status	1 = has savings account; 0 = no savings account

Adapted in part from H. G. Dantes, J. S. Koopman, C. L. Addy, et al., "Dengue Epidemic in Singapore," *Journal of Epidemiology* 17 (1988), pp. 178–86.

1	2	3	4
1	33	1	1
2	35	1	1
3	6	1	1
...	...	...	...
194	31	3	1
195	85	3	1
196	24	2	1

## Data Set C.11 IPO

Private companies often go public by issuing shares in an initial public offering (IPO). A study of 482 IPOs was conducted to determine characteristics of companies that attract venture capital funding. For each company, the following information was recorded: whether or not a company was financed with venture capital funding, the face value of the company, the number of shares offered, and the number of shares sold.