

MS4S09 - Tutorial

For each of the following examples, perform an exploratory data analysis and estimate linear, quadratic and cubic trends.

If the data comes from the TSA package (remember to install it), then use the command `data()` to load it.

For example, the command

```
data(milk)
```

loads the dataset for example 2.

Example 1: Global warming

Global warming refers to an increase in the average temperature of the Earth near-surface air and oceans since the mid-20th century and its projected continuation. The data are annual temperature deviations (1856-1997) in deg C, measured from a baseline average.

- Data file: globaltemps.txt
- There are $n = 142$ observations.
- Measurements are taken each year.
- What are the noticeable patterns?
- Predictions?

Example 2: Milk production data

Commercial dairy farming produces the vast majority of milk in the United States. The data are the monthly U.S. milk production (in millions of pounds) from January, 1994 to December, 2005.

- Data file: milk (TSA)
- There are $n = 144$ observations.
- Measurements are taken each month.
- What are the noticeable patterns?
- Predictions?

Example 3: CREF stock data

TIAA-CREF is the leading provider of retirement accounts and products to employees in academic, research, medical, and cultural institutions. The data are daily values of one unit of the CREF (College Retirement Equity Fund) stock fund from 8/26/04 to 8/15/06.

- Data file: CREF (TSA)
- There are $n = 501$ observations.
- Measurements are taken each trading day.
- What are the noticeable patterns?

- Predictions?

Example 4: Homerun data

The Boston Red Sox are a professional baseball team based in Boston, Massachusetts, and a member of the Major League Baseball American League Eastern Division. The data are the number of homeruns hit by the team each year from 1909 to 2010. Source: Ted Hornback (Spring, 2010).

- Data file: homeruns
- There are $n = 102$ observations.
- Measurements are taken each year.
- What are the noticeable patterns?
- Predictions?

Example 5: Earthquake data

An earthquake occurs when there is a sudden release of energy in the Earth crust. Earthquakes are caused mostly by rupture of geological faults, but also by other events such as volcanic activity, landslides, mine blasts, and nuclear tests. The data are the number of global earthquakes annually (with intensities of 7.0 or greater) during 1900-1998. Source: Craig Whitlow (Spring, 2010).

- Data file: earthquake
- There are $n = 99$ observations.
- Measurements are taken each year.
- What are the noticeable patterns?
- Predictions?

Example 6: Enrollment data

The data are the annual fall enrollment counts for USC (Columbia campus only, 1954-2010). The data were obtained from the USC website <http://www.ipr.sc.edu/enrollment/>, which contains the enrollment counts for all campuses in the USC system.

- Data file: enrollment
- There are $n = 57$ observations.
- Measurements are taken each year.
- What are the noticeable patterns?
- Predictions?

Example 7: Star brightness data.

Two factors determine the brightness of a star: its luminosity (how much energy it puts out in a given time) and its distance from the Earth. The data are nightly brightness measurements (in magnitude) of a single star over a period of 600 nights.

- Data file: star (TSA)
- There are $n = 600$ observations.
- Measurements are taken each night.
- What are the noticeable patterns?

- Predictions?

Example 8: Airline mile data

The Bureau of Transportation Statistics publishes monthly passenger traffic data reflecting 100 percent of scheduled operations for airlines in the United States. The data are monthly U.S. airline passenger miles traveled from 1/1996 to 5/2005.

- Data file: airmiles (TSA)
- There are $n = 113$ observations.
- Measurements are taken each month.
- What are the noticeable patterns?
- Predictions?