

Sample and Populations

The **population** is the set of subjects in which we are interested.

The **sample** is the subset of the population for whom we have (or plan to have) data.

Examples:

- General Social Survey (GSS): sample 4510, population all adult Americans (>200 million)
- California exit poll: sample 2705, population all CA voters (\approx 7 million)

A **census** is a complete enumeration.

Descriptive and Inferential Statistics

Descriptive statistics refers to methods for summarizing the data. The summaries usually consist of graphs and numbers such as averages and percentages.

Inferential statistics refers to methods for making predictions or decisions about a population, based on data obtained from a sample of that population.

Example — confidence intervals. Of 834 people in a survey, 54.0% say they favor handgun control. Using statistical methods, we deduce there is *95% confidence* that the true proportion is between 50.6% and 57.4%.

Parameters and Statistics

A **parameter** is a numerical summary of the population.

A **statistic** is a numerical summary of a sample taken from the population.

Randomness and Variability

- Opinion poll — *randomly* choose the sample
- Clinical trial — *randomly* assign participants to treatment or control

Samples vary because *people* do (or if it's not people we're sampling, whatever it is we are)

Computers and Statistics

- Calculators
- Minitab, SPSS
- Excel

**EXPLORING DATA WITH
GRAPHS AND NUMERICAL
SUMMARIES**

A **variable** is any characteristic that is recorded for subjects in a study.

A variable is called **categorical** if each observation belongs to one of a set of categories.

A variable is called **quantitative** if observations on it take numerical values that represent different magnitudes of the variable.

A quantitative variable is called **discrete** if its possible values form a set of separate numbers, such as 0,1,2,3,.....,

A quantitative variable is **continuous** if its possible values form an interval.

In practice, the method of data analysis depends on the type of variable.

Frequency Tables

Each observation falls in one of several categories.

A frequency table is a table that summarizes the number of observations in each category.

The category with the highest frequency is called the **mode**.

Students in last year's class (who declared a political allegiance):

Party	Frequency	Proportion	Percentage
Democrat	35	0.515	51.5%
Independent	12	0.176	17.6%
Republican	21	0.309	30.9%
Total	68	1.000	100%

The **proportion** of observations that fall in a certain category is the frequency (count) of observations in that category divided by the total number of observations. The **percentage** is the proportion multiplied by 100. Proportions and percentages are also called **relative frequencies**.

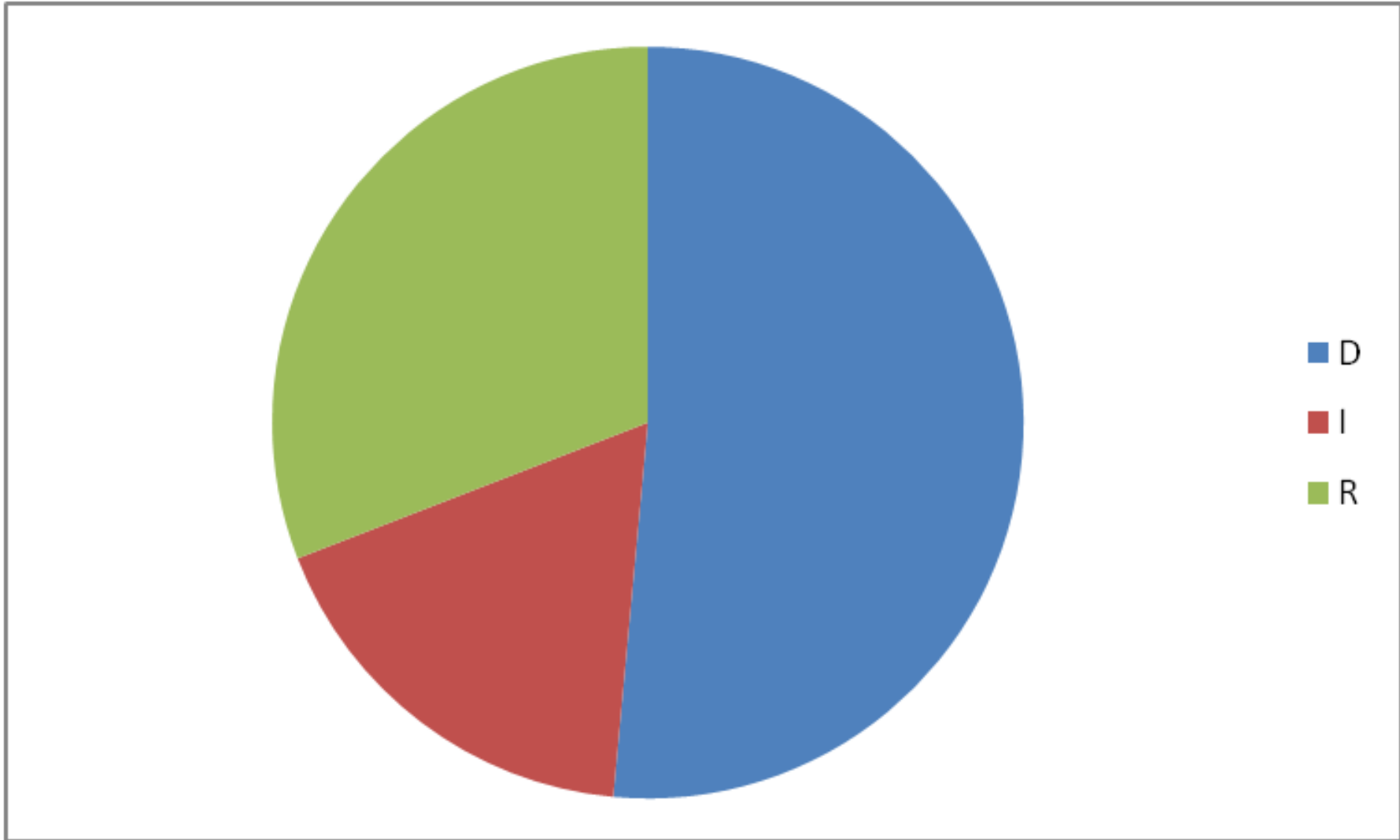
Describing Data Using Graphical Summaries

1. Categorical Data
2. Quantitative Data

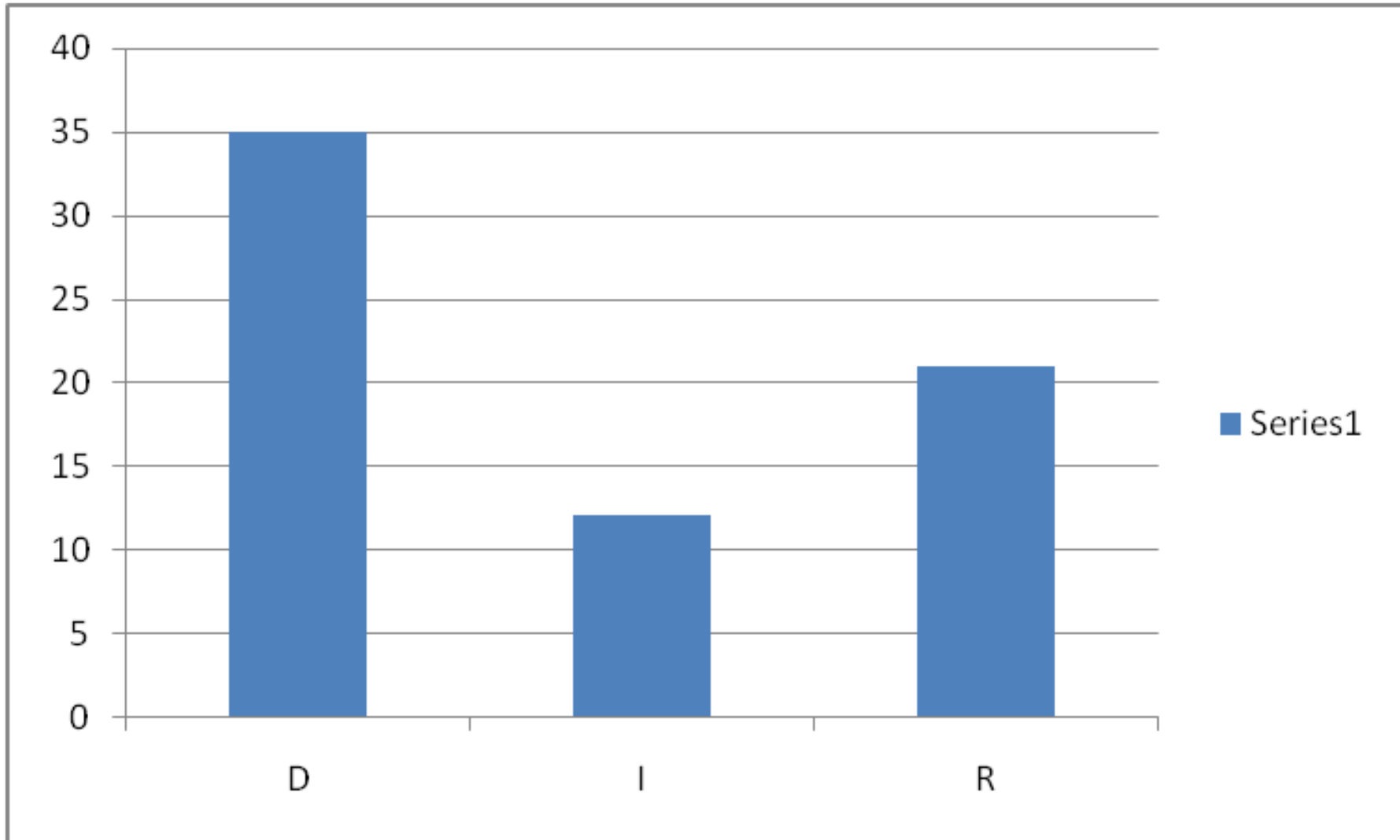
Graphs for Categorical Variables

- Pie charts
- Bar graphs

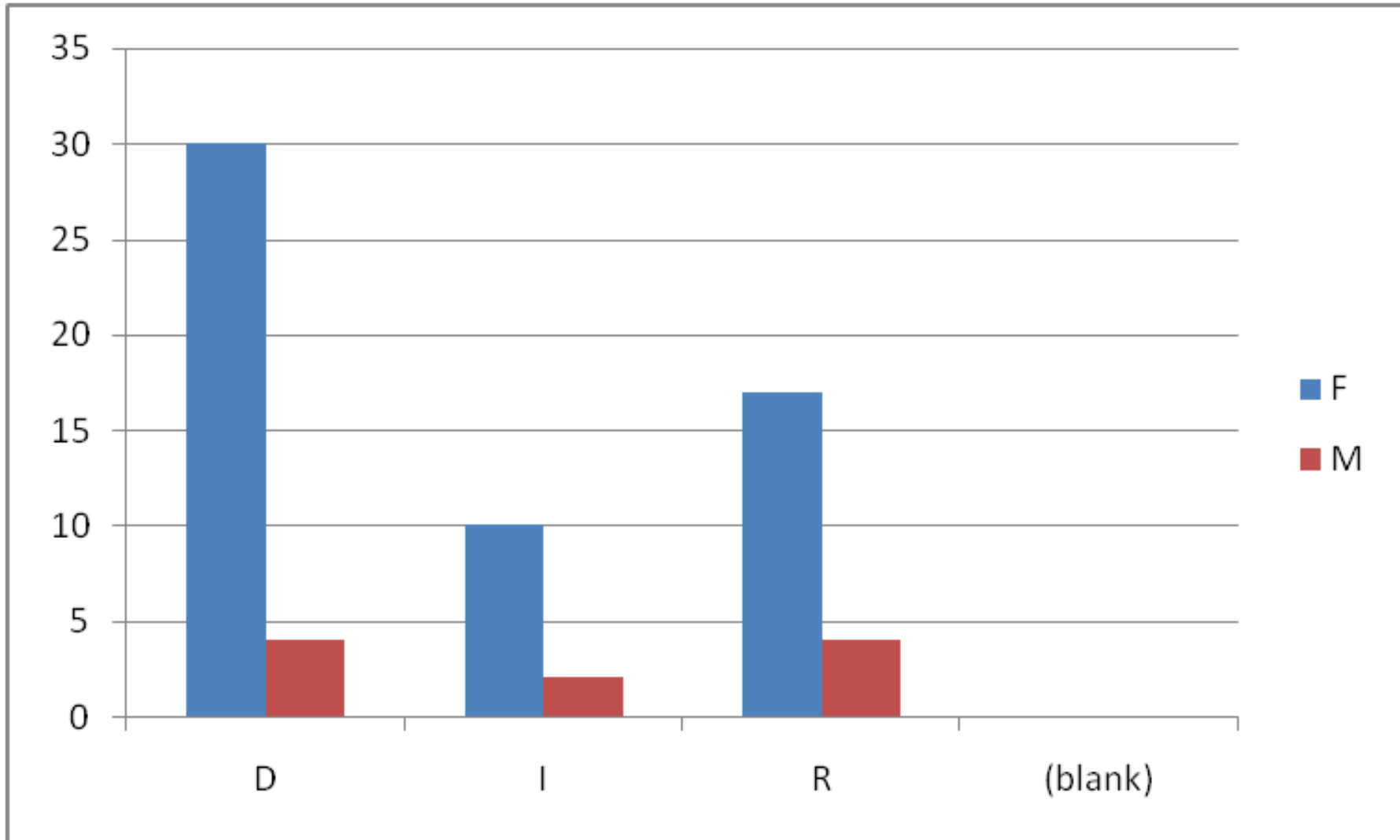
Pie chart



Bar graph



Segmented bar graph



Segmented bar graph expressed as percentages

