Figure 4.6. DATA PRESENTATION AND ANALYSIS: A Numerical Description of Data Distributions

Program 3 in: *Against All Odds: Inside Statistics*

This program shows how numerical measures can give compact descriptions of specific aspects of a distribution. The *median* and the *average* are measures of the centre or location of a distribution. The median is the midpoint of the observations; the average is also called the *mean*. You will learn how these measures differ, in particular that the median is more *resistant* than the average to the influence of extreme observations (or outliers). The *quartiles*, which mark the points 25% and 75% up the ordered list of observations, are used with the median to measure spread. The *interquartile range* or *IQR* (distance between the quartiles) is a single number that measures spread but, as such, it is useful primarily as a description of symmetic distributions. The *five-number summary* gives a fuller description of a distribution; it consists of the median, the two quartiles, and the two extremes (smallest and largest values). The five-number summary leads to a new graphical display for distributions, the *boxplot*. Finally, you will learn that when the average measures centre, spread is usually measured by the *standard deviation*.

The video begins with an examination of some differences between the distributions of incomes for male and female workers in the United States. These distributions are strongly *skewed*. The medians show the incomes of typical male and female workers, providing a numerical comparison. The explanation of the income gap between men and women is controversial.

A comparison of the number of calories in different brands of beef, meat, and poultry hot dogs illustrates the use of five-number summaries and boxplots to compare distributions. The data for this example appear in Table 1.4 on page 29 of the text (see below). Then the video looks at how researchers are using musical tones to help scan the results of medical blood tests – an off-key note alerts the technician that a value falls outside the normal range. The normal range is determined by using the standard deviation to measure the spread about the average as centre.

Text Table 1.4: Calories and sodium in hot dogs by type

Beef		Meat		Poultry	
Calories	Sodium	Calories	Sodium	Calories	Sodium
186	495	173	458	129	430
181	477	191	506	132	375
176	425	182	473	102	396
149	322	190	545	106	383
184	482	172	496	94	387
190	587	147	360	102	542
158	370	146	387	87	359
139	322	139	386	99	357
175	479	175	507	170	528
148	375	136	393	113	513
152	330	179	405	135	426
111	300	153	372	142	513
141	386	107	144	86	358
153	401	195	511	143	581
190	645	135	405	152	588
157	440	140	428	146	522
131	317	138	339	144	545
149	319				
135	298				
132	253				

SOURCE: *U. S. Consumer Reports*, June, 1986, pp. 366-367.

Blank page