Comparing Distributions

Some of the most interesting statistics questions involve comparing two or more groups. Which of two popular diets leads to greater long-term weight loss? Who texts more—males or females? Does the number of people living in a household differ among countries? As the following example suggests, you should always discuss shape, center, spread, and possible outliers whenever you compare distributions of a quantitative variable.

Household Size: U.K. versus South Africa

EXAMPLE

Comparing distributions

How do the numbers of people living in households in the United Kingdom (U.K.) and South Africa compare? To help answer this question, we used CensusAtSchool's "Random Data Selector" to choose 50 students from each country. **Figure 1.12** is a dotplot of the household sizes reported by the survey respondents.

PROBLEM: Compare the distributions of household size for these two countries.

SOLUTION: Don't forget your SOCS! Shape: The distribution of household size for the U.K. sample is roughly symmetric and unimodal, while the distribution for the South Africa sample is skewed to the right and unimodal. Center: Household sizes for the South African students tended to be larger than for the U.K. students. The median household sizes for the two groups are 6 people and 4 people, respectively. Spread: There is more variability (greater spread) in the household sizes for the South African students than for the U.K. students. The range for the South African data is 26 − 3 = 23 people, while the range for the U.K. data is 6 − 2 = 4 people. Outliers: There don't appear to be any potential outliers in the U.K. distribution. The South African distribution has two potential outliers in the right tail of the distribution—students who reported living in households with 15 and 26 people. (The U.K. households with 2 people actually *will be* classified as outliers when we introduce a procedure in the next section.)
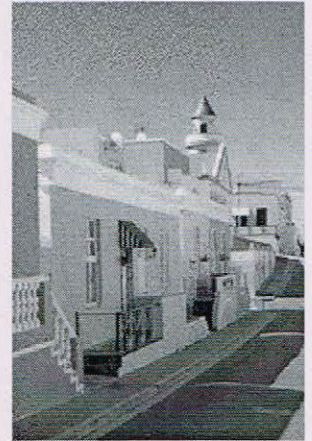
AP EXAM TIP When comparing distributions of quantitative data, it's not enough just to list values for the center and spread of each distribution. You have to explicitly *compare* these values, using words like "greater than," "less than," or "about the same as."
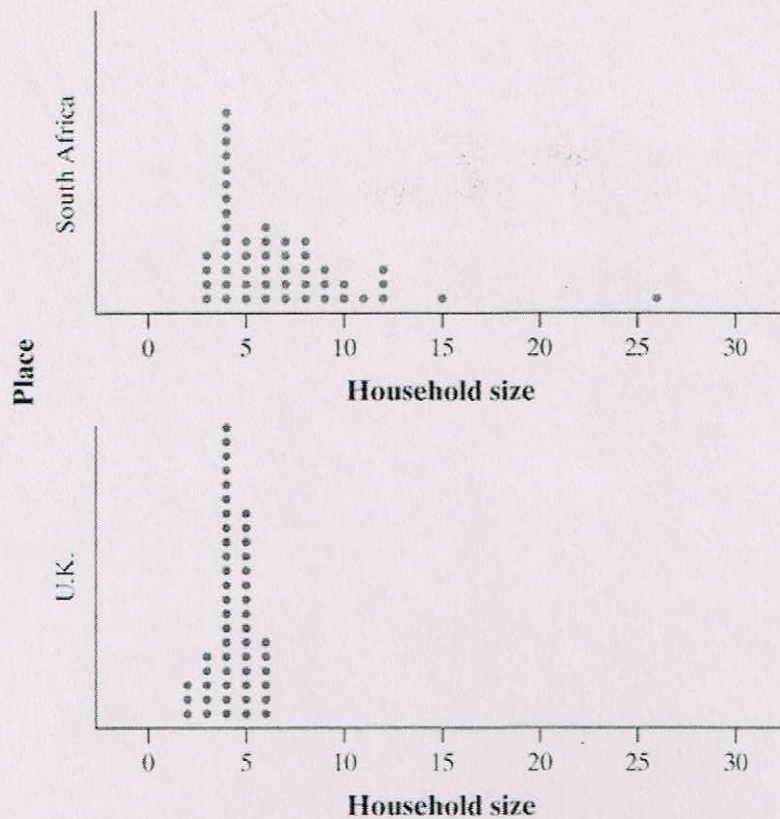
Figure 1.12 Dotplot of household size for random samples of 50 students from the United Kingdom and South Africa.

For Practice *Try Exercise 43*

Notice that we discussed the distributions of household size only for the two *samples* of 50 students in the previous example. We might be interested in whether the sample data give us convincing evidence of a difference in the *population* distributions of household size for South Africa and the United Kingdom. We'll have to wait a few chapters to decide whether we can reach such a conclusion, but our ability to make such an inference later will be helped by the fact that the students in our samples were chosen at random.

◀ **Comparing Distributions** ▶