

Chapter 6

47. Philadelphia Housing These data describe housing prices in the Philadelphia area. Each of the 110 rows of this data table describes a region of the metropolitan area. (Several make up the city of Philadelphia.) One column, labeled Selling Price, gives the median price for homes sold in that area during 1999 in thousands of dollars. Another, labeled Crime Rate, gives the number of crimes committed in that area, per 100,000 residents.

- (a) Make a scatterplot of the selling price on the crime rate. Which observation stands out from the others? Is this outlier unusual in terms of either marginal distribution?
- (b) Find the correlation using all of the data as shown in the prior scatterplot.
- (c) Exclude the distinct outlier and redraw the scatterplot focused on the rest of the data. Does your impression of the relationship between the crime rate and selling price change?
- (d) Compute the correlation without the outlier. Does it change much?
- (e) Can we conclude from the correlation that crimes in the Philadelphia area cause a rise or fall in the value of real estate?

49. Cash Counts There's a special situation in which you can measure dependence using either the correlation or Cramer's V (Chapter 5). Suppose both variables indicate group membership. One variable might distinguish male from female customers, and the other whether the customer uses a credit card, as in this example. These variables define a contingency table for a sample of 150 customers at a department store.

	Sex		Total
	Male	Female	
Pay with cash	50	10	60
Use a credit card	55	35	90
Total	105	45	150

- (a) Compute the value of Cramer's V for this table. You may need to look back at Chapter 5.
- (b) Find the correlation between the numerical variables *Sex* (coded as 1 for male, 0 for female) and *Cash* (coded 1 for those paying with cash, 0 otherwise). Use the counts in the table to manufacture these two columns.
- (c) What's the relationship between Cramer's V and the correlation?