

Unit 13: Two-Way Tables



SUMMARY OF VIDEO

This video deals with analysis of categorical variables (for example, gender, race, occupation) and relationships between categorical variables. The context is a Happiness Survey that was part of Somerville, Massachusetts' 2011 annual census. The video focuses on two of the survey questions, one that asks respondents to rate their current level of happiness and the other that asks them to rate the beauty of Somerville. Happiness ratings are boiled down into three categories: Unhappy, So-So, and Happy. Ratings of Somerville's physical beauty are categorized as Bad, OK, and Good. Results from these two questions are organized into a two-way table with Happiness as the row variable and Physical Beauty as the column variable (see Table 13.1). The marginal totals (bottom row and right-most column) have been added to the two-way table.

		Physical Beauty			Total
		Bad	OK	Good	
Happiness	Unhappy	90	123	62	275
	So-so	555	972	610	2137
	Happy	541	1426	1406	3373
Total		1186	2521	2078	5785

Table 13.1. Results from rating happiness and Somerville's physical beauty.

Notice that 5785 Somerville residents answered both of these questions. (The table only accounts for respondents who have answered both questions.) First, look at the distribution of each variable separately – this is called a marginal distribution. Computations of the marginal distributions of the two variables appear in Tables 13.2 and 13.3. From the marginal distributions we find that slightly more than 58% of respondents reported they were Happy and around 36% of the respondents rated Somerville's physical beauty as Good.

See tables on next page...

		Marginal Distribution
Happiness	Unhappy	$275/5785 \times 100\% \approx 4.75\%$
	So-so	$2137/5785 \times 100\% \approx 36.94\%$
	Happy	$3373/5785 \times 100\% \approx 58.31\%$

Table 13.2. Marginal distribution of Happiness.

		Physical Beauty		
		Bad	OK	Good
Marginal Distribution		$1186/5785 \times 100\% \approx 20.50\%$	$2521/5785 \times 100\% \approx 43.58\%$	$2078/5785 \times 100\% \approx 35.92\%$

Table 13.3. Marginal distribution of Physical Beauty.

Next, we dig even deeper into the two-way table's data by computing conditional distributions, distributions of one variable restricted to a single outcome of another variable. For example, we can investigate how just the Unhappy people rated Somerville's beauty. In this case, we are looking at the distribution of beauty ratings just within the Unhappy group (275 respondents). Here are the calculations:

$$\text{Bad: } 90/275 \times 100\% \approx 32.73\%$$

$$\text{OK: } 123/275 \times 100\% \approx 44.73\%$$

$$\text{Good: } 62/275 \times 100\% \approx 22.55\%$$

Table 13.4 shows the conditional distribution of Physical Beauty for each category of Happiness.

		Physical Beauty			Total
		Bad	OK	Good	
Happiness	Unhappy	32.73%	44.73%	22.55%	100%
	So-so	25.97%	45.48%	28.54%	100%
	Happy	16.04%	42.28%	41.68%	100%

Table 13.4. Conditional distribution of Physical Beauty for each Happiness category.

Notice that only 22.55% of Unhappy people rated Somerville's beauty as Good compared to 41.68% of the Happy people – clearly there is a connection between the Happiness and Physical Beauty variables. The graphic display in Figure 13.1 can help us visualize this linkage.

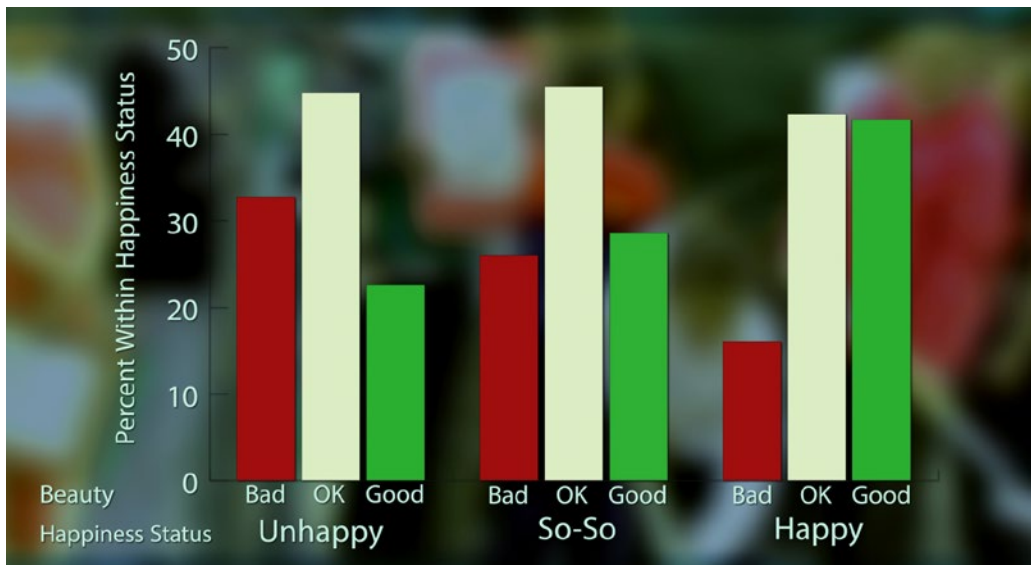


Figure 13.1. Conditional distribution of Physical Beauty for each level of Happiness.

The bar graph in Figure 13.1 shows that as the level of Happiness goes up, the percentage of Bad ratings for Physical Beauty goes down. In addition, as the level of Happiness goes up, the level of Good beauty ratings also goes up. As we know, correlation isn't necessarily causation. However, now that Somerville has identified a link between residents' happiness levels and their thoughts on the city's physical beauty, officials want to dig deeper on the next survey in an effort to improve residents' satisfaction with Somerville.