TAKE IT TO THE MAT

A NEWSLETTER ADDRESSING THE FINER POINTS OF MATHEMATICS INSTRUCTION

Southern Nevada Regional Professional Development Program
January 2004 — High School Edition

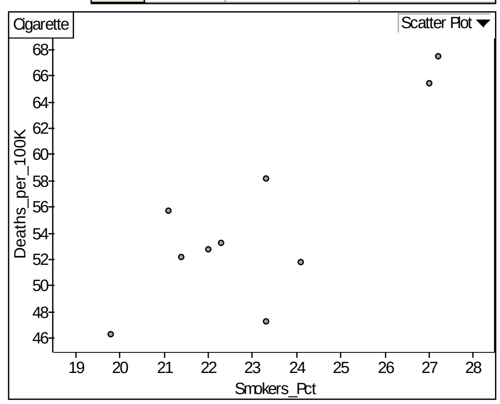
rpdp.ccsd.net

Is there an association between the percentage of the population that smokes and the frequency of deaths due to lung cancer? Our initial response to that question would be an unequivocal, "Yes!" But can we define the relationship; can we quantify it? In this and upcoming issues of *Take It to the MAT*, we will address the use of scatterplots to examine relationships between two variables, the fitting of lines to model them, and making predictions using those models.

Cigarette

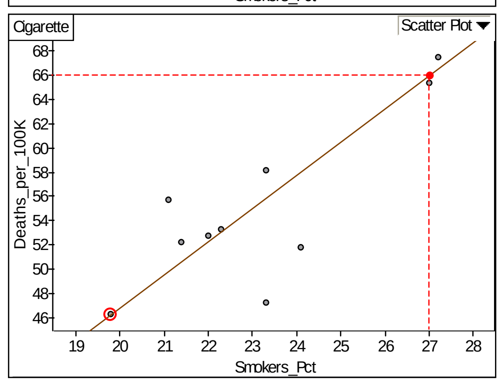
	State	Smokers_Pct	Deaths_per_100K
1	IL	22.3	53.3
2	IN	27	65.4
3	IA	23.3	58.2
4	KS	21.1	55.7
5	MN	19.8	46.3
6	MO	27.2	67.5
7	NE	21.4	52.2
8	ND	23.3	47.3
9	SD	22	52.8
10	WI	24.1	51.8

The table at right shows the percentage of smokers in ten Midwestern states. It also gives the number of deaths from lung cancer per 100,000 population. (Data from *2002 Statistical Abstract of the United States*, www.census.gov/statab/www/). Is there an association between the two variables? Let's make a scatterplot of the data and see.



The scatterplot shows a *positive* association between the rate of smoking and the number of lung cancer deaths per 100,000 population. That is, as the rate of smoking increases, so does the number of deaths. (A *negative* association would show the number of deaths decreasing as the rate of smoking increases.) However, can we define the relationship? Could we predict the number of lung cancer deaths if we knew the percentage of smokers?

In order to make effective predictions, one must have a quantitative description of the relationship—a model. The most common way to model the data is with a line. The pattern of the scatterplot is linear, not strongly, but linear nonetheless.



We want to draw a line that fits the data well. How so? What is our criterion? Should half of the points be above the line and half below? Should the line go through as many data points as possible? Should we use some form of mathematical procedure to do this? When students first attempt this, they should simply drop a line on the plane that looks like it fits the data well—an "eyeball" line. The ultimate question is, "Does our line represent the data well, or as well as we can make it?"

The second graph shows an "eyeball" line. Choosing two points on the line will allow us to compute its equation. Those points may or may not be actual data. In our case it's one of each: (19.8, 46.3), which is Minnesota, and (27, 66). The equation of our eyeball line is $\textit{deaths per 100K} - 66 = \frac{19.7}{7.2}\left(\textit{smoker pct} - 27\right)$ or $\textit{deaths per 100K} \approx 2.7 \cdot \textit{smoker pct} - 7.9$. Writing the equation in this manner preserves context and is preferable to the simple $y \approx 2.7x - 7.9$. Remember, the technique we just used to fit the line was, "It looks good." Did you get the same line?

Now that we have the line, what does it tell us? What can we do with it? That's up for discussion in the next issue.