TAKE IT TO THE MAT

A NEWSLETTER ADDRESSING THE FINER POINTS OF MATHEMATICS INSTRUCTION

Southern Nevada Regional Professional Development Program

February 2004 — High School Edition

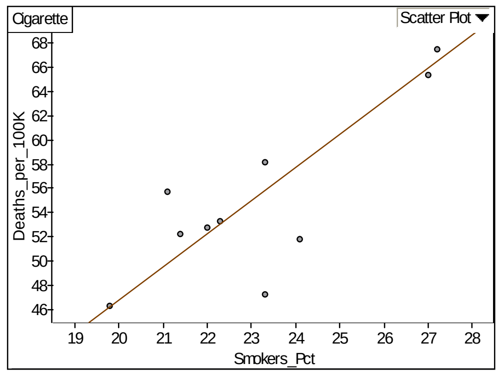
rpdp.ccsd.net

In the January 2004 issue of *Take It to the MAT*, we were presented with data showing the percentage of smokers and lung cancer death rates of ten midwestern states. The questions remain: What does it tell us? What can we do with it? That is where we are going in this edition.

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 foremost reason we fit a line to a set of data is to determine the relationship between the variables. Without some sort of line, we can get lost in the details of the data. That is, by looking at all of the individual points, we don't see the forest for the trees. We need a way to summarize the data to see the "big picture." Finding a model in the form of a line (or curve) helps us find the relationship.



Another thing to note about the line we fit is that it's a type of average. When we speak of averages, we are speaking of where the middle lies. When we fit a line to the data, we are drawing what is an "average" relationship between the two variables.

What exactly is that relationship? The first thing we should look at when describing the relationship is the slope of the line. The equation of the line we drew last month—our "eyeball" line—is *deaths per 100K* $\approx 2.7 \cdot$ *smoker pct* $- 7.9$. The slope of the line is 2.7. Well, that's not quite right. Since the y-variable is deaths per 100,000 due to lung cancer and the x-variable is the percentage of smokers in the state, slope describes the change in death rate as the smoking rate changes.

The slope has units—it has a meaning. In our case, the value of 2.7 should be read as $2.7\,\frac{\text{deaths per 100K}}{\text{percent of smokers}}$. Better yet, it says that for every additional one percent of smokers in a midwestern state, the number of deaths due to lung cancer increases by about 2.7 per 100,000 people. Notice we say *about* 2.7. That's because the slope is itself an average. Some states may go up more, some less, but *on average* an additional 2.7 people per 100K die from lung cancer for every 1% increase in the smoking population.

The intercept of the equation is also an average. Remember that the y-intercept is the value of y when $x = 0$. In this case, the y-intercept is –7.9. That means if nobody in a midwestern state smokes, an *average* of *negative* 7.9 deaths per 100,000 population will be due to lung cancer. Huh? How can the number of deaths be negative? Well, it can't.

There are two things to consider when interpreting the intercept. First, is it within the range of the data? That is, is zero within the spread of the x-values? If not, then the intercept may not have any meaning, since the line is used to fit the data we have, not what's outside the data's range. Second, does it even make sense? In our situation, neither condition is true. Zero is far beyond the spread of our x-values—between about 20 and 27. And, the intercept just doesn't make sense. We need the intercept along with the slope to define the line's equation, but it isn't always meaningful. **The slope always has meaning!**

Next time: A more sophisticated method of finding a line than using our "eyeball."