TAKE IT TO THE MAT

A NEWSLETTER ADDRESSING THE FINER POINTS OF MATHEMATICS INSTRUCTION

Southern Nevada Regional Professional Development Program

April 2004 — High School Edition

rpdp.ccsd.net

For the past few months we've been examining the relationship between the rate of smoking in ten Midwestern states and the rate of death due to lung cancer. In January we fit an "eyeball" line to the scatterplot and estimated the equation of that line. In March we used a more sophisticated technique called the *median-median* line to fit a line. But what good is the line? That's for discussion in this issue of *Take It to the MAT*.

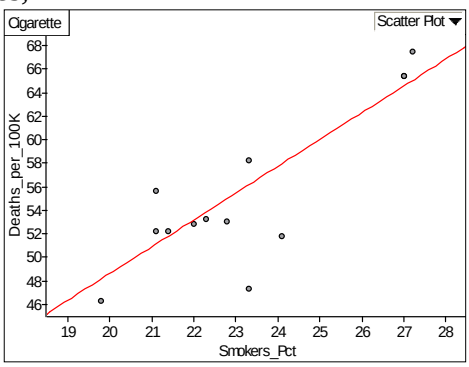
The first thing we must remember is that the line is a type of average. Averages are merely representations of the 'center' or typical value of a data set. When we fit the line, we are creating a function that represents the center of the observations. We recognize that the *y*-variable tends to change as *x* changes and the function reflects this.

In the February 2004 issue we discussed the meaning of the slope and intercept of the line that we fit to the data. The slope and intercept are also averages. The slope tells us how much the *y*-variable changes *on average* if *x* increases by one unit. The intercept gives us an idea of the average *y*-value if $x = 0$.

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

A major reason for finding the line is to make predictions. For example, what would we predict for the death rate from lung cancer in Michigan? (It's alright to make a prediction for Michigan because it's a Midwestern state and our data are from Midwestern states. If we were to ask about California or Virginia, then our prediction has less validity, particularly if lung cancer rates varied by region of the country.)



Anyway, in the absence of any information about smoking rates, one might predict Michigan has a death rate from lung cancer of about 53 per 100,000. Fifty-three is roughly the middle of the death rates for the ten states given. But we have more information we can use—we have a function that relates death rate to smoking rate. (From the March issue, *deaths per 100K* $\approx 2.2 \cdot$ *smoker pct* + 4.0.) We can see there is some relationship between death rate and smoking rate, so we can use smoking rate to make a prediction.

The rate of smoking in Michigan is 24.2 percent of the population. Therefore, using the equation of line, *deaths per 100K* $\approx 2.2\,(24.2) + 4.0$, or *deaths per 100K* ≈ 57.3. Is that a better prediction than simply looking at the middle and estimating 53? Who is to know? But it is better to use more information than less.

One last thing. Since the line is an average, isn't that prediction of 57.3 deaths per 100,000 also an average? Yes, it is indeed. If there were several states with smoking rates of 24.2%, we wouldn't expect all of them to have death rates of 57.3, but the average death rate for those states should be 57.3 according to our model.