1

Smoothers

This primer focuses on smoothers. I assume you have read the section on smoothers in Chapter 6, but I repeat parts of it here to set context. Consider the case where the annual income of a population of adults is thought to be a linear function of the number of years of education the adults have attained. The relationship can be expressed using a traditional linear regression equation:

$$\text{Income}_i = \alpha + \beta \text{ Years Education}_i + \varepsilon_i$$

where α is an intercept, β is the (unstandardized) regression or path coefficient, ε is a disturbance or error term that reflects all factors other than education that independently impact income, and "i" refers to individual i. Suppose our target population has years of education ranging from 10 to 16. We can segregate individuals into those who have 10 years of education, those with 11 years of education, and so on. Suppose for each such segment, we calculate the mean annual income and obtain the following:

Education	Mean Annual Income
10	22,000
11	24,000
12	26,000
13	28,000
14	30,000
15	32,000
16	34,000

It can be seen that the mean income increases by $2,000 for each additional year of education. The value for the regression coefficient in the linear model is thus 2,000. This coefficient provides perspectives on mean changes per unit increase in predictor values and in this case reflects the worth of a year of education. Each of the above means is a *conditional mean*, i.e., the mean value conditional on the predictor equaling a given value. For example, the mean income conditional on 10 years of education is $22,000; the mean income conditional on 11 years of education is $24,000. And so on.

The relationship between predictor values and conditional means does not have to be linear. Here is a different relationship we might observe:

Education	Mean Annual Income
10	22,000
11	22,000
12	22,000
13	24,000
14	26,000
15	28,000
16	30,000

Note that there is a floor effect for mean income, with the mean not changing at the lower end of years of education. This is a non-linear function that traditional linear regression mischaracterizes.

The same concepts apply to dichotomous outcomes. Consider the case where the outcome is the probability adolescents will smoke marijuana in the ensuing year and the predictor is the age of adolescents, ranging from 12 to 17 in units of one. We want to characterize the probability of smoking marijuana for youth who are age 12, for youth who are age 13, for youth who are age 14, and so on. Here are the empirically derived probabilities:

Age	Proportion Smoked Marijuana
12	0.025
13	0.050
14	0.075
15	0.100
16	0.125
17	0.150

The probability of smoking marijuana is 0.025 conditional on age being 12. The probability of smoking marijuana is 0.050 conditional on age being 13. And so on. The regression coefficient is 0.025; for every one unit age increases, the probability of smoking marijuana increases by 0.025 units. This model is called a *linear probability model* because the function is linear. Logistic regression, an alternative analytic approach, assumes the function is non-linear and has the shape of a sigmoid (S) function. With this as background, I now consider smoothers.

SMOOTHERS

When we posit a direct causal relationship between two quantitative variables, we usually have a general sense of whether the relationship is positive or negative, but we may not think through the specific form of the relationship. Traditional regression analyses assume the relationship is linear, but perhaps it is not. A statistical tool to help identify non-linear functional forms is called a smoother (Wilcox, 2017), a graphical and analytic device that makes evident complex functions between variables as compared with more traditional scatterplots. Smoothers plot the conditional means as a function of X and then draw a "line" through the points, a "line" that can be non-linear and irregular in shape. Smoothers are useful because they give insights into how mean Y values change across values of X. With smoothers, one usually does not take every small "bump" or "change in direction" of the smooth literally; rather, it is the general trend suggested by the smoother that is interpreted. See Chapter 11 in the main text for elaboration.

When X is continuous, one can't calculate a Y mean at each value of X because X has, in principle, an infinite number of values (given that it is continuous). Suppose you want to estimate the mean of Y when X = 10, where X is a continuous variable. One popular form of smoothing identifies a span of X scores that are somewhat below X =10 and somewhat above X = 10, and then calculates a predicted mean on the outcome for X = 10 taking all these points into account (using algorithms to weight the various points within the span, with values closer to the point of interest receiving greater weight). The process is repeated for other values of X and then the predicted means are plotted, producing the "smooth." The size of the span to use around a given X can affect the results. Spans are specified by the analyst prior to analysis. It often is useful to examine smooths under different span scenarios. Indeed, for some types of smoothers, choice of the span can be crucial to accurately capturing the true trends in the data. As a general rule, the smaller the span, the better the smoother will characterize the data but at the cost of a more jagged smooth subject to random noise.

Figure 1.1 presents an example of a smooth. We seek to predict student test performance on an exam from a measure of test anxiety. The outcome is measured on a 0 to 100 metric, with higher scores indicating better performance. Test anxiety is a multi-item scale that ranges from -5 to +5, where 0 is a normed average or typical score from a broader student population. Scores greater than 0 reflect increasingly above average test anxiety and scores less than 0 reflect increasingly below average test anxiety.

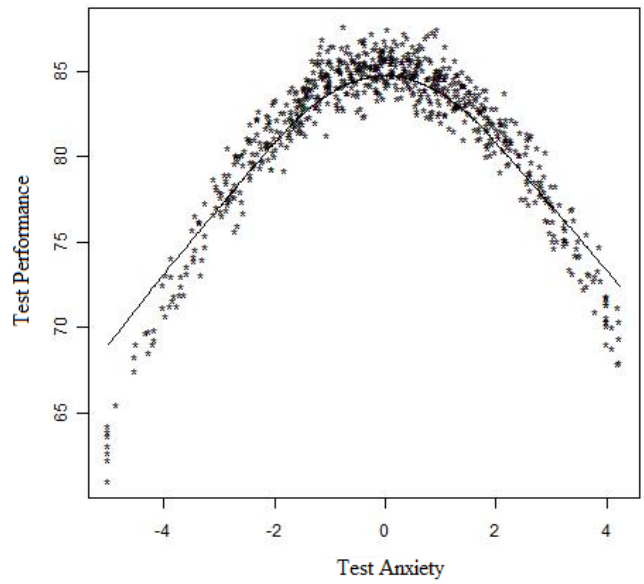

FIGURE 1.1. Smoother for Test Performance and Test Anxiety

The relationship is decidedly non-linear. At the low end of the test anxiety dimension, increases in anxiety are associated with increases in test performance, probably by motivating students to study more. At some point (near a score of 0), however, test anxiety begins to interfere with test performance such that increases in anxiety are associated with decreases in performance.

Smoothers also can be used to double check linearity assumptions in traditional regression analysis. Figure 1.2 shows a smoother predicting the annual salary of professors from the number of times the person's research has been cited by others, where the expectation was the relationship would be linear. The fact that smooth is functionally linear in form affirms the assumption of a linear relationship. I make a routine practice of plotting smooths when conducting regression analyses.

Smoothers traditionally plot conditional means as a function of a predictor X, but they also can plot conditional medians, trimmed means, quantiles, and proportions, among other indices of location. You can do so using the software package shown in the video or using the R programs available in the R package WRS2 by Rand Wilcox.

Smoothers can be extended to scenarios involving multiple predictors, but plots are limited to the case where the number of predictors is 2 or less. With two predictors, three dimensional plots are examined. As well, as discussed in Chapter 11, multiple smoothers can be plotted on the same graph for purposes of gaining insights into interaction effects and moderator analyses.

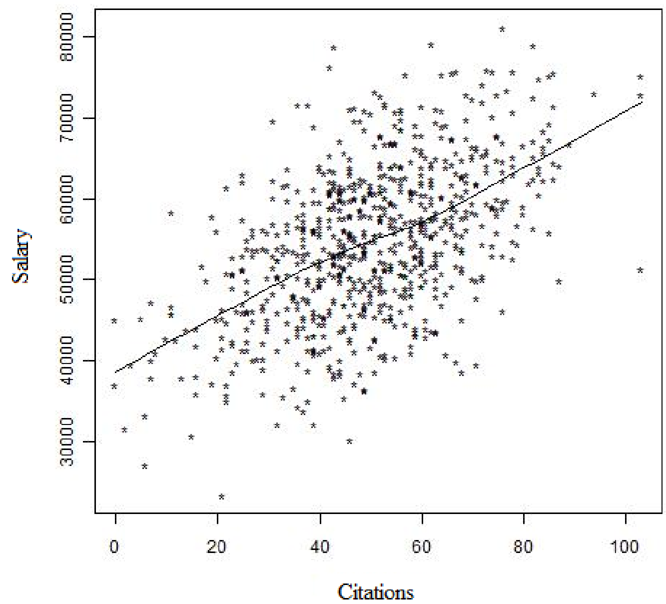

FIGURE 1.2. Smoother for Annual Salary and Research Citations

The type of smoother illustrated in Figure 1.2 is known as a LOWESS (locally weighted scatterplot smoothing) smoother. The span parameter for it ranges in principle from 0 to 1.0. Wilcox (2017) recommends a default value of 0.75, which was used in this example.

SOME USEFUL RESOURCES

A somewhat technical but useful introduction to smoothers is provided in Wilcox (2017). A more introductory treatment is in Wilcox (2010, 2012). A classic but technical introduction to smoothing is by Wand and Jones (1995).

CONCLUDING COMMENTS

Smoothers are a useful tool for exploring the functional form of the relationship between two variables. They can identify non-linear functions and can provide perspectives on the viability of linearity assumptions in statistical tests. When using them, one must be cautious about the choice of the span parameter. Smaller spans better reflect the data but can produce jagged smooths that are difficult to interpret. Larger spans often produce more interpretable smooths but they also can distort the true trends in the data. It sometimes is helpful to plot multiple smooths on the same plot to yield insights into the functional forms of statistical interactions or moderation. As well, you can use smoothers with multiple predictors as a form of non-parametric multiple regression.

REFERENCES

Wand, M.P. & Jones, M.C. (1995). *Kernel smoothing*. New York: Chapman-Hall.

Wilcox, R. (2010). *Fundamentals of modern statistical methods*. New York: Springer.

Wilcox, R. (2012). *Modern statistics for the social and behavioral sciences: An introduction*. New York: CRC Press.

Wilcox, R. (2017). *Introduction to robust estimation and hypothesis testing*. New York: Academic Press (fourth edition).