DigitalCommons@WayneState

Journal of Modern Applied Statistical Methods

Volume 17 | Issue 1 Article 1

2018

An Inferential Method for Determining Which of Two Independent Variables Is Most Important When There Is Curvature

Rand Wilcox
University of Southern California, Los Angeles, rwilcox@usc.edu

Recommended Citation

Wilcox, R. (2018). An Inferential Method for Determining Which of Two Independent Variables Is Most Important When There Is Curvature. Journal of Modern Applied Statistical Methods, 17(1), eP2588. doi: 10.22237/jmasm/1525132920

Journal of Modern Applied Statistical Methods
May 2018, Vol. 17, No. 1, eP2588
doi: 10.22237/jmasm/1525132920
בס"ד

ISSN 1538 − 9472

INVITED ARTICLE

An Inferential Method for Determining Which of Two Independent Variables Is Most Important When There Is Curvature

Rand Wilcox
University of Southern California
Los Angeles, CA

Consider three random variables Y, X_1 and X_2, where the typical value of Y, given X_1 and X_2, is given by some unknown function $m(X_1, X_2)$. A goal is to determine which of the two independent variables is most important when both variables are included in the model. Let τ_1 denote the strength of the association associated with Y and X_1, when X_2 is included in the model, and let τ_2 be defined in an analogous manner. If it is assumed that $m(X_1, X_2)$ is given by $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ for some unknown parameters β_0, β_1 and β_2, a robust method for testing $H_0 : \tau_1 = \tau_2$ is now available. However, the usual linear model might not provide an adequate approximation of the regression surface. Many smoothers (nonparametric regression estimators) were proposed for estimating the regression surface in a more flexible manner. A robust method is proposed for assessing the strength of the empirical evidence that a decision can be made about which independent variable is most important when using a smoother. The focus is on LOESS, but it is readily extended to any nonparametric regression estimator of interest.

Keywords: Smoothers, measures of association, explanatory power, bootstrap methods

Introduction

A common goal when dealing with regression is determining which of two explanatory variables is the most important. For three random variables, say Y, X_1 and X_2, let $m(X_1, X_2)$ denote some unknown function that reflects the typical value

doi: 10.22237/jmasm/1525132920 | Accepted: May 2, 2017; Published: June 7, 2018.
Correspondence: Rand Wilcox, rwilcox@usc.edu

Rand Wilcox is a Professor of Psychology, University of Southern California. His primary interests deal with robust methods for comparing groups and studying associations.

of Y, based on some measure of location, given X_1 and X_2. A way of judging the relative importance X_1 and X_2 is to assume

$$m(X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \tag{1}$$

where β_0, β_1 and β_2 are unknown parameters. Based on (1), many methods have been proposed for making a decision about which of the two independent variables is more important that are known to be unsatisfactory (e.g., Wilcox, 2017, section 11.10). The list of unsatisfactory methods includes stepwise regression, methods based on R^2 (the squared multiple correlation), a homoscedastic approach based on Mallows's (1973) C_p criterion, and ridge regression. There are several alternative methods that provide an estimate of which independent variable is most important. They include cross-validation, the 0.632 bootstrap method (Efron & Tibshirani, 1993), and the nonnegative garrote technique derived by Breiman (1995). Other possibilites are the lasso (Tibshirani, 1996) and least angle regression (see Efron et al., 2004). For a review of the literature dealing with least angle regression, see Zhang and Zamar (2014). But a limitation of all of these methods is that they do not provide an indication of the strength of the empirical evidence that a decision can be made about which independent variable is most important. Inferential methods based on the lasso and least angle regression have been derived (Tibshirani et al., 2016; Lee et al., 2016), but they are not robust: they assume normality and homoscedasticity.

Let ω_1 be some measure of association (e.g., Pearson's correlation or Spearman's rho) between Y and X_1, ignoring X_2, and let ω_2 be defined in an analogous manner. Another approach is to test $\mathrm{H}_0 : \omega_1 = \omega_2$. But a fundamental and well-known concern regarding this approach is that the strength of the association between Y and X_1 can depend on whether X_2 is included in the model.

Another broad approach is to let τ_j be some measure of association between Y and X_j (j = 1,2) when both independent variables are included in the model, and then test

$$H_0 : \tau_1 = \tau_2. \tag{2}$$

The goal is to determine whether there is reasonably strong evidence regarding which of the independent variables has the stronger association. A robust method for accomplishing this goal, still assuming that (1) is true, is described in Wilcox (2018). Also see method IBS in Wilcox (2017, section 11.10.6).

There is, however, a practical concern. The linear model given by (1) might not provide an adequate approximation of the regression surface. A more flexible approach is to include additional parameters in the model. For example, include a quadratic term. But it is known that even this approach can be unsatisfactory, which has led to the development of nonparametric regression estimators, generally called smoothers (e.g., Efromovich, 1999; Eubank, 1999; Fox, 2000; Green & Silverman, 1993; Gyöfri et al., 2002; Härdle, 1990; Hastie & Tibshirani, 1990).

There are numerous examples that smoothers can provide a deeper understanding regarding the association between Y and two independent variables compared to the usual linear model (e.g., Wilcox, 2017).

The goal is a method for testing hypotheses about the relative importance of X_1 and X_2 based on some smoother that provides a flexible approach to curvature. The focus is on the smoother derived by Cleveland and Devlin (1988), but the basic idea is readily extended to any smoother of interest. It is certainly not being suggested that the Cleveland-Devlin estimator dominates other estimators that might be used. Clearly this is not the case. (For comparisons of the Cleveland-Devlin estimator to other smoothers, in terms of mean squared error and bias, see Wilcox, 2005.) When using a smoother, however, rather than the usual linear model given by (1), a refinement of the null hypothesis given by (2) might be needed. Data from the Well Elderly 2 study (Clark et al., 2012) are used to illustrate why.

Generally, the Well Elderly 2 study was designed to assess the effectiveness of an intervention program aimed at improving the physical and emotional wellbeing of older adults. A portion of the study was aimed at understanding the association between a measure of perceived health and wellbeing (SF36) and two independent variables: a measure of depressive symptoms (CESD) and the cortisol awakening response (CAR), which is the difference between cortisol measured upon awakening and again about 30-45 minutes later. The CAR is associated with various measures of stress. Both enhanced and reduced CARs are associated with various psychosocial factors including depression and anxiety disorders (e.g., Bhattacharyya et al., 2008; Pruessner et al., 2003). Here the focus is on measures taken after intervention.

Shown in Figure 1 is the estimated regression surface. It is suggested from the plot the strength of the association between the CAR and SF36 depends on CESD. When CESD is relatively low, for example 7, there appears to be a much stronger association between the CAR and SF36 compared to when CESD is relatively high. The relative importance of CESD, compared to CAR, can depend on both the value of CESD as well as the magnitude of the CAR. More broadly, what is needed is a method that assesses the relative importance of X_1 and X_2, given that $X_1 = x_1$ and $X_2 = x_2$, where x_1 and x_2 are specified values, keeping in mind that the relative importance of X_1 and X_2 can depend on the values of x_1 and x_2. Let $\tau_1(x_2)$ denote some conditional measure of the strength the association between Y and X_1 given that $X_2 = x_2$. In a similar manner, let $\tau_2(x_1)$ denote some conditional measure of the strength the association between Y and X_2 given that $X_1 = x_1$. A natural approach is to test

$$H_0 : \tau_1(x_2) = \tau_2(x_1) \tag{3}$$

and if this hypothesis is rejected, make a decision about which independent variable has the stronger (conditional) association with Y given that $X_1 = x_1$ and $X_2 = x_2$. Of course, estimates of $\tau_1(x_2)$ and $\tau_2(x_1)$ help provide perspective regarding the extent one of the independent variables is more important than the other.

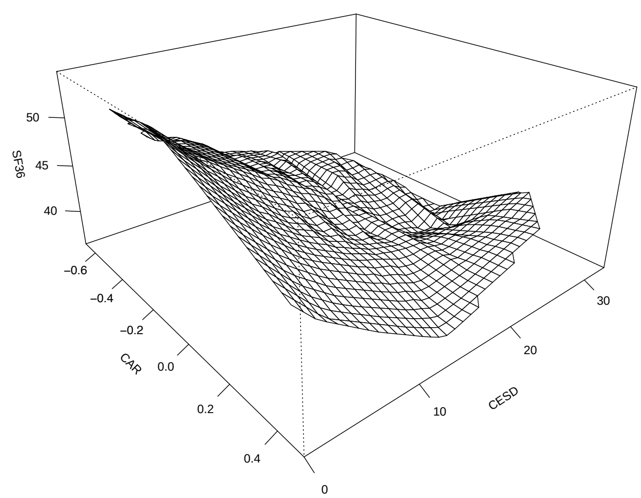

Figure 1. Regression surface predicting the typical SF36 scores as a function of the CAR and CESD.

Another possibility is to focus on $\tau_1(x_2)$ for $J > 1$ values associated with the second independent variable, say $X_{21}, \cdots, X_{2J}$. Of course the same can be done for $\tau_2(x_1)$. Let $\bar{\tau}_1 = \Sigma\tau_1(x_{2j})/J$ and let $\bar{\tau}_2 = \Sigma\tau_2(x_{1j})/J$. One approach might be to test the global hypothesis that the strengths do not differ by testing

$$H_0 : \bar{\tau}_1(x_2) = \bar{\tau}_2(x_1) \tag{4}$$

Another approach is to test

$$H_0 : \tau_1(x_{2j}) = \tau_2(x_{1k}) \tag{5}$$

for each j and k in conjunction with an adjustment that controls the probability of one or more Type I errors among the J^2 tests that are performed. The focus here is on $J = 3$, where the three values for x_{1k} and x_{2j} are estimates of the lower, middle and upper quartiles associated with X_1 and X_2, respectively.

Preliminaries

The immediate goal is to review the Cleveland-Devlin estimator, which is generally known as LOESS. As previously stressed, this is not to suggest that alternative estimators have no practical value. But considering all reasonable choices is extremely difficult, particularly in light of some computational issues described below. A second preliminary issue is choosing some reasonably robust measure that reflects the strength of the association.

Consider the case of a single independent variable, X. Based on the random sample $(X_1, Y_1), \cdots, (X_n, Y_n)$, the smoother derived by Cleveland (1979) is applied as follows. Given X, the method looks for a pre-specified number of points among the X_i values that are close to X. It then scales these distances yielding values in the half open interval [0, 1), and then these scaled values are transformed via the tricube function yielding weights, which in turn yield a weighted mean of the Y values which estimates the mean of Y, given X.

More precisely, let $\delta_i = |X_i - X|$ ($i = 1, \cdots, n$) and let $\delta_{(1)} \leq \cdots \leq \delta_{(n)}$ be the δ_i values written in ascending order. Choose some constant κ, $0 \leq \kappa < 1$, and let K be the value of κn rounded to the nearest integer. Set $Q_i = |X - X_i| / \delta_{(K)}$ and if $0 \leq Q_i < 1$, set $w_i = (1 - Q_i^3)^3$, otherwise $w_i = 0$. Finally, use weighted least squares regression to estimate $m(X_i)$ using w_i as weights.

Consider the more general case dealing with $p \geq 1$ independent variables. Cleveland and Devlin (1988) proceeded as follows. Let $\eta(\boldsymbol{X}, \boldsymbol{X_i})$ be the Euclidean distance between $\boldsymbol{X}$ and $\boldsymbol{X_i} = (X_{i1}, \cdots, X_{ip})$. Let $W(u) = (1 - u^3)^3$, $0 \le u < 1$; otherwise $W(u) = 0$. Let d be the distance of the K^{th}-nearest $\boldsymbol{X_i}$ to $\boldsymbol{X}$. Now $w_i = W(\eta(\boldsymbol{X},\boldsymbol{X_i})/d)$ are used as weights in weighted least squares to compute $m(\boldsymbol{X})$. The R function `loess` performs the computations. Here the span is taken to be $\kappa = 2/3$.

There remains the issue of measuring the strength of the association. Here, τ is taken to be a robust version of explanatory power. Let $\xi^2(Y)$ denote some measure of variation associated with the random variable Y. Let $\hat{Y}$ denote the predicted values of Y, which here are based on LOESS. Then a robust version of explanatory power (e.g., Wilcox, 2017) is

$$\tau^2 = \xi^2(\hat{Y}) \,/\, \xi^2(Y).$$

If $\hat{Y}$ is based on the ordinary least squares estimator, ξ^2 is taken to be the usual variance, and if there is a single independent random variable, then $\tau^2 = \rho^2$, where ρ is Pearson's correlation.

There are many robust measures of variation (e.g., Lax, 1985). For a summary of their relative merits, see Wilcox (2017). Here, the 20% Winsorized variance is used with the understanding that arguments for considering some other measure of variation can be made. Let $g = [0.2n]$, where $[0.2n]$ is the greatest integer less than or equal to $0.2n$. The X_i ($i = 1, \cdots, n$) values written in ascending order are denoted by $X_{(1)} \le \cdots \le X_{(n)}$. The 20% Winsorized values based on X_i ($i = 1, \cdots, n$) are

$$\begin{aligned}
W_i &= X_{(g+1)}, \text{ if } X_i \le X_{(g+1)} \\
W_i &= X_i, \text{ if } X_{(g+1)} < X_i < X_{(n-g)} \\
W_i &= X_{(n-g)}, \text{ if } X \ge X_{(n-g)}.
\end{aligned}$$

The Winsorized sample mean is the mean based on the Winsorized values, and the Winsorized variance is the usual sample variance, again based on the Winsorized values.

Description of the Method

As noted in the previous section, the focus is on the 20% Winsorized variance. For the case of a single independent random variable, τ^2 is readily estimated based on the random sample (X_i,Y_i), $i = 1, \cdots, n$. Simply compute $\hat{Y}_i$ based on LOESS, in which case the numerator of τ^2 is estimated with the Winsorized variance based on the $\hat{Y}_i$ values. And the denominator is estimated via the Winsorized variance of $Y_1, \cdots, Y_n$.

Consider the case where two explanatory variables are included in the model. Let $\xi_1^2(x_2)$ denote the population Winsorized variance of $\hat{Y}$, given that $X_{i2} = x_2$ and note that $\xi_1^2(x_2)$ can be estimated based on the random sample (X_{i1}, X_{i2}, Y_i), $i = 1, \cdots, n$. Let $\tilde{Y}_i$ be the estimate of Y when $\boldsymbol{X_i} = (X_{i1}, x_2)$, in which case the Winsorized variance based on $\tilde{Y}_1, \cdots, \tilde{Y}_n$, say $\hat{\xi}_1^2(x_2)$, estimates $\xi_1^2(x_2)$. In a similar manner, $\hat{\xi}_2^2(x_1)$ estimates $\xi_2^2(x_1)$.

Determining whether X_1 is more important than X_2, rather than testing (3), it suffices to test

$$H_0 : \xi_1^2(x_2) = \xi_2^2(x_1). \tag{6}$$

That is, attention can be focused on the numerator of τ^2, which is the approach taken here henceforth. This distinction was found to make a difference in simulations described in the next section.

The next goal is to describe the bootstrap method that was considered for testing (6). This is followed by an adjustment that was dictated by preliminary simulations.

A basic percentile bootstrap method for testing (6) is applied as follows:

1. Generate a bootstrap sample by resampling with replacement n points from (X_{i1}, X_{i2}, Y_i), $i = 1, \cdots, n$, yielding say $(X_{11}^*, X_{12}^*, Y_1^*), \cdots, (X_{n1}^*, X_{n2}^*, Y_n^*)$.

2. Compute an estimate of $\xi_1^2(x_2)$ and $\xi_2^2(x_1)$ based on this bootstrap sample yielding $\xi_1^*(x_2)$ and $\xi_2^*(x_1)$, respectively, and let $d^* = \xi_1^*(x_2) - \xi_2^*(x_1)$.

3. Repeat steps 1 through 2 B times and let d_b^* $(b = 1, \cdots, B)$ denote the resulting d^* values.

4. Put the d_b^* values in ascending order and label the results $d_{(1)}^* \leq \cdots \leq d_{(B)}^*$.

5. Let $\ell = \alpha B/2$, rounded to the nearest integer and $u = B - \ell$. Then a $1 - \alpha$ confidence interval for $\xi_1^2(x_2) - \xi_2^2(x_1)$ is $(d^*_{(\ell+1)}, d^*_{(u)})$.

From Liu and Singh (1997), a (generalized) p-value is $p = 2\min(\hat{p}, 1 - \hat{p})$, where $\hat{p} = A/B$ and A is the number of d^* values less than zero. The hypothesis given by (4) can be tested in a similar manner.

Here, $B = 500$ was used, which has been found to perform reasonably well, in terms of controlling the Type I error probability, when dealing with other robust estimators (e.g., Wilcox, 2017). However, a larger value for B might increase power (Racine & MacKinnon, 2007; cf. Davidson & MacKinnon, 2000). This will be called method L.

Preliminary simulations based on 2,000 replications indicated that method L performs poorly: the actual probability of a Type I error can be substantially smaller than the nominal level.

This is particularly true when there is no association. The strategy is to assume $Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon$, where the error term ϵ has a standard normal distribution. The idea is to find an adjusted p-value, say p_c, and reject the hypothesis of interest if $p \le p_c$. Then, simulations are used to investigate the impact of non-normality and curvature.

When $\beta_1 = \beta_2 = 0$, the estimate of p_c generally exceeds 0.2, depending on the sample size. An estimate substantially larger than 0.05 was expected based on results in Wilcox (2018). When using a regression estimator based on the usual linear model $Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon$, rather than a smoother, a similar phenomenon was observed. When there is no association, explanatory power makes no distinction between an estimate indicating a slightly positive association and one indicating a slightly negative association. This suggests among the bootstrap samples, if there is no association, the expectation is that $P(d^* < 0)$ will be greater than 0.05, which was found to be the case for the situation at hand.

For convenience, when $\beta_1 = \beta_2$, let β_c denote this common value. As indicated, when $\beta_c = 0$, estimates of p_c generally exceed 0.2. As β_c increases, the estimate of p_c decreases up to a point. For $\beta_c = 0.5$, 1 and 2, the estimates were very similar. The initial strategy was to estimate p_c for $\beta_c = 1$ and sample sizes ranging from 50 to 1,000. Then, the impact of non-normality and curvature is investigated via simulations. The idea is that if there is little or no association, it is relatively unimportant which independent variable is more important. However, if one or both have an association with the dependent variable, the goal is to control the probability of erroneously rejecting.

Estimates of p_c are reported in Table 1. Again, 2,000 replications were used due to the high execution time. With $n = 50$, estimates required approximately two hours, and $n = 500$ required about ten hours. Reported in column two are results when testing (4) and when using three values for both x_1 and x_2, namely, estimates of the lower quartile (Q1), the median, and the upper quartile (Q2). The lower and upper quantiles are estimated with a single order statistic (see, for example, Wilcox, 2017, p. 61) and the population median is estimated with the usual sample median. Column three reports the estimates of p_c when testing (3) and where both x_1 and x_2 are based on the usual sample median. Compiled in column four are results when the median is replaced by an estimate of the lower (or upper) quartile. Finally, in column five, estimates are presented of p_c when testing (5) with $J = 3$ and the values for x_1 and x_2 are again the lower quartile, the median and the upper quartile. Note p_c was determined to control the probability of one or more Type I errors among the nine tests that are performed. Initially the estimates decrease as the sample size increases, but for $n \geq 200$ the estimates change very little.

Table 1. Estimates of the critical *p*-value, p_c. C1 = testing (4), C2 = testing (3) using the median, C3 = testing (3) using the lower quartile, and C4 = testing (5)

n	C1	C2	C3	C4
50	0.082	0.114	0.142	0.042
100	0.076	0.080	0.095	0.021
200	0.067	0.065	0.082	0.024
400	0.057	0.062	0.079	0.026
600	0.064	0.060	0.079	0.026
1000	0.062	0.071	0.072	0.025

Simulation Study

Four types of distributions are considered for the error term: normal, symmetric and heavy-tailed (roughly meaning that outliers tend to be common), asymmetric and relatively light-tailed, and asymmetric and relatively heavy-tailed. Data are generated from *g*-and-*h* distributions (Hoaglin, 1985), which is formed as follows. If Z has a standard normal distribution, then by definition

$$V = \frac{\exp(gZ) - 1}{g} \exp\left(hZ^2 / 2\right), \text{ if } g > 0$$
$$V = Z \exp\left(hZ^2 / 2\right), \text{ if } g = 0$$

has a *g*-and-*h* distribution where *g* and *h* are parameters that determine the first four moments. The four distributions used here were the standard normal ($g = h = 0$), a symmetric heavy-tailed distribution ($h = 0.2$, $g = 0.0$), an asymmetric distribution with relatively light tails ($h = 0.0$, $g = 0.2$), and an asymmetric distribution with heavy tails ($g = h = 0.2$). Compiled in Table 2 are skewness (κ_1) and kurtosis (κ_2) for each distribution. Hoaglin (1985) summarized additional properties of the *g*-and-*h* distributions. The independent variables were generated from a bivariate normal distribution with correlation zero or 0.6.

Table 2. Some properties of the *g*-and-*h* distribution.

g	*h*	K_1	K_2
0.00	0.00	0.00	3.00
0.00	0.20	0.00	21.46
0.20	0.00	0.61	3.68
0.20	0.20	2.81	155.98

For the first set of simulations, data were generated from

$$Y = X_1 + X_2 + \epsilon \tag{7}$$

to check on the ability of the method to control the Type I error probability when the usual linear model holds and where the error term does not have a normal distribution. The second set of simulations were based on the model

$$Y = X_1^2 + X_2^2 + \epsilon \tag{8}$$

to check on how well the methods perform when dealing with a situation where the regression surface is not a plane.

Results

The column headed by G1 in Table 3 contains the estimated probability of a Type I error when testing at the 0.05 level, $n = 50$, Pearson's correlation between the two independent variables is $\rho = 0$, and the goal is to test (4). The column headed by G2 describes results when testing (3) and when both x_{1j} and x_{2j} are the sample medians. The column headed by G3 pertains to when both x_{1j} and x_{2j} are estimates of the lower quartiles and G4 are the results when using the upper quartiles. G5 corresponds to testing (5); the entries are the estimates of the probability of one or more Type I errors.

Table 3. Estimated Type I error probabilities, $n = 50$, $\alpha = 0.05$ and data generated according to (7)

g	*h*	G1	G2	G3	G4	G5
0.000	0.000	0.050	0.050	0.050	0.050	0.050
0.000	0.200	0.066	0.061	0.029	0.032	0.045
0.200	0.000	0.073	0.073	0.038	0.040	0.046
0.200	0.200	0.064	0.059	0.032	0.026	0.041

Although the seriousness of a Type I error can depend on the situation, Bradley (1978) suggested as a general guide when testing at the 0.05 level the actual level should be between 0.025 and 0.075. As indicated in Table 3, all of the estimated Type I error probabilities fall in this range. Using a span equal to 0.75 lowered the estimates slightly. Increasing the correlation between the two independent variables to $\rho = 0.6$, all of the estimates decrease. Most remain above 0.025. But the lowest estimate in Table 3, which occurs for G4 and $g = h = 0.2$, drops from 0.026 to 0.009.

Reported in Table 4 are the estimated Type I error probabilities when generating data based on (8). Pearson’s correlation between the two independent variables is $\rho = 0$. The estimates range between 0.045 and 0.076, with only one instance where the estimates did not satisfy Bradley’s criterion. Increasing the correlation to $\rho = 0.6$ resulted in lower estimates. In some situations the estimates were slightly lower. For G5 and $g = h = 0$, the estimate dropped from 0.03 to 0.01. Increasing the span to 0.75 did not give improved results.

Table 4. Estimated Type I error probabilities, $n = 50$, $\alpha = 0.05$ and data generated according to (8)

g	*h*	G1	G2	G3	G4	G5
0.000	0.000	0.075	0.064	0.069	0.068	0.049
0.000	0.200	0.060	0.045	0.052	0.047	0.034
0.200	0.000	0.076	0.062	0.065	0.059	0.042
0.200	0.200	0.060	0.065	0.057	0.045	0.030

Illustrations

The methods are illustrated using data from two studies. In the first study (conducted by Shelley & Schwartz, n.d.), the dependent variable labeled TOTAGG Score is a sum of peer nomination items that were based on an inventory that included descriptors focusing on adolescents' behaviors and social standing. The peer nomination items were obtained by giving children a roster sheet and asking them to nominate a certain amount of peers who fit particular behavioral descriptors. The independent variables were grade point average (GPA) and a measure of academic engagement (Engage). The sample size was $n = 336$.

Shown in Figure 2 is an estimate of the regression surface with leverage points removed, which reduced the sample size to 323. (Leverage points refer to points for which the independent variables, taken together, are flagged as outliers. Points were flagged as outliers with a projection-type method that takes in account the overall structure of data; see, for example, Wilcox, 2017, section 6.4.9.) Least angle regression indicated GPA is more important than Engage. There appeared to be curvature, particularly for the lower GPA scores. A test of the hypothesis that the regression surface is a plane is significant, $p < 0.001$.

The results based on testing (5) are shown in Table 5. The column headed by str.x1.given.x2 is the estimate of $\tau_1(x_2)$.

Consider, for example, the first row of results in Table 5. The estimate of $\tau_1(x_2)$ is 0.986. This estimate might seem unusually high but it can be explained as follows. The Winsorized standard deviation of the predicted values of TOTAGG given that Engage is 3.43, is 0.2473; the Winsorized standard deviation of the TOTAGG scores is 0.2508, so the estimate of the strength of the association, τ_1 (3.43), is 0.2473/0.2508 = 0.986. With no Winsorizing (the standard deviation is used), the estimate is 0.3747/1.0154 = 0.369. The standard deviation of the TOTAGG scores is about four times as large as the Winsorized standard deviation roughly because the bulk of the points are tightly clustered together. (Also, the distribution of the TOTAGG scores is highly skewed.)

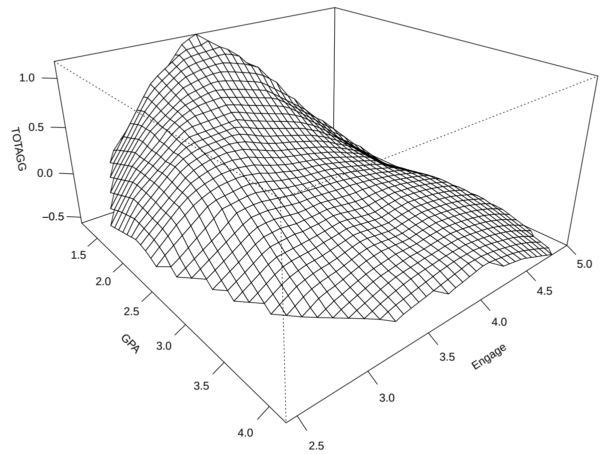

Figure 2. Regression surface predicting the typical TOTAGG score as a function of GPA and a measure of academic engagement.

Table 5. Results when testing (5) based on the data used in Figure 2.

GPA	*Engage*	*p-value*	*str.x1.given.x2*	*str.x2.given.x1*
2.500	3.430	0.028	0.986	0.376
2.500	3.860	0.392	0.649	0.376
2.500	4.140	0.992	0.332	0.376
3.000	3.430	0.020	0.986	0.090
3.000	3.860	0.152	0.649	0.090
3.000	4.140	0.336	0.332	0.090
3.330	3.430	0.024	0.986	0.229
3.330	3.860	0.244	0.649	0.229
3.330	4.140	0.596	0.332	0.229

Indicated in Table 5, GPA always has a stronger association with TOTAGG except when GPA = 2.5 and Engage = 4.14. Controlling the probability of one more Type I errors among the nine tests that were performed, the strongest (significant) evidence that this is the case occurs for two situations. The first is

GPA = 3 and Engage = 3.43, and the second occurs when GPA = 3.33 and again Engage = 3.43. For GPA = 2.50 and Engage = 3.43, the *p*-value = 0.028. Generally, there is evidence GPA is more important than Engage when Engage scores are relatively low. Otherwise, there is no strong indication that this is the case. If the apparent curvature is ignored and the method in Wilcox (2018) is used, GPA is estimated to be more important, but $p = 0.077$. Testing (4), $p = 0.076$.

The next illustration is based on the Well Elderly 2 study data (Clark, 2013). The focus is on measures taken prior to intervention. A portion of the study was aimed at understanding the association between a measure of life satisfaction (LSIZ) and two independent variables: a measure of meaningful activities (MAPA) and a measure of interpersonal support (PEOP). An estimate of the regression surface is shown in Figure 3.

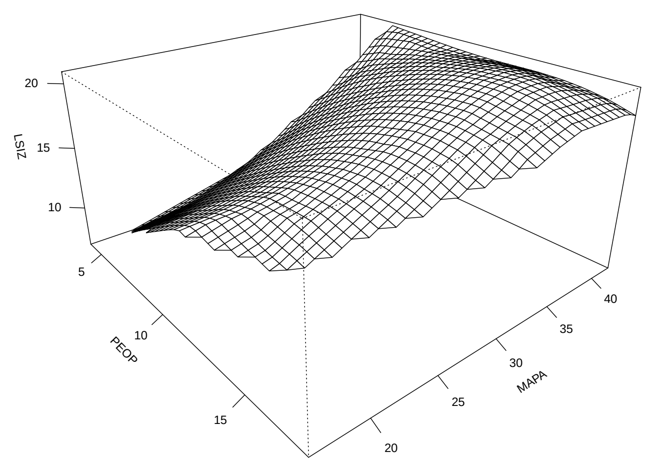

Figure 3. Regression surface predicting life satisfaction (LSIZ), prior to intervention, based on measures of meaningful activities (MAPA) and personal support (PEOP).

If the usual linear model is assumed, least angle regression indicates that PEOP is more important than MAPA. Using the robust method in Wilcox (2018), now MAPA is found to be more important, $p = 0.032$. But both of these methods are suspect due to the apparent curvature. Testing the hypothesis that the regression surface is a plane, $p < 0.001$. Testing (4), $p = 0.068$.

The results based on testing (5) are shown in Table 6. The strength of the association between MAPA and LSIZ, given PEOP is equal to 9, is 0.591. The association between PEOP and LSIZ, given MAPA = 28, is 0.330. PEOP is more important, the strongest evidence occurring when the strength of PEOP given that MAPA = 28, is compared to the strength of MAPA given that PEOP = 13.

Table 6. Results when testing (5) based on the data used in Figure 3.

PEOP	*MAPA*	*p-value*	*str.x1.given.x2*	*str.x2.given.x1*
9.000	28.000	0.036	0.330	0.591
9.000	32.000	0.248	0.310	0.591
9.000	36.000	0.688	0.221	0.591
11.000	28.000	0.044	0.330	0.524
11.000	32.000	0.176	0.310	0.524
11.000	36.000	0.480	0.221	0.524
13.000	28.000	0.028	0.330	0.446
13.000	32.000	0.032	0.310	0.446
13.000	36.000	0.108	0.221	0.446

The results for the data in Figure 1 are shown in Table 7. CESD is the more important independent variable. The evidence is particularly strong when focusing on the median value of CESD. The same is true when both CAR and CESD are taken to be the lower quartiles as well as when both are taken to be upper quartiles. Testing (4), the $p = 0.006$.

Table 7. Results when testing (5) based on the data used in Figure 1.

CAR	*CESD*	*p-value*	*str.x1.given.x2*	*str.x2.given.x1*
−0.174	4.000	0.024	0.127	0.462
−0.174	9.000	0.004	0.169	0.462
−0.174	16.000	0.044	0.153	0.462
−0.029	4.000	0.080	0.127	0.586
−0.029	9.000	0.024	0.169	0.586
−0.029	16.000	0.056	0.153	0.586
0.072	4.000	0.088	0.127	0.502
0.072	9.000	0.004	0.169	0.502
0.072	16.000	0.004	0.153	0.502

Conclusion

In the illustrations, the hypothesis that the regression surface is a plane was rejected. It is not being suggested, however, that if this test fails to reject, it would now be reasonable to use the usual linear model. It is unclear when such a test has enough power to detect situations where curvature is a practical concern. All indications are that the proposed methods avoid Type I errors well above the nominal level. There is room for improvement, however, because as the correlation among the independent variables increases, situations are found where the actual level is well below the nominal level. The R function `lplotcomBCI` tests the hypotheses given by (3) and (4), and `lplotcomBCI9` tests the hypotheses indicated by (5). Both of these functions are being added to the R package `WRS`.

References

Bhattacharyya, M. R., Molloy, G. J., & Steptoe, A. (2008). Depression is associated with flatter cortisol rhythms in patients with coronary artery disease. *Journal of Psychosomatic Research, 65*(2), 107-113. doi: 10.1016/j.jpsychores.2008.03.012

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31*(2), 144-152. doi: 10.1111/j.2044-8317.1978.tb00581.x

Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics*, *37*(4), 373-384. doi: 10.2307/1269730

Clark, F. (2013). *Well Elderly 2, Los Angeles, California, 2004-2008*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013-05-28. https://doi.org/10.3886/ICPSR33641.v1

Clark, F., Jackson, J., Carlson, M., et al. (2012). Effectiveness of a lifestyle intervention in promoting the well-being of independently living older people: results of the Well Elderly 2 randomised controlled trial. *Journal of Epidemiology and Community Health, 66*(9), 782-790. doi: 10.1136/jech.2009.099754

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*(368), 829-836. doi: 10.2307/2286407

Cleveland, W.S., & Devlin, S. J. (1988). Locally-weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association, 83*(403), 596-610. doi: 10.2307/2289282

Davidson, R., & MacKinnon, J. G. (2000). Bootstrap tests: How many bootstraps? *Econometric Reviews, 19*(1), 55-68. doi: 10.1080/07474930008800459

Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology, 45*(2), 265-282. doi: 10.1111/j.2044-8317.1992.tb00992.x

Efromovich, S. (1999). *Nonparametric curve estimation: methods, theory and applications.* New York: Springer-Verlag.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004) Least angle regression (with discussion and rejoinder). *Annals of Statistics, 32*(2), 407-499. doi: 10.1214/009053604000000067

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York: Chapman and Hall.

Eubank, R. L. (1999). *Nonparametric regression and spline smoothing*. New York: Marcel Dekker.

Fox, J. (2000). Multiple and generalized nonparametric regression. Thousands Oaks, CA: Sage. doi: 10.4135/9781412985154

Green, P. J., & Silverman, B. W. (1993). *Nonparametric regression and generalized linear models: a roughness penalty approach.* Boca Raton, FL: CRC Press. doi: 10.1201/b15710

Györfi, L., Kohler, M., Krzyzk, A., & Walk, H. (2002). *A distribution-free theory of nonparametric regression* (Springer Series in Statistics). New York: Springer Verlag. doi: 10.1007/b97848

Härdle, W. (1990). *Applied Nonparametric Regression* (Econometric Society Monographs No. 19). Cambridge, UK: Cambridge University Press.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized Additive Models*. New York: Chapman and Hall.

Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distribution. In D. Hoaglin, F. Mosteller & J. Tukey (Eds.) *Exploring Data Tables Trends and Shapes*. New York: Wiley, pp. 461-511. doi: 10.1002/9781118150702.ch11

Lax, D. A. (1985). Robust estimators of scale: Finite-sample performance in long-tailed symmetric distributions. *Journal of the American Statistical Association, 80*(391), 736-741. doi: 10.1080/01621459.1985.10478177

Lee, J., Sun, D., Sun, Y., & Taylor, J. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics, 44*(3), 907-927. doi: 10.1214/15-aos1371

Liu, R. G., & Singh, K. (1997). Notions of limiting *p* values based on data depth and bootstrap. *Journal of the American Statistical Association, 92*(437), 266-277. doi: 10.2307/2291471

Mallows, C. L. (1973). Some comments on C_p. *Technometrics, 15*(4), 661-675. doi: 10.1080/00401706.1973.10489103

Pruessner, J. C., Hellhammer, D. H., & Kirschbaum, C. (1999). Burnout, perceived stress, and cortisol responses to awakening. *Psychosomatic Medicine, 61*(2), 197-204. doi: 10.1097/00006842-199903000-00012

Racine, J., & MacKinnon, J. G. (2007). Simulation-based tests than can use any number of simulations. *Communications in Statistics - Simulation and Computation, 36(*2), 357-365. doi: 10.1080/03610910601161256

Shelley, T., & Schwartz, D. (n.d.). [Peer nomination data for particular behavioral descriptors]. Unpublished raw data. Retrieved from https://dornsife.usc.edu/assets/sites/239/docs/shelley.csv.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 58,* 267-288.

Tibshirani, R. J., Taylor, J., Lockhart, R., & Tibshirani, R. (2016). Exact post-selection inference for sequential regression procedures. *Journal of the American Statistical Association, 111*(514), 600-620. doi: 10.1080/01621459.2015.1108848

Wilcox, R. R. (2005). A comparison of six smoothers when there are multiple predictors. *Statistical Methodology, 2*(1), 49-57. doi: 10.1016/j.stamet.2004.11.004

Wilcox, R. R. (2017). *Introduction to Robust Estimation and Hypothesis Testing* (4th Ed.*)*. San Diego, CA: Academic Press.

Wilcox, Rand R. (2018). Robust regression: an inferential method for determining which independent variables are most important. *Journal of Applied Statistics, 45*(1), 100-111. doi: 10.1080/02664763.2016.1268105

Zhang, H. & Zamar, R. H. (2014). Least angle regression for model selection. *Wiley Interdisciplinary Reviews: Computational Statistics, 6*(2), 116-123. doi: 10.1002/wics.1288.