ROBUST ANCOVA: HETEROSCEDASTIC CONFIDENCE INTERVALS THAT HAVE SOME SPECIFIED SIMULTANEOUS PROBABILITY COVERAGE

Rand R. Wilcox [1*]

[1] Dept. of Psychology University of Southern California

Abstract: The paper deals with robust ANCOVA when there are one or two covariates. Let Mj (Y |X) = β0j + β1j X1 + β2j X2 be some conditional measure of location associated with the random variable Y , given X, where β0j , β1j and β2j are unknown parameters. A basic goal is testing the hypothesis H0: M1(Y |X) = M2(Y |X). A classic ANCOVA method is aimed at addressing this goal, but it is well known that violating the underlying assumptions (normality, parallel regression lines and two types of homoscedasticity) create serious practical concerns. Methods are available for dealing with heteroscedasticity and non-normality, and there are well-known techniques for controlling the probability of one or more Type I errors. But some practical concerns remain, which are reviewed in the paper. An alternative approach is suggested and found to have a distinct power advantage.

Key words: prediction intervals, heteroscedasticity, robust regression, analysis of covariance

1. Introduction

A classic problem is comparing two independent groups based on some outcome measure Y while taking into account one or two covariates. For the jth group, let Mj (Y |X) be some conditional population measure of location associated with the random variable Y given X = (X_1, X_2). It is assumed that

$$M_j(Y|\mathbf{X}) = \beta_{0j} + \beta_1 X_{1j} + \beta_1 X_{2j},$$ (1)

where β0j , β1j and β2j are unknown parameters that are estimated based on the random sample $(X_{1j}, Y_{1j}), \ldots, (Xnj, Ynj)$. Further assume that given X,

$$Y_j = \beta_{0j} + \beta_{1j} X_{1j} + \beta_{2j} X_{2j} + \lambda(\mathbf{X}_j)\epsilon_j,$$ (2)

---

[*] Corresponding author.

where s_j has some unknown distribution with $M(s_j) = 0$ and $\lambda(\mathbf{X}_j)$ is some unknown function of X_j used to model heteroscedasticity. As is evident, a fundamental goal is computing a confidence interval for $M_1(Y|\mathbf{X}) - M_2(Y|\mathbf{X})$.

For convenience, momentarily focus on the single covariate case. The classic and best-known approach is to assume s_j has a normal distribution with $M(s_j) = E(s_j) = 0$ and $\lambda(X_j) \equiv 1$ (within group homoscedasticity). It is furthermore assumed that $\beta_{11} = \beta_{12}$ (parallel regression lines), $\sigma^2_1 = \sigma^2_2$ (between group homoscedasticity), and the unknown slope and intercept are estimated via the ordinary least squares (OLS) estimator. Based on these assumptions, it suffices to compare the intercepts. It is well known, however, that violating any of these assumptions is a serious practical concern (e.g., Wilcox, 2017). More broadly, OLS is not robust (e.g., Staudte and Sheather, 1990; Maronna et al., 2006; Heritier et al., 2007; Hampel et al., 1986; Huber and Ronchetti, 2009), which can result in a misleading summary of the data as well relatively poor power.

For a single value of the covariate X, there is a simple method for testing

$$H_0 : M_1(Y|X) = M_2(Y|X) \tag{3}$$

that effectively deals with non-normality, outliers, non-parallel regression lines and both types of heteroscedasticity. (Details are described in section 2.) But focusing on a single value of the covariate can miss important details regarding how and where the regression lines differ. A way of dealing with this issue is to test (3) using K covariate values and control the probability of one or more Type I errors using some improvement on the Bonferroni inequality (e.g., Hochberg, 1988; Hommel, 1988). Another approach is to use a Studentized maximum modulus distribution with infinite degrees of freedom, assuming that a robust regression estimator is used. If for example K = 5 covariate values are chosen, these approaches appear to perform fairly well in terms of controlling the probability of one or more Type I errors (Wilcox, 2013). But again there is the concern that this does not provide enough detail about where and how the regression lines differ. In particular, when K is small, power can be negatively impacted as will be illustrated.

Consider instead the goal of testing

$$H_0 : M_1(Y|X = z_k) = M_2(Y|X = z_k) \tag{4}$$

for each k = 1, . . . , K, where $z_1, \ldots, z_K$ are K values evenly space between $\min(X_{ij})$ and $\max(X_{ij})$, the minimum and maximum values among all of the covariate values that were observed. If, for example, K = 25 is used, it is fairly evident that this raises concerns about power when using the methods derived by Hochberg (1988) or Hommel (1988) to control the probability of one or more Type I errors.

Consider, for example, Hochberg's sequentially rejective method, which is applied as follows. Let $p_1, \ldots, p_K$ be the p-values associated with the K tests, put these p-values in descending order, and label the results $p_{[1]} \geq p_{[2]} \geq \cdots \geq p_{[K]}$. Beginning with k = 1, reject all hypotheses if

$$p_{[k]} \leq \alpha/k.$$

That is, reject all hypotheses if the largest p-value is less than or equal to α. If $p_{[1]} > \alpha$, proceed as follows:
  1. Increment k by 1. If

$$p_{[k]} \leq \frac{\alpha}{k},$$

  stop and reject all hypotheses having a p-value less than or equal $p_{[k]}$
  2. If $p_{[k]} > \alpha/k$, repeat step 1.
  3. Repeat steps 1 and 2 until all *K* hypotheses have been tested.

Because at the kth step, hypotheses are tested at the *α/k* level, this suggests that Hochberg's method can have relatively low power when *K* is even moderately large. Simulation results in section 3 illustrate that this is indeed the case and that the same is true when using Hommel's method.

When there are two covariates and the goal is to test H_0: $M_1(Y |X) = M_2(Y |X)$, again an extant method can be used when dealing with a single covariate point. Using a small number of covariate points, the probability of one or more Type I errors can again be controlled using say Hochberg's method. But using only a small number of covariate points can miss important differences, so again there is the issue how to control the probability of one or more Type I errors when testing the null hypothesis for even a moderately large number of points.

The goal in this paper is to suggest a robust heteroscedastic method for testing (4)   when K is relatively large, say greater than or equal 25, which has a power advantage over using Hochberg's or Hommel's method for controlling the probability of one or more Type I errors. Section 2 describes the proposed method when dealing with a single covariate. Section 3 reports simulation results on how well the method performs.   Section 4 describes a generalization of the method in section 2 to situations where there are two covariates and section 5 reports simulation results on the ability of the method to control the probability of one or more Type I errors. Results on power are reported as well.

2. Description of the Proposed Method for a Single Covariate

The proposed method, when dealing with a single covariate, mimics to some degree the approach used by the two-sample version of Student's T test. First, determine an appropriate critical value, based on an obvious test statistic, assuming normality and homoscedasticity and in conjunction with a regression estimator of interest. Then study the impact of non- normality and heteroscedasticity via simulations.

Momentarily, focus on a single value for the covariate, X = x. Let $\tau 2$ denote the squared standard error of $\overline{Y}= b_0 + b_1 x$, where b_0 and b_1 are estimates of β_0 and β_1, respectively based on some regression estimator to be determined. A basic percentile bootstrap method   is used to estimate τ^2 (e.g., Efron & Tibshirani, 1993). More precisely, generate a bootstrap sample by

randomly sampling with replacement n pairs of points from $(X_1, Y_1),\ldots, (X_n, Y_n)$ yielding $(X^*_1, Y^*_1),\ldots,(X^*_n, Y^*_n)$. Based on this bootstrap sample, estimate the intercept and slope and label the results b^*_0 and b^*_1, which yields $\widehat{Y}^* = b^*_0 + b^*_1 X$, Repeat this B times yielding $\widehat{Y}^*_1,\ldots, \widehat{Y}^*_B$, in which case an estimate of τ^2 is

$$\hat{\tau}^2 = \frac{1}{B-1}\sum(\hat{Y}^*_b - \bar{Y}^\bullet)^2,$$

where $\bar{Y}^* = \sum \hat{Y}^*_b / B$ In terms of controlling the probability of a Type I error, B = 100 appears to suffice, at least for the situations considered in sections 3 and 5, which is consistent with related results summarized in Wilcox (2017).

For the jth group, let $\hat{\tau}^2_j$ be the estimate of the squared standard error of $\widehat{Y}j = b_{0j} + b_{1j} X$.

Then, of course, the hypothesis given by (3) can be tested with

$$W = \frac{\hat{Y}_1 - \hat{Y}_2}{\sqrt{\hat{\tau}^2_1 + \hat{\tau}^2_2}}$$

once an appropriate critical value has been determined.

Momentarily assume that W has a standard normal distribution, in which case a p- value can be determined for each $X = x_k$, $k = 1, \ldots, K$. Denote the resulting p-values by $p_1, \ldots, p_K$ and let pm = min$(p_1, \ldots, p_K)$. As is evident, if the α quantile of pm can be determined, say pα, the probability of one or more Type I errors can be controlled simply by rejecting the kth hypothesis if and only if pk $\leq$ pα. And in addition, confidence intervals for each $M_1(Y |X = z_k) - M_2(Y |X = z_k)$ can be computed that have simultaneous probability coverage $1 - \alpha$.

The distribution of pm is approximated in the following manner. Assume both the error term s and X have a standard normal distribution and without loss of generality, consider the case $\beta_{0j} = \beta_{1j} = 0$ (j = 1,2). Then a simulation can be performed yielding an estimate of the α quantile of pm. That is, generate data in the manner just indicated, compute pm and repeat this process A times yielding $P_{m1}, \ldots P_{mA}$. Put these A values in ascending order yielding $P_{m(1)} \leq \ldots \leq P_{m(A)}$ and let A = αA rounded to the nearest integer. Then the α quantile of pm, pα, is estimated with $\widehat{Y}\alpha = Pm(l)$. This will be called method JN henceforth (because it represents a robust generalization of the method derived by Johnson and Neyman, 1936).

2.1 Choosing a Regression Estimator

As previously indicated, it is known that OLS is not robust. Some practical consequences are that outliers can result in a misleading summary of the association among the bulk of the points. Also, outliers among the dependent variable can result in a relatively high standard error, which in turn can mean relatively low power compared to various robust estimators that have been derived. Numerous robust estimators have been proposed that are aimed at dealing with these concerns (e.g., Wilcox, 2017, chapter 10). No single estimator is optimal, but one that performs relatively well is the Theil (1950) and Sen (1964) estimator. Peng et al. (2008) established that it can be super efficient when the error term is discontinuous. Its efficiency compares well to OLS under normality and there are situations where the Theil– Sen estimator

has a substantially smaller standard error (Wilcox, 2017). Accordingly, results based on the Theil–Sen estimator are reported plus results based on the quantile regression estimator derived by Koenker and Bassett (1978).

For a single independent variable, the Theil–Sen estimator is computed as follows. For any i < i', for which Xi $\neq$ Xi' , let

$$S_{ii'} = \frac{Y_i - Y_{i'}}{X_i - X_{i'}}.$$

The Theil–Sen estimate of the slope is b1ts, the median of all the slopes represented by S_{iir} . Two strategies for estimating the intercept have been proposed. The first is

$$M_y - b_1 M_x,$$

where My and Mx are the usual sample medians of the Y and X values, respectively. The second estimates the intercept with the median of $Y_1 - b_1 X_1$, . . . , $Y_n - b_1 X_n$. Wilcox (2017) summarizes various strategies for extending the Theil–Sen estimator to more than one independent variable.

As for the Koenker and Bassett (1978) estimator, let

$$\psi_q(u) = u(q - I_{u<0}),$$

where *I* is the indicator function. Then the slope and intercept of the regression line are determined by minimizing

$$\sum \psi_q(r_i),$$

where r_1, . . . , r_n are the usual residuals. For example, if the goal is to predict the .5 quantile of Y , given X, the parameters are estimated by minimizing$\sum$ .5|r_i|, which just the well- known least absolute value regression estimator. If the goal is to predict the .75 quantile of Y given X, this is done by minimizing . W_i|r_i|, where W_i = .25 if r_i < 0 and .75 otherwise.

2.2 Some Estimates of Pα

Table 1 shows some estimates of pα, α = .05, and K = 25 and 100, based on A = 4000 replications and sample sizes n = n1 = n2 ranging from 20 to 500. As can be seen, the

Table 1: Estimates of pα based on 4000 replications

n	K	TS	QREG
20	25	.018	.021
20	100	.019	.022
30	25	.015	.016
30	100	.015	.017
50	25	.014	.015
50	100	.015	.013
100	25	.014	.014
100	100	.015	.015
200	25	.016	.014
200	100	.015	.015
300	25	.015	.013
300	100	.014	.013
500	25	.011	.013
500	100	.014	.016

TS=Theil–Sen

QREG=Quantile Regression

estimates are very stable for n ≥ 30. All indications are that for n ≥ 30 and K ≥ 25, p.05 is approximately equal to .015 when using either the Theil–Sen estimator or the Koenker– Basset quantile regression estimator. (Different results are obtained when using OLS. The estimates are smaller.)

3. Simulation Results

Simulations were used to study the small-sample properties of the method in section 2 with n = 30 and α = .05. Estimated Type I error probabilities were based on 4000 replications.

Four types of distributions were used: normal, symmetric and heavy-tailed, asymmetric and light-tailed, and asymmetric and heavy-tailed.   More precisely, both the error term and

Table 2: Some properties of the g-and-h distribution

g	h	κ_1	κ_2
0.0	0.0	0.00	3.0
0.0	0.2	0.00	21.46
0.2	0.0	0.61	3.68
0.2	0.2	2.81	155.98

the distribution of the independent variable were taken to be one of four g-and-h distributions (Hoaglin, 1985) that contain the standard normal distribution as a special case. If Z has a standard normal distribution, then

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

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 = 0.0), 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). Table 2 shows the skewness (κ_1) and kurtosis (κ_2) for each distribution. Additional properties of the g-and-h distribution are summarized by Hoaglin (1985).

Three choices for λ were used: $\lambda(X) = 1$, $\lambda(X) = |X| + 1$ and $\lambda(X) = 1/(|X| + 1)$. For convenience, these three choices are denoted by variance patterns (VP) 1, 2 and 3. As is evident, VP 1 corresponds to the usual homoscedasticity assumption.

Table 3 summarizes the simulation results. Although the seriousness of a Type I error depends on the situation, Bradley (1978) has suggested that as a general guide, when testing at the .05 level, at a minimum the actual level should be between .025 and .075, and ideally it should be between .045 and .055. All of the estimates are less than .053, but when dealing with heavy-tailed distributions (h = .2) and VP 3, estimates drop below .025. The lowest estimate using Theil–Sen is .02 and the lowest estimate using the quantile regression estimator is .014.

As previously indicated, another approach to controlling the probability of one or more Type I errors is to use the improvements on the Bonferroni method derived by Hochberg

Table 3: Estimated probability of one or more Type I errors, α = .05, n = 30, K = 25 covariate values

g	h	VP	TS	QREG
0.0	0.0	1	.050	.050
0.0	0.0	2	.041	.053
0.0	0.0	3	.032	.032
0.0	0.2	1	.032	.035
0.0	0.2	2	.046	.043
0.0	0.2	3	.023	.017
0.2	0.0	1	.041	.043
0.2	0.0	2	.056	.056
0.2	0.0	3	.032	.033
0.2	0.2	1	.032	.032
0.2	0.2	2	.041	.041
0.2	0.2	3	.020	.014

TS=Theil–Sen

QREG=Quantile Regression

Table 4: Estimated power, α = .05, n = 50, K = 25 covariate values using Theil–Sen

g	h	SMM	Hochberg	JN
0.0	0.0	.673	.669	.825
0.0	0.2	.769	.763	.894
0.2	0.0	.679	.675	.852
0.2	0.2	.742	.740	.894

(1988) or a Studenized maximum modulus (SMM) distribution might be used with infinite degrees of freedom. Next, simulations are used to contrast the power of these methods with method JN in section 2. Consider the situation where both X and the error term have one of the four g-and-h distributions used in Table 3 and the error term is homoscedastic. For the first group both the slope and intercept are zero, the intercept for the second group is also zero but the slope is .8. Table 4 reports the probability of one or more significant results when n = 50 and K = 25. As can be seen, there is little separating the power estimates using the Studentized maximum modulus distribution and Hochberg's method. (Results based on Hommel's method are virtually identical to those based on Hochberg's method.) But clearly method JN can provide a substantial increase in power.

At least in principle, situations can arise where the Hommel and Hochberg methods reject in contrast to the method JN. If, for example, all of the p-values range between .02 and .049, then the Hommel and Hochberg methods reject all K hypotheses at the .05 level, but method JN would fail to reject. But in practice, perhaps such a situation is highly unlikely.

4. Dealing with Two Covariates

This section considers the case of two covariates. The basic strategy aimed at controlling the probability of one or more Type I errors is similar to the approach used in section 2. That is, momentarily assume normality and homoscedasticity and use simulations to estimate pα. But now, the number of covariate points that are used (the number of hypotheses tested) impacts the critical value, as will be illustrated. Consequently, this complicates how best to choose the covariate points where comparisons are to be made. Two strategies are used here with the understanding that many variations of these strategies are reasonable.

Before describing these strategies, first note that there are various methods for measuring the distance of the point X from the center of a data cloud (e.g., Wilcox, 2017), the best- known being Mahalanobis distance. As suggested by Liu and Singh (1997), if U (X) is some distance measure, the depth of X is taken to D(X) = 1/(1 + U (X)). Here a particular variation of a so-called projection-type method is used to measure the depth of a point. Complete computational details are described in Wilcox (2017, section 6.2.5). For brevity, only an outline of the method is described.

For convenience, focus on (X_{1i}, X_{2i}), i = 1, . . . , n1, the covariate points associated with the first group. Let $\hat{\theta}$ be some robust measure of location based on these n1 points. Here the marginal medians are used but there are several other reasonable choices (e.g., Wilcox, 2017, section 6.3). For fixed k, project all n1 points onto the line determined by (X1k, X2k) and $\hat{\theta}$. Denote these points by V_{ik} (i = 1, . . . , n1). Let U_{ik} = $H_{ik}/(q_{ik2} - q_{ik1})$, where H_{ik} is the Euclidean distance between the projection of (X1i, X2i) and $\hat{\theta}$, and where q_{ik2} and q_{ik1} are the upper and lower quartiles based on V_{ik} (i = 1, . . . , n_1). Then the projection distance of (X_{1i}, X_{2i}) is taken to be Gi = max U_{ik}, the maximum taken over k = 1, . . . , n_1, and the corresponding depth is Di = 1/(1 + Gi).

Here, the first strategy for choosing the covariate points is to pick the deepest half of the covariate points among the cloud of all n1 covariate points associated with the first group. That is, use all points corresponding to D_i satisfying $D_i > M_d$, the median of D_1, . . . D_{n1} . The second is to use all of the covariate points associated with the first group. The first strategy is considered here because the estimates of p_α are generally larger compared to   the estimates based on the latter strategy, which raises the issue of whether the deepest half might result in higher power. But an obvious disadvantage of the first strategy is that significant and possibly important differences might be missed due to ignoring half of the covariate points.

The value of pα is more sensitive to the number of hypotheses tested compared to the single covariate case. The focus here is on the quantile regression estimator simply because simulations based on the Theil–Sen estimator result in substantially higher execution times. For the equal sample size case, let n denote the sample size for both groups. For α = .05 and n = 50, and when using the deepest half of the covariate points, the estimate of pα is .013. In contrast, when using all of the covariate points associated with the first group, the estimate is .008.

The value of pα can be estimated via an R function described in the final section of this paper. With a multicore processor and n = 50, execution time is approximately 3.2 minutes

using a quad core 2.5 GHz MacBook pro and the quantile regression estimator in conjunction with the R package parallel. (When using the Theil–Sen estimator, or when a multi-core processor is not available, execution time increases substantially.) To the extent this is a deterrent to using the method, it is noted that for $50 \leq n \leq 300$, the estimate of pα ranges from .013 to .008 when using the deepest half of the covariate points. If, for example, pα = .012 is used when n = 300, the actual probability of one or more Type I errors is estimated to be .066. If this is deemed to be reasonably satisfactory, pα = .012 might be used to reduce execution time. As for using all of the covariate points, the estimate of pα ranges between .008 and .005. A more refined approximation is obtained by noting that for n = 100, 200 and 300, the estimates of pα using the deepest half are .01, .007 and .008, respectively. Using all of the covariate points the corresponding estimates are .008, .005 and.005.

5. Simulation Results for Two Covariates

As was the case for a single covariate, once pα has been estimated, there is the issue of how non-normality and heteroscedasticity impact the probability of one or more Type errors. Here, the results are limited to the quantile regression estimator due to substantially higher execution time associated with the Theil–Sen estimator. Now heteroscedasticity is modeled using $\lambda(X_i) = |X_{i1}| + 1$ (VP 2) and $\lambda(X_i) = 1/(|X_{i1}| + 1)$ (VP 3). As before VP 1 refers to the homoscedastic case. Table 5 reports estimates of the actual probability of one or more Type I errors for n = 50 when using pα = .012 for the deepest half of the covariate points and pα = .007 using all of the points. As can be seen, the estimates never exceed .050. The lowest estimates occur for VP 3 and when dealing with heavy-tailed distributions (h = .2). The lowest estimate is .02.

To provide at least some perspective on power, first consider the case where for the first group $\beta_{11} = \beta_{21} = .7$ and $\beta_{01} = 0$. For the second group $\beta_{12} = \beta_{22} = \beta_{02} = 0$. Under

Table 5: Estimated probability of one or more Type I errors, α = .05, n = 50

g	h	VP	DEEP	C
0.0	0.0	1	.047	.043
0.0	0.0	2	.046	.043
0.0	0.0	3	.041	.037
0.0	0.2	1	.034	.032
0.0	0.2	2	.031	.033
0.0	0.2	3	.023	.021
0.2	0.0	1	.046	.049
0.2	0.0	2	.045	.044
0.2	0.0	3	.035	.032
0.2	0.2	1	.030	.033
0.2	0.2	2	.032	.030
0.2	0.2	3	.023	.020

DEEP=Deepest half used

C=All covariate points used

normality and homoscedasticity and with n = 50, power using Hochberg's method based on the deepest half of the covariate points was estimated to be .444. Using instead the extension of method JN to two covariates, as described in section 4, power was estimated to be .662. (Hommel's method and the Studentized maximum modulus distribution have virtually the same amount of power as Hochberg's method.) Using all of the covariate points, the corresponding estimates were .646 and .806 . The superiority of using all of the covariate points is not too surprising because the smallest differences occur among the deepest points.

Now consider the case where $\beta_{11} = \beta_{21} = 0$ and $\beta_{01} = .6$ and again $\beta_{12} = \beta_{22} = \beta_{02} = 0$. Using only the deepest half of the covariate points, the estimates are now .286 (Hochberg) and .514 (the extension of method JN). So again the extension of method JN has a distinct advantage. Using all of the covariate points, the estimates are .222 and .486. So now, using the extension of method JN in conjunction with the deepest half of the covariate values yields higher power compared to using all of the covariate points, but it might be argued that the increase in power is not that striking.

6. Illustrations

The methods are illustrated using data from the Well Elderly 2 study (Clark et al., 2012) that dealt with an intervention program aimed at improving the physical and emotional wellbeing of older adults. A portion of the study focused on comparing a control group (n = 179) to a group receiving intervention (n = 101) based on a measure of meaningful activities (MAPA). Moreover, a goal was to take into account the cortisol awakening response (CAR), which refers to the change in cortisol concentration that occurs 30-60 minutes after waking from sleep. (Past studies indicate that the CAR is associated with various measures of stress.)

Figure 1 shows a plot of the data. The solid line is the estimate of the regression line for the control group. (Leverage points were removed.) As can be seen, there appears to be little or no difference in MAPA scores when the CAR is close to zero. When the CAR is $-.390$ or $.385$, a significant difference is found when testing at the .05 level.

This last illustration is repeated, only now a second covariate is added: a measure of personal support (PEOP). Figure 2 shows a plot of the covariate points that were used. Covariate points where a significant difference was found are indicated by a +. So there is a significant result at seven covariate points that roughly occur when the CAR is negative and simultaneously the PEOP score is relatively low. (The median PEOP measure was estimated to be 11 and the lower .25 quantile was 9.)

7. Concluding Remarks

There are, of course, many other ways the covariate points might be chosen. The main point is that extant strategies based on robust regression estimators are based on a relatively small number of points, which raises obvious concerns about about detecting and understanding any differences that might exist. All indications are that the probability of one or more Type I errors can be controlled fairly well based on the estimate of $p\alpha$ when using the the deepest half or even all of the covariate points associated with the first group. Moreover, in terms of power, using $p\alpha$ offers a practical advantage over using Hochberg's method, Hommel's
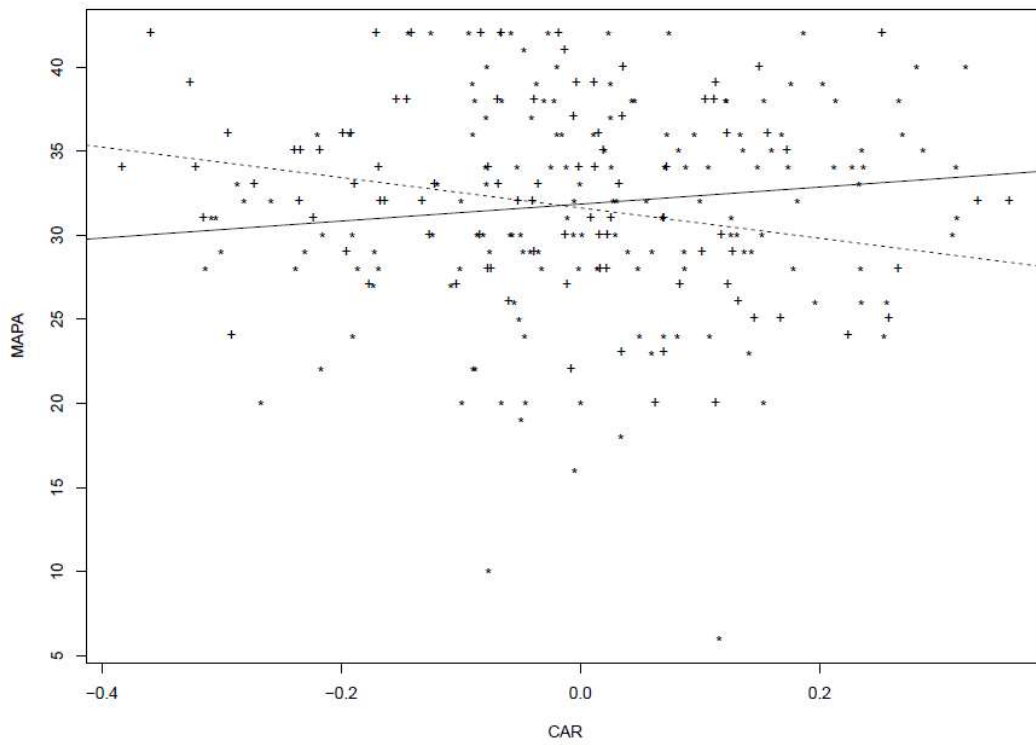
Figure 1: The solid line is the estimated regression line for the control group. Significant differences are found at CAR values −.390 or .385 when testing at the .05 level.
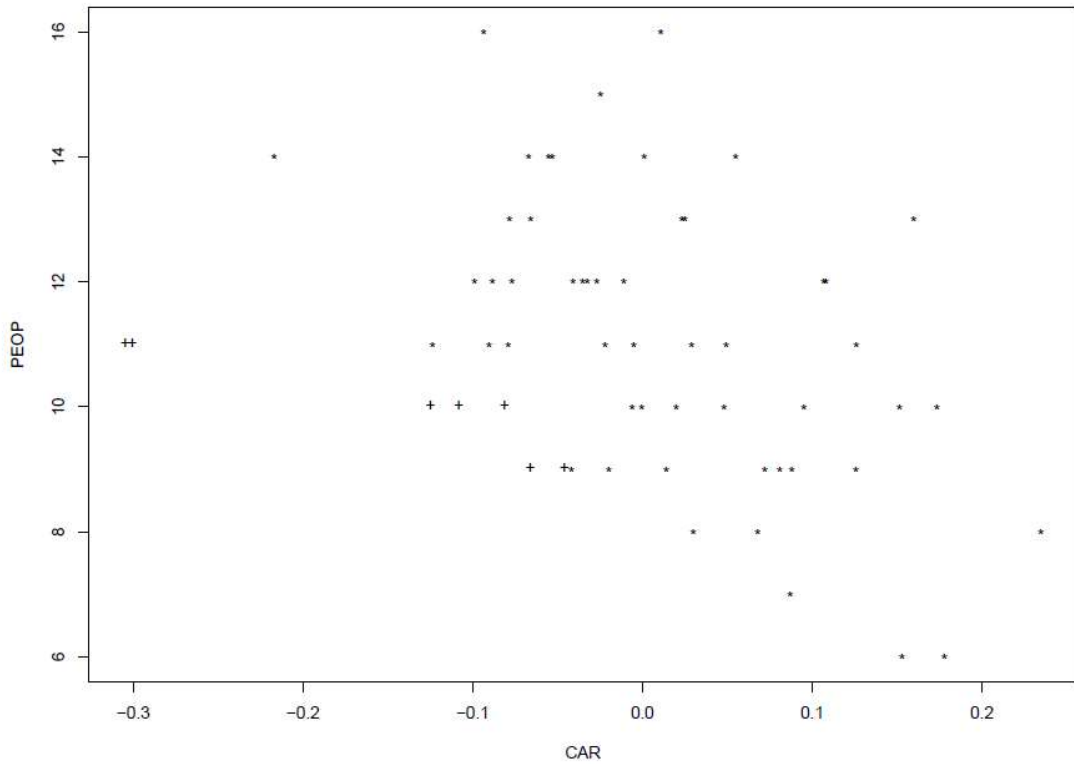
Figure 2: Shown are the covariate points used when comparing the two groups. Significant differences in MAPA scores were obtained for the points indicated by +.

method or the Studentized maximum modulus distribution. In terms of maximizing power, situations can be constructed where using the deepest half of the covariate points is a bit better than using all of the covariate points. But there are situations where using all of the covariate points offers a striking advantage. Consequently, a speculation is that in most situations, using all of the covariate points offers the highest power despite the smaller $p\alpha$ that is used, but more experience with the method is needed to determine the extent this is the case.

Finally, the R functions ancJN and ancJNmp apply the proposed methods and are being added to the R package WRS. The function ancJN is designed for the single covariate case and ancJNmp deals with two covariates. For the two covariate case, the R function ancJNmpcp can be used to estimate $p\alpha$.

References

[1] Bradley, J. V. (1978) Robustness? British Journal of Mathematical and Statistical Psychology, 31, 144–152.

[2] Clark, F., Jackson, J., Carlson, M., Chou,C.-P., Cherry, B. J., Jordan-Marsh, M., Knight, B.G., Mandel, D. Blanchard, J., Granger, D. A., Wilcox, R. R., Lai, M. Y., White, B., Hay, J., Lam, C., Marterella, A., & Azen, S. P. (2012). Effectiveness of a lifestyle intervention in promoting the well-being of independently living older people: results of the Well Elderly 2 Randomise Controlled Trial. Journal of Epidemiology and Community Health, 66, 782–790. doi:10.1136/jech.2009.099754

[3] Efron, B. & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. New York: Chapman & Hall.

[4] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (1986). Robust Statistics. New York: Wiley..

[5] Heritier, S., Cantoni, E, Copt, S. & Victoria-Feser, M.-P. (2009). Robust Methods in Biostatistics. New York: Wiley.

[6] Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distribution. In D. C. Hoaglin, F. Mosteller & J. W. Tukey (Eds.) Exploring Data Tables Trends and Shapes. New York: Wiley, pp. 461–514.

[7] Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. Biometrika, 75, 800–802.

[8] Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. Biometrika, 75, 383–386.

[9] Huber, P. J. & Ronchetti, E. (2009). Robust Statistics, 2nd Ed. New York: Wiley.

[10] Johnson, P. O. & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. Statistical Research Memoirs, 1, 57–93.

[11] Koenker, R. & Bassett, G. (1978). Regression quantiles. Econometrika, 46, 33–50.

[12] Liu, R. G. & Singh, K. (1997). Notions of limiting P values based on data depth and bootstrap. Journal of the American Statistical Association, 92, 266–277.

[13] Maronna, R. A., Martin, D. R. & Yohai, V. J. (2006). Robust Statistics: Theory and Methods. New York: Wiley. Koenker, R.W. and Hallock, K.F. (2001). Quantile regression, J*ournal of Economic Perspectives* **15**, 143-156.

[14] Peng, H., Wang, S. & Wang, X. (2008). Consistency and asymptotic distribution of the Theil– Sen estimator. Journal of Statistical Planning and Inference, 138, 1836–1850.

[15] Rousseeuw, P. J. & Leroy, A. M. (1987). Robust Regression & Outlier Detection. New York: Wiley.

[16] Sen, P. K. (1968). Estimate of the regression coefficient based on Kendall's tau. Journal of the American Statistical Association, 63, 1379–1389.

[17] Staudte, R. G. & Sheather, S. J. (1990). Robust Estimation and Testing. New York: Wiley.

[18] Theil, H. (1950). A rank-invariant method of linear and polynomial regression analysis. Indagationes Mathematicae, 12, 85–91.

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

[20] Wilcox, R. R. (2013). A heteroscedastic method for comparing regression lines at specified design points when using a robust regression estimator. Journal of Data Science, 11, 281– 291.