



April 2004

# Measuring the Effects of Multiplicity: A Study of Multiple Comparison Procedures in Multiple Linear Regression

Paul Borochin

*University of Pennsylvania*, [pborochi@wharton.upenn.edu](mailto:pborochi@wharton.upenn.edu)

Follow this and additional works at: [http://repository.upenn.edu/wharton\\_research\\_scholars](http://repository.upenn.edu/wharton_research_scholars)

---

Borochin, Paul, "Measuring the Effects of Multiplicity: A Study of Multiple Comparison Procedures in Multiple Linear Regression" (2004). *Wharton Research Scholars*. 2.

[http://repository.upenn.edu/wharton\\_research\\_scholars/2](http://repository.upenn.edu/wharton_research_scholars/2)

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/wharton\\_research\\_scholars/2](http://repository.upenn.edu/wharton_research_scholars/2)

For more information, please contact [libraryrepository@pobox.upenn.edu](mailto:libraryrepository@pobox.upenn.edu).

---

# Measuring the Effects of Multiplicity: A Study of Multiple Comparison Procedures in Multiple Linear Regression



**Measuring the Effects of Multiplicity:  
A Study of Multiple Comparison  
Procedures in Multiple Linear Regression**

By –Paul Borochin
Advisor – Abba Krieger
Wharton Undergraduate Research Scholars WH-299-301
April 2004

## **1. Introduction**

The simultaneous testing of multiple hypotheses arises in all forms of modern research. For example, a drug manufacturer is considering the effect of several chemical compounds on the observed outcome, such as the curing of a disease or the stimulation of an organ. One can statistically measure the effect of a factor on an outcome through hypothesis testing, which checks whether the factor is statistically different from the null hypothesis, which is usually zero, connoting no effect. The drug manufacturer would perform this test for all the chemicals used and look for ones that are statistically different from the null hypothesis, that is, ones that have a significant effect on the observed outcome. The ones that are significantly different could be the next medical breakthrough, so the importance of the accuracy of this test is paramount.

A marketer could look for his target audience by studying which demographics drive consumer interest in a specific product. He would test the effect of income, family size, age, gender, and any other possible demographic for significant effect in the consumption of a product, and look for statistically significant ones which would drive his future efforts.

Multiple hypothesis testing is used in a multiple linear regression setting in which the researcher attempts to predict the observed dependent variable  $Y$  by a set of independent factors  $X$ . Each  $X$  is evaluated for significance in a statistical hypothesis test that measures the likelihood of each factor being related to the observed outcome of interest  $Y$ .

The literature, specifically in the social and medical sciences, contains many applications of multiple regression, logistic regression, general linear models or analysis

of variance; for our purposes the issue of multiple comparisons arises in all models of this type. For example, the Framingham study attempts to determine risk factors for heart attacks. These risk factors include weight, smoking and hypertension. Other factors such as age, gender, and family history are included as well. Determining whether each of these factors is related to heart disease results in many hypothesis tests. Since there are many factors, the issue of multiple tests arises.

We consider below a multiple regression problem in the context of a financial analysis. Multiple linear regressions as a subset of multiple simultaneous hypothesis testing are used very frequently in the development of new financial models that seek to explain returns on assets in terms of a number of factors. For example, some of the multifactor models that have come to replace the CAPM and lead to multiple hypothesis testing as different factors such as firm size, book/market value, market risk premium, and many others are evaluated for contributing to the performance of returns.

In all these cases, the process involves the collection of data on the possible factors, observation of the outcomes (e.g. market returns), and statistical hypothesis tests for each of the factors considered. Each factor can be described by one of two hypotheses: the null, if the factor doesn't contribute to the outcome of the model, and the alternative if it does have some effect. The statistical test considers whether each individual factor is statistically significant, thereby rejecting the null hypothesis that asserts that this variable has no relation at all (or in linear regression, no linear relation) to the observed outcome.

In the next section, we describe the mathematical setting. There are two subsections, one describing the issue of multiplicity and the other providing background

on regression. Section 3 motivates the research by describing a case study. The simulation is described in Section 4. The results of the simulation are presented in Section 5. Conclusions and areas of future research are presented in the final section. Details of the simulation and output, as well as the C code used to generate the results appear in the appendix.

## **2. Mathematical Setting**

Any research effort that seeks to test multiple possible factors for evidence of contributing to the observed effect, from drug manufacturers evaluating the effects of different chemicals to economists testing various predictors, has two objectives in its hypothesis tests: it must control their testing procedures to avoid false positives, factors that appear to be significant by chance alone, and yet be able to reliably reject the null hypothesis for truly significant factors. In the former case, failing to retain the null hypothesis for a factor that has no connection with the model is called a Type I error whereas in the latter case, failing to reject the null hypothesis for a factor that contributes to the model's outcome is referred to as a Type II error. The probability of not making a Type II error is called the statistical power of a procedure. There is a tradeoff between a liberal procedure that makes it easy to reject null hypotheses for all factors, thus minimizing the likelihood of making a Type II error at the cost of increasing the likelihood of false discoveries and thus a higher probability of a Type I error, and a conservative procedure that minimizes false discoveries at the cost of reducing the number of true ones as well. The performance of a multiple comparison procedures at controlling these errors can be represented in a 2x2 plot as follows:

	True Hypothesis state	
	H0	H1
Retain H0	<b>S</b> (Correct)	<b>T</b> (Type II error)
Reject H0	<b>U</b> (Type I error)	<b>V</b> (Correct)

An effective multiple comparison procedure will retain as many true nulls  $H_o$  as possible, which is represented by S, and reject as many nulls as possible in the case of a true alternative hypothesis,  $H_a$ , which is represented by V. The values T and U are the number of Type II and Type I errors, respectively, and should be minimized. By looking at these four summaries one can easily evaluate the performance of a multiple comparison procedure since one wants to maximize the number of correct classifications of hypotheses as alternative and null (V and S respectively) and minimize the number of classification errors that are made.

One area in which the testing of multiple hypotheses takes place is in multiple linear regression, which predicts the outcome based on a set of factors. Multiple linear regression is an extension of simple linear regression which attempts to predict an outcome based on the input of factors, by  $E(Y | X_1 X_2 X_3 \dots X_p)$  which the expected value of the outcome is given the respective X. The model for multiple linear regression is

$$Y_i = \alpha + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \dots + \beta_p * X_{pi} + \epsilon_i$$

where the betas are the coefficients assigned to the factors X (i.e. the beta reflects the average increase in Y for each additional unit increase in X holding all other Xs constant), and alpha is the intercept term (i.e. the average value of Y when all of the X variables are zero). That part of Y which is not fit by the X factors is captured by the disturbance term that represents the difference between the theoretical regression line,

$E(Y | X_1 X_2 X_3 \dots X_p)$ , and the true value of Y which may perhaps be due to other factors that are not considered in the model since the equation of the model can be rewritten as



$$Y_i - (\alpha + \beta_1 * X_{1i} + \beta_2 * X_{2i} + \dots + \beta_p * X_{pi}) = \varepsilon_i$$

For a linear regression to be valid, it has to satisfy a set of assumptions. The error terms all have to be independent of each other. If the error terms follow a pattern it is indicative that the model is missing an important factor that isn't capturing all the variability in the outcome or the relationship between the factors and the outcome is not linear. The error terms also have to be homoscedastic, meaning that they should have the same variance  $\sigma^2$  for all error terms. If this is not so, it also means that the model isn't perfect, perhaps because of non-linearities which can be fixed with a transformation of the factors and outcomes using any of a number of functions (the natural logarithm being a common choice). The error terms also have to be normally distributed about 0. The violation of one or some of these assumptions doesn't immediately invalidate all the results of the regression, but these assumptions are essential in our context as we are going to consider the performances of tests for the significance of coefficients. The resulting t-tests rely heavily on the above assumptions.

Once the data are collected the parameters of the model, namely the intercept, slopes and variance of the error term can be estimated using classical results. Specifically, for an X matrix (an N by P+1 matrix where N is the number of observations and P+1 is the number of factors considered including a column of 1s for the intercept term) the coefficients of the slope and intercept are found from

$$\hat{\beta} = (X'X)^{-1} X'Y$$

Once  $\hat{\beta}$  is known, then the predicted value of  $Y_i$ ,  $\hat{Y}_i = \hat{\beta}_i X_i$  can be found. Ultimately, an estimate of  $\sigma^2$  is calculated from

$$S_{error}^2 = \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N - (P + 1)}$$

Multivariable regressions seek to identify factors that contribute to the outcome being observed and separate them from the factors that have no effect on the outcome. A measure of significance of a variable in a regression is commonly obtained from the t-statistics derived from the coefficients of the regression. This t-statistic measures the number of estimated standard deviations that the estimated coefficient  $\hat{\beta}$  is from the hypothesized value under the null, typically taken to be zero. Since it is well-known that

$$Var(\hat{\beta}_i) = \sigma^2 c_{ii}$$

where  $c_{ii}$  is the  $i, i^{\text{th}}$  element of  $(X'X)^{-1}$ , and the estimated value is then  $S_{error}^2 c_{ii}$  and hence

$$T_i = \frac{\hat{\beta}_i}{S_{error} \sqrt{c_{ii}}}$$

Under the null hypothesis this is known to have a T-distribution with  $N-(P+1)$  degrees of freedom. Since the T-distribution has known density the p-value can be determined. In the context of multiple regression, the p-value for a variable  $X$  can be interpreted as the probability of observing the estimated slope  $\hat{\beta}_i$  or one larger in magnitude under the null hypothesis (i.e. the slope is actually zero). If the p-value is less than a specified  $\alpha$  we reject the null hypothesis in favor of the claim that  $X$  has some effect on the outcome  $Y$  (see Johnson pp. 121-170 for a more thorough discussion of inference in multiple regression). However, since in many cases a significantly large number of variables are being considered simultaneously, the problem of statistical multiplicity arises. As the

number of factors simultaneously examined increases, so does the probability of making a Type I error; it becomes far larger than the probability implied by the alpha-level at which the test is nominally carried out. For example, if 20 simultaneous comparisons are carried at alpha-level .05 then the probability of making a Type I error is 5% per variable but for the whole model the expected number of variables found to be significant would be

$$E(\text{significant}) = 20 * \frac{1}{20} = 1$$

assuming no variable is significant. Over fitting a model by including too many variables can lead to mistaken conclusions due to this effect. For example, if one were to fit the NYSE daily returns by rainfall amounts from twenty or more locations worldwide at a 5% alpha-level one could expect to find at least some statistically significant variables even though they are of course not related in any way. However, if the statistical tests are followed without consideration for this effect one would have little choice but admit that rainfall in the Amazon may well be a predictor of the market in New York.

A more realistic example can be drawn from the study of genetics in which there are thousands of possible gene locations on a strand of DNA that could be responsible for the genetic feature being studied. If a simultaneous hypothesis test is performed for such a large set of factors, treating each test singly and simply comparing it to a desired overall alpha-level, say 5%, will invariably lead to a great number of false discoveries due to simple chance. When comparing large numbers of factors simultaneously, as is often necessary to create models complex enough to model the observed phenomenon accurately, it is easy to find false significant factors. This problem can lead researchers

to make false statements about the significance of certain variables they are examining if it is not accounted for.

There are several criteria for evaluating the performance of a simultaneous hypothesis test. The ones most commonly used in the literature are the Familywise Error Rate, the False Discovery Rate, and the False Acceptance Rate (Hochberg and Tamhane). The Familywise Error Rate (FWER) is the probability of making any false discovery, the likelihood of  $U$  being greater than zero for the set of multiple hypothesis tests. The False Discovery Rate (FDR) is the rate at which these false discoveries are made in relation to total discoveries,  $\frac{U}{U+V}$ . This represents the importance of making real discoveries, since a few false ones may be a fair price to pay for a greater number of true discoveries. Thus, a hypothesis test that makes two false discoveries in 10 (FDR=.2) is clearly worse than one that makes two false discoveries in 100 (FDR=.02). Most of the time an experiment with a low rate of false discoveries, as the one in the second case, would seem to be very much worth conducting since the proportion of false to true discoveries is quite attractive due to the high number of true discoveries made.

Analogous to the FDR is the FAR which looks at a similar measure with a focus on the retained hypotheses. Specifically, retaining a small fraction of null hypotheses erroneously  $\left(\frac{T}{T+S}\right)$  is also acceptable if the test retains many others correctly. A test that fails to reject 2 significant hypotheses out of a total of 10 retained ones would have an FAR of .2 while one that failed to find 2 out of 100 retained would have an FAR of .02 and would thus be vastly preferable. Note that by definition the FDR and FAR are

assumed to be zero when no hypotheses are rejected ( $\mathbf{U}+\mathbf{V}$ ) =0 or no hypotheses are retained ( $\mathbf{T}+\mathbf{S}$ ) =0 respectively.

There are several approaches to handling multiple simultaneous hypothesis testing, the most frequent of which is ignoring the effects of multiplicity. This approach takes t-statistics and their corresponding p-values at face value when testing for significance, and is attractive to researchers because it does not decrease statistical power of the tests by adjusting the alpha-level down as other multiple comparison procedures do. Of course, this approach provides no protection against false discoveries, and thus can lead to claims of statistical significance when none is present if large numbers of factors are examined simultaneously as in the case of a genetic study, or if the model is over fit as in the case of regressing NYSE returns by rainfall patterns.

Another well-known approach is the Bonferroni p-value adjustment which controls the FWER. This approach imposes a limit on the FWER by dividing the alpha level by the number of variables being considered. As in the previous case when 20 variables are simultaneously considered, each of the variables' p-value would be tested not at the alpha-level, but at

$$\alpha/20$$

thus making the overall FWER at alpha-level. Unfortunately, there is a tradeoff between false positive control and the power of the statistical method, or the type II error. By increasing the rejection region by dividing the alpha-level by the number of factors one increases the likelihood that a non-significant factor will be rejected due to the heightened threshold. However, by making it difficult to find effects that aren't really there, the approach also makes it difficult to find true effects as well. If the number of

factors is truly large, such as it would be in a genetics study where upwards of 10000 possible gene locations may be examined for contribution to a certain trait, the alpha-level would be so drastically low as to virtually eliminate the likelihood of finding any significant effects whatsoever. Since Bonferroni has a strong detrimental effect on the statistical power of the multiple comparison, it is seldom used in research in these contexts.

A newer approach to controlling false positives is accomplished by focusing on the FDR. This rate is the same as the FWER in the case that all null hypotheses are true, that is, if there is no correlation between the factors and the effect, and is smaller or equal to the FWER if one or more factors really vary with the effect. The FDR does not control the probability of making a false discovery directly, but instead focuses on the relative proportion of false discoveries, false rejections of the null hypothesis, as compared to the total number of null hypotheses rejected. An example of an FDR controlling procedure is the linear step-up procedure which ranks the p-values corresponding to the variables being tested from lowest to highest and compares them against progressively higher, and thus more tolerant, alpha levels. The lowest p-value is compared against a Bonferroni-equivalent critical level of  $\frac{\alpha}{P}$  where P is the number of tests, however the next lowest p-value is compared against a critical level of  $2 * \frac{\alpha}{P}$ , the third against  $3 * \frac{\alpha}{P}$ , and so on until the last is compared against  $P * \frac{\alpha}{P}$  or just the regular alpha level. All the variables ranked lower than the one with the highest p-value to pass this test are considered significant in a one-step step-up approach. The one-step approach controls the FDR at  $\frac{m_o}{m} * \alpha$  where  $m_o$  is the number of true null hypotheses and  $m$  is the total

number of hypotheses (Benjamini and Hochberg 5). Thus a one-step approach with 50% true nulls would actually control the FDR at  $\frac{\alpha}{2}$ , a more conservative level than would be otherwise desired. The two-step approach is an improvement upon the one-step, continuing from where it left off by adjusting the alpha level by the index of the highest p-value found significant in the one-step case. The adjusted alpha is obtained using the index K of the highest significant p-value

$$\alpha' = \frac{P * \alpha}{P - K}$$

and the linear step-up procedure is redone again using the new, and still more tolerant, alpha level  $\alpha'$ . It in effect uses the one-step method's results to approximate the number of true nulls and readjusts the alpha-level to the level desired for the overall test. Thus, in the case of 50% true nulls, the alpha-level used in the two-step procedure would be .10 which would control the FDR at a level of .05 as desired. Because of the nature of the FDR approach it doesn't reduce statistical power as greatly as FWER approaches and yet provides a way to handle the problem of multiplicity. However, it doesn't give any indication as to which of the discoveries are the false ones, just the proportion of false to true ones, making it ill-suited for use in situations when one has to be sure that a variable is significant.

### 3. Case Study

A good case study for the importance of the appropriate use of multiple comparison procedures may be had in the paper “Arbitrage Risk and the Book-to-Market Anomaly” (Ashiq Ali, Lee-Seok Hwang, and Mark A. Trombley, Journal of Financial Economics, Aug 2003) about the book-to-market (B/M) effect being greater for stocks with higher return volatility, higher transaction costs, and lower investor sophistication, as well as the book-to-market effect’s property of being higher for high-volatility stocks than for the low-volatility ones. The book-to-market ratio is the relationship between the actual value of a firm’s equity and the value that is implied by its market price which can be used to identify overvalued or undervalued stocks. They test for the statistical significance of these effects, and also for the significance of cross-correlation terms between the individual effects and the B/M effect to see if any are positively correlated. The large number of cross-terms in addition to the factors themselves produces a sizable multiple regression with a total of 15 factors tested simultaneously. To demonstrate the relationship between B/M and stock returns, the researchers ran a regression on stock returns with the following variables:



Variable	Coefficient	T-Statistic	Level of Sig.	P-Value	Bonferroni .05	Step- Up .05
Beta	0.174	4.879	0.01	0.00000107	0.003333	0.003333
Intercept	-0.355	-4.865	0.01	0.00000114	0.003333	0.006667
Ln(Volume)*(10 <sup>-2</sup> )	-7.758	-4.855	0.01	0.0000012	0.003333	0.01
Price*(10 <sup>-2</sup> )	0.45	3.16	0.01	0.001578	0.003333	0.013333
Ln(ME)*(10 <sup>-2</sup> )	4.429	2.978	0.01	0.002901	0.003333	0.016667
B/M	0.078	2.799	0.01	<b>0.005126</b>	<b>0.003333</b>	<b>0.02</b>
B/M*(Involatility <sup>(-1)</sup> )*(10 <sup>-2</sup> )	-0.147	-2.589	0.01	0.009626	0.003333	0.023333
Analysts*(10 <sup>-2</sup> )	0.655	2.498	0.01	0.01249	0.003333	0.026667
B/M*Ln(ME)*(10 <sup>-2</sup> )	2.152	1.746	0.1	0.080811	0.003333	0.03
(Involatility <sup>(-1)</sup> )*(10 <sup>-2</sup> )	0.183	1.68	.	0.092957	0.003333	0.033333
B/M*Analysts*(10 <sup>-2</sup> )	-0.357	-1.305	.	0.191893	0.003333	0.036667
(Zerofreq <sup>(-1)</sup> )*(10 <sup>-2</sup> )	-0.287	-0.842	.	0.399788	0.003333	0.04
B/M*(Price)*(10 <sup>-2</sup> )	0.119	0.787	.	0.431282	0.003333	0.043333
B/M*(Zerofreq <sup>(-1)</sup> )*(10 <sup>-2</sup> )	-0.413	-0.581	.	0.56124	0.003333	0.046667
B/M*Ln(Volume)*(10 <sup>-2</sup> )	-0.054	-0.049	.	0.960919	0.003333	0.05

The researchers then looked at the resulting p-values individually and asserted, most importantly, that B/M was a statistically significant factor at the 99% confidence level, and thus was virtually certain to affect stock returns. However, this was done without taking the effects of multiplicity into consideration and so this conclusion could suffer from the overfitting problem described earlier in which one could get significant-seeming variables purely by chance if enough are used. Their conclusion would vary if multiple

comparison procedures were used to analyze their results. If the classical FWER Bonferroni procedure was used, then B/M would not be judged as significant even at the lower 95% confidence level since there were fifteen parameters and so the target p-value would be  $\frac{.05}{15} = .0033$ . At the 5% confidence level Bonferroni would reject the null for only the five factors with the lowest p-values. This illustrates the problem with such approaches and the reason that social science researchers are reluctant to use them as the decrease in power is just too great. If the one-step linear step-up procedure controlling the FDR at the .05 level were used, the critical alpha-levels would be

$$\frac{.05}{15}, \frac{2 * .05}{15}, \frac{3 * .05}{15} \dots \frac{15 * .05}{15}$$

for the factors sorted by lowest p-value. This approach would reject the null hypothesis for seven factors with the lowest p-values including B/M achieving the same statistical power as the original approach of ignoring multiplicity while at the same time controlling the rate of false discoveries and thus accounting for the problem of multiplicity, demonstrating the difference between the FWER and FDR approaches. The thesis of this paper depends on whether the book-to-market effect is found to be significant or not, which varies greatly depending on which multiple comparison procedure is used. It is significant at a confidence level of 1% if multiplicity is ignored, at 5% if the False Discovery Rate procedure is implemented and not significant at either if the Bonferroni method for controlling the Familywise Error Rate is used. Because of the importance that multiplicity and multiple comparison procedures have on the outcomes of research it is important to observe the performance of the different methods under experimental

conditions in order to determine the optimal course of action under specific sets of circumstances.

#### 4. Simulation

In this section, we describe a simulation that considers various regression problems and describes how the approaches indicated above perform in rejecting alternative hypotheses and retaining null hypotheses. We begin the section with a description of how the regression problems are simulated. We then discuss how the p-values are calculated. Finally the various approaches and criteria are provided.

As mentioned above, a regression problem begins with an N by P matrix of predictors X. Since the performance of various methods depend on the independence or lack thereof of the resulting p-values it is important to consider cases where there is some collinearity among the Xs.

In our simulation three parameters govern the generation of our X matrix:

- 1) N – the number of observations, which is chosen from a set {100, 500, 1000, 2000}
- 2) P – the number of variables in the regression, taken from a set {8, 16, 32, 64}
- 3) The correlation  $\rho$  between the Xs selected from {0, 1/4, -1/4, 3/4, -3/4}

The correlation between the Xs for each observation is created by generating a common term  $X_0$  from a N (0, 1) distribution and for each  $X_i$  also from N (0, 1) establishing the following relationship:

$$X_i = \sqrt{\rho}X_0 + \sqrt{1-\rho}X_i \text{ if } \rho > 0$$

and

$$X_i = -1^{i-1}\sqrt{\rho}X_0 + \sqrt{1-\rho}X \text{ if } \rho < 0$$

More critically, Y needs to be generated to exhibit certain properties. First of all, the number of null and alternative hypotheses is controlled by the fourth parameter

4) F = fraction of true null hypotheses of the total number of hypotheses tested, which is drawn from  $\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\}$ . The first P(1-F) Xs in each observation are made the true alternatives and they contribute to the generated Y while the remaining PF Xs are true null and have no effect on the generation of the Y.

In addition, the size of the effect is controlled by a coded parameter:

5) The coefficient parameter is selected from a set  $\{1,2\}$  where 1 stands for effects of equivalent magnitude for all alternative factors and 2 stands for linearly increasing effects for the same. In the first case  $\beta_1 = \beta_2 = \dots = \beta_{P(1-F)} = 1$  while in the second case  $\beta_i = 2i$  for  $i=1, 2, \dots, P(1-F)$ . The coefficient is always zero for true null factors by definition.

Finally, the strength of the regression is controlled by  $R^2$ . We let  $R^2$  vary according to

6) The  $R^2$  parameter is selected from a set of  $\{10\%, 30\%, 50\%, 70\%, 90\%\}$  and affects the regression through  $\sigma^2$  using the formula

$$\sigma = \sqrt{\frac{C(1-R^2)}{R^2(N-P)}}$$

where

$$C = \sum_{i=1}^P \sum_{j=1}^P \beta_i \beta_j X_{ij}$$

For any configuration (i.e. setting of the above six parameters, see appendix 2 for details) we generate the X and Y matrices 100 times. For each replication we compute

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (\text{as indicated above})$$

and

$$\hat{Y} = \sum_{j=1}^P \hat{\beta}_j X_{ij}$$

$$S_{error} = \sum_{i=1}^N \sqrt{\frac{(Y_i - \hat{Y}_i)^2}{N - (P + 1)}}$$

Finally, we compute for each variable X:

$$t = \frac{\hat{\beta}_i}{S_{error} * c_{ii}}$$

where  $c_{ii}$  is the (i, i)<sup>th</sup> entry in  $(X'X)^{-1}$ .

An algorithm from Abramowitz and Stegun describes how to convert the t-value into a p-value for performing the hypothesis test:

$$p\text{-value}(t | v) = \begin{cases} \frac{2}{\pi} \{ \theta + \sin \theta [\cos \theta + \frac{2}{3} \cos^3 \theta + \dots + \frac{2 * 4 * \dots * (v-3)}{1 * 3 * \dots * (v-2)} \cos^{v-2} \theta] \} & \text{for } (v > 1 \quad \text{and} \quad \text{odd}) \\ \frac{2}{\pi} \theta & \text{for } (v = 1) \end{cases}$$

and

$$p\text{-value}(t | v) = \begin{cases} \sin \theta \{ 1 + \frac{1}{2} \cos^2 \theta + \frac{1 * 3}{2 * 4} \cos^4 \theta + \dots + \frac{1 * 3 * 5 * \dots * (v-3)}{2 * 4 * 6 * \dots * (v-2)} \cos^{v-2} \theta \} & \text{for } (v \quad \text{even}) \end{cases}$$

where v is the measure of the degrees of freedom, N-(P+1) and

$$\theta = \arctan\left(\frac{t}{\sqrt{v}}\right)$$

The result of each regression run will be a set of p-values equal in number to the number of predictors. We then apply the four methods discussed above that determine the statistical significance of each predictor based on the p-values provided:

- 1) Ignoring Multiplicity
- 2) Bonferroni
- 3) One-Step Adaptive Approach
- 4) Two-Step Adaptive Approach

Since we know which hypotheses are true null ( $\beta = 0$ ) and which are alternative ( $\beta \neq 0$ ) we can fill in the 2x2 table mentioned above:

	True Hypothesis state		TOTAL
	H0	H1	
Retain H0	<b>S</b>	<b>T</b>	<b>A</b>
Reject H0	<b>U</b>	<b>V</b>	<b>R</b>
	$m_0=PF$	$m_1=P(1-F)$	

These statistics are aggregated across the 100 replications to produce estimates of the following criteria:

- 1)  $FWER = P(U > 0)$
- 2)  $FDR = E\left(\frac{U}{R}\right)$  where  $FDR=0$  if  $R=0$
- 3)  $FAR = E\left(\frac{T}{A}\right)$  where  $FAR=0$  if  $A=0$
- 4) Fraction of nulls rejected  $E\left(\frac{U}{m_0}\right)$

5) Fraction of alternatives retained  $E\left(\frac{\mathbf{T}}{m_1}\right)$

Based on these criteria the performance of each of the four multiple comparison procedures can be evaluated in each of the 20 combinations of parameters representing 20 different regression problems.

## 5. Results of Simulation

The results from the simulation can be used to make some powerful statements about the performance of multiple comparison procedures relative to each other, and to illustrate the importance of using multiple comparison procedures instead of simply ignoring the effects of multiplicity. The case for multiple comparison procedures is well illustrated by run number 9 of 20, in which the regression was performed using 500 observations with 32 possible factors of which 24 were true nulls, a correlation of  $-\frac{3}{4}$  between the factors, linearly increasing coefficients, and an  $R^2$  of 50%. This simulation has relatively few observations, an average  $R^2$  and a sizeable number of factors under consideration, meaning it can be likened to an exploratory study of a phenomenon in order to establish a future model. The output of the model shows the shortcomings of ignoring multiplicity:

MCP	U/R	T/A	Avg U/m0	Avg T/m1	Avg FWER
Ignore Multiplicity	0.122663	0.139725	0.032917	0.473750	0.42
Bonferroni	0.008333	0.212037	.000833	0.810000	0.02
Step1	0.023667	0.196953	.00375	0.738750	0.08
Step2	0.025119	0.195979	.004583	0.733750	0.08

Ignoring multiplicity results in many false discoveries being made by pure chance, giving a high FDR (**U/R**) and a high FWER for ignoring multiplicity. Around 12% of the discoveries that would be made using the approach would be false ones, too large for most research considerations. Although more true factors are also discovered under this approach, resulting in the lowest FAR (**T/A**) of the set, this benefit is offset by the disproportionately large number of false discoveries. By making it easier to discover true



effects this approach also discovers false ones far more readily, invalidating the results of the experiment since more than one in ten of its discoveries are false. The Bonferroni approach performs as expected, displaying the lowest FWER but paying for such precision in Type I error control through an increase in Type II error seen in its FAR of .212, the highest of the set. The Two-Step approach provides a desirable middle ground, controlling the rate of false discoveries at a very acceptable, if lower than expected, .025 level, and yet discovering far more true effects than the Bonferroni approach. Making many false discoveries in this case would lead the researcher to pursue false leads and come up with faulty models, illustrating the importance of avoiding Type I errors.

Run 8 of 20 provides further evidence in support of the Two-Step approach over Bonferroni. This run was conducted with 100 observations, 8 factors of which 2 were true nulls, a correlation of  $-\frac{3}{4}$ , equal coefficients for false-null factors, and an  $R^2$  of .9. This could be likened to a more developed study in which the model has been narrowed to a few variables which explain the outcome well, resulting in a high  $R^2$ . The difference between Bonferroni and the Two-Step method can be seen from the results of this simulation:

MCP	U/R	T/A	Avg U/m0	Avg T/m1	Avg FWER
Ignore					
Multiplicity	0.007143	0.016667	0.025	.00833	0.05
Bonferroni	0	0.087667	0.00	0.05	0
Step1	0.005714	0.021667	0.02	0.0117	0.04
Step2	0.034286	0.006667	0.125	0.0033	0.21

The Bonferroni approach is too conservative in this case, failing to reject almost 10% of the false nulls. Again its tradeoff between Type I and II is evident, given the approach's perfect control of FWER at the expense of FAR. The Two-Step approach has an FAR that is an order of magnitude less than that of the Bonferroni approach yielding more true discoveries while controlling the FDR below the desired 5%. In this case the Two-Step would allow the researcher to test the model far more efficiently by discovering the true effects far better than under Bonferroni, which would be essential when testing a carefully-built model with a few variables, while handily controlling the rate of false discoveries. Ignoring multiplicity performs well in this case too, given the small number of variables under consideration. However, its performance would degenerate rapidly if more factors were added to the model.

Run 15 of 20 further provides evidence in support of the Two-Step method's ability to provide an alternative that accounts for both Type I and Type II errors with good accuracy and reliability. This regression was run with 500 observations and 64 factors of which 32 were true null and the rest had coefficients equal to one with no correlation between the factors and an  $R^2$  of .9. This regression's makeup places it in the middle between the earlier two since it incorporates a large number of factors of which a large fraction are extraneous but has little background noise suggesting that the model is well-developed.

MCP	U/R	T/A	Avg U/m0	Avg T/m1	Avg FWER
Ignore					
Multiplicity	0.040544	0.050894	0.04156	0.0528	0.68
Bonferroni	0	0.27206	0.00	0.3778	0
Step1	0.020241	0.08591	0.0194	0.095	0.43
Step2	0.038825	0.053012	0.04	0.05594	0.63

In this regression the Bonferroni method shows dramatically poor performance with a FAR of just a little under 30% since the alpha-level is adjusted down strongly due to the large number of variables considered. The Two-Step method still controls FDR at below the target 5% and at the same time provides a much more preferable FAR of 5.3%.

## 6. Conclusions and Future Research Improvements

The Two-Step method thus appears to be a multiple comparison procedure that is capable of balancing Type I and II errors and thus avoiding the pitfalls of specializing in either. Its focus on the ratio of true discoveries to false makes it preferable to exploratory studies where the discovery of a true effect is more valuable than the mistake of claiming a false one to be significant. In tests where it is important to be sure that all discoveries made are true ones, such as in a refined model with few factors all of which are of interest it may be preferable to use an FWER approach like the Bonferroni method.

In the future, this study could be refined by a more thorough calibration of parameters. It was only conducted with 20 combinations, but it may be of interest to investigate the performance of the MCPs by altering the combinations already examined or creating new ones. For example, both runs 8 and 15 discussed here were performed at an  $R^2$  of .9 which is a very high level of precision. It would be informative to observe how the performance of the MCPs would change if the noise in the regression were higher.

These regressions were run with 100 replications, but it also may be more informative to run them with more replications to increase precision. 1000 replications would be preferable, though it wasn't feasible under the time constraints of this project and the performance of the C routine. To increase the number of replications in a reasonable time the algorithm would have to be optimized for better performance.

Given the data obtained from the simulation, it may be worthwhile to perform data analysis to determine possible connections between the parameters and the criteria of evaluation. For example, it would be interesting to find which parameter is most

influential in determining an MCP's performance in controlling the FWER, or the FDR, or the FAR. Ideally it would be optimal to develop a formal proof tying a parameter to a criterion, but this is certainly beyond my capabilities. Furthermore, it is interesting to further examine the influence of multiplicity on social science research as was done in the sample study presented in this paper. Re-evaluating the findings of social science research in light of the different multiple comparison procedures could further expose the importance of the problem of multiplicity and draw attention to the need to control its effects.

## References

- 1) Benjamini, Y. and Hochberg Y. Royal Statistical Society “Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing”, 1995
- 2) Benjamini, Y., Krieger, A., and Yekutieli, D. (2003) Adaptive Linear Step-up Procedures that control the False Discovery Rate. 1-7.
- 3) Ashiq Ali, Lee-Seok Hwang, and Mark A. Trombley, Journal of Financial Economics, “Arbitrage Risk and the Book-to-Market Anomaly”, Aug. 2003
- 4) Johnston, J. *Econometric Methods*, 2<sup>nd</sup> edition. McGraw-Hill Book Company, NY, 1963.
- 5) Abramowitz and Stegun. *Handbook of Mathematical Functions*. US Dept. of Commerce, National Bureau of Standards, Applied Mathematics Series, 1964.
- 6) Hochberg, Y. and Tamhane A. *Multiple Comparison Procedures*. John Wiley and Sons, 1987.
- 7) Devore, J. *Probability and Statistics, 5<sup>th</sup> Ed.* Duxbury, 2000

## Appendix A: C Source Code for Simulation Routine

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>

#define SEED 4847634

#define N 500
#define P 33
#define FNull 4
#define Corr 5
#define Bet 2
#define RSQ 3
#define RUNS 100

double* normal (int, int, double[]);
double* xfun(int);
double pvalue(double);
double* Step1(double[], double, int, double[]);
double* Step2(double[], double, int, double[]);

main( int argc, char *argv[] )
{
    double X[N][P];
    double tempArr1[N];
    double tempArr2[P];
    double A[P][2*P+1];
    double xbar[P];
    double AA[P][P];
    double obsY[N]; //stores predicted Y
    double obsB[P]; //stores Beta observed
    double tstat[P]; //stores t-Statistic
    double pVal[P]; //stores corresponding P-values
    double bonfPVal[P-1];
    double Step1PVal[P-1];
    double Step2PVal[P-1];
    double Beta[P-1];
    double* Step1Data;
    double* Step2Data;
    double tempData[4];
    double* temp;
    double* Y;
    double B, TST;
    int seed=SEED;
    int i, j, k, l, m, NTRN;
    int IgnoreH0=0;
    int IgnoreHa=0;
    int BonfH0=0;
    int BonfHa=0;
    int Step1H0=0;
    int Step1Ha=0;
    int Step2H0=0;
    int Step2Ha=0;
    int RealHa=0;
    int RealH0=0;
```

```

double IgnoreT=0;
double IgnoreV=0;
double IgnoreU=0;
double IgnoreS=0;
int runIgT=0;
int runIgV=0;
int runIgU=0;
int runIgS=0;
double IgnoreUR=0;
double IgnoreTA=0;
double IgnoreFWER=0;
double BonfT=0;
double BonfV=0;
double BonfU=0;
double BonfS=0;
int runBoT=0;
int runBoV=0;
int runBoU=0;
int runBoS=0;
double BonfUR=0;
double BonfTA=0;
double BonfFWER=0;
double Step1T=0;
double Step1V=0;
double Step1U=0;
double Step1S=0;
double Step1UR=0;
double Step1TA=0;
double Step1FWER=0;
double Step2T=0;
double Step2V=0;
double Step2U=0;
double Step2S=0;
double Step2UR=0;
double Step2TA=0;
double Step2FWER=0;
double rsquar;
double frac;
int sum=0;
int num;
double t;
double Serror=0;
double SSE=0;
double dsum;
double run;

switch(RSQ){
case 1:
    rsquar=.1;
    break;
case 2:
    rsquar=.3;
    break;
case 3:
    rsquar=.5;
    break;
case 4:

```



```

    rsquar=.7;
    break;
case 5:
    rsquar=.9;
    break;
}

switch(FNull){
case 1:
    frac=0;
    break;
case 2:
    frac=.25;
    break;
case 3:
    frac=.5;
    break;
case 4:
    frac=.75;
    break;
case 5:
    frac=1;
    break;
}

for(run=1;run<=RUNS;run++){

    Y=normal(N, seed+1234, tempArr1);
    for(j=0;j<N;j++){
        temp=xfun(seed);
        seed+=1543;
        for(k=0;k<P;k++){
            X[j][k]=temp[k];
        }
    }

    for(i=0; i<P; i++){
        for(j=0;j<2*P+1; j++){
            A[i][j]=0;
        }
    }

    for(i=0;i<P;i++){
        for(j=0;j<P;j++){
            for(k=0;k<N;k++){
                A[i][j]=A[i][j]+X[k][i]*X[k][j]; //X transpose
            }
        }
    }

    num=(P-1)*(1-frac);
    RealHa=num;
    RealH0=P-1-num;

    if(Bet==1){
        for(i=0;i<P-1;i++){
            if(i<num){

```

```

        Beta[i]=1;
    }
    else{
        Beta[i]=0;
    }
}
}

else if(Bet==2){
    for(i=0;i<P-1;i++){
        if(i<num){
            Beta[i]=(i+1)*2;
        }
        else{
            Beta[i]=0;
        }
    }
}

for(i=0;i<P;i++){
    xbar[i]=0;
    for(j=0;j<N;j++){
        xbar[i]=xbar[i]+X[j][i];
        //xbar[i]=xbar[i]/N;
    }
    xbar[i]=xbar[i]/N;
}

for (i=0;i<P;i++){
    for(j=0;j<P;j++){
        AA[i][j]=0;
        for (k=0;k<N;k++){
            AA[i][j]=AA[i][j]+(X[k][i]-xbar[i])*(X[k][j]-xbar[j]);
        }
    }
}

dsum=0;
for(i=0;i<P-1;i++){
    for(j=0;j<P-1;j++){
        dsum=dsum+Beta[i]*Beta[j]*AA[i+1][j+1];
    }
}

t=sqrt(fabs((dsum*(1-rsuar))/(rsuar*(N-P))));

for(i=0;i<N;i++){
    if(t!=0){
        Y[i]=Y[i]*t;
    }
    for(j=1;j<P;j++){
        Y[i]=Y[i]+X[i][j]*Beta[j-1];
    }
}

for(i=0;i<P;i++){
    A[i][P+i]=1; //identity matrix

```

```

    for(j=0;j<N;j++){
        A[i][2*P]=A[i][2*P]+X[j][i]*Y[j]; //Y vector
    }
}

NTRN=2*P+1;

for(k=0;k<P;k++){
    if(fabs(A[k][k])>=0.0){
        goto fifty;
    }
    for(l=k+1;l<P;l++){
        if(fabs(A[l][k])>=0.0){
            goto thirty;
        }
    }
}

thirty:

    for(m=k;m<NTRN;m++){
        TST=A[k][m];
        A[k][m]=A[l][m];
        A[l][m]=TST;
    }

fifty:

    B=A[k][k];
    for(m=k;m<NTRN;m++){
        A[k][m]=A[k][m]/B;
    }

    for(l=0;l<P;l++){
        if(l!=k){
            //goto eighty;

            B=A[l][k];
            for(m=k;m<NTRN;m++){
                A[l][m]=A[l][m]-B*A[k][m];
            }
        }
    }

for(i=0;i<P;i++){
    obsB[i]=A[i][2*P];
}

for(i=0;i<N;i++){
    dsum=0;
    for(j=0;j<P;j++){
        dsum=dsum+X[i][j]*obsB[j];
    }
    obsY[i]=dsum;
}

```

```

for(i=0;i<N;i++){
    Serror=Serror+(Y[i]-obsY[i])*(Y[i]-obsY[i]);
}

dsum=0;
for(i=0;i<P;i++){
    for(j=0;j<P;j++){
        //new A matrix here
        dsum=dsum+obsB[i]*obsB[j]*AA[i][j];
    }
}

Serror=Serror/(N-P);

SSE=sqrt(fabs(Serror));

for(i=0;i<P;i++){
    tstat[i]=obsB[i]/(SSE*sqrt(fabs(A[i][P+i])));
}

for(i=0;i<P;i++){
    pVal[i]=1-pvalue(tstat[i]);
}

//ignoring multiplicity

for(i=1;i<=num;i++){
    if(pVal[i]<.05){
        runIgV++;
    }
    else{
        runIgT++;
    }
}

for(i=num+1;i<P;i++){
    if(pVal[i]<.05){
        runIgU++;
    }
    else{
        runIgS++;
    }
}

if((runIgU+runIgV)>0){
    IgnoreUR+=(double)runIgU/(runIgU+runIgV);
}

if((runIgS+runIgT)>0){
    IgnoreTA+=(double)runIgT/(runIgS+runIgT);
}

if(runIgU>0){
    IgnoreFWER++;
}

IgnoreT+=runIgT;
IgnoreS+=runIgS;

```

```

IgnoreU+=runIgU;
IgnoreV+=runIgV;
runIgT=0;
runIgS=0;
runIgU=0;
runIgV=0;

//Bonferroni

for(i=0;i<P;i++){
  if(pVal[i]<(.05/(P-1))){
    BonfHa++;
  }
}

BonfH0+=P-1-BonfHa;

for(i=1;i<=num;i++){
  if(pVal[i]<(.05/(P-1))){
    runBoV++;
  }
  else{
    runBoT++;
  }
}

for(i=num+1;i<P;i++){
  if(pVal[i]<(.05/(P-1))){
    runBoU++;
  }
  else{
    runBoS++;
  }
}

if((runBoU+runBoV)>0){
BonfUR+=(double)runBoU/(runBoU+runBoV);
}
if((runBoS+runBoT)>0){
BonfTA+=(double)runBoT/(runBoS+runBoT);
}
if(runBoU>0){
  BonfFWER++;
}
BonfT+=runBoT;
BonfS+=runBoS;
BonfU+=runBoU;
BonfV+=runBoV;
runBoT=0;
runBoS=0;
runBoU=0;
runBoV=0;

//Step-up One Level
Step1Data=Step1(pVal, .05, num, tempData);
if((Step1Data[2]+Step1Data[3])>0){

```

```

Step1UR+=(double) Step1Data[2]/(Step1Data[2]+Step1Data[3]);
}
if((Step1Data[0]+Step1Data[1])>0){
Step1TA+=(double) Step1Data[1]/(Step1Data[0]+Step1Data[1]);
}
if(Step1Data[2]>0){
Step1FWER++;
}
Step1S+=Step1Data[0];
Step1T+=Step1Data[1];
Step1U+=Step1Data[2];
Step1V+=Step1Data[3];

//Step-up Two Level
Step2Data=Step2(pVal, .05, num, tempData);
if((Step2Data[2]+Step2Data[3])>0){
Step2UR+=(double) Step2Data[2]/(Step2Data[2]+Step2Data[3]);
}
if((Step2Data[0]+Step2Data[1])>0){
Step2TA+=(double) Step2Data[1]/(Step2Data[0]+Step2Data[1]);
}
if(Step2Data[2]>0){
Step2FWER++;
}
Step2S+=Step2Data[0];
Step2T+=Step2Data[1];
Step2U+=Step2Data[2];
Step2V+=Step2Data[3];
}

printf("\nNumber of Observations: %d", N);
printf("\nRSquared: %f", rsquar);
if(Bet==1){
printf("\nBetas are all equal to 1");
}
else{
printf("\nBetas are linear: 2,4,6,8");
}
switch(Corr){
case 1:
printf("\nCorrelation between Xs is 0");
break;
case 2:
printf("\nCorrelation between Xs is 1/4");
break;
case 3:
printf("\nCorrelation between Xs is -1/4");
break;
case 4:
printf("\nCorrelation between Xs is 3/4");
break;
case 5:
printf("\nCorrelation between Xs is -3/4");
break;
}
printf("\nTrue Nulls: %d False Nulls: %d", P-1-num, num);
printf("\nIgnore Multiplicity");

```

```

    printf("\nS: %f T: %f A: %f", IgnoreS/RUNS, IgnoreT/RUNS,
(IgnoreS+IgnoreT)/RUNS);
    printf("\nU: %f V: %f R: %f", IgnoreU/RUNS, IgnoreV/RUNS,
(IgnoreU+IgnoreV)/RUNS);
    printf("\nU/R: %f", IgnoreUR/RUNS);
    printf("\nT/A: %f", IgnoreTA/RUNS);
    if((P-1-num)>0){
printf("\nAvg U/m0: %f", (IgnoreU/RUNS)/(P-1-num));
    }
    else{
        printf("\nAvg U/m0: %f", 0);
    }
    if(num>0){
printf("\nAvg T/m1: %f", (IgnoreT/RUNS)/num);
    }
    else{
printf("\nAvg T/m1: %f", 0);
    }
    printf("\nAvg FWER: %f", IgnoreFWER/RUNS);
    printf("\n");

    printf("\nBonferroni");
    printf("\nS: %f T: %f A: %f", BonfS/RUNS, BonfT/RUNS,
(BonfS+BonfT)/RUNS);
    printf("\nU: %f V: %f R: %f", BonfU/RUNS, BonfV/RUNS,
(BonfU+BonfV)/RUNS);
    printf("\nU/R: %f", BonfUR/RUNS);
    printf("\nT/A: %f", BonfTA/RUNS);
    if((P-1-num)>0){
printf("\nAvg U/m0: %f", (BonfU/RUNS)/(P-1-num));
    }
    else{
        printf("\nAvg U/m0: %f", 0);
    }
    if(num>0){
printf("\nAvg T/m1: %f", (BonfT/RUNS)/num);
    }
    else{
        printf("\nAvg T/m1: %f", 0);
    }
    printf("\nAvg FWER: %f", BonfFWER/RUNS);
    printf("\n");

    printf("\nStep-Up 1 Stage");
    printf("\nS: %f T: %f A: %f", Step1S/RUNS, Step1T/RUNS,
(Step1S+Step1T)/RUNS);
    printf("\nU: %f V: %f R: %f", Step1U/RUNS, Step1V/RUNS,
(Step1U+Step1V)/RUNS);
    printf("\nU/R: %f", Step1UR/RUNS);
    printf("\nT/A: %f", Step1TA/RUNS);
    if((P-1-num)>0){
printf("\nAvg U/m0: %f", (Step1U/RUNS)/(P-1-num));
    }
    else{
        printf("\nAvg U/m0: %f", 0);
    }
    if(num>0){

```

```

printf("\nAvg T/m1: %f", (Step1T/RUNS)/num);
}
else{
    printf("\nAvg T/m1: %f", 0);
}
printf("\nAvg FWER: %f", Step1FWER/RUNS);
printf("\n");

    printf("\nStep-Up 2 Stage");
    printf("\nS: %f T: %f A: %f", Step2S/RUNS, Step2T/RUNS,
(Step2S+Step2T)/RUNS);
    printf("\nU: %f V: %f R: %f", Step2U/RUNS, Step2V/RUNS,
(Step2U+Step2V)/RUNS);
    printf("\nU/R: %f", Step2UR/RUNS);
    printf("\nT/A: %f", Step2TA/RUNS);
    if((P-1-num)>0){
printf("\nAvg U/m0: %f", (Step2U/RUNS)/(P-1-num));
    }
else{
    printf("\nAvg U/m0: %f", 0);
}
    if(num>0){
    printf("\nAvg T/m1: %f", (Step2T/RUNS)/num);
}
else{
    printf("\nAvg T/m1: %f", 0);
}
printf("\nAvg FWER: %f", Step2FWER/RUNS);
printf("\n");
}

double* Step1(double allp[], double alpha, int m1, double tempData[]){
    double* data=tempData;
    int index[P-1];
    double varp[P-1];
    int i;
    int kay=-1;
    double S=0;
    double T=0;
    double U=0;
    double V=0;
    double A=0;
    double R=0;
    int sigindex=0;

    for(i=1;i<P;i++){
        varp[i-1]=allp[i];
        index[i-1]=i;
    }

    sort(varp, index);

    for(i=1;i<=P-1;i++){
        if(varp[i-1]<=(i*alpha)/(P-1)){
            sigindex=i;
        }
    }
}

```



```

if(sigindex==0){      //none rejected
  data[0]=P-1-m1;
  data[1]=m1;
  data[2]=0;
  data[3]=0;
}
else{
  for(i=1;i<=sigindex;i++){
    if(index[i-1]<=m1){
      V++;
    }
  }
  R=sigindex;
  A=P-1-R;
  T=m1-V;
  S=A-T;
  U=R-V;
  data[0]=S;
  data[1]=T;
  data[2]=U;
  data[3]=V;
}
return data;
}

double* Step2(double allp[], double alpha, int m1, double tempData[]){
  double* data=tempData;
  int index[P-1];
  double varp[P-1];
  double adjalph;
  int i;
  int kay=-1;
  int adjindex=0;
  double S=0;
  double T=0;
  double U=0;
  double V=0;
  double A=0;
  double R=0;
  int sigindex=0;

  for(i=1;i<P;i++){
    varp[i-1]=allp[i];
    index[i-1]=i;
  }

  sort(varp, index);

  for(i=1;i<=P-1;i++){
    if(varp[i-1]<=(i*alpha)/(P-1)){
      sigindex=i;
    }
  }

  adjalph=((P-1)*alpha)/(P-1-sigindex);

```

```

for(i=1;i<=P-1;i++){
    if (varp[i-1]<=(i*adjalph)/(P-1)){
        adjindex=i;
    }
}

if(sigindex=0){ //none rejected
    data[0]=P-1-m1; //stop
    data[1]=m1;
    data[2]=0;
    data[3]=0;
}
else{
    for(i=0;i<adjindex;i++){
        if(index[i]<=m1){
            V++;
        }
    }
    R=adjindex;
    A=P-1-adjindex;
    T=m1-V;
    S=A-T;
    U=R-V;
    data[0]=S;
    data[1]=T;
    data[2]=U;
    data[3]=V;
}
return data;
}

void sort(double pvalue[], int index[]){
    int i,j;
    for ( i=0;i<P-2; i++)
        for (j=P-2;j>i; j--)
            if ( pvalue[j-1] > pvalue[j] ){
                dswap( &pvalue[j-1], &pvalue[j]);
                swap( &index[j-1], &index[j]);
            }
}

void dswap( double *x, double *y)
{
    double tmp;

    tmp = *x;
    *x = *y;
    *y = tmp;
}

void swap( int *x, int *y)
{
    int tmp;

    tmp = *x;
    *x = *y;
    *y = tmp;
}

```

```

}

double pvalue(double t){
    double Theta, PI, CS, SS, XC, TOT, RES, Tstat;
    int i, ND, NT;

    ND=N-P;
    Tstat=fabs(t);
    Theta=atan(Tstat/sqrt(ND+0.0));
    CS=cos(Theta);
    SS=sin(Theta);
    NT=ND/2;
    NT=NT*2;
    if(NT==ND){
        goto ten;
    }

    PI=4.0*atan(1.0);
    NT=(ND+1)/2;
    XC=1.0;
    TOT=0;
    for(i=1;i<=NT-1;i++){
        TOT=TOT+XC*pow(CS,2.0*i-1.0);
        XC=XC*(2.0*i/(2*i+1));
    }
    RES=(TOT*SS+Theta)*(2.0/PI);
    goto fifty;

ten:
    TOT=1;
    XC=1.0;
    for(i=1;i<=NT-1;i++){
        XC=XC*(2*i-1)/(2.9*i);
        TOT=TOT+XC*pow(CS,2.0*i);
    }
    RES=TOT*SS;

fifty:
    return RES;
}

double* xfun(int seed){
    double tempArr[P];
    double* zArr;
    double finArr[P];
    double* XARR=finArr;
    int i;

    zArr=normal(P, seed, tempArr);
    switch(Corr){
    case 1:
        XARR[0]=1.0;
        for(i=1;i<P;i++){
            XARR[i]=zArr[i];
        }
        break;
    case 2:

```

```

    XARR[0]=1.0;
    for(i=1;i<P;i++){
        XARR[i]=sqrt(1/4)*zArr[0]+sqrt(1-1/4)*zArr[i];
    }
    break;
case 3:
    XARR[0]=1.0;
    for(i=1;i<P;i++){
        XARR[i]=pow(-1,i-1)*sqrt(1/4)*zArr[0]+sqrt(1-1/4)*zArr[i];
    }
    break;
case 4:
    XARR[0]=1.0;
    for(i=1;i<P;i++){
        XARR[i]=sqrt(3/4)*zArr[0]+sqrt(1-3/4)*zArr[i];
    }
    break;
case 5:
    XARR[0]=1.0;
    for(i=1;i<P;i++){
        XARR[i]=pow(-1,i-1)*sqrt(3/4)*zArr[0]+sqrt(1-3/4)*zArr[i];
    }
    break;
}

return XARR;
}

double* normal(int n, int iseed, double arr[]){
    double* array=arr;
    int nn=0;
    int i=0;
    int temp;
    int j;
    double v1, v2, r, fac, x1, x2;
    int nt=(n+1)/2;
    temp=iseed;
    srand(temp);
    while(i<nt){
        do{
            temp++;
            srand(temp);
            x1=(double)rand()/((double)RAND_MAX);
            x2=(double)rand()/((double)RAND_MAX);
            v1=2.0*x1-1.0;
            v2=2.0*x2-1.0;
            r=v1*v1+v2*v2;
        }while((r>=1)|| (r<=0));
        fac=sqrt((double) -2.0*log((double)r)/r);
        array[nn]=v1*fac;
        nn+=1;
        array[nn]=v2*fac;
        nn+=1;
        i+=1;
    }
    return array;
}

```

## Appendix B: Parameter Settings and Simulation Output for the 20 Regression

### Problems Considered

1)

Number of Observations: 1000

RSquared: 0.300000

Betas are all equal to 1

Correlation between Xs is  $-1/4$

True Nulls: 6 False Nulls: 2

Ignore Multiplicity

S: 5.810000 T: 0.000000 A: 5.810000

U: 0.190000 V: 2.000000 R: 2.190000

U/R: 0.054333

T/A: 0.000000

Avg U: 0.190000

Avg T: 0.000000

Avg FWER: 0.140000

Bonferroni

S: 5.990000 T: 0.000000 A: 5.990000

U: 0.010000 V: 2.000000 R: 2.010000

U/R: 0.003333

T/A: 0.000000

Avg U: 0.010000

Avg T: 0.000000

Avg FWER: 0.010000

Step-Up 1 Stage

S: 5.930000 T: 0.000000 A: 5.930000

U: 0.070000 V: 2.000000 R: 2.070000

U/R: 0.020000

T/A: 0.000000

Avg U: 0.070000

Avg T: 0.000000

Avg FWER: 0.050000

Step-Up 2 Stage

S: 5.890000 T: 0.000000 A: 5.890000

U: 0.110000 V: 2.000000 R: 2.110000

U/R: 0.029333

T/A: 0.000000

Avg U: 0.110000

Avg T: 0.000000

Avg FWER: 0.070000

**2)**

Number of Observations: 100

RSquared: 0.100000

Betas are linear: 2,4,6,8

Correlation between Xs is 3/4

True Nulls: 48 False Nulls: 16

Ignore Multiplicity

S: 45.490000 T: 15.420000 A: 60.910000

U: 2.510000 V: 0.580000 R: 3.090000

U/R: 0.628328

T/A: 0.253471

Avg U: 2.510000

Avg T: 15.420000

Avg FWER: 0.720000

Bonferroni

S: 47.990000 T: 15.990000 A: 63.980000

U: 0.010000 V: 0.010000 R: 0.020000

U/R: 0.010000

T/A: 0.249921

Avg U: 0.010000

Avg T: 15.990000

Avg FWER: 0.010000

Step-Up 1 Stage

S: 47.990000 T: 15.990000 A: 63.980000

U: 0.010000 V: 0.010000 R: 0.020000

U/R: 0.010000

T/A: 0.249921

Avg U: 0.010000

Avg T: 15.990000

Avg FWER: 0.010000

Step-Up 2 Stage

S: 47.990000 T: 15.990000 A: 63.980000

U: 0.010000 V: 0.010000 R: 0.020000

U/R: 0.010000

T/A: 0.249921

Avg U: 0.010000

Avg T: 15.990000

Avg FWER: 0.010000

**3)**

Number of Observations: 2000

RSquared: 0.300000

Betas are linear: 2,4,6,8

Correlation between Xs is 1/4

True Nulls: 16 False Nulls: 48



Ignore Multiplicity

S: 15.000000 T: 31.190000 A: 46.190000

U: 1.000000 V: 16.810000 R: 17.810000

U/R: 0.055870

T/A: 0.673161

Avg U: 1.000000

Avg T: 31.190000

Avg FWER: 0.590000

Bonferroni

S: 15.930000 T: 44.280000 A: 60.210000

U: 0.070000 V: 3.720000 R: 3.790000

U/R: 0.012333

T/A: 0.735172

Avg U: 0.070000

Avg T: 44.280000

Avg FWER: 0.070000

Step-Up 1 Stage

S: 15.730000 T: 39.130000 A: 54.860000

U: 0.270000 V: 8.870000 R: 9.140000

U/R: 0.024038

T/A: 0.710915

Avg U: 0.270000

Avg T: 39.130000

Avg FWER: 0.200000

Step-Up 2 Stage

S: 15.660000 T: 38.450000 A: 54.110000

U: 0.340000 V: 9.550000 R: 9.890000

U/R: 0.026419

T/A: 0.707550

Avg U: 0.340000

Avg T: 38.450000

Avg FWER: 0.220000

**4)**

Number of Observations: 500

RSquared: 0.100000

Betas are all equal to 1

Correlation between Xs is 1/4

True Nulls: 0 False Nulls: 8

Ignore Multiplicity

S: 0.000000 T: 6.780000 A: 6.780000

U: 0.000000 V: 1.220000 R: 1.220000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 6.780000

Avg FWER: 0.000000

Bonferroni

S: 0.000000 T: 7.790000 A: 7.790000

U: 0.000000 V: 0.210000 R: 0.210000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 7.790000

Avg FWER: 0.000000

Step-Up 1 Stage

S: 0.000000 T: 7.730000 A: 7.730000

U: 0.000000 V: 0.270000 R: 0.270000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 7.730000

Avg FWER: 0.000000

Step-Up 2 Stage

S: 0.000000 T: 7.730000 A: 7.730000

U: 0.000000 V: 0.270000 R: 0.270000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 7.730000

Avg FWER: 0.000000

**5)**

Number of Observations: 2000

RSquared: 0.700000

Betas are linear: 2,4,6,8

Correlation between Xs is 0

True Nulls: 8 False Nulls: 0

Ignore Multiplicity

S: 7.820000 T: 0.000000 A: 7.820000

U: 0.180000 V: 0.000000 R: 0.180000

U/R: 0.150000

T/A: 0.000000

Avg U: 0.180000

Avg T: 0.000000

Avg FWER: 0.150000

Bonferroni

S: 8.000000 T: 0.000000 A: 8.000000

U: 0.000000 V: 0.000000 R: 0.000000

U/R: 0.000000

T/A: 0.000000

Avg U: 0.000000

Avg T: 0.000000

Avg FWER: 0.000000

Step-Up 1 Stage

S: 8.000000 T: 0.000000 A: 8.000000

U: 0.000000 V: 0.000000 R: 0.000000

U/R: 0.000000

T/A: 0.000000

Avg U: 0.000000

Avg T: 0.000000

Avg FWER: 0.000000

Step-Up 2 Stage

S: 8.000000 T: 0.000000 A: 8.000000

U: 0.000000 V: 0.000000 R: 0.000000

U/R: 0.000000

T/A: 0.000000

Avg U: 0.000000

Avg T: 0.000000

Avg FWER: 0.000000

**6)**

Number of Observations: 1000

RSquared: 0.500000

Betas are linear: 2,4,6,8

Correlation between Xs is 3/4

True Nulls: 4 False Nulls: 4

Ignore Multiplicity

S: 3.900000 T: 0.090000 A: 3.990000

U: 0.100000 V: 3.910000 R: 4.010000

U/R: 0.020333

T/A: 0.019000

Avg U: 0.100000

Avg T: 0.090000

Avg FWER: 0.090000

Bonferroni

S: 3.990000 T: 0.250000 A: 4.240000

U: 0.010000 V: 3.750000 R: 3.760000

U/R: 0.002500

T/A: 0.050500

Avg U: 0.010000

Avg T: 0.250000

Avg FWER: 0.010000

Step-Up 1 Stage

S: 3.950000 T: 0.150000 A: 4.100000

U: 0.050000 V: 3.850000 R: 3.900000

U/R: 0.010500

T/A: 0.030500

Avg U: 0.050000

Avg T: 0.150000

Avg FWER: 0.050000

Step-Up 2 Stage

S: 3.890000 T: 0.130000 A: 4.020000

U: 0.110000 V: 3.870000 R: 3.980000

U/R: 0.021833

T/A: 0.026500

Avg U: 0.110000

Avg T: 0.130000

Avg FWER: 0.100000

7)

Number of Observations: 500

RSquared: 0.300000

Betas are all equal to 1

Correlation between Xs is 3/4

True Nulls: 32 False Nulls: 0

Ignore Multiplicity

S: 31.010000 T: 0.000000 A: 31.010000

U: 0.990000 V: 0.000000 R: 0.990000

U/R: 0.510000

T/A: 0.000000

Avg U: 0.990000

Avg T: 0.000000

Avg FWER: 0.510000

Bonferroni

S: 31.980000 T: 0.000000 A: 31.980000

U: 0.020000 V: 0.000000 R: 0.020000

U/R: 0.020000

T/A: 0.000000

Avg U: 0.020000

Avg T: 0.000000



Avg FWER: 0.020000

Step-Up 1 Stage

S: 31.980000 T: 0.000000 A: 31.980000

U: 0.020000 V: 0.000000 R: 0.020000

U/R: 0.020000

T/A: 0.000000

Avg U: 0.020000

Avg T: 0.000000

Avg FWER: 0.020000

Step-Up 2 Stage

S: 31.980000 T: 0.000000 A: 31.980000

U: 0.020000 V: 0.000000 R: 0.020000

U/R: 0.020000

T/A: 0.000000

Avg U: 0.020000

Avg T: 0.000000

Avg FWER: 0.020000

**8)**

Number of Observations: 100

RSquared: 0.900000

Betas are all equal to 1

Correlation between Xs is  $-3/4$

True Nulls: 2 False Nulls: 6

Ignore Multiplicity

S: 1.950000 T: 0.050000 A: 2.000000

U: 0.050000 V: 5.950000 R: 6.000000

U/R: 0.007143

T/A: 0.016667

Avg U: 0.050000

Avg T: 0.050000

Avg FWER: 0.050000

Bonferroni

S: 2.000000 T: 0.300000 A: 2.300000

U: 0.000000 V: 5.700000 R: 5.700000

U/R: 0.000000

T/A: 0.087667

Avg U: 0.000000

Avg T: 0.300000

Avg FWER: 0.000000

Step-Up 1 Stage

S: 1.960000 T: 0.070000 A: 2.030000

U: 0.040000 V: 5.930000 R: 5.970000

U/R: 0.005714

T/A: 0.021667

Avg U: 0.040000

Avg T: 0.070000

Avg FWER: 0.040000

Step-Up 2 Stage

S: 1.750000 T: 0.020000 A: 1.770000

U: 0.250000 V: 5.980000 R: 6.230000

U/R: 0.034286

T/A: 0.006667

Avg U: 0.250000

Avg T: 0.020000

Avg FWER: 0.210000

9)

Number of Observations: 500

RSquared: 0.500000

Betas are linear: 2,4,6,8

Correlation between Xs is -3/4

True Nulls: 24 False Nulls: 8

Ignore Multiplicity

S: 23.210000 T: 3.790000 A: 27.000000

U: 0.790000 V: 4.210000 R: 5.000000

U/R: 0.122663

T/A: 0.139725

Avg U: 0.790000

Avg T: 3.790000

Avg FWER: 0.420000

Bonferroni

S: 23.980000 T: 6.480000 A: 30.460000

U: 0.020000 V: 1.520000 R: 1.540000

U/R: 0.008333

T/A: 0.212037

Avg U: 0.020000

Avg T: 6.480000

Avg FWER: 0.020000

Step-Up 1 Stage

S: 23.910000 T: 5.910000 A: 29.820000

U: 0.090000 V: 2.090000 R: 2.180000

U/R: 0.023667

T/A: 0.196953

Avg U: 0.090000

Avg T: 5.910000

Avg FWER: 0.080000

Step-Up 2 Stage

S: 23.890000 T: 5.870000 A: 29.760000

U: 0.110000 V: 2.130000 R: 2.240000

U/R: 0.025119

T/A: 0.195979

Avg U: 0.110000

Avg T: 5.870000

Avg FWER: 0.080000

**10)**

Number of Observations: 100

RSquared: 0.300000

Betas are linear: 2,4,6,8

Correlation between Xs is 0

True Nulls: 0 False Nulls: 16

Ignore Multiplicity

S: 0.000000 T: 15.310000 A: 15.310000

U: 0.000000 V: 0.690000 R: 0.690000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 15.310000

Avg FWER: 0.000000

Bonferroni

S: 0.000000 T: 15.980000 A: 15.980000

U: 0.000000 V: 0.020000 R: 0.020000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 15.980000

Avg FWER: 0.000000

Step-Up 1 Stage

S: 0.000000 T: 15.980000 A: 15.980000

U: 0.000000 V: 0.020000 R: 0.020000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 15.980000

Avg FWER: 0.000000

Step-Up 2 Stage

S: 0.000000 T: 15.980000 A: 15.980000

U: 0.000000 V: 0.020000 R: 0.020000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 15.980000

Avg FWER: 0.000000

## 11)

Number of Observations: 500

RSquared: 0.700000

Betas are linear: 2,4,6,8

Correlation between Xs is -1/4

True Nulls: 4 False Nulls: 12

Ignore Multiplicity

S: 3.920000 T: 4.860000 A: 8.780000

U: 0.080000 V: 7.140000 R: 7.220000

U/R: 0.010635

T/A: 0.548457

Avg U: 0.080000

Avg T: 4.860000

Avg FWER: 0.080000

Bonferroni

S: 4.000000 T: 8.240000 A: 12.240000

U: 0.000000 V: 3.760000 R: 3.760000

U/R: 0.000000

T/A: 0.670953

Avg U: 0.000000

Avg T: 8.240000

Avg FWER: 0.000000

Step-Up 1 Stage

S: 3.980000 T: 6.040000 A: 10.020000

U: 0.020000 V: 5.960000 R: 5.980000

U/R: 0.002679

T/A: 0.598587

Avg U: 0.020000

Avg T: 6.040000

Avg FWER: 0.020000

Step-Up 2 Stage

S: 3.940000 T: 5.250000 A: 9.190000

U: 0.060000 V: 6.750000 R: 6.810000

U/R: 0.007718

T/A: 0.564290



Avg U: 0.060000

Avg T: 5.250000

Avg FWER: 0.060000

**12)**

Number of Observations: 2000

RSquared: 0.500000

Betas are all equal to 1

Correlation between Xs is 3/4

True Nulls: 12 False Nulls: 4

Ignore Multiplicity

S: 11.650000 T: 0.000000 A: 11.650000

U: 0.350000 V: 4.000000 R: 4.350000

U/R: 0.061143

T/A: 0.000000

Avg U: 0.350000

Avg T: 0.000000

Avg FWER: 0.240000

Bonferroni

S: 11.990000 T: 0.000000 A: 11.990000

U: 0.010000 V: 4.000000 R: 4.010000

U/R: 0.002000

T/A: 0.000000

Avg U: 0.010000

Avg T: 0.000000

Avg FWER: 0.010000

Step-Up 1 Stage

S: 11.900000 T: 0.000000 A: 11.900000

U: 0.100000 V: 4.000000 R: 4.100000

U/R: 0.019333

T/A: 0.000000

Avg U: 0.100000

Avg T: 0.000000

Avg FWER: 0.090000

Step-Up 2 Stage

S: 11.860000 T: 0.000000 A: 11.860000

U: 0.140000 V: 4.000000 R: 4.140000

U/R: 0.026667

T/A: 0.000000

Avg U: 0.140000

Avg T: 0.000000

Avg FWER: 0.120000

**13)**

Number of Observations: 1000

RSquared: 0.700000

Betas are all equal to 1

Correlation between Xs is  $-3/4$

True Nulls: 0 False Nulls: 64

Ignore Multiplicity

S: 0.000000 T: 15.320000 A: 15.320000

U: 0.000000 V: 48.680000 R: 48.680000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 15.320000

Avg FWER: 0.000000

Bonferroni

S: 0.000000 T: 47.460000 A: 47.460000

U: 0.000000 V: 16.540000 R: 16.540000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 47.460000

Avg FWER: 0.000000

Step-Up 1 Stage

S: 0.000000 T: 18.480000 A: 18.480000

U: 0.000000 V: 45.520000 R: 45.520000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 18.480000

Avg FWER: 0.000000

Step-Up 2 Stage

S: 0.000000 T: 7.350000 A: 7.350000

U: 0.000000 V: 56.650000 R: 56.650000

U/R: 0.000000

T/A: 0.930000

Avg U: 0.000000

Avg T: 7.350000

Avg FWER: 0.000000

**14)**

Number of Observations: 2000

RSquared: 0.900000

Betas are linear: 2,4,6,8

Correlation between Xs is  $-1/4$

True Nulls: 0 False Nulls: 32

Ignore Multiplicity

S: 0.000000 T: 4.120000 A: 4.120000

U: 0.000000 V: 27.880000 R: 27.880000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 4.120000

Avg FWER: 0.000000

Bonferroni

S: 0.000000 T: 8.340000 A: 8.340000

U: 0.000000 V: 23.660000 R: 23.660000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 8.340000

Avg FWER: 0.000000

Step-Up 1 Stage

S: 0.000000 T: 4.330000 A: 4.330000

U: 0.000000 V: 27.670000 R: 27.670000

U/R: 0.000000

T/A: 1.000000

Avg U: 0.000000

Avg T: 4.330000

Avg FWER: 0.000000

Step-Up 2 Stage

S: 0.000000 T: 1.620000 A: 1.620000

U: 0.000000 V: 30.380000 R: 30.380000

U/R: 0.000000

T/A: 0.900000

Avg U: 0.000000

Avg T: 1.620000

Avg FWER: 0.000000

**15)**

Number of Observations: 500

RSquared: 0.900000

Betas are all equal to 1

Correlation between Xs is 0

True Nulls: 32 False Nulls: 32

Ignore Multiplicity

S: 30.670000 T: 1.690000 A: 32.360000

U: 1.330000 V: 30.310000 R: 31.640000

U/R: 0.040544

T/A: 0.050894

Avg U: 1.330000

Avg T: 1.690000

Avg FWER: 0.680000

Bonferroni

S: 32.000000 T: 12.090000 A: 44.090000

U: 0.000000 V: 19.910000 R: 19.910000

U/R: 0.000000

T/A: 0.272060

Avg U: 0.000000

Avg T: 12.090000

Avg FWER: 0.000000

Step-Up 1 Stage

S: 31.380000 T: 3.040000 A: 34.420000

U: 0.620000 V: 28.960000 R: 29.580000

U/R: 0.020241

T/A: 0.085910

Avg U: 0.620000

Avg T: 3.040000

Avg FWER: 0.430000

Step-Up 2 Stage

S: 30.720000 T: 1.790000 A: 32.510000

U: 1.280000 V: 30.210000 R: 31.490000

U/R: 0.038825

T/A: 0.053012

Avg U: 1.280000

Avg T: 1.790000

Avg FWER: 0.630000

**16)**

Number of Observations: 1000

RSquared: 0.100000

Betas are all equal to 1

Correlation between Xs is 0

True Nulls: 8 False Nulls: 24

Ignore Multiplicity

S: 7.840000 T: 22.320000 A: 30.160000

U: 0.160000 V: 1.680000 R: 1.840000

U/R: 0.053000

T/A: 0.739785

Avg U: 0.160000



Avg T: 22.320000

Avg FWER: 0.110000

Bonferroni

S: 8.000000 T: 23.900000 A: 31.900000

U: 0.000000 V: 0.100000 R: 0.100000

U/R: 0.000000

T/A: 0.749194

Avg U: 0.000000

Avg T: 23.900000

Avg FWER: 0.000000

Step-Up 1 Stage

S: 8.000000 T: 23.870000 A: 31.870000

U: 0.000000 V: 0.130000 R: 0.130000

U/R: 0.000000

T/A: 0.748941

Avg U: 0.000000

Avg T: 23.870000

Avg FWER: 0.000000

Step-Up 2 Stage

S: 8.000000 T: 23.870000 A: 31.870000

U: 0.000000 V: 0.130000 R: 0.130000

U/R: 0.000000

T/A: 0.748941

Avg U: 0.000000

Avg T: 23.870000

Avg FWER: 0.000000

**17)**

Number of Observations: 100

RSquared: 0.500000

Betas are all equal to 1

Correlation between Xs is -1/4

True Nulls: 64 False Nulls: 0

Ignore Multiplicity

S: 60.940000 T: 0.000000 A: 60.940000

U: 3.060000 V: 0.000000 R: 3.060000

U/R: 0.800000

T/A: 0.000000

Avg U: 3.060000

Avg T: 0.000000

Avg FWER: 0.800000

Bonferroni

S: 63.980000 T: 0.000000 A: 63.980000

U: 0.020000 V: 0.000000 R: 0.020000

U/R: 0.020000

T/A: 0.000000

Avg U: 0.020000

Avg T: 0.000000

Avg FWER: 0.020000

Step-Up 1 Stage

S: 63.980000 T: 0.000000 A: 63.980000

U: 0.020000 V: 0.000000 R: 0.020000

U/R: 0.020000

T/A: 0.000000

Avg U: 0.020000

Avg T: 0.000000

Avg FWER: 0.020000

Step-Up 2 Stage

S: 63.980000 T: 0.000000 A: 63.980000

U: 0.020000 V: 0.000000 R: 0.020000

U/R: 0.020000

T/A: 0.000000

Avg U: 0.020000

Avg T: 0.000000

Avg FWER: 0.020000

**18)**

Number of Observations: 100

RSquared: 0.700000

Betas are all equal to 1

Correlation between Xs is 1/4

True Nulls: 16 False Nulls: 16

Ignore Multiplicity

S: 15.180000 T: 14.670000 A: 29.850000

U: 0.820000 V: 1.330000 R: 2.150000

U/R: 0.267655

T/A: 0.491655

Avg U: 0.820000

Avg T: 14.670000

Avg FWER: 0.440000

Bonferroni

S: 15.990000 T: 15.950000 A: 31.940000

U: 0.010000 V: 0.050000 R: 0.060000

U/R: 0.010000

T/A: 0.499333

Avg U: 0.010000

Avg T: 15.950000

Avg FWER: 0.010000

Step-Up 1 Stage

S: 15.990000 T: 15.920000 A: 31.910000

U: 0.010000 V: 0.080000 R: 0.090000

U/R: 0.010000

T/A: 0.498816

Avg U: 0.010000

Avg T: 15.920000

Avg FWER: 0.010000

Step-Up 2 Stage

S: 15.990000 T: 15.920000 A: 31.910000

U: 0.010000 V: 0.080000 R: 0.090000

U/R: 0.010000

T/A: 0.498816

Avg U: 0.010000

Avg T: 15.920000

Avg FWER: 0.010000

**19)**

Number of Observations: 2000

RSquared: 0.100000

Betas are linear: 2,4,6,8

Correlation between Xs is  $-3/4$

True Nulls: 8 False Nulls: 8

Ignore Multiplicity

S: 7.820000 T: 5.050000 A: 12.870000

U: 0.180000 V: 2.950000 R: 3.130000

U/R: 0.041119

T/A: 0.390534

Avg U: 0.180000

Avg T: 5.050000

Avg FWER: 0.120000

Bonferroni

S: 8.000000 T: 7.310000 A: 15.310000

U: 0.000000 V: 0.690000 R: 0.690000

U/R: 0.000000

T/A: 0.476275

Avg U: 0.000000

Avg T: 7.310000

Avg FWER: 0.000000

Step-Up 1 Stage

S: 7.960000 T: 7.090000 A: 15.050000

U: 0.040000 V: 0.910000 R: 0.950000

U/R: 0.013333

T/A: 0.468788

Avg U: 0.040000

Avg T: 7.090000

Avg FWER: 0.040000

Step-Up 2 Stage

S: 7.950000 T: 7.050000 A: 15.000000

U: 0.050000 V: 0.950000 R: 1.000000

U/R: 0.018333

T/A: 0.467348

Avg U: 0.050000

Avg T: 7.050000

Avg FWER: 0.050000

**20)**

Number of Observations: 1000

RSquared: 0.900000

Betas are linear: 2,4,6,8

Correlation between Xs is 1/4

True Nulls: 16 False Nulls: 0

Ignore Multiplicity

S: 15.530000 T: 0.000000 A: 15.530000

U: 0.470000 V: 0.000000 R: 0.470000

U/R: 0.290000

T/A: 0.000000

Avg U: 0.470000

Avg T: 0.000000

Avg FWER: 0.290000

Bonferroni

S: 15.990000 T: 0.000000 A: 15.990000

U: 0.010000 V: 0.000000 R: 0.010000

U/R: 0.010000

T/A: 0.000000

Avg U: 0.010000

Avg T: 0.000000

Avg FWER: 0.010000

Step-Up 1 Stage

S: 15.990000 T: 0.000000 A: 15.990000

U: 0.010000 V: 0.000000 R: 0.010000

U/R: 0.010000



T/A: 0.000000

Avg U: 0.010000

Avg T: 0.000000

Avg FWER: 0.010000

Step-Up 2 Stage

S: 15.990000 T: 0.000000 A: 15.990000

U: 0.010000 V: 0.000000 R: 0.010000

U/R: 0.010000

T/A: 0.000000

Avg U: 0.010000

Avg T: 0.000000

Avg FWER: 0.010000