



Virginia Rovnyak, Ph.D.
2008

WALD TESTS FOR LINEAR MODELS IN SAS, SPSS, AND STATA

**This work was supported by National Institutes of Health grant
1 RO1 MH066293-01A1.**

Introduction

In a regression situation one often wishes to test joint hypotheses involving the regression coefficients. For example, one might wish to test the joint hypothesis that several levels of a categorical predictor variable all have the same effect on the dependent variable. Joint linear tests such as that can be carried out with Wald tests (Judge, et al, 1985, 182-187). The purpose of this paper is to present the syntax needed to carry out certain joint linear tests in multiple linear regression and in logistic regression, in each of the three statistical packages: SAS, SPSS, and Stata. Stern, et al (2008) determined that the results of such tests might offer important assistance to a researcher who wishes to choose the best measure of rurality to use in studies of physical and mental health and health care. We are presenting the syntax necessary for these tests as an aid to such analyses.

We give examples below of the syntax needed for two kinds of joint tests: (1) an “overall” test of the effect of a categorical variable in a given model, which tests the joint hypothesis that every level of the variable has the same effect on the dependent variable, and therefore the predictor does not explain any of the variation in the dependent variable and so is not a significant predictor in the model; and (2) an “equality” test, which tests the hypothesis that two or more, but not necessarily all, of the levels for a categorical predictor in a given model have the same effect on the dependent variable (see Stern, et al (2008), Sec. 3.3).

Multiple Linear Regression

The examples will cover four different statistical packages or procedures: SAS PROC REG, SAS PROC GLM, SPSS, and Stata. To illustrate the syntax we will regress one quantitative dependent variable on two quantitative predictors and two categorical predictors. The dependent variable is *restless*, a self-rated measure of how often the person felt restless or fidgety in the past month. The Wald tests will concern the effects on *restless* of the different levels of the categorical variable *region*, which provides the U.S. region of residence of the individual (Northeast, Midwest, South, or West). The other predictors in the regression are the categorical variable *mental* (3 levels of self-rated mental health), and the quantitative variables *age* and *poverty*, the latter being a measure of the level of poverty in the person’s locality. Some of the data used come from Medical Expenditure Panel Surveys, but the data have been altered for purposes of the demonstration and do not represent any real population.

For SAS PROC REG and Stata, the predictors *region* and *mental* cannot go directly into the model. Each of those multi-level variables is replaced in the model by a set of dichotomous variables that indicate its levels, one of each set being omitted so that the regression coefficients for the remaining indicators will be uniquely determined. Let *midwest*, *south*, *west*, and *northeast* denote the indicators for those four regions and let *goodmh*, *medmh*, and *poormh* indicate the upper, middle, and lower levels of *mental*, respectively. The *northeast* and *goodmh* variables will be the omitted indicator variables; omitting them from the model simply means that we have set their coefficients to 0.

For SAS GLM and SPSS we enter the multi-level variables *region* and *mental* into the regression and carry out the tests by means of estimable functions. In this situation we do not set the coefficient of some level to zero to insure that the rest of the coefficients will be uniquely determined, and so there are infinitely many solutions for estimates of the coefficients of the levels of *region*. We are therefore limited in the tests we can carry out--we cannot test, for example, whether the effect of some particular level is equal to 0. That is, if β_1, \dots, β_4 are regression coefficients for the four levels of *region*, then we cannot test $\beta_1 = 0$, say, because the value of β_1 is not uniquely determined. We can, however, always test whether two levels have the same effect--whether $\beta_1 - \beta_3 = 0$, say. The function $\beta_1 - \beta_3$ is an example of a function of the coefficients whose value is the same, regardless of which set of coefficients we might use. Since the value of the function is constant, we have the possibility of estimating it. Such a function is called an “estimable” function. All of the functions of interest to us are of the form $\beta_i - \beta_k$, so they are all estimable. See Green, et al (1999) for an extensive discussion of estimable functions, with many examples.

Another issue in SAS GLM and SPSS is the coding of the multi-level variables. In the demonstration dataset, the variables *region* and *mental* have numerical values. The levels of *region* are: 1,2,3,4, where 1=Northeast, 2=Midwest, 3=South, 4=West., and the levels of *mental* are: 1,2,3, where 1=good mental health, 2=medium mental health, 3=poor mental health. We will see later that the type of the codes (numeric vs. character), as well as the codes themselves, affect the test syntax.

We will carry out the following tests in all four statistical packages/procedures: (1) an “overall” test of the effect of *region* in our model on the outcome *restless*, which tests the joint hypothesis that every one of the four levels of *region* has the same effect on *restless*, and therefore *region* does not explain any of the variation in *restless*, and so it is not a significant predictor of *restless* in the model; and (2) an “equality” test, which tests the hypothesis that the effects on *restless* of residing in the Midwest, in the South, or in the West, respectively, are all equal to each other in this model, but not necessarily equal to the effect of living in the Northeast.

SAS PROC REG

PROC REG assumes that all variables are quantitative, so the categorical predictors *region* and *mental* must be replaced in the model by indicator variables: indicators for 3 of the 4 levels of *region* and for 2 of the 3 levels of *mental*.

If the name of the dataset is “demoreg”, then the regression and test commands are:

```
proc reg data=demoreg;
  model restless=age poverty midwest south west medmh poormh;
  overall: test midwest=0, south=0, west=0;
  equality: test midwest=south, midwest=west;
run;
```

where the words “overall” and “equality” are user-supplied names that SAS will use to identify the output of the tests, and where the coefficients of the indicators for the Northeast region and for good mental health have been set to zero. In the test commands here, and also in Stata, the variable name stands in for the coefficient of the variable. Thus the “overall” test tests the joint hypothesis that the coefficients of *midwest*, *south*, and *west* are all three equal to zero. Since the coefficient of *northeast* was set to zero, it is a test that the four regions have the same effect on *restless* in this model. The “equality” test tests the hypothesis that the coefficients of *midwest*, *south*, and *west* are all equal to each other.

The SAS 9.1.3 for Windows output from the test commands is:

Test overall Results for Dependent Variable RESTLESS				
Source	DF	Mean Square	F Value	Pr > F
Numerator	3	3.96009	4.18	0.0057
Denominator	37331	0.94731		

Test equality Results for Dependent Variable RESTLESS				
Source	DF	Mean Square	F Value	Pr > F
Numerator	2	3.41803	3.61	0.0271
Denominator	37331	0.94731		

We see that the “overall” test is significant at the $\alpha=.05$ level: we reject the hypothesis that all four regions have the same effect on restlessness in our model. We also reject the “equality” hypothesis that the three regions outside the Northeast: the Midwest, South and West, all have the same effect on restlessness in the model.

SAS PROC GLM

With the GLM procedure, we regress *restless* on *age*, *poverty*, *region* (with 4 levels) and *mental* (with 3 levels). If the name of the dataset is “demoglm”, then the regression and test commands in PROC GLM are:

```
proc GLM data=demoglm;
  class region mental;
  model restless=age poverty region mental /e;
  contrast 'equality' region 0 -1 1 0,region 0 -1 0 1;
  contrast 'overall' region 1 -1 0 0,region 1 0 -1 0,region 1 0 0 -1 ;
run;
```

(In all of the packages, it does not matter whether the “equality” or the “overall” test command comes first.)

The “class” command tells SAS that *region* and *mental* are categorical variables.

The “contrast” statements encode the test commands. As in PROC REG, the words 'equality' and 'overall' are user-supplied names that SAS will use to identify the output of the tests. The boldface quadruples of numbers in the code are linear vectors with an entry corresponding to each level of the multi-level variable named just before the vector. For example, the code “`region 0 -1 1 0`” in the first contrast statement, instructs SAS to test the equality of the effects of the second and third levels of the variable *region*. More specifically, if β_1, \dots, β_4 are regression coefficients for the four levels of *region*, then the code “`region 0 -1 1 0`” instructs SAS to test whether $0 * \beta_1 + (-1) * \beta_2 + 1 * \beta_3 + 0 * \beta_4 = 0$, i.e., whether $-\beta_2 + \beta_3 = 0$, or, equivalently, whether $\beta_2 = \beta_3$. In order to know what vector to use in the contrast command, one must know the order in which SAS has arranged the levels of the variable; that is, one must know which level of *region* SAS considers to be the first level, which the second level, and so on. The “/e” option at the end of the model command tells SAS to produce a table of estimable functions; in that table we can see how SAS has ordered the levels of *region* and *mental*.

Selected SAS 9.1.3 for Windows output from running the program above:

General Form of Estimable Functions					
Effect	Coefficients				
Intercept	L1				
AGE	L2				
poverty	L3				
region	1	L4			
region	2	L5			
region	3	L6			
region	4	L1-L4-L5-L6			
mental	1	L8			
mental	2	L9			
mental	3	L1-L8-L9			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
AGE	1	295.606149	295.606149	312.05	<.0001
poverty	1	1.684111	1.684111	1.78	0.1824
region	3	11.880277	3.960092	4.18	0.0057
mental	2	1382.574106	691.287053	729.74	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
equality	2	6.83606248	3.41803124	3.61	0.0271
overall	3	11.88027676	3.96009225	4.18	0.0057

The first section of the output tells us that SAS has put the codes for the levels of *region* in numerical order: 1,2,3,4. Therefore the syntax: `region 0 -1 1 0`, which tells SAS to test the equality of the effects of what SAS considers to be the second and third levels of *region*, is telling SAS to test the equality of the effects of *region* = 2 and *region* = 3. Similarly, the syntax:

`region 0 -1 0 1`, which tells SAS to test the equality of the effects of what SAS considers to be the second and fourth levels of *region*, is telling SAS to test the equality of the effects of *region* = 2 and *region* = 4. The entire “equality” test command:

```
contrast 'equality' region 0 -1 1 0,region 0 -1 0 1;
```

asks SAS to jointly test:

- └equality of the effects of *region* = 2 and *region* = 3
- └equality of the effects of *region* = 2 and *region* = 4.

Therefore it tests whether the effects of regions 2, 3, and 4 are all the same.

The “overall” test command:

```
contrast 'overall' region 1 -1 0 0,region 1 0 -1 0,region 1 0 0 -1 ;
```

asks SAS to jointly test:

- └equality of the effects of *region* = 1 and *region* = 2
- └equality of the effects of *region* = 1 and *region* = 3
- └equality of the effects of *region* = 1 and *region* = 4.

Therefore it is a test that all four regions have the same effect. In that situation, *region* behaves simply as a constant in the model and is therefore not a significant predictor of restlessness.

The test results are in the last two lines of the output above and are identical to the results from SAS PROC REG.

NOTE ON LEVEL ORDER IN SAS GLM: In order to define the vectors appearing in the contrast commands, one must know how SAS has ordered the levels of the given multi-level variable. By using the “/e” option in the model statement, we can learn from the `Estimable Functions` list the ordering that SAS has used. For numeric-valued categorical variables like *region* and *mental*, SAS puts the levels in numerical order. With character values, SAS puts them in alphabetical order. If, instead of numeric values, the values of the levels of *region* had been “northeast”, “midwest”, “south”, and “west”, then SAS would have put “midwest” first. The list of estimable functions for *region* would have been:

Effect		Coefficients
region	midwest	L4
region	northea	L5
region	south	L6
region	west	L1-L4-L5-L6

and the test commands would need to have been:

```
contrast 'equality' region -1 0 1 0,region -1 0 0 1;
contrast 'overall' region 1 -1 0 0,region 1 0 -1 0,region 1 0 0 -1 ;
```

The syntax for the “overall” test is the same as before, since it is simply testing whether all four of the levels of *region* have the same effect on *restless*, so their order doesn’t matter.

NOTE ALSO: As explained above, the “overall” test is simply a test of the significance of the *region* variable as a whole in the model, since it is testing whether all of the levels have the same effect. The overall significance of *region* as a predictor of *restless* in this model can be seen in the Type III sum of squares output, as well as in the output of the “overall” contrast statement. The sum of squares, mean square, F-value, and P-value are the same in the Type III sum of squares output as in the “overall” test output farther down. So, in fact, one could omit the

“overall” contrast command and simply look at the significance of the multi-level variable of interest in the Type III SS output.

SPSS

It is necessary to use syntax to carry out the tests in SPSS. To get a syntax editor, at the top of an SPSS Output window or Data Editor choose File→New→Syntax.

As with SAS GLM, we use the multi-level variables *region* and *mental* as predictors and specify the tests by means of linear vectors. We ran the following code in an SPSS syntax window:

```
GLM
  restless BY region mental WITH age poverty
  /METHOD = SSTYPE(3)
  /PRINT = GEF
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /LMATRIX = "OVERALL" region 1 -1 0 0; region 1 0 -1 0; region 1 0 0 -1
  /LMATRIX = "EQUALITY" region 0 -1 1 0; region 0 -1 0 1
  /DESIGN = region mental age poverty .
```

The categorical variables in the model are listed as BY variables, while the quantitative variables in the model are listed as WITH variables. The `METHOD = SSTYPE(3)` command specifies the sum of squares type. As in SAS GLM, with `SSTYPE(3)` we will be able to see the results of the overall test of significance of *region* in model output as well as in the output from the corresponding `LMATRIX` test command. The `PRINT = GEF` command tells SPSS to print the General Estimable Function list; from that one gets the order of the levels that one needs in order to specify the linear vectors. The `LMATRIX` commands set up the tests. The vectors are chosen exactly as in SAS PROC GLM. The `DESIGN` command specifies the predictors in the model, while the dependent variable, *restless*, appears at the beginning of the first line.

Selected output using SPSS 15.0.1 for Windows:

Tests of Between-Subjects Effects

Dependent Variable: restless

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1536.706(a)	7	219.529	231.741	.000
Intercept	12643.648	1	12643.648	13346.954	.000
region	11.880	3	3.960	4.180	.006
mental	1382.574	2	691.287	729.740	.000
age	295.606	1	295.606	312.049	.000
poverty	1.684	1	1.684	1.778	.182
Error	35363.876	37331	.947		
Total	138044.000	37339			
Corrected Total	36900.581	37338			

a R Squared = .042 (Adjusted R Squared = .041)

General Estimable Function(a)

Parameter	Contrast							
	L1	L2	L3	L4	L6	L7	L9	L10
Intercept	1	0	0	0	0	0	0	0
[region=1.00]	0	1	0	0	0	0	0	0
[region=2.00]	0	0	1	0	0	0	0	0
[region=3.00]	0	0	0	1	0	0	0	0
[region=4.00]	1	-1	-1	-1	0	0	0	0
[mental=1.00]	0	0	0	0	1	0	0	0
[mental=2.00]	0	0	0	0	0	1	0	0
[mental=3.00]	1	0	0	0	-1	-1	0	0
age	0	0	0	0	0	0	1	0
poverty	0	0	0	0	0	0	0	1

a Design: Intercept+region+mental+age+poverty

Custom Hypothesis Tests Index

1	Contrast Coefficients (L' Matrix)	LMATRIX Subcomm and 1: OVERALL
	Transformation Coefficients (M Matrix)	Identity Matrix
	Contrast Results (K Matrix)	Zero Matrix
2	Contrast Coefficients (L' Matrix)	LMATRIX Subcomm and 2: EQUALITY
	Transformation Coefficients (M Matrix)	Identity Matrix
	Contrast Results (K Matrix)	Zero Matrix

Custom Hypothesis Tests #1

Test Results

Dependent Variable: restless

Source	Sum of Squares	df	Mean Square	F	Sig.
Contrast	11.880	3	3.960	4.180	.006
Error	35363.876	37331	.947		

Custom Hypothesis Tests #2

Test Results

Dependent Variable: restless

Source	Sum of Squares	df	Mean Square	F	Sig.
Contrast	6.836	2	3.418	3.608	.027
Error	35363.876	37331	.947		

In the General Estimable Function list, we see that SPSS has put the levels of *region* and *mental* in numerical order. The Custom Hypothesis Tests Index gives the test output number for the tests that we named “OVERALL” and “EQUALITY” in the LMATRIX commands. The results of the tests are identical to those from SAS. As in SAS GLM, we can also see the results of a test of the overall effect of the *region* variable in the Tests of Between-Subjects Effects section, those results again being identical to the overall test specified in the first LMATRIX command.

Stata

Testing with the `regress` command in Stata is very similar to testing with SAS PROC REG. Indicator functions are used in the model instead of multi-level functions. The regression and test commands in Stata are:

```
regress restless age poverty midwest south west medmh poormh
test midwest=south=west=0
test midwest=south=west
```

The first command specifies the model, with the dependent variable listed first. Next are the test commands, which do not need labels since Stata includes the test command with the test output. The command syntax given here for the tests is in a shorthand style, but alternative forms work just as well. The test command for the first test could have been written out this way, listing the individual tests that are being jointly tested:

```
test (midwest=0) (south=0) (west=0)
```

and the test command for the second test could have been written:

```
test (midwest=south) (midwest=west)
```

and the results would have been the same as with the shorthand syntax. Stata 9.2 for Windows produced the following test output:

```
. test midwest=south=west=0

( 1)  midwest - south = 0
( 2)  midwest - west  = 0
( 3)  midwest  = 0

      F( 3, 37331) =    4.18
      Prob > F    =    0.0057
```

```
. test midwest=south=west

( 1)  midwest - south = 0
( 2)  midwest - west  = 0

      F(  2, 37331) =    3.61
      Prob > F    =    0.0271
```

The test results from Stata are the same as from the other statistical packages.

Logistic Regression

The examples for logistic regression will cover the same three statistical packages: SAS, SPSS, and Stata. We will demonstrate the syntax using the same data as in the multiple linear regression case, except that the dependent variable will be a dichotomized version of the variable, *restless*. The new dependent variable, *restlss2*, has values 0 and 1, where *restlss2*=0 for respondents who said that they had no times in the past month during which they felt restless or fidgety, and *restlss2*=1 for respondents who acknowledged at least some restless or fidgety time in the past month (recall, though, that the data do not represent any real population). The predictors are the same: *region*, *mental*, *age* and *poverty*. In two of the programs we will use the indicator variables defined earlier for the levels of the categorical predictors: *midwest*, *south*, *west*, and *northeast* indicate the levels of *region*, and *goodmh*, *medmh*, and *poormh* indicate the levels of *mental*.

We will carry out: (1) an “overall” test of the effect of *region* in a logistic model for the outcome: *restlss2*=1, which tests the joint hypothesis that every one of the four levels of *region* has the same effect on the probability of that outcome, and therefore *region* does not explain any of the variation in the probability of the outcome, and so it is not a significant predictor in the model; and (2) an “equality” test, which tests the hypothesis that the effects in this logistic model of residing in the Midwest, in the South, or in the West, respectively, are all equal to each other, but not necessarily equal to the effect of living in the Northeast.

SAS PROC LOGISTIC

The model and test syntax for PROC LOGISTIC is almost identical to that in PROC REG. We use the indicator variables for the levels of the categorical predictors.

If the name of the dataset is “demologit”, then the regression and test commands are:

```
proc logistic data=demologit descending;
model restlss2=age poverty midwest south west medmh poormh;
overall: test midwest=0, south=0, west=0;
equality: test midwest=south, midwest=west;
run;
```

As before, the words “overall” and “equality” are user-supplied names that SAS will use to identify the output of the tests, and the coefficients of the indicators for the Northeast region and for good mental health have been set to zero.

The option "descending" tells SAS that you want it to model the logit of the probability that $\text{restlss2}=1$ (i.e., that there was some restlessness in the past 30 days). The default method, when the dependent variable has the values 0 and 1, is that SAS models the logit of the probability that the value is 0.

The SAS 9.1.3 for Windows output from the test commands is:

Linear Hypotheses Testing Results			
Label	Wald Chi-Square	DF	Pr > ChiSq
overall	70.8312	3	<.000
equality	58.8828	2	<.0001

As in the linear model, the “overall” test is significant at the $\alpha=.05$ level: we reject the hypothesis that all four regions have the same effect in our model on the probability of restlessness. We also reject the “equality” hypothesis that the three regions outside the Northeast: the Midwest, South and West, all have the same effect in this logistic model.

SPSS

One must use syntax in SPSS in order to carry out the joint tests for a logistic regression; see the linear model notes above for SPSS for information on how to obtain a syntax window in SPSS. As in the SPSS linear model, we use the multi-level variables *region* and *mental* as predictors and specify the tests by means of linear vectors. We ran the following code in an SPSS syntax window:

```
NOMREG
  restlss2 (BASE=FIRST ORDER=ASCENDING) BY region mental WITH age poverty
  /TEST = "OVERALL" region 1 -1 0 0; region 1 0 -1 0; region 1 0 0 -1
  /TEST = "EQUALITY" region 0 -1 1 0; region 0 -1 0 1.
```

The procedure NOMREG is the SPSS procedure for multinomial logistic regression, but it can also be used when there are only two levels of the outcome variable. There is a separate binary logistic procedure in SPSS, but it does not have an option for carrying out joint tests.

If your outcome variable has values 0 and 1 and you want to model the logit of the probability that the outcome is 1, you can set BASE=FIRST and ORDER=ASCENDING. Then SPSS will order the values of the outcome in ascending order--0,1--and will use the first one (zero) as the reference value. As before, the categorical variables in the model are listed as BY variables, while the quantitative variables in the model are listed as WITH variables. The syntax for the two tests is exactly the same as in the linear model case, except that here the test command is

“TEST”, instead of “LMATRIX”. [Many other options for NOMREG are available, of course, that will produce additional tables, output sets, etc.; see the SPSS documentation for NOMREG.] The test commands in the syntax above produced the output below (using SPSS 15.0.1 for Windows). In the first table, one can verify that the three linear vectors in the “Overall” test command produced the desired contrasts. For example, in the “Overall” joint test, the table below says that the first hypothesis tested is that *region=1* and *region=2* have the same effect (1 times the effect of *region=1*, plus (-1) times the effect of *region=2*, is zero); that is, we are testing that the Northeast region and the Midwest region have the same effect in the model, which is exactly what we wanted to test. The second table is produced automatically by SPSS; it tests the individual hypotheses in each TEST statement separately. We are not interested in those results in this paper. The third table gives the results of the joint test.

Custom Hypothesis Tests 1 (OVERALL)

Contrast Coefficients

	C1	C2	C3
Intercept	0	0	0
age	0	0	0
poverty	0	0	0
[region=1.00]	1	1	1
[region=2.00]	-1	0	0
[region=3.00]	0	-1	0
[region=4.00]	0	0	-1
[mental=1.00]	0	0	0
[mental=2.00]	0	0	0
[mental=3.00]	0	0	0

Contrast Results

restlss2	Contrasts	Estimate	Std. Error	Test value	Wald	df	Sig.	Exp(Estimate)	95% Confidence Interval for Exp(Estimate)	
									Lower Bound	Upper Bound
1.00	C1	-.224	.034	0	42.314	1	.000	.799	.747	.855
	C2	-.004	.031	0	.013	1	.908	.996	.937	1.059
	C3	-.125	.034	0	13.280	1	.000	.882	.825	.944

Test Results

restlss2	Contrasts	Wald	df	Sig.
1.00	All	70.851	3	.000

Custom Hypothesis Tests 2 (EQUALITY)

Contrast Coefficients

	C1	C2
Intercept	0	0
age	0	0
poverty	0	0
[region=1.00]	0	0
[region=2.00]	-1	-1
[region=3.00]	1	0
[region=4.00]	0	1
[mental=1.00]	0	0
[mental=2.00]	0	0
[mental=3.00]	0	0

Contrast Results

restlss2	Contrasts	Estimate	Std. Error	Test value	Wald	df	Sig.	Exp(Estimate)	95% Confidence Interval for Exp(Estimate)	
									Lower Bound	Upper Bound
1.00	C1	-.221	.029	0	56.857	1	.000	.802	.757	.849
	C2	-.099	.032	0	9.331	1	.002	.906	.850	.965

Test Results

restlss2	Contrasts	Wald	df	Sig.
1.00	All	58.897	2	.000

These test results differ in the 4th place from the ones produced by SAS, perhaps due to the particular convergence algorithm used.

Stata

The commands for logistic regression in Stata are exactly the same as for linear regression, except for the name of the procedure. Indicator functions are used in the model instead of multi-level functions.

Stata has two logistic regression procedures: logit and logistic. They perform the same analyses and give exactly the same results for joint tests, but they produce different output for the effects of the predictors. The “logit” procedure reports the regression coefficients from the logistic regression, whereas the “logistic” procedure reports the odds ratios for the predictors.

We are only interested here in the results of the joint tests, so it does not matter which procedure we use. We ran the following program in Stata 9.2 for Windows:

```
logistic restlss2 age poverty midwest south west medmh poormh
test midwest=south=west=0
test midwest=south=west
```

The output from running that program contained the following joint test results:

```
. test midwest=south=west=0

( 1) midwest - south = 0
( 2) midwest - west = 0
( 3) midwest = 0

      chi2( 3) =    70.85
Prob > chi2 =    0.0000

.
. test midwest=south=west

( 1) midwest - south = 0
( 2) midwest - west = 0

      chi2( 2) =    58.90
Prob > chi2 =    0.0000
```

The “logit” command:

```
logit restlss2 age poverty midwest south west medmh poormh
test midwest=south=west=0
test midwest=south=west
```

produced exactly the same results for the joint tests.

Alternative forms of the test commands are also available in the logistic regression case, e.g.,

```
test (midwest=0) (south=0) (west=0)
```

and

```
test (midwest=south) (midwest=west).
```

The results from Stata are the same as the results from SPSS.

CONCLUSION

All three statistical packages provide ways to carry out joint linear tests of hypotheses concerning the effects of different levels of categorical predictors on a dependent variable, in both a multiple linear regression model and a logistic regression model. Tests such as these were found by Stern, et al (2008) to potentially provide important information to researchers looking for the best measure of rurality to use in analyses of health care and physical and mental health.

REFERENCES

Green, S. B., Marquis, J. G., Hershberger, S. L., Thompson, M. S., McCollam K. M. (1999). "The Overparameterized Analysis of Variance Model." *Psychological Methods*. 4(2): 214-233.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lütkepohl, H., Lee, T.-C. (1985) *Theory and Practice of Econometrics, 2nd Ed.*, John Wiley and Sons, New York.

Stern, S., Merwin, E., Hauenstein, E., Hinton, I., Rovnyak, V., Wilson, M., Williams, I., and Mahone, I. (2008, in submission) "The Effect of Rurality on Mental and Physical Health".

Additional information may be found in the documentation for each of the statistical packages.