ONE-WAY ANOVA IN SPSS

Richard Lee Rogers

One-way ANOVA

- ANalysis Of VAriance
- Bivariate statistic
  - One numeric variable
  - One categorical variable with two or more categories
- A form of the General Linear Model

Example

Descriptives

LnViolentR

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Northeast	9	5.5271	.45220	.15073	5.1795	5.8747	4.82	6.06
Midwest	12	5.7229	.25994	.07504	5.5578	5.8881	5.40	6.10
South	16	6.0068	.33469	.08367	5.8285	6.1851	5.29	6.41
West	13	5.7929	.39609	.10985	5.5535	6.0322	5.28	6.41
Total	50	5.7967	.38777	.05484	5.6865	5.9069	4.82	6.41

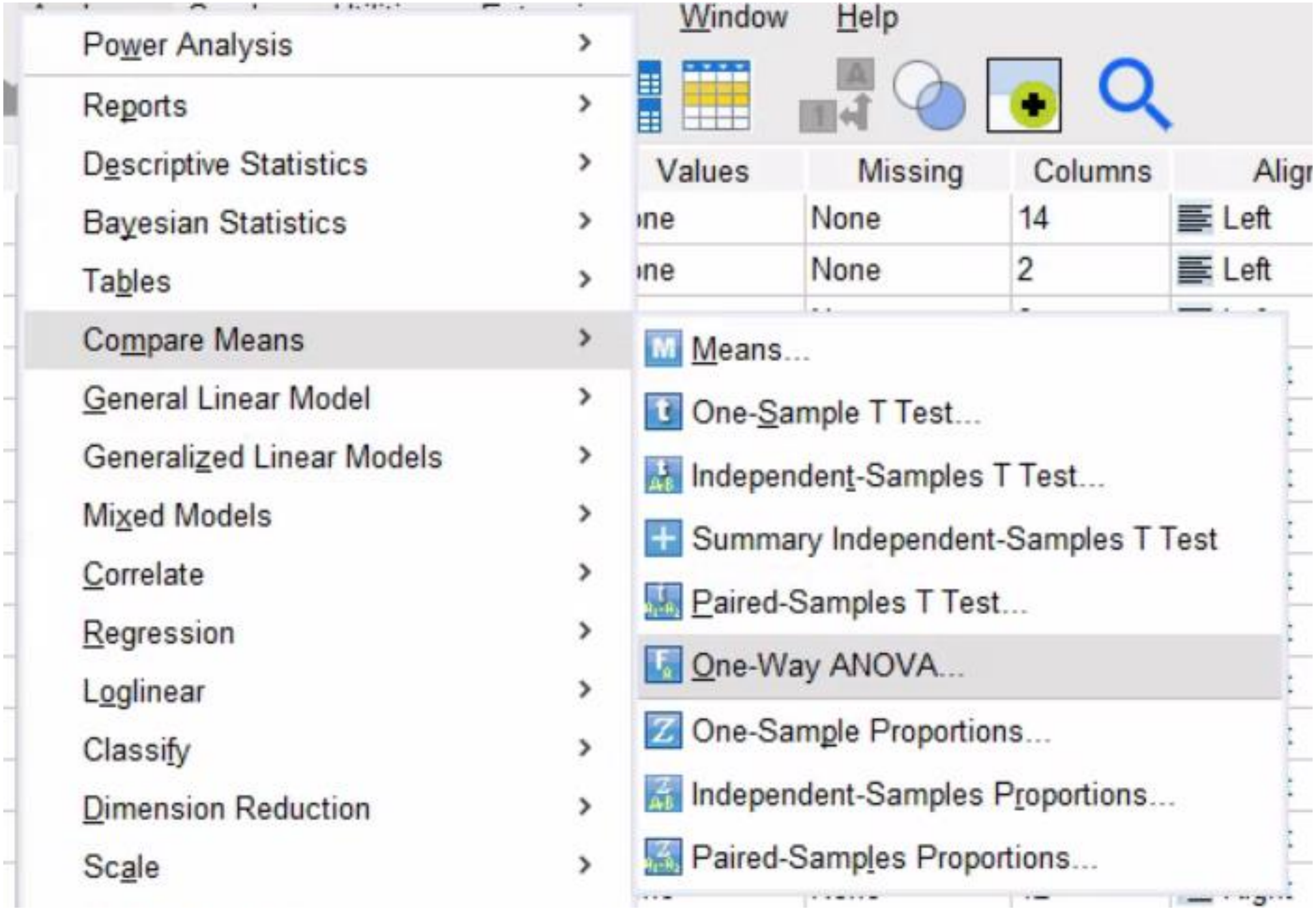
Example

Descriptives

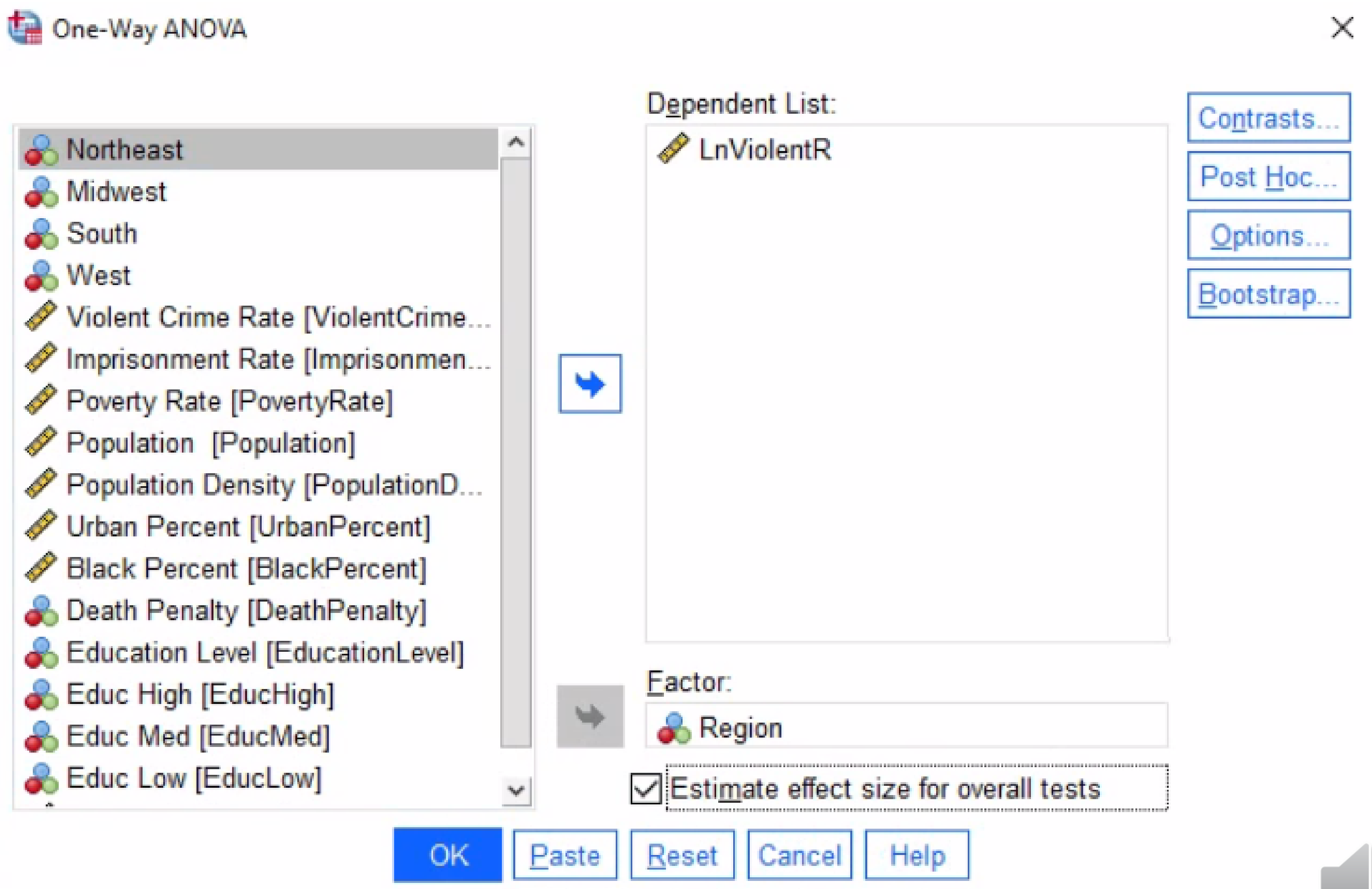
LnViolentR

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Northeast	9	5.5271	.45220	.15073	5.1795	5.8747	4.82	6.06
Midwest	12	5.7229	.25994	.07504	5.5578	5.8881	5.40	6.10
South	16	6.0068	.33469	.08367	5.8285	6.1851	5.29	6.41
West	13	5.7929	.39609	.10985	5.5535	6.0322	5.28	6.41
Total	50	5.7967	.38777	.05484	5.6865	5.9069	4.82	6.41

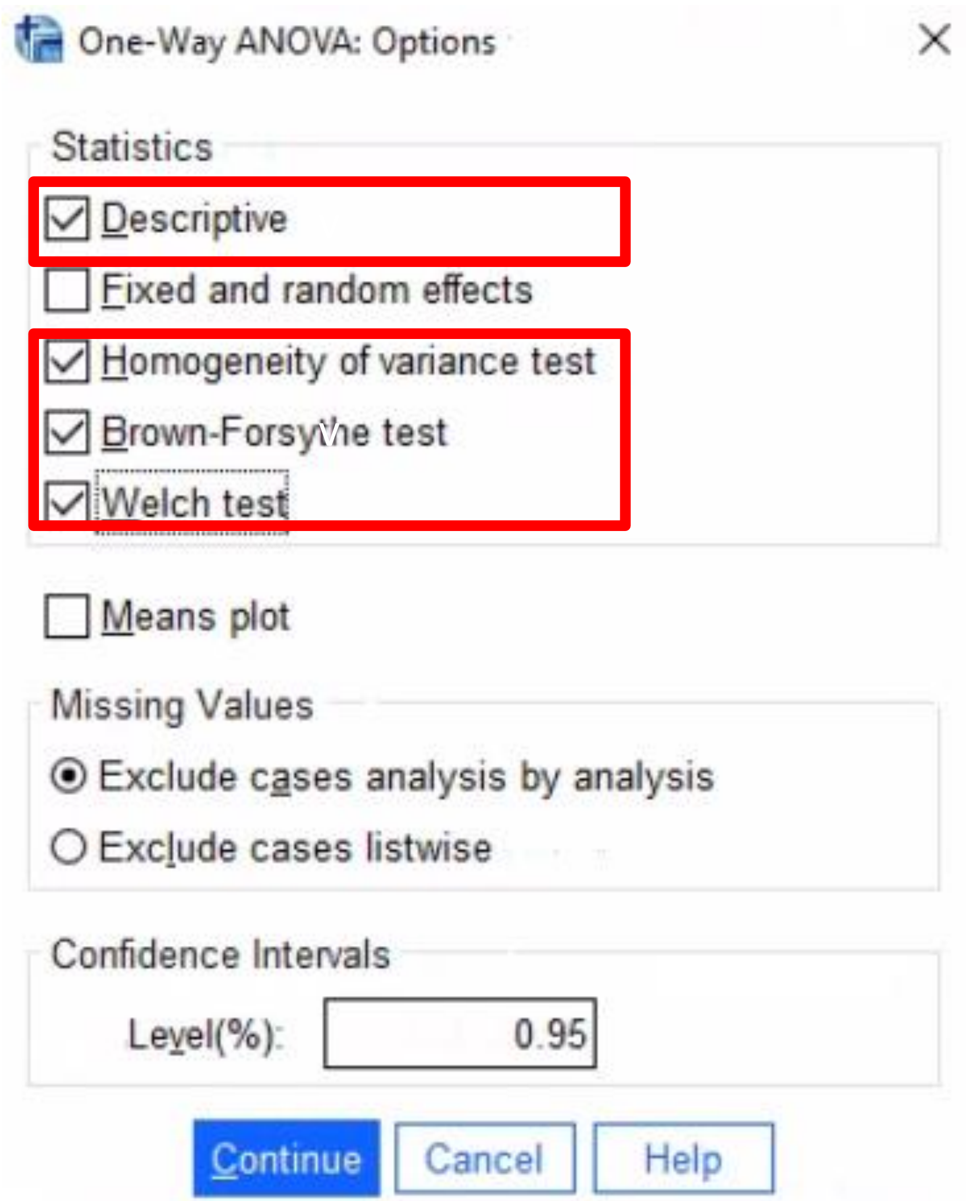
Analyze > Compare Means > One-way ANOVA



Command Dialog Box



Options


One-Way ANOVA: Options
Statistics
Descriptive
Fixed and random effects
Homogeneity of variance test
Brown-Forsythe test
Welch test
Means plot
Missing Values
Exclude cases analysis by analysis
Exclude cases listwise
Confidence Intervals
Level(%):
0.95
Continue
Cancel
Help

Example

Descriptives

LnViolentR

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Northeast	9	5.5271	.45220	.15073	5.1795	5.8747	4.82	6.06
Midwest	12	5.7229	.25994	.07504	5.5578	5.8881	5.40	6.10
South	16	6.0068	.33469	.08367	5.8285	6.1851	5.29	6.41
West	13	5.7929	.39609	.10985	5.5535	6.0322	5.28	6.41
Total	50	5.7967	.38777	.05484	5.6865	5.9069	4.82	6.41

Test of Homogeneity of Variances

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
LnViolentR	Based on Mean	1.246	3	46	.304
	Based on Median	.778	3	46	.513
	Based on Median and with adjusted df	.778	3	40.523	.513
	Based on trimmed mean	1.204	3	46	.319

Test of Homogeneity of Variances

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
LnViolentR	Based on Mean	1.246	3	46	.304
	Based on Median	.778	3	46	.513
	Based on Median and with adjusted df	.778	3	40.523	.513
	Based on trimmed mean	1.204	3	46	.319

Null hypothesis: Variances are similar for each category (homogeneity)

Goodness of Fit and Effect Size

ANOVA

LnViolentR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.426	3	.475	3.679	.019
Within Groups	5.942	46	.129		
Total	7.368	49			

ANOVA Effect Sizes[a,b]

		Point Estimate	95% Confidence Interval	
			Lower	Upper
LnViolentR	Eta-squared	.194	.005	.346
	Epsilon-squared	.141	-.060	.303
	Omega-squared Fixed-effect	.138	-.059	.299
	Omega-squared Random-effect	.051	-.019	.125

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

Goodness of Fit

ANOVA

LnViolentR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.426	3	.475	3.679	.019
Within Groups	5.942	46	.129		
Total	7.368	49			

Total SS=$\Sigma(y_i-\bar{y})^2$=Distance of every point from the overall mean.

Goodness of Fit

ANOVA

LnViolentR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.426	3	.475	3.679	.019
Within Groups	5.942	46	.129		
Total	7.368	49			

Total SS=$\Sigma(y_i-\bar{y})^2$=Distance of every point from the overall mean.

Regression SS=$\Sigma(\hat{y}-\bar{y})^2$=Distance of the predicted value from the overall mean.

Goodness of Fit

ANOVA

LnViolentR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.426	3	.475	3.679	.019
Within Groups	5.942	46	.129		
Total	7.368	49			

Total SS=$\Sigma(y_i-\bar{y})^2$=Distance of every point from the overall mean.

Regression SS=$\Sigma(\hat{y}-\bar{y})^2$=Distance of the predicted value from the overall mean.

Residual sum of squares=$\Sigma(\hat{y}-\bar{y})^2$=Distance of every point from its group mean.

Testing the Null Hypothesis

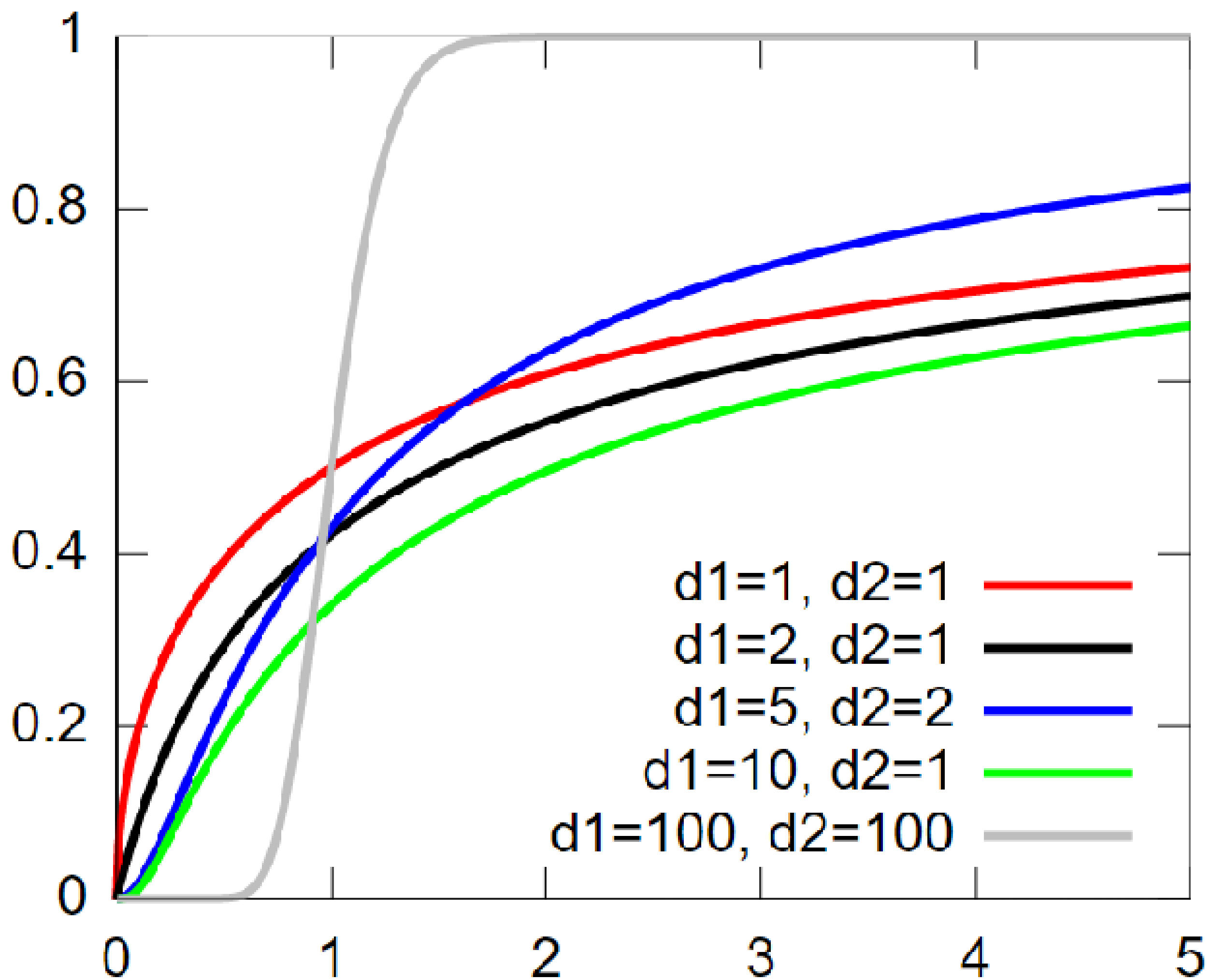
ANOVA

LnViolentR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.426	3	.475	3.679	.019
Within Groups	5.942	46	.129		
Total	7.368	49			

Null hypothesis: $\bar{\mu}_1 = \bar{\mu}_2 = \bar{\mu}_3 = \bar{\mu}_4$

Cumulative Distribution Function for F



Source: Wikipedia. Graph is licensed under Creative Commons.

Testing the Null Hypothesis

ANOVA

LnViolentR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.426	3	.475	3.679	.019
Within Groups	5.942	46	.129		
Total	7.368	49			

Null hypothesis: $\bar{\mu}_1 = \bar{\mu}_2 = \bar{\mu}_3 = \bar{\mu}_4$

Alternatives for Heterogeneity of Variances

Robust Tests of Equality of Means

LnViolentR

	Statistic[a]	df1	df2	Sig.
Welch	3.280	3	22.638	.039
Brown-Forsythe	3.466	3	32.135	.027

a. Asymptotically F distributed.

Goodness of Fit and Effect Size

ANOVA

LnViolentR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.426	3	.475	3.679	.019
Within Groups	5.942	46	.129		
Total	7.368	49			

ANOVA Effect Sizes[a,b]

		Point Estimate	95% Confidence Interval	
			Lower	Upper
LnViolentR	Eta-squared	.194	.005	.346
	Epsilon-squared	.141	-.060	.303
	Omega-squared Fixed-effect	.138	-.059	.299
	Omega-squared Random-effect	.051	-.019	.125

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

Goodness of Fit and Effect Size

ANOVA

LnViolentR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.426	3	.475	3.679	.019
Within Groups	5.942	46	.129		
Total	7.368	49			

ANOVA Effect Sizes[a,b]

		Point Estimate	95% Confidence Interval	
			Lower	Upper
LnViolentR	Eta-squared	.194	.005	.346
	Epsilon-squared	.141	-.060	.303
	Omega-squared Fixed-effect	.138	-.059	.299
	Omega-squared Random-effect	.051	-.019	.125

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

Goodness of Fit and Effect Size

ANOVA

LnViolentR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.426	3	.475	3.679	.019
Within Groups	5.942	46	.129		
Total	7.368	49			

ANOVA Effect Si[a,b]

		Point Estimate	95% Confidence Interval	
			Lower	Upper
LnViolentR	Eta-squared	.194	.005	.346
	Epsilon-squared	.141	-.060	.303
	Omega-squared Fixed-effect	.138	-.059	.299
	Omega-squared Random-effect	.051	-.019	.125

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

Interpretation of η^2

Small	.0196
Medium	.13
Large	.26

Goodness of Fit and Effect Size

ANOVA

LnViolentR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.426	3	.475	3.679	.019
Within Groups	5.942	46	.129		
Total	7.368	49			

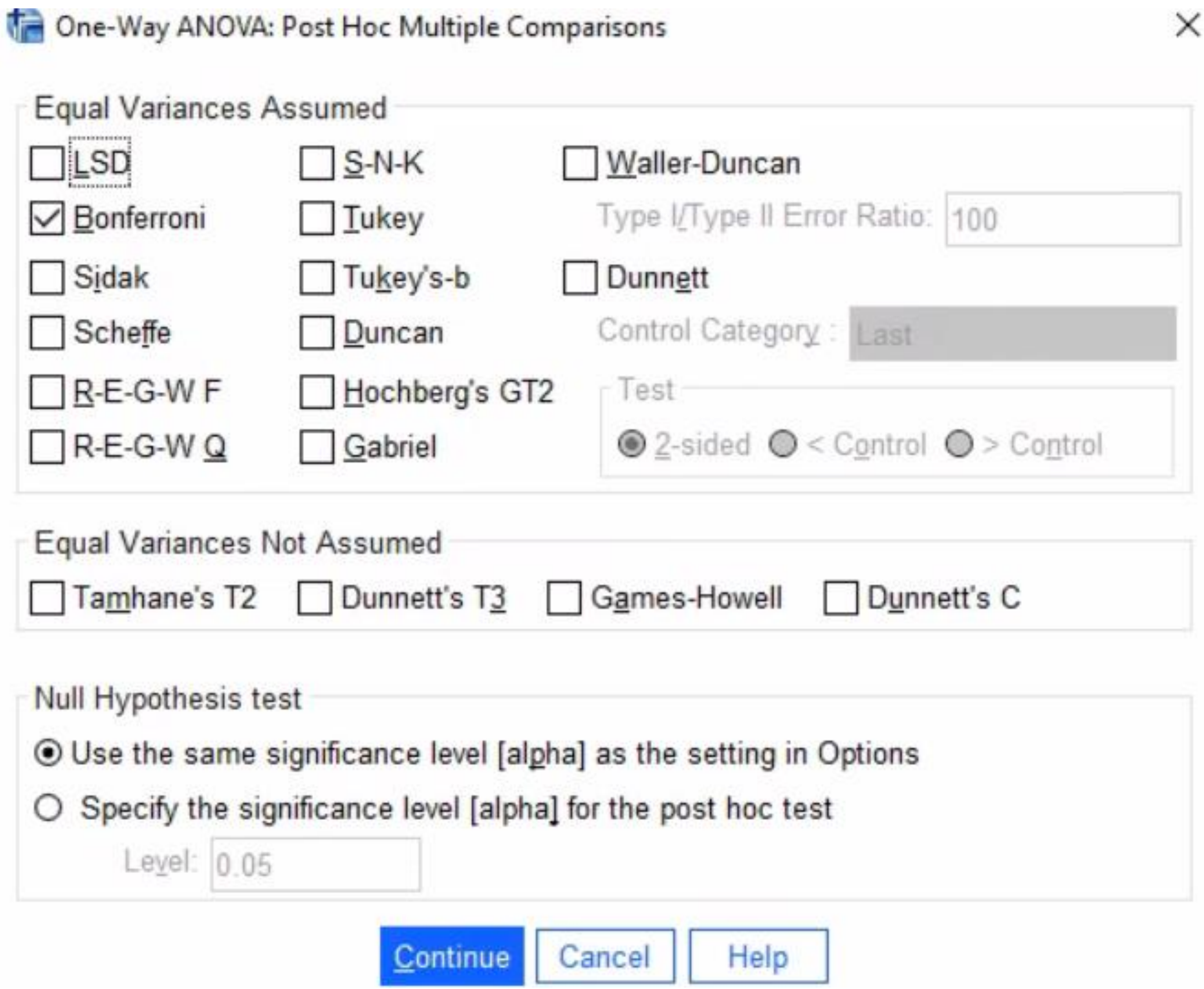
ANOVA Effect Sizes[a,b]

		Point Estimate	95% Confidence Interval	
			Lower	Upper
LnViolentR	Eta-squared	.194	.005	.346
	Epsilon-squared	.141	-.060	.303
	Omega-squared Fixed-effect	.138	-.059	.299
	Omega-squared Random-effect	.051	-.019	.125

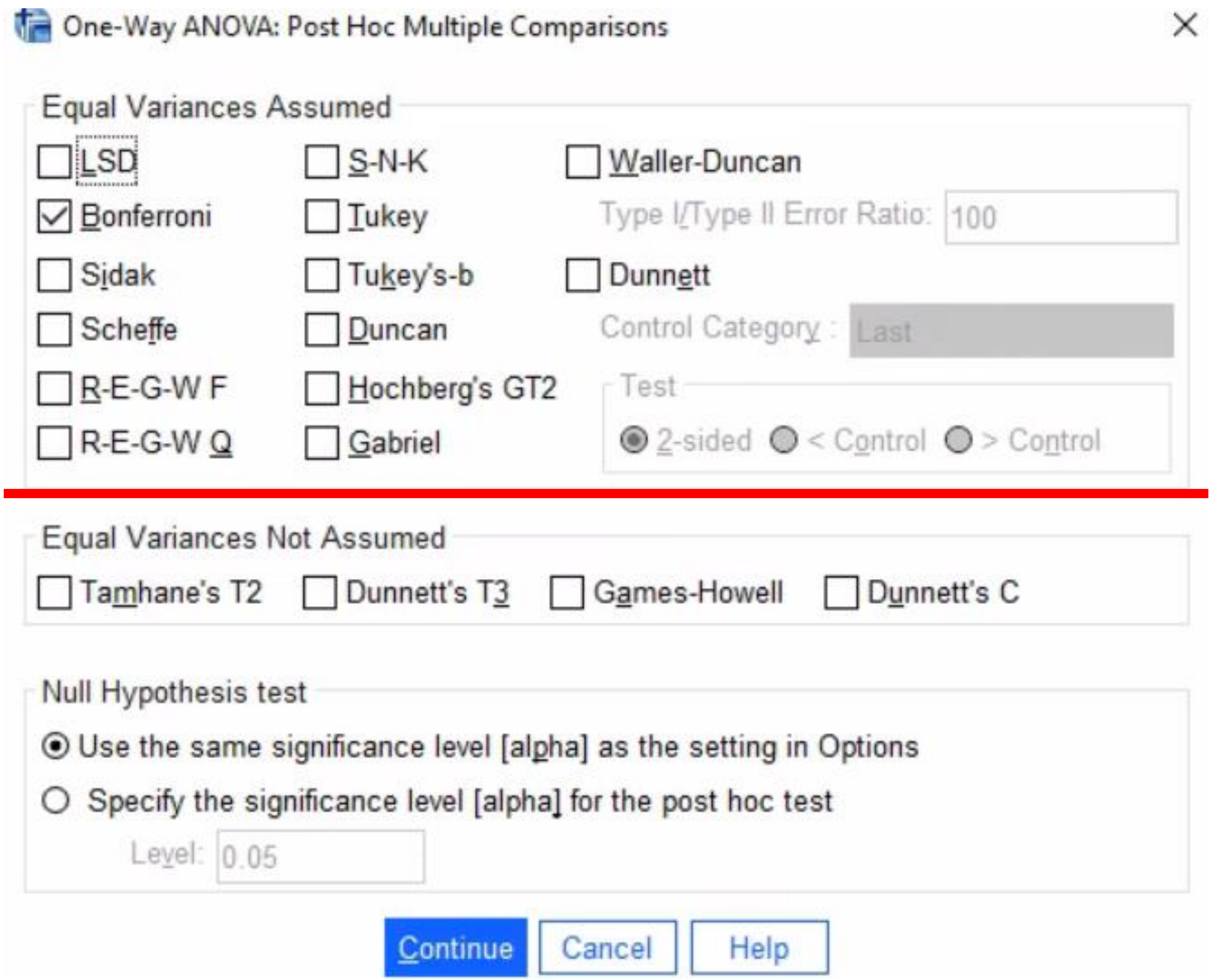
a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

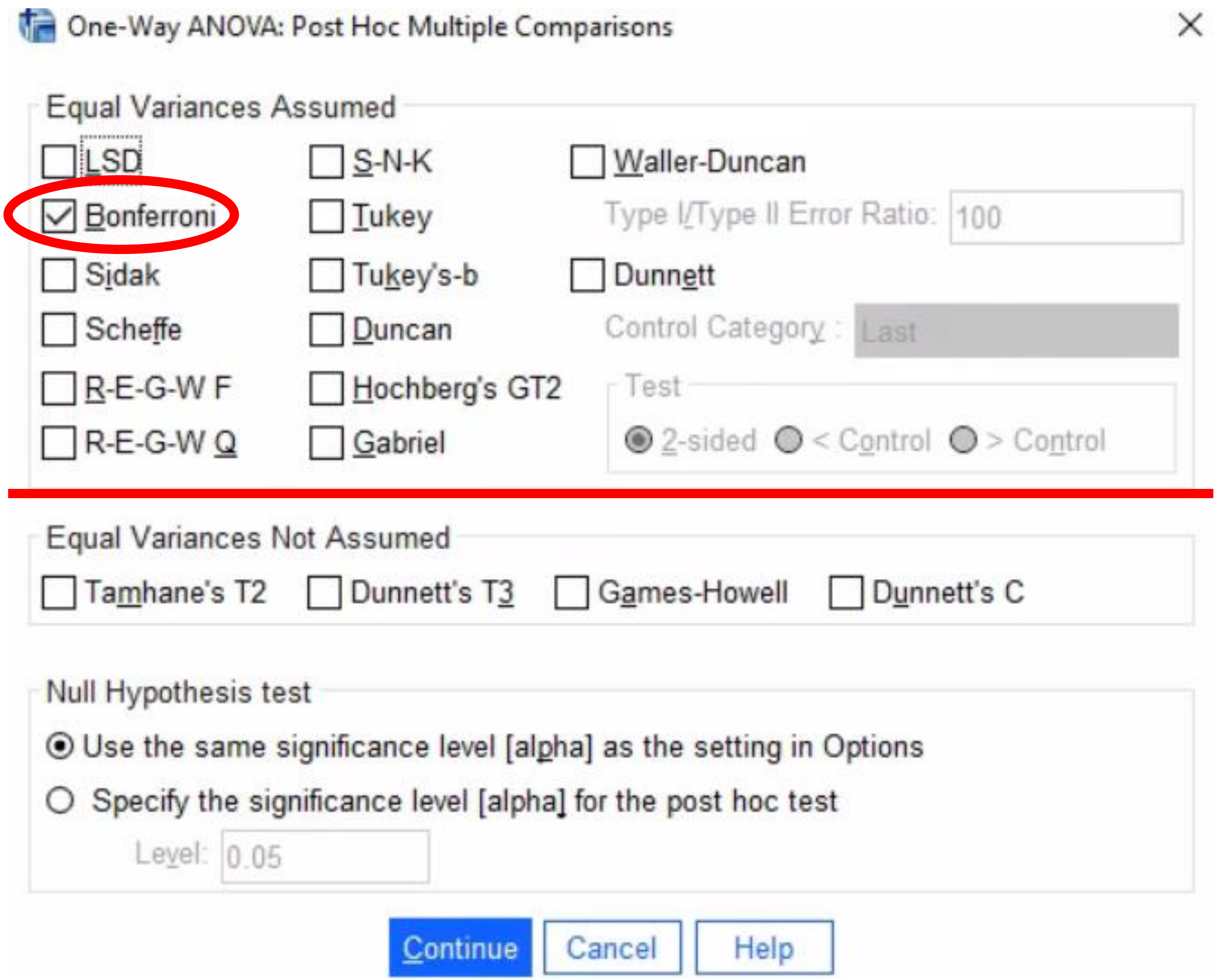
Comparison Plots



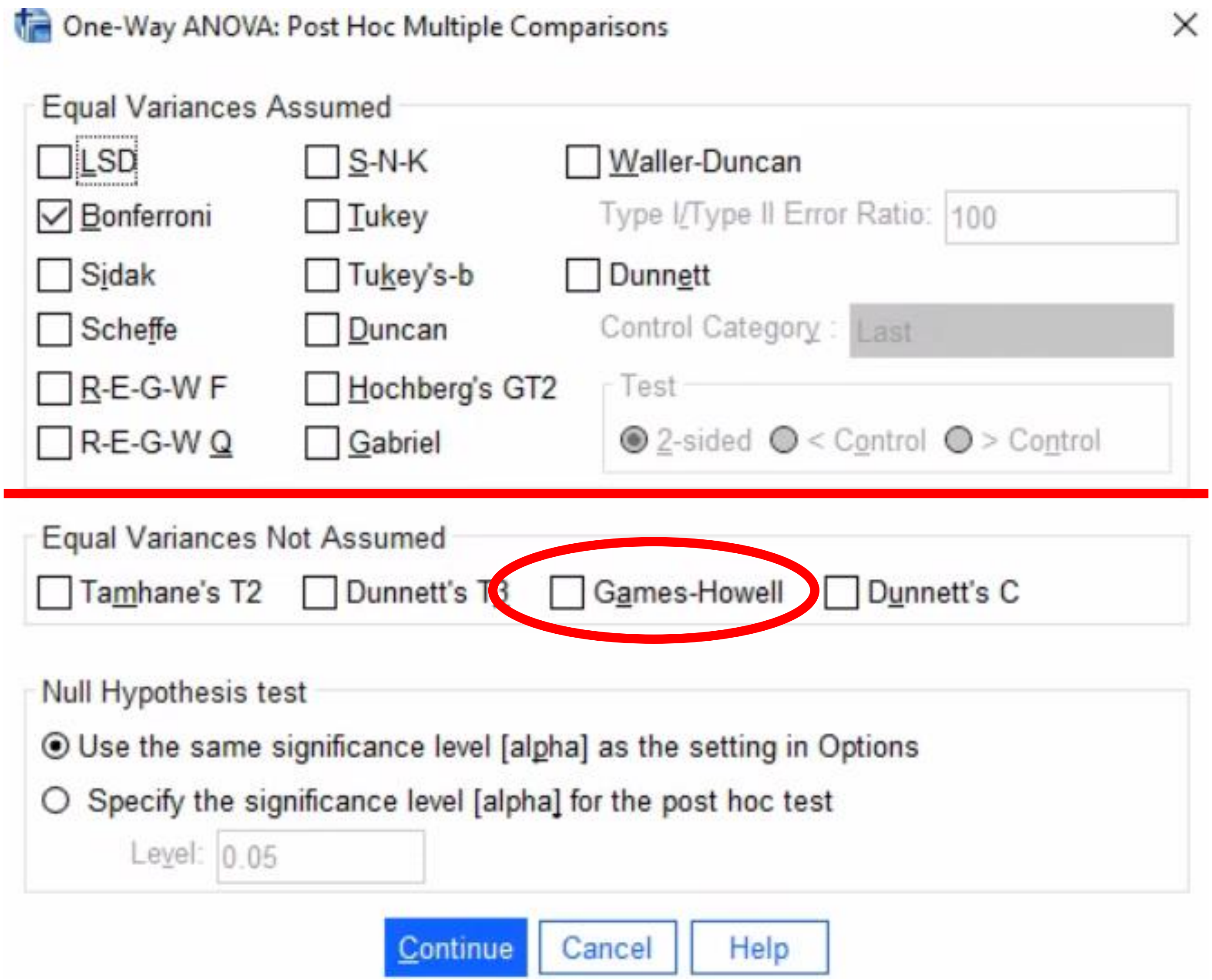
Comparison Plots



Comparison Plots



Comparison Plots



Comparison Plots

Post Hoc Tests

Multiple Comparisons

Dependent Variable: LnViolentR

Bonferroni

(I) Region	(J) Region	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Northeast	Midwest	-.19578	.15848	1.000	-.6328	.2412
	South	-.47966*	.14975	.015	-.8926	-.0668
	West	-.26574	.15585	.570	-.6954	.1640
Midwest	Northeast	.19578	.15848	1.000	-.2412	.6328
	South	-.28388	.13725	.266	-.6623	.0946
	West	-.06996	.14388	1.000	-.4667	.3267
South	Northeast	.47966*	.14975	.015	.0668	.8926
	Midwest	.28388	.13725	.266	-.0946	.6623
	West	.21391	.13420	.707	-.1561	.5839
West	Northeast	.26574	.15585	.570	-.1640	.6954
	Midwest	.06996	.14388	1.000	-.3267	.4667
	South	-.21391	.13420	.707	-.5839	.1561

*. The mean difference is significant at the 0.05 level.

Comparison Plots

Post Hoc Tests

Multiple Comparis s

Dependent Variable: LnViolentR

Bonferroni

(I) Region	(J) Region	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Northeast	Midwest	-.19578	.15848	1.000	-.6328	.2412
	South	-.47966*	.14975	.015	-.8926	-.0668
	West	-.26574	.15585	.570	-.6954	.1640
Midwest	Northeast	.19578	.15848	1.000	-.2412	.6328
	South	-.28388	.13725	.266	-.6623	.0946
	West	-.06996	.14388	1.000	-.4667	.3267
South	Northeast	.47966*	.14975	.015	.0668	.8926
	Midwest	.28388	.13725	.266	-.0946	.6623
	West	.21391	.13420	.707	-.1561	.5839
West	Northeast	.26574	.15585	.570	-.1640	.6954
	Midwest	.06996	.14388	1.000	-.3267	.4667
	South	-.21391	.13420	.707	-.5839	.1561

*. The mean difference is significant at the 0.05 level.

Comparison Plots

Post Hoc Tests

Multiple Comparis s

Dependent Variable: LnViolent

Bonferroni

(I) Region	(J) Region	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Northeast	Midwest	-.19578	.15848	1.000	-.6328	.2412
	South	-.47966*	.14975	.015	-.8926	-.0668
	West	-.26574	.15585	.570	-.6954	.1640
Midwest	Northeast	.19578	.15848	1.000	-.2412	.6328
	South	-.28388	.13725	.266	-.6623	.0946
	West	-.06996	.14388	1.000	-.4667	.3267
South	Northeast	.47966*	.14975	.015	.0668	.8926
	Midwest	.28388	.13725	.266	-.0946	.6623
	West	.21391	.13420	.707	-.1561	.5839
West	Northeast	.26574	.15585	.570	-.1640	.6954
	Midwest	.06996	.14388	1.000	-.3267	.4667
	South	-.21391	.13420	.707	-.5839	.1561

*. The mean difference is significant at the 0.05 level.

Syntax

```
ONEWAY LnViolentR BY Region
  /ES=OVERALL
  /STATISTICS DESCRIPTIVES HOMOGENEITY BROWNFORSYTHE WELCH
  /MISSING ANALYSIS
  /CRITERIA=CILEVEL(0.95)
  /POSTHOC=BONFERRONI ALPHA(0.05).
```

Issues in Use of One-way ANOVA

Possible Uses

- Definitely use with a categorical variable with 3 or more categories
- Use judiciously with a categorical variable with 2 or more categories
- Use to reduce comparison plots to reduce the number of categories in a categorical variable. (Reduces degrees of freedom / makes analysis simpler)

Controversies

- Reservations about the test for homogeneity
- Use with binary dependent variables

The End

- ANalysis Of VAriance

- Bivariate statistic
  - One numeric variable
  - One categorical variable with two or more categories

- A form of the General Linear Model