## Weighted Least-Squares Regression

For regressions with cross-section data (where the subscript "i" denotes a particular individual or firm at a point in time), it is usually safe to assume the errors are uncorrelated, but often their variances are not constant across individuals. This is known as the problem of heteroskedasticity (for "unequal scatter"); the usual assumption of constant error variance is referred to as homoskedasticity. Although the mean of the dependent variable might be a linear function of the regressors, the variance of the error terms might also depend on those same regressors, so that the observations might "fan out" in a scatter diagram.

## Approaches to Dealing with Heteroskedasticity

. For known heteroskedasticity (e.g., grouped data with known group sizes), use weighted least squares (WLS) to obtain efficient unbiased estimates;

. Test for heteroskedasticity of a special form using a squared residual regression;

. Estimate the unknown heteroskedasticity parameters using this squared residual regression, then use the estimated variances in the WLS formula to get efficient estimates of regression coefficients (known as feasible WLS); or

. Stick with the (inefficient) least squares estimators, but get estimates of standard errors which are correct under arbitrary heteroskedasticity.

In this procedure, the "residualization" method is used to obtain weights that will reduce the effect of heteroskedastic values. The method consists of four stages:

Step 1. Perform an Ordinary Least Squares (OLS) regression and obtain the residuals and squared residuals where the residual is the difference between the observed dependent variable and the predicted dependent variable value for each case.

Step 2. Regress the values of the squared residuals on the independent variables using OLS. The F test for the model is an indication of heteroskedasticity in the data.

Step 3. Obtain the reciprocal of the square root of the absolute squared residuals. These weights are then multiplied times all of the variables of the regression model.

Step 4. Obtain the OLS regression of the weighted dependent variable on the weighted independent variables. One can obtain the regression through the origin. If elected, each variable's values are converted to deviations from their respective mean before the OLS analysis is performed.

As an alternative, the user may use weights he or she has derived. These should be similar to the reciprocal values obtained in step 3 above. When these weights are used, they are multiplied times the values of each variable and step 4 above is completed.

Shown below is the dialog box for the Weighted Least Squares Analysis and an analysis of the cansas.OS4 data file.

Weighted Least Squares Regression	1		X			
Variables:	Dependent jumps Independent(s) weight waist pulse chins situps	Exploratory Weighted Least Squares Regression You may complete an Didnary Least Squares Regression analysis and save the residuals and squared residuals from that analysis. You may also complete a regression of these squared residuals on the independent variables and obtain the residuals and squared residuals from that analysis. The square root of the reciprocal of the absolute squared residuals of this last analysis may be used as weights to reduce the heteroscedasticity in your data. If this option is chosen, an OLS regression of the weighted variables is conducted. This may be conducted through the origin.				
▶	User's Weight (Optional)	Image: Second save squared residuals    Image: Plot Squared Residuals Against All Indep Variables    Image: Regress Squared Residuals on Independent Variables    Image: Save Estimated Weights in the Grid    Image: Apply Weights to Variables and Obtain in WLS Reg.    Image: Through the Drigin    Image: Use Weights Entered by the User    Image: Through the Drigin    Reset  Cancel    Compute  Return	1			

## Figure 1 Weighted Least Squares Regression Dialog

OLS REGRESSION RESULTS									
Means									
Variables	weight	waist	pulse	chins	situps	jumps			
	178.600	35.400	56.100	9.450	145.550	70.300			
Standard Devi:	ations								
Standard Devia	10115								
Variables	weight	waist	pulse	chins situ		jumps			
	24.691	3.202	7.210	5.286	62.567	51.277			
No. of valid cases = 20									
CORRELATION MATRIX									
weight waist pulse	VARIABLE weight 1.000 0.870 -0.366	waist 0.870 1.000 -0.353	pulse -0.366 -0.353 1.000	chins -0.390 -0.552 0.151	situps -0.493 -0.646 0.225	jumps -0.226 -0.191 0.035			

-0.390-0.5520.1511.0000.6960.496-0.493-0.6460.2250.6961.0000.669-0.226-0.1910.0350.4960.6691.000 chins situps jumps Dependent variable: jumps tiableBetaBStd.Err.tProb.>tVIFTOLweight-0.588-1.2210.704-1.7340.1054.4240.226waist0.98215.7186.2462.5170.0255.8570.171pulse-0.064-0.4531.236-0.3660.7201.1640.859 Variable -0.453 
 chins
 0.201
 1.947
 2.243

 situps
 0.888
 0.728
 0.205

 tercept
 0.000
 -366.967
 183.214
0.868 0.400 2.059 0.486 3.546 0.003 2.413 0.414 -2.003 0.065 Intercept SOURCE DF Prob.>F SS MS ч F Prop./r 4.901 0.0084 
 Regression
 5
 31793.741
 6358.748

 Residual
 14
 18164.459
 1297.461

 Total
 19
 49958.200
R2 = 0.6364, F = 4.90, D.F. = 5 14, Prob>F = 0.0084 Adjusted R2 = 0.5066Standard Error of Estimate = 36.02 REGRESSION OF SQUARED RESIDUALS ON INDEPENDENT VARIABLES Means Variables weight waist pulse chins situps ResidSqr 178.600 35.400 56.100 9.450 145.550 908.196 Standard Deviations waist chins Variables weight pulse situps ResidSqr 7.210 62.567 2086.828 24.691 3.202 5.286 No. of valid cases = 20 CORRELATION MATRIX VARIABLE 
 waist
 pulse
 chins
 situps
 ResidSqr

 0.870
 -0.366
 -0.390
 -0.493
 -0.297

 1.000
 -0.353
 -0.552
 -0.646
 -0.211
weight waist weight 1.000 waist 0.870 
 -0.353
 1.000
 0.151

 -0.552
 0.151
 1.000

 -0.646
 0.225
 0.696
0.225 -0.366 pulse -0.049 -0.390 chins 0.696 0.441 -0.646 -0.493 0.478 1.000 situps -0.297 -0.211 -0.049 0.441 1.000 ResidSqr 0.478 Dependent variable: ResidSqr Deca B Std.Err. t -0.768 -64.916 36.077 .ir TOL 4.424 
 Beta
 B
 Std.Err.
 t
 Prob.>t
 VIF
 TOL

 -0.768
 -64.916
 36.077
 -1.799
 0.094
 4.424
 0.226

 0.887
 578.259
 320.075
 1.807
 0.092
 5.857
 0.171
Variable weight waist pulse -0.175 -50.564 63.367 -0.798 0.438 1.164 0.859 0.316 124.826 114.955 1.086 0.296 2.059 0.486 0.491 16.375 10.515 1.557 0.142 2.413 0.414 chins situps 0.000 -8694.402 9389.303 -0.926 0.370 Intercept F Prob.>F 2.056 0.1323 DF SOURCE SS MS Regression 5 3503623.363 7007250.673 Residual 14 47705927.542 3407566.253 19 82742180.905 Total R2 = 0.4234, F = 2.06, D.F. = 5 14, Prob>F = 0.1323

Adjusted R2 = 0.2175 Standard Error of Estimate = 1845.96 X versus Y Plot X = ResidSqr, Y = weight from file: C:\Documents and Settings\Owner\My Documents\Projects\Clanguage\OpenStat4\cansaswls.OS4 Variable Mean Variance Std.Dev. ResidSqr 908.20 4354851.63 2086.83 weight 178.60 609.62 24.69 Correlation = -0.2973, Slope = -0.00, Intercept = 181.79 Standard Error of Estimate = 23.57

Number of good cases = 20



WLS	REGRESSION	RESULTS
Mear	ıs	

Variables	weight	waist	pulse	chins	situps	jumps
	-0.000	0.000	-0.000	0.000	-0.000	0.000
Standard Dev	viations					
Variables	weight	waist	pulse	chins	situps	jumps
	7.774	1.685	2.816	0.157	3.729	1.525

No. of valid cases = 20

CORRELATION MATRIX

	VARTARIE					
	weight	waist	pulse	chins	situps	jumps
weight	1.000	0.994	0.936	0.442	0.742	0.697
waist	0.994	1.000	0.965	0.446	0.783	0.729
pulse	0.936	0.965	1.000	0.468	0.889	0.769
chins	0.442	0.446	0.468	1.000	0.395	0.119
situps	0.742	0.783	0.889	0.395	1.000	0.797
jumps	0.697	0.729	0.769	0.119	0.797	1.000

Dependent variable: jumps

Variable	E	Beta	В		Std.Er	r. t		Prob.>t	VIF	TOL
weight	-2	2.281	-0.	448	0.414	-1.0	)82	0.298	253.984	0.004
waist	3	3.772	З.	415	2.736	1.2	248	0.232	521.557	0.002
pulse	-1	L.409	-0.	763	0.737	-1.0	)35	0.318	105.841	0.009
chins	- (	0.246	-2.	389	1.498	-1.5	594	0.133	1.363	0.734
situps	(	0.887	Ο.	363	0.165	2.2	202	0.045	9.258	0.108
Intercept	(	0.000	-0.	000	0.197	-0.0	000	1.000		
SOURCE	DF		SS	MS		F	Prob.	>F		
Regression	5	33.	376	6.675	8.	624	0.000	7		
Residual	14	10.	337	0.774						
Total	19	44.2	212							

R2 = 0.7549, F = 8.62, D.F. = 5 14, Prob>F = 0.0007 Adjusted R2 = 0.6674

Standard Error of Estimate = 0.88

