Violation of Assumptions: Heteroskedastic errors

Allison's example: Apparent differences in effects across groups may be an artifact of differences in residual variability

Table 1: Results of Logit Regressions Predicting Promotion to Associate Professor for Male and Female Biochemists (Adapted from Allison 1999, p. 188)

	Men		*Women*		*Ratio of*	*Chi-Square*
Variable	*Coefficient*	*SE*	*Coefficient*	*SE*	*Coefficients*	*for Difference*
Intercept	-7.6802***	.6814	-5.8420***	.8659	.76	2.78
Duration	1.9089***	.2141	1.4078***	.2573	.74	2.24
Duration squared	-0.1432***	.0186	-0.0956***	.0219	.67	2.74
Undergraduate selectivity	0.2158***	.0614	0.0551	.0717	.25	2.90
Number of articles	*0.0737****	*.0116*	*0.0340***	*.0126*	*.46*	*5.37**
Job prestige	-0.4312***	.1088	-0.3708*	.1560	.86	0.10
Log likelihood	-526.54		-306.19			
Error variance	3.29		3.29			

**p < .05, **p < .01, *** p < .001*

Allison's solution: Add delta to adjust for differences in residual variability

Table 2: Logit Regressions Predicting Promotion to Associate Professor for Male and Female Biochemists, Disturbance Variances Unconstrained (Adapted from Allison 1999, p. 195)

	All Coefficients Equal		*Articles Coefficient Unconstrained*	
Variable	*Coefficient*	*SE*	*Coefficient*	*SE*
Intercept	-7.4913***	.6845	-7.3655***	.6818
Female	-0.93918**	.3624	-0.37819	.4833
Duration	1.9097***	.2147	1.8384***	.2143
Duration squared	-0.13970***	.0173	-0.13429***	.01749
Undergraduate selectivity	0.18195**	.0615	0.16997***	.04959
Number of articles	0.06354***	.0117	0.07199***	.01079
Job prestige	-0.4460***	.1098	-0.42046***	.09007
δ	*-0.26084**	*.1116*	*-0.16262*	*.1505*
Articles x Female			*-0.03064*	*.0173*
Log likelihood	-836.28		-835.13	

**p < .05, **p < .01, *** p < .001*

In an ordinal regression model, the model for the underlying y* can be written as

$$y_i^* = \alpha_0 + \alpha_1 x_{i1} + ... + \alpha_K x_{iK} + \sigma\varepsilon_i$$

As Allison (1999, citing Amemiya 1985:269) notes, the αs and the βs are related this way:

$$\beta_k = \alpha_k / \sigma \quad k = 1, \dots, K$$

This now leads us to a potential problem with the ordered logit/probit model. When σ is the same for all cases – residuals are homoskedastic – the ratio between the βs and the αs is also the same for all cases. But, when σ differs across cases – there is heteroskedasticity – the ratio also differs (Allison 1999). As Hoetker (2004, p. 17) notes, "in the presence of even fairly small differences in residual variation, naive comparisons of coefficients [across groups] can indicate differences where none exist, hide differences that do exist, and even show differences in the opposite direction of what actually exists."

Case 1: Underlying alphas are equal, residual variances differ

αs & σ for group 0	X1 + X2 + X3	$\alpha_1 = \alpha_2 = \alpha_3 = 1$, $\sigma_0 = 1$
αs & σ for group 1	X1 + X2 + X3	$\alpha_1 = \alpha_2 = \alpha_3 = 1$, $\sigma_1 = 2$
βs for group 0	X1 + X2 + X3	$\beta_1 = \beta_2 = \beta_3 = 1$
βs for group 1	.5X1 + .5X2 + .5X3	$\beta_1 = \beta_2 = \beta_3 = .5$

In Case 1, the underlying αs are equal. But, because the residual variances differ, the βs will only be half as large for group 1 as for group 0. Naive comparisons of coefficients can indicate differences where none exist.

Case 2: Underlying alphas differ, residual variances differ

αs & σ for group 0	X1 + X2 + X3	$\alpha_1 = \alpha_2 = \alpha_3 = 1$, $\sigma_0 = 1$
αs & σ for group 1	2X1 + 2X2 + 2X3	$\alpha_1 = \alpha_2 = \alpha_3 = 2$, $\sigma_1 = 2$
βs for group 0	X1 + X2 + X3	$\beta_1 = \beta_2 = \beta_3 = 1$
βs for group 1	X1 + X2 + X3	$\beta_1 = \beta_2 = \beta_3 = 1$

In Case 2, the αs are twice as large in group 1 as in group 0. But, because the residual variances also differ, the βs are the same. Differences in residual variances obscure the differences in the underlying effects. Naive comparisons of coefficients can hide differences that do exist.

Case 3: Underlying alphas differ, residual variances differ even more

αs & σ for group 0	X1 + X2 + X3	$\alpha_1 = \alpha_2 = \alpha_3 = 1$, $\sigma_0 = 1$
αs & σ for group 1	2X1 + 2X2 + 2X3	$\alpha_1 = \alpha_2 = \alpha_3 = 2$, $\sigma_1 = 3$
βs for group 0	X1 + X2 + X3	$\beta_1 = \beta_2 = \beta_3 = 1$
βs for group 1	2/3X1 + 2/3X2 + 2/3X3	$\beta_1 = \beta_2 = \beta_3 = 2/3$

In Case 3, the αs are again twice as large in group 1 as in group 0. But, because of the large differences in residual variances, the βs are smaller for group 0 than group 1. Differences in residual variances make it look like the Xs have smaller effects on group 1 when really the effects are larger. Naive comparisons of coefficients can even show differences in the opposite direction of what actually exists.