

Comparing Logit and Probit Coefficients Across Groups

Richard Williams, University of Notre Dame, <https://academicweb.nd.edu/~rwilliam/>

Last revised November 5, 2024

Allison's example: Apparent differences in effects across groups may be an artifact of differences in residual variability, i.e. error variances are heteroskedastic.

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

Variable	Men		Women		Ratio of Coefficients	Chi-Square for Difference
	Coefficient	SE	Coefficient	SE		
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)

Variable	All Coefficients Equal		Articles Coefficient Unconstrained	
	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
Articles x Female			-0.03064	.0173
δ	-0.26084*	.1116	-0.16262	.1505
Log likelihood	-836.28		-835.13	

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

Alternative solution 1: Modify the model and make the hetero go away

```
. use https://academicweb.nd.edu/~rwilliam/statafiles/xtenure, clear
(Gender differences in receipt of tenure (Scott Long 06Jul2006))
. logit tenure i.female year c.year#c.year select articles prestige i.female#c.articles , nolog
```

```
Logistic regression                                Number of obs   =       2797
                                                    LR chi2(7)      =       414.17
                                                    Prob > chi2     =       0.0000
Log likelihood = -835.74584                        Pseudo R2      =       0.1986
```

tenure	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female						
Female	.0099733	.2011598	0.05	0.960	-.3842927	.4042392
year	1.720148	.1642749	10.47	0.000	1.398175	2.042121
c.year#c.year	-.1252837	.0141564	-8.85	0.000	-.1530297	-.0975376
select	.1521132	.0460391	3.30	0.001	.0618783	.2423481
articles	.0721718	.0113192	6.38	0.000	.0499866	.0943571
prestige	-.3935079	.0885858	-4.44	0.000	-.5671329	-.2198829
female#c.articles						
Female	-.0375456	.015789	-2.38	0.017	-.0684914	-.0065998
_cons	-7.000433	.5373974	-13.03	0.000	-8.053712	-5.947154

```
. logit tenure i.female year c.year#c.year select articles prestige i.female#c.articles
c.articles #c.articles , nolog
```

```
Logistic regression                                Number of obs   =       2797
                                                    LR chi2(8)      =       439.04
                                                    Prob > chi2     =       0.0000
Log likelihood = -823.30695                        Pseudo R2      =       0.2105
```

tenure	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female						
Female	-.2743559	.2317365	-1.18	0.236	-.7285511	.1798393
year	1.64929	.1651839	9.98	0.000	1.325536	1.973045
c.year#c.year	-.1198692	.0142307	-8.42	0.000	-.1477607	-.0919776
select	.1584052	.0466861	3.39	0.001	.0669021	.2499083
articles	.1479075	.0234248	6.31	0.000	.1019958	.1938193
prestige	-.4388466	.0897686	-4.89	0.000	-.6147898	-.2629035
female#c.articles						
Female	-.0045154	.019011	-0.24	0.812	-.0417763	.0327455
c.articles#c.articles	-.0024989	.000686	-3.64	0.000	-.0038435	-.0011544
_cons	-7.085517	.5411367	-13.09	0.000	-8.146125	-6.024909

Alternative solution 2: Heterogeneous Choice Models

With heterogeneous choice (aka Location-Scale) models, the dependent variable can be ordinal or binary. For a binary dependent variable, the model (Keele & Park, 2006) can be written as

$$\Pr(y_i = 1) = g\left(\frac{x_i\beta}{\exp(z_i\gamma)}\right) = g\left(\frac{x_i\beta}{\exp(\ln(\sigma_i))}\right) = g\left(\frac{x_i\beta}{\sigma_i}\right)$$

In the above formula,

- g stands for the link function (in this case logit; probit is also commonly used, and other options are possible, such as the complementary log-log, log-log and cauchit).
- x is a vector of values for the i th observation. The x 's are the explanatory variables and are said to be the determinants of the choice, or outcome.
- z is a vector of values for the i th observation. The z 's define groups with different error variances in the underlying latent variable. The z 's and x 's need not include any of the same variables, although they can.
- β and γ are vectors of coefficients. They show how the x 's affect the choice and the z 's affect the variance (or more specifically, the log of σ).
- The numerator in the above formula is referred to as the choice equation, while the denominator is the variance equation. These are also referred to as the location and scale equations. Also, the choice equation includes a constant term but the variance equation does not.
- The conventional logit and probit models, which do not have variance equations, are special cases of the above, where $\sigma_i = 1$ for all cases.
- Allison's model is a special case of a heterogeneous choice model, where the dependent variable is a dichotomy and both the variance and choice equations include the same dichotomous grouping variable.

In Stata, heterogeneous choice models can be estimated via the user-written routine `oglm`.

```
. * oglm replication of Allison's Table 2, Model 2 with interaction added
. use https://academicweb.nd.edu/~rwilliam/statafiles/xtenure, clear
(Gender differences in receipt of tenure (Scott Long 06Jul2006))
. oglm tenure i.female year c.year#c.year select articles prestige i.female#c.articles,
het(female )
```

```
Heteroskedastic Ordered Logistic Regression      Number of obs   =      2797
                                                LR chi2(8)      =      415.39
                                                Prob > chi2     =      0.0000
Log likelihood = -835.13347                    Pseudo R2       =      0.1992
```

	tenure	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
tenure							
	female						
	Female	-.3780592	.4500197	-0.84	0.401	-1.260082	.5039633
	year	1.838257	.202949	9.06	0.000	1.440484	2.23603
	c.year#c.year	-.1342829	.017024	-7.89	0.000	-.1676492	-.1009165
	select	.169966	.0516643	3.29	0.001	.0687058	.2712262
	articles	.0719821	.0114106	6.31	0.000	.0496178	.0943464
	prestige	-.4204743	.0961206	-4.37	0.000	-.6088671	-.2320814
female#c.articles							
	Female	-.0304836	.0187427	-1.63	0.104	-.0672185	.0062514
lnsigma							
	female	.1774195	.1627084	1.09	0.276	-.1414831	.4963221
	/cut1	7.365287	.6547118	11.25	0.000	6.082075	8.648498

```
. display "Allison's delta = " (1 - exp(.1774193)) / exp(.1774193)
Allison's delta = -.16257142
```

Problem with heterogeneous choice models: Radically different interpretations of the same results are possible

Example: Hauser & Andrew's (Sociological Methodology 2006) Logistic Response Model with Partial Proportionality Constraints.

To the surprise of many, Mare found that the effect of socio-economic status declined across educational transitions, e.g. SES had more of an impact on whether someone made the transition from grade school to high school than it did on whether or not someone made the transition from high school to college. Hauser and Andrew replicated and extended Mare's analysis of school continuation. They argued that the relative effects of some (but not all) background variables are the same at each transition, and that multiplicative scalars express proportional change in the effect of those variables across successive transitions. Specifically, Hauser & Andrew estimate two new types of models.

<i>logistic response model with proportionality constraints (LRPC):</i>	<i>logistic response model with partial proportionality constraints (LRPPC):</i>
$\log_e \left(\frac{p_{ij}}{1 - p_{ij}} \right) = \beta_{j0} + \lambda_j \sum_k \beta_k X_{ijk}$	$\log_e \left(\frac{p_{ij}}{1 - p_{ij}} \right) = \beta_{j0} + \lambda_j \sum_{k=1}^{k'} \beta_k X_{ijk} + \sum_{k'+1}^K \beta_{jk} X_{ijk}$

Hauser & Andrew used their model to examine educational transitions. But, let's see what happens when their model (and their program) are applied to the Allison Biochemist data:

```
. capture program drop lrpc02
. * Hauser & Andrew's original LRPC program
. * Code has been made more efficient and readable,
. * but results are the same.
. program define lrpc02
1.     tempvar theta
2.     version 8
3.     args lnf intercepts lambdaminus1 betas
4.     gen double `theta' = `intercepts' + `betas' + (`lambdaminus1' * `betas')
5.     quietly replace `lnf' = ln(exp(`theta')/(1+exp(`theta')))) if $ML_y1==1
6.     quietly replace `lnf' = ln(1/(1+exp(`theta')))) if $ML_y1==0
7. end

. * Hauser & Andrews original LRPC parameterization used with Allison's data
. ml model lf lrpc02 ///
>     (intercepts: tenure = male female, nocons) ///
>     (lambdaminus1: female, nocons) ///
>     (betas: year yearsq select articles prestige, nocons), max nolog
```

. ml display

Log likelihood = -836.28235

Number of obs = 2797
Wald chi2(2) = 180.60
Prob > chi2 = 0.0000

	tenure	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
intercepts							
male		-7.490506	.659663	-11.36	0.000	-8.783421	-6.19759
female		-6.23096	.6205867	-10.04	0.000	-7.447287	-5.014632
lambdaminus1							
female		-.2608322	.1080501	-2.41	0.016	-.4726066	-.0490579
betas							
year		1.909544	.1996936	9.56	0.000	1.518152	2.300936
yearsq		-.1396868	.0169425	-8.24	0.000	-.1728935	-.1064801
select		.1819201	.0526572	3.45	0.001	.0787139	.2851264
articles		.0635345	.010219	6.22	0.000	.0435055	.0835635
prestige		-.4462073	.096904	-4.60	0.000	-.6361356	-.256279

Compare that to Allison's model where he added delta:

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		
Variable	Coefficient	SE
Intercept	-7.4913***	.6845
Female	-0.93918**	.3624
Duration	1.9097***	.2147
Duration squared	-0.13970***	.0173
Undergraduate selectivity	0.18195**	.0615
Number of articles	0.06354***	.0117
Job prestige	-0.4460***	.1098
δ	-0.26084*	.1116
Articles x Female		
Log likelihood	-836.28	

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

- The similarities are obvious: other than the intercepts, which the two programs parameterize differently, the coefficient estimates are identical.
- Most critically, Allison's δ , which his program estimated and which he reported in his paper, is exactly identical to Hauser and Andrew's $\lambda - 1$, which their program estimated and which they reported in their paper.
- Hauser and Andrew's software is, in fact, a generalization of Allison's software for when there are two or more groups.
- But, the theoretical concerns that motivated their models and programs lead to radically different interpretations of the results.

Alternative solution 3: Compare predicted probabilities across groups

Long (2009) says “While regression coefficients are affected by the identifying assumption for the variance of the errors, the predicted probabilities are not... Since predicted probabilities are not affected by group differences in residual variation, you can compare groups by testing the equality of predicted probabilities at substantively interesting values of the independent variables.” Long therefore suggests estimating models in which all coefficients are free to differ by gender, and then testing whether and where predicted probabilities differ by gender. He starts with a simple model that has only gender and # of articles in the model. (Gender differences diminish as more variables are added to the model but continue to exist.)

```
. use https://academicweb.nd.edu/~rwilliam/statafiles/xtenure, clear
(Gender differences in receipt of tenure (Scott Long 06Jul2006))
```

```
. logit tenure ib1.female c.articles ib1.female#c.articles, nolog
```

```
Logistic regression                                Number of obs   =       2797
                                                    LR chi2(3)      =       121.58
                                                    Prob > chi2     =       0.0000
Log likelihood = -982.04029                        Pseudo R2       =       0.0583
```

	tenure	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
female							
Male		-.2198428	.1853876	-1.19	0.236	-.5831959	.1435102
articles		.0471351	.0104974	4.49	0.000	.0265605	.0677097
female#c.articles							
Male		.0552514	.0148436	3.72	0.000	.0261585	.0843444
_cons		-2.501162	.140056	-17.86	0.000	-2.775667	-2.226657

```
. quietly margins , dydx(female) at(articles=(0(1)50))
```

```
. marginsplot
```

