More on High School and Beyond

Covariances of Parameter Estimates

	(Intercept)	x
(Intercept)	.48271	-.00183
x	-.00183	7.0E-006

Dependent Variable: program
Model: (Intercept), x

The SE^2 for $\hat{\alpha} + \hat{\beta}x$ (needed for the confidence interval for $\pi(x)$) is

$$\hat{Var}(\hat{\alpha}) + x^2\hat{Var}(\hat{\beta}) + 2x\hat{Cov}(\hat{\alpha}, \beta).$$

Therefore a $(1-\alpha) \times 100\%$ confidence interval for $logit(\pi(x))$ is given by

$$logit(\hat{\pi}(x)) \pm z_{\alpha/2}\sqrt{\hat{Var}(\hat{\alpha}) + x^2\hat{Var}(\hat{\beta}) + 2x\hat{Cov}(\hat{\alpha}, \hat{\beta})}$$

A 95% confidence interval for $logit(\pi(100))$

$$logit(\hat{\pi}(100)) \pm 1.96\sqrt{.48271 + 100^2 0.000007 + 2(100)(-0.00183)}$$

From this we can find a confidence interval for the odds(100), or $\pi(100)$ by transforming the bounds.

Model with categorical factors

X = SES (socio economic status, categories: 1 = low, 2 = middle,3 = high)
Y = academic program (no=0, yes=1)
The dummies:

I	Variable
1	$d_1 = \begin{cases} 1 & \text{if } X = 1 \\ 0 & \text{otherwise} \end{cases}$
2	$d_2 = \begin{cases} 1 & \text{if } X = 2 \\ 0 & \text{otherwise} \end{cases}$

The model

$$logit(\pi(x)) = \alpha + \beta_1 d_1 + \beta_2 d_2$$

The estimates from SPSS:

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	.956	.1754	.612	1.299	29.673	1	.000
[ses=1.00]	-1.725	.2530	-2.221	-1.229	46.488	1	.000
[ses=2.00]	-.989	.2101	-1.401	-.577	22.152	1	.000
[ses=3.00]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						

Dependent Variable: y
Model: (Intercept), ses

Therefore the estimated model equation is

$$logit(\hat{\pi}(x)) = 0.956 - 1.725d_1 - .989d_2$$

The odds for being in an academic program for students with low SES (=1) is $e^{-1.725} = .178$ times the odds for students with high SES (=3).

(odds(1)/odds(3)=.178)

According to the output we can reject $H_0 : \beta_1 = 0$ at significance level of 5% ($\chi^2_0 = 46.49, df = 1$) and conclude that the odds for being in an academic program are significantly different for students with low and high SES.

Model with two factors

Again the response will be the variable identifying the program (1=academic, 0 = general or vocational),

The factors considered will be gender (0=female, 1=male) and school type (0=public, 1=private).

Because gender and school type are 0/1 variables and therefore already dummy type variables and can be included with the model as coded.

The model:

$$logit(\pi(gender, school)) = \alpha + \beta_1 gender + \beta_2 school$$

The SPSS output:

Goodness of Fit[b]

	Value	df	Value/df
Deviance	.321	1	.321
Scaled Deviance	.321	1	
Pearson Chi-Square	.318	1	.318
Scaled Pearson Chi-Square	.318	1	
Log Likelihood[a]	-396.377		
Akaike's Information Criterion (AIC)	798.754		
Finite Sample Corrected AIC (AICC)	.		
Bayesian Information Criterion (BIC)	796.913		
Consistent AIC (CAIC)	799.913		

Dependent Variable: program
Model: (Intercept), gender, school

a. The full log likelihood function is displayed and used in computing information criteria.

b. Information criteria are in small-is-better form.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
38.596	2	.000

Dependent Variable: program
Model: (Intercept), gender, school

a. Compares the fitted model against the intercept-only model.

The omnibus test indicates that at least one slope is different from 0. At significance level of 5% the data provide sufficient evidence ($\chi_0^2 = 38.569, df = 2, P-value < 0.001$) that either gender or school type or both have an effect on the chance of a student being in an academic program.

The Deviance is reported being 0.321 with df=1 indicating a very good model fit.

The estimates:

Parameter Estimates

Parameter			95% Profile Likelihood Confidence Interval		Hypothesis Test		
	B	Std. Error	Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	-.191	.1184	-.424	.041	2.596	1	.107
[gender=1.00]	.070	.1692	-.261	.402	.173	1	.678
[gender=.00]	0[a]	.	.	.	.	.	.
[school=1.00]	1.533	.2720	1.020	2.092	31.791	1	.000
[school=.00]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						

Dependent Variable: program
Model: (Intercept), gender, school

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

The estimated model equation:

$$logit(\hat{\pi}(gender, school)) = -.191 + 0.070 gender + 1.533 school$$

The gender has at significance level of 5% no effect on the chance of a student being in an academic program when controlling for school type ($\chi_0^2 = 0.173, df = 1, P = 0.678$). (At significance level of 5% gender and being in an academic program are conditionally independent (conditioned on the school type)).
But at significance level of 5% the data provide sufficient evidence that the proportion of students in academic programs is different ($\chi_0^2 = 31.79, df = 1, P < 0.001$) in private and public schools, when correcting for gender. Since the slope is positive this indicates that at the proportion for the higher level of school type (which is private schools) is higher than for the lower level (which is public schools).

The odds for being in an academic program are $e^{1.533} = 4.632$ times higher for students in private schools than for students in public schools.

Comment: The model introduced above does not include interaction terms. I.e. the effect of X on Y is the same for all levels of Z and the the effect of Z on Y is the same for all levels of X.
Thus this model implies homogeneous associations between X and Y given Z and between Z and Y given X.