Handout: Model Building

How many predictors?
Guideline: You need at least 10 of each outcome for every predictor.

y

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	292	48.7	48.7	48.7
	1.00	308	51.3	51.3	100.0
	Total	600	100.0	100.0	

For the HSB model we could use up to 29 variables according to this guideline, we are lucky because the data is balanced.

Example 1
HSB
The variable X was calculated as the total of science writing, civic and math, therefore
x=math + science + writing + civic
The four variables are collinear and including them all in one model would make the parameter estimates undefined

Example 2
HSB with the following predictors:
SES
school type
gender
x, academic success
reading
writing
math
science
civics

Categorical Variables Codings

			Parameter coding	
		Frequency	(1)	(2)
ses	1.00	139	1.000	.000
	2.00	299	.000	1.000
	3.00	162	.000	.000
gender	.00	327	.000	
	1.00	273	1.000	
school	.00	506	.000	
	1.00	94	1.000	

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	139.082	1	.000
	Block	139.082	1	.000
	Model	139.082	1	.000
Step 2	Step	28.161	1	.000
	Block	167.243	2	.000
	Model	167.243	2	.000
Step 3	Step	13.506	1	.000
	Block	180.749	3	.000
	Model	180.749	3	.000
Step 4	Step	12.806	2	.002
	Block	193.555	5	.000
	Model	193.555	5	.000

Variables in the Equation

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	x	.027	.003	106.411	1	.000	1.028	1.022	1.033
	Constant	-7.055	.695	103.104	1	.000	.001		
Step 2[b]	school(1)	1.421	.288	24.337	1	.000	4.143	2.355	7.287
	x	.027	.003	99.251	1	.000	1.027	1.022	1.033
	Constant	-7.160	.714	100.606	1	.000	.001		
Step 3[c]	school(1)	1.449	.291	24.699	1	.000	4.257	2.404	7.537
	x	.040	.005	72.679	1	.000	1.041	1.031	1.051
	science	-.063	.018	12.879	1	.000	.939	.907	.972
	Constant	-7.308	.725	101.481	1	.000	.001		
Step 4[d]	ses			12.506	2	.002			
	ses(1)	-.926	.289	10.256	1	.001	.396	.225	.698
	ses(2)	-.731	.238	9.413	1	.002	.481	.302	.768
	school(1)	1.374	.295	21.738	1	.000	3.952	2.218	7.043
	x	.039	.005	65.371	1	.000	1.039	1.030	1.049
	science	-.065	.018	13.502	1	.000	.937	.905	.970
	Constant	-6.193	.786	62.014	1	.000	.002		

a. Variable(s) entered on step 1: x.

b. Variable(s) entered on step 2: school.

c. Variable(s) entered on step 3: science.

d. Variable(s) entered on step 4: ses.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	692.268[a]	.207	.276
2	664.107[a]	.243	.324
3	650.601[b]	.260	.347
4	637.795[b]	.276	.368

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

b. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Example 3
HSB, compare models
M_1 with predictors school type, x, and science and
M_2 with predictors school type, x, science and SES.

Model	AIC
M_1	658.601
M_2	649.795

Therefore model M_2 gives a better fit than model M_1, even when correcting for the number for parameters in the model.

Example 4
HSB for "best" model

Classification Table[a]

Observed			Predicted		
			y		
			.00	1.00	Percentage Correct
Step 1	y	.00	208	84	71.2
		1.00	74	234	76.0
		Overall Percentage			73.7

a. The cut value is .500

Example 5
HSB for "best" model

Correlations

		y	Predicted probability
y	Pearson Correlation	1	.533**
	Sig. (2-tailed)		.000
	N	600	600
Predicted probability	Pearson Correlation	.533**	1
	Sig. (2-tailed)	.000	
	N	600	600

**. Correlation is significant at the 0.01 level (2-tailed).

R^2=0.284.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	639.691[a]	.273	.365

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.