MULTIPLE LOGISTIC REGRESSION

Example

- Suppose adding number of drinks mom had while pregnant as a predictor will help us predict odds survival?

Model:

$$\text{Log(odds surviving)} = \beta_0 + \beta_1\text{Birthweight} + \beta_2\text{Drinks}$$

$$\text{Probability of surviving } \hat{\pi} = \frac{e^{\beta 0 + \beta 1\text{Birthweight} + \beta 2\text{Drinks}}}{1+e^{\beta 0 + \beta 1\text{Birthweight} + \beta 2\text{Drinks}}}$$

Assumptions

- Linearity:
  - Box-Tidwell test (for each individual predictor)
- Randomness:
  - Random selection or random assignment?
- Independence:
  - No pairing or clustering of the data in space or time (no time/space order)

No e so don't have to worry about normality or equal variance

Linearity

- Box-Tidwell test for each predictor

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Birthweight	-.100	.050	4.060	1	.044	.905
	xbylogx	.012	.006	4.094	1	.043	1.012
	Constant	20.907	10.995	3.616	1	.057	1201354381

a. Variable(s) entered on step 1: Birthweight, xbylogx.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Drinks_1	33.842	49778.796	.000	1	.999	4.983E+14
	xbylogx	-21.916	21711.177	.000	1	.999	.000
	Constant	-11.743	76457.433	.000	1	1.000	.000

a. Variable(s) entered on step 1: Drinks_1, xbylogx.

Linearity

Birthweight:

H_0: β_2=0

H_a: $\beta_2 \neq 0$

Since the $p<0.05$, we can reject our null hypothesis. This suggests that we do not meet the assumption of linearity.

Drinks:

H_0: β_2=0

H_a: $\beta_2 \neq 0$

Since the $p>0.05$, we fail to reject our null hypothesis. This suggests that we meet the assumption of linearity.

Testing our Betas

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Drinks	.165	.103	2.543	1	.111	0.179
	Birthweight	4.676	1.642	8.115	1	.004	1.07
	Constant	-22.373	6.454	12.017	1	.001	.000

H_0: $\beta_1=0$

H_a: $\beta_1 \neq 0$

Since the $p>0.05$, we fail to reject our null hypothesis. This suggests that number of drinks doesn't significantly predict whether the baby will survive or not, **after accounting** for birthweight

Testing our Betas

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Drinks	.165	.103	2.543	1	.111	0.179
	Birthweight	4.676	1.642	8.115	1	.004	1.07
	Constant	-22.373	6.454	12.017	1	.001	.000

H_0: $\beta_2=0$

H_a: $\beta_2 \neq 0$

Since the $p<0.05$, we can reject our null hypothesis. This suggests that there is a significant log-linear relationship between birthweight and whether the baby will survive or not, **after accounting** for number of drinks

Interpretation:

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Drinks	.165	.103	2.543	1	.111	0.179
	Birthweight	4.676	1.642	8.115	1	.004	1.07
	Constant	-22.373	6.454	12.017	1	.001	.000

Drinks:

For each additional drink, the odds of surviving decrease by a factor of 0.179, **after accounting for** birthweight

Birthweight:

For each additional gram in birthweight, the odds of surviving increase by a factor of 1.07 **after accounting for** number of drinks

Maybe there is an interaction?

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Drinks	-7.475	25191.924	.000	1	1.000	.001
	Birthweight	.014	21.843	.000	1	1.000	1.014
	Birthweight_Drinks	-.004	11.808	.000	1	1.000	.996
	Constant	13.183	44296.780	.000	1	1.000	531500.232

a. Variable(s) entered on step 1: Drinks, Birthweight, Birthweight_Drinks.

Yikes! Take out the interaction term

H_0: $\beta_3=0$

H_a: $\beta_3 \neq 0$

Since the $p>0.05$, we fail to reject our null hypothesis. This suggests that the interaction term doesn't significantly predict whether the baby will survive or not, **after accounting** for birthweight and number of drinks

Assessing the model

Omnibus test

H_0: $\beta_1 = \beta_2 = 0$
H_a: at least one beta is not 0

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	30.498	3	.000
	Block	30.498	3	.000
	Model	30.498	3	.000

Since the $p<0.05$, we can reject our null hypothesis. This suggests that the baby birthweight and number of drinks together are useful in predicting whether a baby will survive or not

Comparing models

- -2 log Likelihood is a measure of how well the data fit the model
  - Unexplained variability
  - Want a smaller number

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	28.120[a]	.102	.137

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

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	10.000[a]	7.50	1.500

a. Estimation terminated at iteration number 20 because maximum iterations has been reached.

Single predictor

Multiple variables

SPSS - Q10.24