Week 6 – Thursday Session

Logistic Regression

EPI202 – Epidemiologic Methods II
Murray A. Mittleman, MD, DrPH
Department of Epidemiology, Harvard TH Chan School of Public Health

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Logistic Regression as Extension of 2x2 Tables

- Like contingency table analyses and χ^2 tests, logistic regression allows the analysis of **dichotomous** or **binary** outcomes with 2 mutually exclusive levels

- In addition, logistic regression
  - permits the use of continuous or categorical predictors and
  - provides the ability to adjust for multiple predictors

- This makes logistic regression especially useful for analysis of observational data when adjustment is needed to reduce the potential confounding resulting from differences in the groups being compared

Logistic Regression Coefficients

$$logit(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

- p is the probability that the outcome event Y occurs, Pr[Y=1|x_1, x_2, … , x_k]
- p/(1-p) is the odds of the outcome
- ln[p/(1-p)] is the log odds of the outcome, or "logit(p)"
- β_0 = log (odds) when all X_i=0
- $e^{\beta 0}$ = odds when all X_i=0
- β_i = change in log odds per unit change in X_i holding all other X's constant

  = log (odds ratio) per unit change in X_i
- $e^{\beta i}$ = odds ratio per unit increase in X_i holding all other X's constant

Odds Ratios (OR)

- The difference in the logarithms of 2 values is equal to the logarithm of the ratio of the 2 values, so by taking the exponential of β_i, we obtain the ratio of the odds (the odds ratio) corresponding to a 1-unit change in X.

$$\operatorname{logit}(p) = \beta_0 + \beta_1(\text{smoke})$$

$$\ln(\text{odds}_{\text{exposed}}) - \ln(\text{odds}_{\text{unexposed}}) = \ln\left[\frac{\text{odds}_{\text{exposed}}}{\text{odds}_{\text{unexposed}}}\right] = \ln(\text{OR})$$

$$\begin{aligned}\ln(\text{OR}) &= \ln(\text{odds}_{\text{smoke}}) - \ln(\text{odds}_{\text{no smoke}}) \\ &= [\beta_0 + \beta_1(1)] - [\beta_0 + \beta_1(0)] = \beta_1\end{aligned}$$

$$\text{OR} = \frac{\text{odds}_{\text{exposed}}}{\text{odds}_{\text{unexposed}}} = \frac{e^{\beta_0 + \beta_1(1)}}{e^{\beta_0 + \beta_1(0)}} = \frac{e^{\beta_1}}{1} = e^{\beta_1}$$

What does exp(β1) represent in a logistic regression model?

- Proportion
- Odds
- Log odds
- Odds ratio
- Do not know

Total Results: 0

In a prospective closed cohort study of the association between high stress levels vs. low stress levels and the occurrence of cancer, the reported odds ratio is 1.25. Which of the following interpretations is correct?

- There is a 25% lower odds of cancer among people with high stress than people with low stress.
- There is a 25% higher risk of cancer among people with high stress than people with low stress.
- There is a 25% higher odds of cancer among people with high stress than people with low stress.
- Do not know

Total Results: 0

Framingham Heart Study
Glucose and Death, Adjusted for High Blood Pressure

- Some software can report odds ratios directly

- Continuous Glucose: logit(p) = $\beta_0 + \beta_1$***highBP** + β_2* **glucose**

death	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
highbp	2.689838	.1863263	14.28	0.000	2.348351	3.080983
glucose1	1.013414	.0018089	7.47	0.000	1.009875	1.016966
_cons	.1234172	.0187604	-13.76	0.000	.0916191	.1662515

- Categories of Glucose: logit(p) = $\beta_0 + \beta_1$***highBP** + β_2* **glucose$_{76\text{-}99}$** + β_3* **glucose$_{100\text{-}126}$** +β_4* **glucose$_{>126}$**

death	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
highbp	2.723607	.1887925	14.45	0.000	2.377616	3.119947
gluccat						
2	1.009428	.0732079	0.13	0.897	.8756745	1.163612
3	1.479099	.1956121	2.96	0.003	1.141367	1.916768
4	6.84843	1.845605	7.14	0.000	4.038295	11.61406
_cons	.3404762	.0206828	-17.74	0.000	.3022589	.3835256

Framingham Heart Study
High Blood Pressure and Death, By Gender

- Interaction: logit(p) = $\beta_0 + \beta_1$***highBP** + β_2***women+** β_3 * **highBP*women**

death	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
highbp	.889004	.0967952	9.18	0.000	.699289	1.078719
women	-.8345291	.0871063	-9.58	0.000	-1.005254	-.6638039
highbpwomen	.2888108	.1336518	2.16	0.031	.0268581	.5507635
_cons	-.5937747	.0591273	-10.04	0.000	-.709662	-.4778874

$$\ln(OR_{\text{high BP}|\text{men}}) = \ln(odds_{\text{high BP},\text{men}}) - \ln(odds_{\text{normal BP},\text{men}})$$

$$= \ln(odds_{\text{highBP},\text{men}}) = -.5937747 + .889004(1) - .8345291(0) + .2888108(1 * 0)$$

$$= \ln(odds_{\text{normal BP},\text{men}}) = -.5937747 + .889004(0) - .8345291(0) + .2888108(0 * 0)$$

$$= .889004(1)$$

$$OR_{\text{high BP}|\text{men}} = e^{.889004} = 2.43$$

$$\ln(OR_{\text{high BP}|\text{women}}) = \ln(odds_{\text{high BP},\text{women}}) - \ln(odds_{\text{normal BP},\text{women}})$$

$$= \ln(odds_{\text{highBP},\text{women}}) = -.5937747 + .889004(1) - .8345291(1) + .2888108(1 * 1)$$

$$= \ln(odds_{\text{normal BP},\text{women}}) = -.5937747 + .889004(0) - .8345291(1) + .2888108(1 * 0)$$

$$= .889004(1) + .2888108(1) = 1.178$$

$$OR_{\text{high BP}|\text{women}} = e^{1.178} = 3.25$$

Week 6 Logistic Regression Exercise

HAVE A GOOD WEEKEND