Unit 10 Biostatistics Recitation

Ryan Wei

2023-07-07

Problem 4

You decide to conduct a case-control study of your own to test if smoking status during pregnancy (1 = never, 2 = past, 3 = current) is associated with infant mortality (0 = did not die within one year of birth, 1 = died within one year of birth).

Problem 4a

► Should you use linear or logistic regression to answer this question? Explain your reasoning.

Problem 4a

- ► Should you use linear or logistic regression to answer this question? Explain your reasoning.
- **▶** Binary outcome → Logistic regression

Problem 4b

. logistic mo	rtality i.smol	king, nolog					
Logistic regr	ession			Number o	of obs	-	120
				LR chi2	(2)	=	8.96
				Prob > 0	chi2	=	0.0114
Log likelihoo	d = -78.69982	2	Pseudo R2			-	0.0538
mortality	Odds Ratio	Std. Err.	z	P> z	[95%	Conf.	Interval]
smoking							
2	2.284615	1.045482	1.81	0.071	.931	7246	5.601942
3	3.994083	1.919169	2.88	0.004	1.55	7443	10.24288
_cons	.4814815	.1625384	-2.17	0.030	.248	4455	.9330997
Note: _cons e	stimates base:						
		5,8					
Logistic regr	ession			Number		=	120
				LR chi2		=	8.96
				Prob > 0		=	0.0114
Log likelihoo	d = -78.69982	2		Pseudo I	R2	=	0.0538
mortality	Coef.	Std. Err.	z	P> z	[95%	Conf.	Interval]
smoking							
2	.8261977	.4576185	1.81	0.071	07	0718	1.723113
3	1.384814	.4805031	2.88	0.004	.443	0453	2.326583
_cons	7308875	.3375798	-2.17	0.030	-1.39		0692433

▶ What is the difference between the two outputs above and how are they related?

Problem 4b: Answer

► The first output provides the odds ratios comparing the odds of infant mortality in each group with the odds of infant mortality among babies born to mothers who were never smokers at the time of pregnancy.

Problem 4b: Answer

- The first output provides the odds ratios comparing the odds of infant mortality in each group with the odds of infant mortality among babies born to mothers who were never smokers at the time of pregnancy.
- ► The second output provides the coefficients used in the regression equation.

Problem 4b: Answer

- ▶ The first output provides the odds ratios comparing the odds of infant mortality in each group with the odds of infant mortality among babies born to mothers who were never smokers at the time of pregnancy.
- ► The second output provides the coefficients used in the regression equation.
- The odds ratios can be calculated by exponentiating the coefficients: $e^{\hat{\beta}_2} = \widehat{OR}_2$.
- The coefficients can be calculated by taking the natural log of the odds ratios $\hat{\beta}_2 = \ln(\widehat{OR}_2)$.

```
exp(1.384814);log(2.284615)
```

```
## [1] 3.994083
```

[1] 0.8261975

Problem 4c

Write the fitted regression model. Be sure to define your variables.

							ng, nolog	tality i.smol	. logistic mor
12		=	obs	of	Number			ssion	Logistic regre
8.9		=		2(2)	LR chi				
0.011		=	2	chi	Prob >				
0.053		-	Pseudo R2				Log likelihood = -78.699822		
erval	1	Conf.	[95%		P> z	z	Std. Err.	Odds Ratio	mortality
	_								smoking
60194		/246	.9317		0.071	1.81	1.045482	2.284615	2
.2428		443	1.557		0.004	2.88	1.919169	3.994083	3
33099		455	. 2484		0.030	-2.17	.1625384	.4814815	_cons
							ne odds.	timates base	Note: _cons es
							nolog	ity i.smoking	. logit mortal
12		-	obs	of	Number			ssion	Logistic regre
8.9		=		2(2)	LR chi				
0.011		=	2	chi	Prob >				
0.053		=		R2	Pseudo			= -78.69982	Log likelihood
erval	1	Conf.	[95%		P> z	z	Std. Err.	Coef.	mortality
									smoking
72311		718	076		0.071	1.81	.4576185	.8261977	2
32658		1453	.4436		0.004	2.88	.4805031	1.384814	3
69243		2532	1.392	-	0.030	-2.17	.3375798	7308875	_cons
8. 9.01 9.05 601 9.24 33309 11 8.0 9.05 9.05 13 13 13 13 13 13 13 13 13 13	. 1	= = = = = = = = = = = = = = = = = = =	2 [95% .9317 1.557 .2484 obs 2 [95% 076 .4436	of 2(2) 2(2) 2(2) 2(2) 2(2) 2(2)	LR chi Prob > Pseudo Pseudo P> z a.071 a.084 Number LR chi Prob > Pseudo Pseudo a.084 a.084 a.084	1.81 2.88 -2.17	Std. Err. 1.045482 1.919169 .1625384 ne odds. nolog Std. Err. .4576185 .4805031	= -78.69982: Odds Ratio 2.284615 3.994983 4814815 :imates base: tty 1.smoking sion = -78.69982: Coef8261977 1.384814	i ti

Problem 4c: Answer

- ► Model: $logit(\hat{p}) = ln\left(\frac{\hat{p}}{1-\hat{p}}\right) = -0.73 + 0.83x_2 + 1.38x_3$

 - \triangleright p = the estimated probability of death within the first year of birth

Problem 4d

Based on what you know about the study, state which values of the first column in the Stata output (i.e. Odds Ratio/Coef.) are interpretable and interpret those values.

Logistic regr	ession			Number o	of obs	=	120
				LR chi2((2)	=	8.96
				Prob > 0	hi2	=	0.0114
Log likelihood = -78.699822				Pseudo F	R2	-	0.0538
mortality	Odds Ratio	Std. Err.	z	P> z	[95%	Conf.	Interval]
smoking							
2	2.284615	1.045482	1.81	0.071	.931	7246	5.601942
3	3.994083	1.919169	2.88	0.004	1.55	7443	10.24288
_cons	.4814815	.1625384	-2.17	0.030	.248	4455	.9330997
Note: _cons e							
. logit morta	lity i.smokin			Northead			420
_	lity i.smokin			Number o		=	120
. logit morta	lity i.smokin			LR chi2	(2)	=	8.96
. logit morta	lity i.smokin	g, nolog			(2) :hi2		
. logit morta	lity i.smokin	g, nolog	z	LR chi2(Prob > c	(2) chi2 R2	=	8.96 0.0114
. logit morta Logistic regr	lity i.smoking ession d = -78.69982	g, nolog	z	LR chi2(Prob > c Pseudo F	(2) chi2 R2	=	8.96 0.0114 0.0538
. logit morta Logistic regr Log likelihoo mortality	lity i.smoking ession d = -78.69982	g, nolog	2	LR chi2(Prob > c Pseudo F	(2) thi2 R2	=	8.96 0.0114 0.0538
. logit morta Logistic regr Log likelihoo mortality smoking	lity i.smoking ession d = -78.69982	g, nolog 2 Std. Err.		LR chi2(Prob > c Pseudo F	[95% 07	= = = Conf.	8.96 0.0114 0.0538 Interval]

Problem 4d: Answer

➤ Since this is a case-control study, we cannot estimate the probability or odds of the outcome, infant mortality. However, we can estimate the odds ratio for this outcome. Therefore, we cannot interpret the intercept (i.e. the value in the row for _cons in the Stata output), but we can interpret the slopes (i.e. the values in the rows for smoking in the Stata output).

Problem 4d: Answer

➤ Since this is a case-control study, we cannot estimate the probability or odds of the outcome, infant mortality. However, we can estimate the odds ratio for this outcome. Therefore, we cannot interpret the intercept (i.e. the value in the row for _cons in the Stata output), but we can interpret the slopes (i.e. the values in the rows for smoking in the Stata output).

Interpretation

- $ightharpoonup \widehat{OR}_2$: The estimated odds ratio comparing the odds of infant mortality among babies born to mothers who were past smokers during pregnancy is 2.28 times (or 128% higher than) the odds of infant mortality among babies born to mothers who were never smokers during pregnancy.
- ▶ \widehat{OR}_3 : The estimated odds ratio comparing the odds of infant mortality among babies born to mothers who were current smokers during pregnancy is 3.99 times (or 299% higher than) the odds of infant mortality among babies born to mothers who were never smokers during pregnancy.