Biostatistics 140.654
Fourth Term, 2021
April 12, 2021

Quiz 1 SOLUTION

The purpose of this quiz is to assess your knowledge of the course materials covered during the first two weeks of class and covered in Problem Set 1.

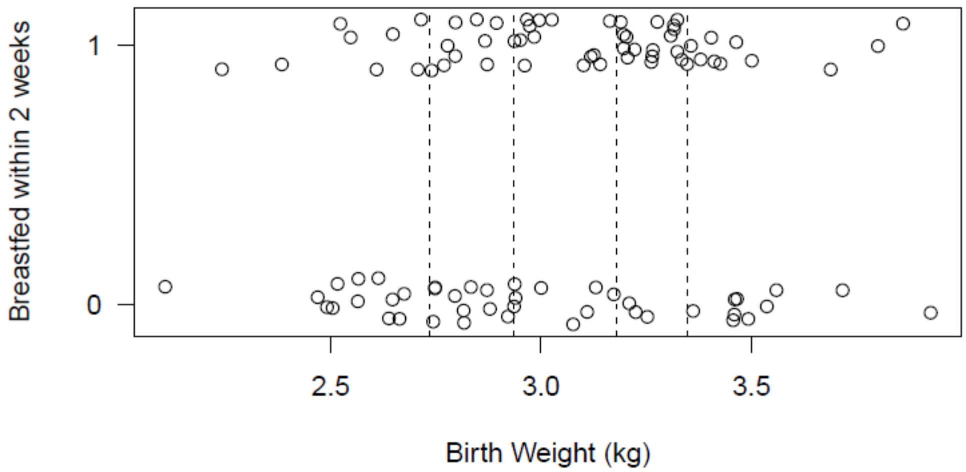
Instructions:

- **This is an open book quiz; you may consult your course notes and handouts.**
- **You should not discuss this quiz with any other student during Monday April 12th.**
- **This quiz is designed to be completed in 20-30 minutes.**
- **You can use calculators or R on your computer for arithmetic. But you should NOT use the 'glm' function in R to compute estimates of logistic regression coefficients.**
- **You may provide your solution by editing the word version of this quiz, annotating the pdf version of this quiz or writing your solution on paper and submitting a picture of your solution.**

By signing my name, I enter agree to abide by the instructions above and the Johns Hopkins University School of Public Health Academic Code:

Name (Print): ____

Signature: ____



1. **In the figure above, you will find a display of data from a set of 100 Nepali infants showing whether each infant began breastfeedling within the first 2 weeks (1 – yes, 0 – no; jittered) against the child's birth weight (kg). Vertical lines are drawn at roughly the quintiles of birth weight. Use these data to estimate the coefficients in a simple logistic regression model.**

   Report:

   a. **The logistic regression equation and your approximate estimates of the coefficients in your model.**

Strata	BW (midpoint)	N	# Y = 1	Pr(Y=1)	Log(Pr(Y=1)/(1-Pr(Y=1))
1	2.42	20	8	0.40	-0.41
2	2.84	20	9	0.45	-0.20
3	3.06	20	13	0.65	0.62
4	3.26	20	17	0.85	1.73
5	3.64	20	10	0.5	0.00

Data simulated as: Log(Pr(Y=1)/(1-Pr(Y=1)) = 0 + 1 (BW – 3)

Using the data above: Log(Pr(Y=1)/(1-Pr(Y=1)) = 0.3 + $\boxed{0.8}$ (BW – 3)

$$.8 = \log\left[Pr(Y=1 \mid BW=x)/Pr(Y=0 \mid BW=x)\right] - \log\left[Pr(Y=1 \mid BW=x-1)/Pr(Y=0 \mid BW=x-1)\right]$$

$\exp(.8) = 2.2 = \dfrac{Pr(Y=1 \mid BW=x) / Pr(Y=0 \mid BW=x)}{Pr(Y=1 \mid BW=x-1) / Pr(Y=0 \mid BW=x-1)}$

120%

b. **The approximate predicted probability of breast feeding within 2 weeks for a child with birth weight of 2 kg**

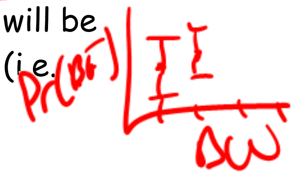
$8/20 = .4$

Exp(0.3 + 0.8 (2 – 3)) / (1 + Exp(0.3 + 0.8 (2 – 3)) = 0.38

c. **Your findings in a sentence or two for a public health journal. Be numerate, eliminate jargon to the extent possible.**

log odds

In a sample of Nepali newborns with birthweights ranging from roughly 2 to 4kg, we observe a positive relationship between being breastfed and birthweight. For newborns who weight 2 kg, we estimate that 38% will be breastfed and the odds of breastfeeding increase by a factor of 2 (i.e. exp(0.8) = 2.22) per kg increase in birthweight.



2. **Below find two 2x2 tables showing: whether or not a person spent more than $1000 on medical services (Y), whether the person has a major smoking cause disease (mscd=1) or not (mscd=0), and age group.**

Age < 65

Y	MSCD 0	MSCD 1
0	5436	119
1	2028	323

Age ≥ 65

Y	MSCD 0	MSCD 1
0	2647	280
1	1878	881

$age = \begin{cases} 1 & \geq 65 \\ 0 & < 65 \end{cases}$

* **The scientific question is whether the mscd effect on risk of an expenditure above $1,000 is the same for persons younger than 65 vs. older.**

$\text{logit}[Pr(Y=1 \mid mscd, age)]$

a. **Conduct an analysis to answer this question.**

$(1) = \beta_0 + \beta_1 mscd$

$*(2) = \alpha_0 + \alpha_1 mscd + \alpha_2 age$

$(3) = \gamma_0 + \gamma_1 mscd + \gamma_2 age + \gamma_3 mscd \times age$

Age < 65:

log(OR) = log[(323*5436) / (119*2028)] = 1.98

Var(log(OR)) = 1/323 + 1/5436 + 1/119 + 1/2028 = 0.012

Age ≥ 65

Log(OR) = log[(881*2647) / (280*1878)] = 1.58 1.48

Var(log(OR)) = 1/881 + 1/2647 + 1/280 + 1/1878 = 0.0056

Generate 2 95% CIs for log OR ⇒ no overlap

Diff LogORs → Z test $H_0: = 0$

generate CI for Diff LogOR

H_0: Diff in Log OR = 0 $\quad H_A$: $\neq 0$

Diff: 1.98 – 1.58 = 0.40
Var(DIFF) = 0.012 + 0.0056
Se(DIFF) = 0.13
95% CI: 0.40 – 2 * 0.13, 0.40 + 2 * 0.13 -> (0.14, 0.66)

$$z = \frac{\text{obs diff} - 0}{se(\text{diff})} = \frac{.4 - 0}{.13} \approx 3$$

b. Write a sentence or two to report your findings to a public health audience. Be numerate. Avoid jargon!

Among persons at least 65 years of age, the odds of a big expenditure are an estimated 4.85 times greater for persons with a MSCD vs. those without. Whereas, for persons under 65 years of age, the odds of a big expenditure are an estimated 7.24 times greater for persons with a MSCD vs. those without. Accounting for the sample size and variation in the data, these estimates indicate that the odds ratio of a big expenditure comparing persons with and without a MSCD is greater for persons under 65 years of age (ratio of odds ratios: 1.49, 95% 1.15 – 1.93).

CI's not overlapping

z-test