DALITE Q8 - Logistic Regression. Solutions.

EPIB607 - Inferential Statistics[a]

[a]Fall 2018, McGill University

This version was compiled on November 19, 2018

This DALITE quiz will cover logistic regression.

Two sample proportions | Odds ratio | Logit transformation | Odds | Log odds | Risk calculator

FINRISK Calculator

There are now several risk calculators derived from epidemiological studies, such as the Framingham Heart Study. See, e.g., https://www.uptodate.com/home/medical-calculators and https://qxmd.com/calculate/.

Many of the online ones for 10 year risks are based on the proportional hazards model. One risk equation that is based on logistic regression (see Rothman chapter 12, 2012 or Chapter 10, 2002) is that described in the article Predicting Coronary Heart Disease and Stroke The FINRISK Calculator by Erkki Vartiainen et al. GLOBAL HEART, VOL. 11, NO. 2, 2016 June 2016: 213-216 [article in myCourses]

By applying the equation in Table 2 of the article for the 10 year risk of coronary heart disease for 55 year old females who don't smoke, but whose cholesterol is 6, systolic blood pressure is 140, LDL is 0.5, are NOT diabetic, and do not have a family history of MI, one obtains 9.3%.

The same equation applied to females with the same profile, except that they ARE diabetic, yields a 10 year risk of 22.3%.

Thus, with respect to the No Diabetes (ref. category) vs. Diabetes (index category) contrast, say if the following statement is TRUE or FALSE and explain why in your rationale:

1. The risk difference is 13%

a. **TRUE (Correct)**
b. FALSE

1.1. Correct rationales.

- When we want to calculate the risk difference we have to subtract the two risks from either group so in this case that would be 22.3% - 9.3%= 13%

1.2. Incorrect rationales.

2. The odds ratio is 2.4

a. TRUE
b. **FALSE (Correct)**

2.1. Correct rationales.

- Estimate for the risk ratio rather than the odds.
- 2.4 is the risk ratio (when we divide 22.3/9.3) This is not the odds ratio

2.2. Incorrect rationales.

- You divide the risk of the exposed over the risk of the unexposed to get the odds ratio. In this case 22.3/9.3 = 2.40, answer A) True.

3. The risk ratio is 2.8

a. TRUE
b. **FALSE (Correct)**

3.1. Correct rationales.

- The odds ratio is 2.8 ((0.223)/(1-0.223)/(0.093/(1-0.093)

3.2. Incorrect rationales.

- (1) Find odds first Odds (no diabetes) = 0.093 / (1 - 0.093) = 0.1025 Odds (diabetes) = 0.223 / (1 - 0.223) = 0.2870 (2) Exponentiate the odds No diabetes: exp(0.1025) = 1.107937 Diabetes: exp(0.2870) = 1.332424 Odds ratio = 1.352424 / 1.107937 = 1.20 Therefore it is false.

4. The two risks can be connected via the equation: risk(%) = 22.3 - 13 * Diabetes

a. TRUE
b. **FALSE (Correct)**

4.1. Correct rationales.

- Since no diabetes is the reference category (X= 0); which is risk(%) = 9.3 + 13*Diabetes

4.2. Incorrect rationales.

5. The two risks can be connected via the equation: log[risk] = -2.38 + 0.87 * Diabetes

a. **TRUE (Correct)**
b. FALSE

5.1. Correct rationales.

- The regression model:

$$\log(risk) = \log(\mu_0) + \log(\theta) * Diabetes$$

Then

$$\log(\mu_0) = \log(0.093) = -2.38$$

and

$$\log(\theta) = \log(risk\ ratio) = \log(0.223/.093) = 0.87$$

5.2. Incorrect rationales.

6. The two risks can be connected via the equation: log[risk/(1-risk)] = -2.28 + 1.03 * Diabetes

a. **TRUE (Correct)**
b. FALSE

6.1. Correct rationales.

-

$$\log[R1/(1-R1)] = \log[R0/(1-R0)] + \log[R1/(1-R1)]/[R0/(1-R0)] \cdot diabetes$$
$$\log[R1/(1-R1)] = \log(9.3/90.7) + \log(2.8) \times diabetes$$

6.2. Incorrect rationales.

7. The two risks can be connected via the equation: log[risk/(1-risk)] = 1.03 - 2.28 * Diabetes

a. TRUE
b. **FALSE (Correct)**

7.1. Correct rationales.

- Coefficients reversed from previous question

7.2. Incorrect rationales.

8. 4 year risk for intermittent claudication I

Using the coefficients in his Table 12.2, Rothman (2012) calculated a 4 year risk for intermittent claudication of 6.7% if the 70-year old man had the example profile, which included normal blood pressure, AND diabetes.

If the profile did NOT include diabetes, but was otherwise unchanged, then (ignoring small rounding errors) the single equation that connects/expresses/summarizes the risks for those without (reference category) and with diabetes (Index category) can be written as (select all that apply and justify your choices in the rationale)

a. **%Risk[Diabetes] = 2.7 * (2.5 ^ Diabetes) (Correct)**
b. **Odds[Diabetes] = (2.7/97.3) * (((6.7/93.3)/(2.7/97.3)) ^ Diabetes) (Correct)**
c. **Odds[Diabetes] = (2.7/97.3) * (2.59 ^ Diabetes) (Correct)**
d. **log[Odds[Diabetes]] = -3.58 + 0.95 * Diabetes (Correct)**
e. log[Risk[Diabetes]] = -3.58 + 0.95 * Diabetes

8.1. Correct rationales.

- A, B, C, and D are correct. If diabetes = 1 and no-diabetes = 0, the equation of A would work: 2.7 * (2.5 ˆ Diabetes) 2.7 * (2.5 ˆ 0) 2.7 (1) = 2.7 = the risk among the unexposed, it's the intercept 2.7 * (2.5 ˆ Diabetes) 2.7 * (2.5 ˆ 1) 2.7 (1) = 6.7 = the risk among the exposed E is incorrect because when you calculate the log of theta, we get a value that is NOT 0.95.
- A, B, C, and D are correct A is correct because when diabetes =0, % risk = 2.7, and when diabetes =1, % risk = 6.75% which can be rounded down to 6.7% E is wrong because even though Risk when no diabetes = 0.02717253 ln0.02717253 = 3.60 ln (risk ratio) = ln(6.7/2.7) = 0.909. This is very far off from 0.95 D is correct because: Odds when no diabetes = 0.0279 Ln odds no diabetes = -3.58 Odds (diabetes =1) = 0.0718 , this aligns with the values from B (0.0718) and C(0.0719) for diabetes =1. Ln odds ratio = ln (0.0718/ 0.0279)=0.95

8.2. Incorrect rationales.

9. 4 year risk for intermittent claudication II

Using the coefficients in his Table 12.2, Rothman (2012) calculated a 4 year risk for intermittent claudication of 6.7% if the 70-year old man had the example profile, which included normal blood pressure, AND diabetes.

If we reversed the outcome and worked with an equation for STAYING FREE of intermittent claudication, rather than developing it, the logistic equations would

a. **have the same coefficients but all their signs would be reversed (Correct)**
b. have the same-sign intercept but the other coefficients would have their signs reversed
c. have the opposite-sign intercept but the same-sign other coefficients

9.1. Correct rationales.

- A is correct because we are predicting the opposite effect. B is incorrect because when all other predictors are 0, the risk of developing intermittent claudication cannot be the same as staying free of intermittent claudication. C is incorrect because the slopes for the predictors cannot be the same sign if we are predicting the opposite outcome.

9.2. Incorrect rationales.