DPH112 Biostatistics 2
Final Report

Elmer V Villanueva

Elmer.Villanueva@xjtlu.edu.cn

1 Background

In this final report, you will replicate and extend the analyses conducted by Sun and colleagues in their paper

Sun J, Guo X, Zhang J, Wang M, Jia C, Xu A. Incidence and fatality of serious suicide attempts in a predominantly rural population in Shandong, China: a public health surveillance study. *BMJ Open* 2015;5:e006762. doi:10.1136/bmjopen-2014-006762.

You will need to download and read the original paper. The data used in the study is found in the file called *Sun.csv*.

You must perform the regression techniques described by Sun and colleagues in their original article. In particular, you must replicate the results presented in Table 3 of their paper. In addition, you must perform three additional sets of analyses that extends their work: (1) you must test the assumptions of the original multivariable model presented by Sun and colleagues; (2) you must rerun the analysis, this time using a Poisson regression model; (3) you must test the assumptions of the Poisson regression model. You must present your R commands in an RMD file. In a written report, you must describe your methods and findings.

2 Instructions

Prepare a written report of no more than 3,000 words, at least three and no more than seven figures and at least three and no more than seven tables. The report must show that you are able to

1. describe the methods you've employed (10 marks);

2. summarise the results of the regression analyses (20 marks);

3. summarise the results of the diagnostic tests (20 marks);

4. interpret the findings in the context of the research question (10 marks); and

5. use technical English appropriately (5 marks)

The R file must show that you are able to

1. import data from Excel to R (5 points);

2. perform a logistic regression analysis (10 points);

3. perform a Poisson regression analysis (10 points);

4. perform tests of the validity of the models (10 points)

3   Submission

You must submit *two* files.

- File 1 is the research report in PDF format. The file must be named
  $< StudentID > .pdf$. For example, 123456789.pdf.

- File 2 is the RMD file containing your R commands. The file must be
  named $< StudentID > .R$. For example, 123456789.Rmd.

You must submit the files via ICE. All penalties for late or incomplete submissions apply.

4 Marking Rubric

Criterion	Level 1	Level 2	Level 3	Level 4
Report: Methods	Not done (0 marks)	Major mistakes or omissions (3 marks)	Minor mistakes or omissions (6 marks)	No mistakes or omissions (10 marks)
Report: Results - Logistic regression	Not done (0 marks)	Major mistakes or omissions (3 marks)	Minor mistakes or omissions (6 marks)	No mistakes or omissions (10 marks)
Report: Results - Poisson regression	Not done (0 marks)	Major mistakes or omissions (3 marks)	Minor mistakes or omissions (6 marks)	No mistakes or omissions (10 marks)
Report: Diagnostic tests - Logistic regression	Not done (0 marks)	Major mistakes or omissions (3 marks)	Minor mistakes or omissions (6 marks)	No mistakes or omissions (10 marks)
Report: Diagnostic tests - Poisson regression	Not done (0 marks)	Major mistakes or omissions (3 marks)	Minor mistakes or omissions (6 marks)	No mistakes or omissions (10 marks)
Report: Interpretation	Not done (0 marks)	Major mistakes or omissions (3 marks)	Minor mistakes or omissions (6 marks)	No mistakes or omissions (10 marks)
Report: Technical English	Major mistakes (0 marks)	Minor mistakes (2 marks)	Appropriate (5 marks)	
R: Data import	Not done (0 marks)	Done (5 marks)		
R: Logistic regression	Not done (0 marks)	Inappropriate (4 marks)	Appropriate (10 marks)	
R: Poisson regression	Not done (0 marks)	Inappropriate (4 marks)	Appropriate (10 marks)	
R: Model validity	Not done (0 marks)	Inappropriate (4 marks)	Appropriate (10 marks)	