DS560 Categorical Data Analysis
Fall 2022 Applied Statistics and Decision Sciences

Homework Assignment 4

Submit your work: (1) You need to present your work in the format of a PDF file. (2) Name the file as: *Yourname_Homework4.pdf*, for example, mine would be RongZheng_Homework4.pdf. (3) Submit the PDF via Western Online. **# You do NOT need to save the workspace.**

NOTE: All course materials are course properties. You can use them for self-learning purpose. Most of the lecture notes, examples, and codes are summarized from the textbook. Sharing or posting materials onto public websites without the instructor's permission will be considered as Academic Integrity Violation.

To prepare the PDF file: Please present your work in a professional format: (1) You can use *R Markdown* (highly recommended) to edit your homework, including sentences, formulas, R codes and outputs, etc., and Knit it into a PDF document. (2) You can edit your work in a *Microsoft Word* document, then save the file as a PDF document. (3) You can try *Latex* or *Overleaf* if you want to learn a new editor tool. Nowadays, many academic papers, posters, and slides are created with *Latex* or *Overleaf*. (4) You can write down your work in a clear and tidy way, and merge a scanned copy of your writing, codes, and outputs into a single PDF file.

R Markdown tutorial (recommended):
https://www.youtube.com/watch?v=DNS7i2m4sB0
Latex tutorial:
https://www.youtube.com/watch?v=fCzF5gDy60g
Overleaf template for homework:
https://www.overleaf.com/latex/templates/simple-math-homework-template/tbszsswsndrz
Installation of R/R Studio:
https://mirror.las.iastate.edu/CRAN/
https://rstudio.com/products/rstudio/download/
Import Data into R Studio:
https://www.youtube.com/watch?v=cnD1op2Oo3M

Homework and Exams are individual work. You can discuss it with your instructor if you have confusions or encounter tech issues.

Most of the questions come from the textbook *"An Introduction to Categorical Data Analysis"* (Agresti. A., 2006) with link below:
https://mregresion.files.wordpress.com/2012/08/agresti-introduction-to-categorical-data.pdf

1. Hastie and Tibshirani (1990, p. 282) described a study to determine risk factors for kyphosis, which is severe forward flexion of the spine following corrective spinal surgery. The age in months at the time of the operation for the 18 subjects for whom kyphosis was present were 12, 15, 42, 52, 59, 73, 82, 91, 96, 105, 114, 120, 121, 128, 130, 139, 139, 157 and for the 22 subjects for whom kyphosis was absent were 1, 1, 2, 8, 11, 18, 22, 31, 37, 61, 72, 81, 97, 112, 118, 127, 131, 140, 151, 159, 177, 206.

(a). Fit a logistic regression model using age as a predictor of whether kyphosis is present. Test whether age has a significant effect.
(b). Plot the data. Note the difference in dispersion of age at the two levels of kyphosis.
(c). Fit the model $\text{logit}[\pi(x)] = \alpha + \beta_1 x + \beta_2 x^2$. Test the significance of the squared age term, plot the fit, and interpret.

2. The table below shows the result of cross classifying a sample of people from the MBTI Step II National Sample (collected and compiled by CPP, Inc.) on whether they report drinking alcohol frequently (1= yes, 0 = no) and on the four binary scales of the Myers–Briggs personality test: Extroversion / Introversion (E/I), Sensing / Intuitive (S/N), Thinking / Feeling (T/F) and Judging / Perceiving (J/P). The 16 predictor combinations correspond to the 16 personality types: ESTJ, ESTP, ESFJ, ESFP, ENTJ, ENTP, ENFJ, ENFP, ISTJ, ISTP, ISFJ, ISFP, INTJ, INTP, INFJ, INFP.

(a). Fit a model using the four scales as predictors of π = the probability of drinking alcohol frequently. Report the prediction equation, specifying how you set up the indicator variables.
(b). Find $\hat{\pi}$ for someone of personality type ESTJ.
(c). Based on the model parameter estimates, explain why the personality type with the highest $\hat{\pi}$ is ENTP.

Extroversion/Introversion		E				I			
Sensing/iNtuitive		S		N		S		N	
		Alcohol Frequently							
Thinking/Feeling	Judging/Perceiving	Yes	No	Yes	No	Yes	No	Yes	No
T	J	10	67	3	20	17	123	1	12
	P	8	34	2	16	3	49	5	30
F	J	5	101	4	27	6	132	1	30
	P	7	72	15	65	4	102	6	73

Source: Reproduced with special permission of CPP Inc., Mountain View, CA 94043. Copyright 1996 by CPP, Inc. All rights reserved. Further reproduction is prohibited without the Publisher's written consent.

3. Refer to model with width and color effects for the horseshoe crab data. Using the data:
(a). Fit the model, treating color as nominal-scale but with weight instead of width as x. Interpret the parameter estimates.
(b). Using models that treat color in a quantitative manner with scores {1, 2, 3, 4}, repeat the analyses in (a).