Unit 4 – Categorical Data Analysis

Practice Problems

Due Friday March 12, 2021

Last Date to submit for credit (-20 points): March 19, 2021

#1. *Source: Fisher LD and VanBelle G. Biostatistics: A Methodology for the Health Sciences New York: John Wiley, 1993. Chapter 6 Problem #12, page 234.*

Peterson et al (1979) studied the patterns of infant deaths, especially SIDS, in King County Washington during the years 1969-1977. They compared the SIDS deaths with a 1% sample of all births during the specified time period. Tables relating the occurrence of SIDS with maternal age less than or equal to 19 years of age, and to birth order greater than one, follow. These are data for singleton births.

	Child	
Birth Order	**SIDS**	**Control**
> 1	201	689
=1	92	626

	Child	
Maternal Age	**SIDS**	**Control**
≤ 19	76	164
> 19	217	1151

	Child	
	SIDS	**Control**
Birth order > 1 and Maternal Age ≤ 19	26	17
Birth order = 1 **OR** Maternal Age > 19	267	1298

	Child	
	SIDS	**Control**
Birth order > 1 and Maternal Age ≤ 19	26	17
Birth order = 1 **AND** Maternal Age > 19	42	479

A. Compute odds ratios and 95% confidence intervals for the four tables. **Tip!** **I encourage you to go straight to the solutions here (with a hard copy, actually) and follow along.**

B. Which table of the last two do you think reflects best the risk of both risk factors at once? Comment. There is no single right answer here.

#2. *Source: Fisher LD and VanBelle G. Biostatistics: A Methodology for the Health Sciences New York: John Wiley, 1993. Chapter 6 Problem #14, page 235.*

Hello Stata Users - I have already created the Stata data set for you to use here. Download from the course website:

Stata Users:
hw_categorical.dta

R Users – The solutions illustrate how to enter these data "by hand"

Stratum = NEVER SMOKED

Cups Coffee per day (0/1)	**MI** (micase=1)	**Control** (micase=0)
1 = ≥ 5	7	31
0 = < 5	55	2691

Stratum = FORMER SMOKER

Cups Coffee per day	**MI** (micase=1)	**Control** (micase=0)
≥ 5	7	18
< 5	20	112

Stratum = 1-14 CIGARETTES/DAY

Cups Coffee per day	**MI** (micase=1)	**Control** (micase=0)
≥ 5	7	24
< 5	33	11

Stratum = 15-24 CIGARETTES/DAY

Cups Coffee per day	**MI** (micase=1)	**Control** (micase=0)
≥ 5	40	45
< 5	88	172

Stratum = 25-34 CIGARETTES/DAY

Cups Coffee per day	**MI** (micase=1)	**Control** (micase=0)
≥ 5	34	24
< 5	50	55

Stratum = 35-44 CIGARETTES/DAY

Cups Coffee per day	**MI** (micase=1)	**Control** (micase=0)
≥ 5	27	24
< 5	55	58

Stratum = 45+ CIGARETTES/DAY (micase=0)

Cups Coffee per day	**MI** (micase=1)	**Control**
≥ 5	30	17
< 5	34	17

A. Compute the Mantel Haenszel estimate of the odds ratio. **Tip!** **Again … I encourage you to go straight to the solutions here.**

B. Compute the appropriate chi square test for association.

C. In 1-2 sentences, interpret your findings to a client who is not an expert in biostatistics.

#3. *Source: Fisher LD and VanBelle G. Biostatistics: A Methodology for the Health Sciences New York: John Wiley, 1993. Chapter 6 Problem #15, page 236.*

The paper of Remein and Wilkerson (1961) considers screening tests for diabetes. The Somogyi-Nelson (venous) blood test (data at one hour after a test meal and using 130 mg/100 ml as the blood sugar cutoff gives the following table:

Test	**Diabetic**	**Non-diabetic**	**Total**
Positive	59	48	107
Negative	11	462	473
Total	70	510	580

A. Compute the sensitivity, specificity, predictive value of a positive test and predictive value of a negative test.

B. ***OPTIONAL***

Dear class – This is a "*for fun*" suggestion and, by all means, just follow along the solutions to get a feel for things! I encourage you to give it a try but please don't stress!

Using the sensitivity and specificity values you got in part "A", consider the following various values of prevalence: .01, .05, .10, .20, .30, .40, .50, .60, .70, .80, .90, .95. Using these prevalence values, construct a plot of the predictive value of a positive test on the vertical axis versus prevalence on the horizontal axis. What you will have constructed is a plot of the probability of diabetes given a positive test result as a function of prevalence.