

IGP 304 (Spring 2007) Homework 4

1. A study was performed looking at the risks of fractures in three rural Iowa communities according to whether their drinking water was “higher calcium” or “control” as determined by water samples. The following table presents the data comparing the rate of fractures over 5 years between the higher calcium versus the control communities for women ages 20-35 and 55-80, respectively.

Age 20-35	Total number of women	Number of women with fractures	Age 55-80	Total number of women	Number of women with fractures
Control	37	3	Control	121	11
Higher calcium	33	1	Higher calcium	148	21

- What test can be used to compare the fracture rates in these two communities while controlling for age?
 - Implement the test and report a p-value (two-sided).
 - Estimate the odds ratio relating higher calcium and fractures while controlling for age.
 - Provide a 95% confidence interval for the estimate obtained.
 - Is there an alternative test?
2. A longitudinal study was conducted among children in the Greater Boston Otitis Media Study. Based on all doctor visits during the first year of life, children were classified as having ≥ 1 episodes versus 0 episodes of otitis media (OTM). A separate classification was performed for the right and left ears. Several risk factors were studied as possible predictors of OTM. One such risk factor was a sibling history of ear infection, with relevant data displayed in the following table.

Sib ear infection = yes			Sib ear infection = no		
Right ear	Left ear	n	Right ear	Left ear	n
-	-	76	-	-	115
+	-	21	+	-	20
-	+	20	-	+	18
+	+	77	+	+	91
Total		194	Total		244

- Assess whether a sib history of ear infection is associated with OTM incidence in the first year of life.
- Provide a 95% confidence interval for the true difference in incidence rates for those children with sibs between those children with and without a sib history of ear infection.

3. Improving control of blood glucose levels in an important motivation for the use of insulin pumps by diabetic patients. However, certain side effects have been reported with pump therapy. The following table provides data on the occurrence of diabetic ketoacidosis (DKA) in patients before and after start of pump therapy.

After pump therapy	Before pump therapy	
	No DKA	DKA
No DKA	128	7
DKA	19	7

- What is the appropriate procedure to test if the rate of DKA is different before and after start of pump therapy?
- Perform the significance test and report a p-value.

NOTE: Below is not part of the homework. I leave it here just for those who are interested.

In a study on esophageal cancer, researchers collected data on 975 subjects. The variables collected were age, alcohol and tobacco usage, and esophageal cancer status. The data were tallied into a four-way table with 6 age groups, 4 alcohol usage levels, 4 tobacco usage levels, and 2 cancer status. In the data “esophageal.csv”, the "patients" column has the counts for all four-way combinations. The "heavy" variable is an indicator variable for heavy alcohol consumption. The values and their meanings are:

- age: 1 (25-34), 2 (35-44), 3 (45-54), 4 (55-64), 5 (65-74), 6 (≥ 75)
- alcohol: 1 (0-39), 2 (40-79), 3 (80-119), 4 (≥ 120)
- tobacco: 1 (0-9), 2 (10-19), 3 (20-29), 4 (≥ 30)
- cancer: 0 (No), 1 (Yes)
- heavy: 0 (< 80 gm), 1 (≥ 80 gm)

Cancer is the outcome variable. Age group, alcohol usage, and tobacco usage are input variables.

- Understanding the relationships among the input variables is an important part of statistical analysis. It allows you to gain insight into how the variables correlate with each other and if the results on some variables could be influenced by inclusion of some other variables in the analysis. Explore the relationship among age group, alcohol usage, and tobacco usage.
- For each input variable, create a two-way table between the input variable and the outcome variable, and carry out logistic regression analysis using the input variable as the only regression variable. Summarize results.
- Carry out logistic regression analysis using all three input variables. Are the results similar or different from those in the last question? If they are different, why?
- Explore the interaction effects among the input variables on the risk of esophageal cancer.