

Department of Statistics

COURSE STATS 330/762

Assignment 5, 2012

Instructions: Hand in your completed assignment to the Student Resource Centre by **4pm Monday October 15th**.

Leukoplakia are patches on the tongue, in the mouth, or on the inside of the cheek that occur in response to long-term irritation. Leukoplakia mainly affects the mucus membranes of the mouth. Irritation in the mouth may be caused by:

- Rough teeth
- Rough places on dentures, fillings, and crowns
- Smoking or other tobacco use (smoker's keratosis), especially pipes
- Holding chewing tobacco or snuff in your mouth for a long period of time

The disorder is most common in elderly persons. The table below shows data from a study which sought to identify risk factors for this condition.

There is no data set on the web for this assignment; you must create it for yourselves from the table given below.

Daily alcohol intake	Leukoplakia?			
	Yes		No	
	Smoker?		Smoker?	
	Yes	No	Yes	No
None	26	8	10	8
0-40 gm	38	43	8	24
41-80 gm	4	14	1	17
>80 gm	1	3	0	7

1. Type the data into R, and make a suitable data frame with factor levels "Yes", "No" for smoking and leukoplakia, and levels "None", "0-40 gm", "41-80 gm", ">80 gm" for alcohol intake. Print out the data frame. Also print out a cross-tab like the table above.
[10 marks]
2. Fit a suitable Poisson regression model to the contingency table counts. State what your model means in terms of conditional independence, homogeneous association etc.
[10 marks]

3. Fit a logistic regression model to the data using Leukoplakia as the response. Do both smoking and Alcohol have a relationship with the response? (Note that this is modeling the conditional distribution of Leukoplakia, given smoking and alcohol, whereas the model fitted in part 2 is modelling the joint distribution of all three variables. [10 marks])
4. Do you notice any similarities between the two analyses? [5 marks]
5. Can you explain these (Hard!!!) [5 marks]

R hint: Use the functions `expand.grid` and `xtabs` for Q1.