

Department of Statistics

COURSE STATS 330

Assignment 5, 2007

Instructions: Hand in your completed assignment to the Student Resource Centre by 4pm on Thursday 11 October.

Q1. The data in Table 1 relate to motor vehicle accidents in Florida in 1988. All persons injured in motor accidents in Florida in 1988 are classified according to three factors, namely

Injury: either Nonfatal or Fatal;

Seatbelt: either Yes if a seatbelt was worn, or No if a seatbelt was not worn;

Ejected: either Yes if the person was ejected from the vehicle, or No if not.

Table 1. Florida motor vehicle accident data.

		Injury	
Seatbelt	Ejected	Nonfatal	Fatal
Yes	Yes	1,105	14
	No	411,111	483
No	Yes	4,624	497
	No	157,342	1,008

1. Type the data into R and construct a data frame suitable for fitting log-linear models. Print out the data frame. [5 marks]
2. Fit a suitable log-linear model to the data. Are there any cells that are not fitted well by the model? [10 marks]
3. Using the model you fitted in 2, compute the conditional odds ratio between seatbelt use and injury, conditional on Ejected. Do these odds ratios depend on being ejected from the car? Compute a 95% confidence interval for the conditional odds ratio. What effect does seatbelt use have on the type of injury? [10 marks]

PTO

Q2. The data in Table 2 refer to suicides in England and Wales over the period 1982-1996. The persons committing suicide in that period are classified according to gender and the month in which the suicide took place.

Table 2. Suicides in England and Wales, 1982-1996.

Gender	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Male	3755	3251	3777	3706	3717	3660	3669	3626	3481	3590	3650	3392
Female	1362	1244	1496	1452	1448	1376	1370	1301	1337	1351	1416	1226

Type the data into R and make a barplot showing the relative frequencies of suicides across the months, separately for males and females. Comment on the plot.

Do you think the distribution across the months is the same for each month (i.e. are the observed differences due to chance, or is there a systematic difference between the months? If there is a difference, is the pattern the same for males and females? Fit a suitable log-linear model or models to answer this part. [15 marks]