

STAT 506, Spring 2017: Extra Credit Homework 6

- **2×2 table.** The following 2×2 table cross-classifies mortality results for drivers and passengers in auto accidents in Florida in 2008; each person's injury was classified as fatal or nonfatal, and it was recorded whether the person was wearing a seat-belt or not.

Seat-belt use	Injury	
	Fatal	Nonfatal
No	1085	55623
Yes	703	441239

- (a) Use the Fisher's exact test function `fisher.test` to obtain an odds ratio and 95% CI in R. Interpret the odds ratio and its significance as shown in the notes. In particular, what do you conclude about wearing a seat-belt?
 - (b) Obtain an estimate and 95% CI for the difference in probability of fatality across the two groups (those wearing seat-belts and those not). Interpret.
- **Testing independence in a 2×3 table.** Consider data on race and political party identification from the 2006 General Social Survey.

Race	Party identification		
	Democrat	Independent	Republican
Black	192	75	8
White	459	586	471

Test that political party identity is independent of race in this population at the 5% level using `chisq.test`. If you reject no association between race and party affiliation, use the following code to explain how party membership changes from blacks to whites.

```
m=matrix(c(192,459,75,586,8,471),nrow=2)
colnames(m)=c("Democrat","Independent","Republican")
rownames(m)=c("Black","White")
barplot(prop.table(t(m),2),ylab="proportion",legend.text=T)
```

- **A $2 \times 2 \times 8$ table.** Consider data from a clinical trial with eight centers. The study compared two cream preparations, an active drug and a control, on the success of curing an infection. This is a typical example of a pharmaceutical designed experiment, comparing two treatments on a binary response with observations from several strata. The following R code reads in the data and then collapses the data over center.

```
count=c(11,10,25,27,16,22,4,10,14,7,5,12,2,1,14,16,
6,0,11,12,1,0,10,10,1,1,4,8,4,6,2,1)
trt=factor(rep(c(1,2),16))
out=factor(rep(c(1,1,2,2),8))
ctr=factor(rep(1:8,each=4))
m=array(count,dim=c(2,2,8),
dimnames=list(c("drug","cont"),c("success","failure"),
c("A","B","C","D","E","F","G","H")))
m # look at the 8 2x2 tables, one for each center
mc=m[, ,1]+m[, ,2]+m[, ,3]+m[, ,4]+m[, ,5]+m[, ,6]+m[, ,7]+m[, ,8]
# mc is collapsed table, ignoring the center effect
fisher.test(mc)
mantelhaen.test(m)
```

- Interpret the results from the Cochran-Mantel-Haenszel approach. In particular, what is the odds ratio being reported? Is it significant?
 - What happens to the estimate of the odds ratio when we ignore the strata “center?” Is it significant?
- **Logistic regression.** Data on the results of whether a patient had a sore throat after surgery under general anesthesia (0 or 1, no or yes) were recorded for 35 patients. Also recorded was the number of minutes the operation took and the type of device used to secure the airway (0 = laryngeal mask airway, 1 = tracheal tube). Use logistic regression to determine how the odds of having a sore throat changes when (a) the number of minutes is increased by one, (b) how the odds of having a sore throat changes when the number of minutes is increased by 10, and (c) how the odds of having a sore throat changes when a tracheal tube is used instead of the laryngeal mask airway.

```
sore=c(0,0,1,1,1,1,1,1,1,1,1,1,0,1,0,1,0,1,1,1,1,0,1,0,0,1,0,1,0,1,1,0,1,0,0)
minutes=c(45,15,40,83,90,25,35,65,95,35,75,45,50,75,30,25,20,60,70,
30,60,61,65,15,20,45,15,25,15,30,40,15,135,20,40)
type=c(0,0,0,1,1,1,0,0,0,0,0,1,1,1,0,0,1,1,1,0,0,0,0,1,1,0,1,0,1,0,0,1,1,1,1)
sore=factor(sore)
levels(sore)=c("no","yes")
type=factor(type)
levels(type)=c("laryngeal.mask","tracheal.tube")
f=glm(sore~minutes+type,family="binomial")
exp(f$coef) # odds ratios
exp(confint(f)) # CI for odds ratios
```

- **Survival data.** Use the Cox proportional hazards model to quantify how increasing the temperature one degree Celsius affects the hazard (instantaneous risk) of failure in the next hour for the resin-bond data.

```
library(survival)
resin=read.table("http://users.stat.umn.edu/~gary/book/fcdae.data/exmpl3.2",
  header=T)
colnames(resin)=c("temp","logTime") # column names now match
resin[,1]=c(rep(175,8),rep(194,8),rep(213,8),rep(231,7),rep(250,6))
resin$time=10^resin$logTime
attach(resin) # now can use time and temp as variables in coxph
boxplot(time~temp,xlab="temperature",ylab="hours")
```