STA 312f12 Assignment Eleven[1]

Please bring your R printouts to the quiz. Also, **bring a calculator to the quiz**. The final formula sheet will be provided.

1. The Florida Death Penalty data are available from the data sets link on the course home page. The rows are convicted murderers in Florida jails. There are three categorical variables, one in each column:

   - Prisoner's race: 1 = White 2 = Black
   - Victim's race: 1 = White 2 = Black
   - Death penalty: 1 = Yes 2 = No

   (a) Test independence with a likelihood ratio test in all three marginal two-way tables. Be ready to state conclusions in plain language, guided by the $\alpha = 0.05$ significance level without mentioning it.

   (b) Controlling for Victim's race, is the Prisoner's race related to whether he got the death penalty? Carry out a likelihood ratio test to answer this question. If the answer is Yes, be able to describe the relationship.

   (c) Controlling for Prisoner's's race, is the Victim's race related to whether the prisoner got the death penalty? If the answer is Yes, be able to describe the relationship.

   (d) Is $(PR, VR)(DP)$ an improvement over the model of complete independence?

   (e) Does $(PR, VR)(DP)$ fit adequately by a Pearson chi-squared test?

   (f) Is there a non-saturated model that fits the data adequately by a Pearson chi-squared test? If so, find the simplest one you can, and be able to describe the relationships in the model using plain language. I found one that just barely fit using a Pearson test, and just barely didn't fit using a likelihood ratio test.

2. Planning another study of race and the death penalty, assume there are approximately equal numbers of Black and White prisoners, which is the case with the Florida data. In a 2×2 table of race by death penalty, suppose the probability of a White prisoner getting the death penalty is 10%, while the probability of a Black prisoner getting the death penalty is 15%.

   (a) What is the power of the Pearson chi-square test if the total sample size is $n = 300$?

   (b) What sample size would be required for a power of .50? Power of 0.80?

[1] Copyright information is at the end of the last page.

3. Actuaries at an insurance company are interested in the associations among gender (G), being on the honour roll (H), and being in a reported auto accident (A) among high school students who drive. In particular, they are interested in possible conditional independence of Accident and Horour roll given Gender. The idea is that if honour students get in fewer accidents, maybe it's just because more of them are girls.

   So they will test conditional independence in the usual way by testing the fit of (AG)(GH), but what about power? What sample size do they need?

   Suppose there is a two-way (negative) association between Honour roll and Accidents, so the correct model is (AG)(GH)(AH). Here is a set of reasonable π_{ijk} values:

```
Gender = Male

         Honour
Accident       Yes       No
     Yes 0.005777638 0.144224
     No  0.024222362 0.325776

Gender = Female

         Honour
Accident       Yes         No
     Yes 0.004222372  0.04577688
     No  0.065777628  0.38422312
```

   Your job is to find the smallest sample size so the test of conditional independence has a power of at least 0.90 at $\alpha = 0.05$. Don't worry; actuaries have access to large samples.
