Stat 5421 (Geyer) Fall 2022
Homework Assignment 4
Due Monday, November 21, 2022

Our problem 4.1. Agresti problem 8.28 except just analyze as a contingency table. Do not treat any variable as the "response." These data are in R package 'CatDataAnalysis'.

```
library(CatDataAnalysis)
data(exercise_8.28)
sapply(exercise_8.28, class)

## Satisfaction      Contact    Influence      Housing       counts
##     "factor"     "factor"     "factor"     "factor"    "numeric"
```

Our problem 4.2. Agresti problem 8.28 as stated by Agresti (so this is a redo of the preceding problem treating `Satisfaction` as the "response").

Our problem 4.3. Agresti problem 8.8 except rather than use the analysis in Agresti do your own analysis using the R function `polr` in the R package `MASS`. The data can be read into R as follows as follows

```
foo <-
    read.table(url("http://www.stat.umn.edu/geyer/5421/data/table-8.18.txt"),
        header = TRUE)
sapply(foo, class)

##      status   seat.belt    location      gender      counts
## "character" "character" "character" "character"   "integer"

lapply(foo, unique)

## $status
## [1] "not injured"                   "injured, no ER"
## [3] "injured, ER, not hospitalized" "hospitalized did not die"
## [5] "died"
##
## $seat.belt
## [1] "No"  "Yes"
##
## $location
```

```
## [1] "Urban" "Rural"
##
## $gender
## [1] "Female" "Male"
##
## $counts
##  [1]  7287   175   720    91    10 11587   126   577    48     8  3246    73
## [13]   710   159    31  6134    94   564    82    17 10381   136   566    96
## [25]    14 10969    83   259    37     1  6123   141   188    45  6693    74
## [37]   353    12
```

Note that `status` is not an ordered factor. You have to handle that yourself.

Also note that `counts` cannot be the "response" variable on the left-hand side of the formula, because the "response" variable is the ordered factor, in this case `status` (or whatever variable name you call `status` after you turn it into an ordered factor). So where to you tell R function `polr` what the counts are? One of the examples on the help for the `polr` function shows how to use the `weights` argument for that.

Find the best model you can. R generic function `anova` function produces P-values for tests of model comparison. It has a method for objects of class `"polr"` produced by R function `polr`.

I'm not sure exactly what Agresti meant for (b). Just make a confidence interval for the regression coefficient for gender.

In part (c) the *cumulative odds ratio* Agresti asks about is discussed on page 302 in Agresti.

For part (c) interpret all of the regression coefficients. What does it mean that they have the signs they do (if anything). In interpreting an interaction term (like the location by seat belt interaction Agresti asks about), figure out what the size of the effects are for all four classes (for this interaction, rural without seat belt, rural with seat belt, urban without seat belt, urban with seat belt).

It may help to read the details section of the on-line help for the `polr` function. It may also help to look at the probabilities produced by the method of R generic function `predict` applied to R objects of class `"polr"` produced by R function `polr`. It is weird that this function has no on-line help but has arguments unlike the method of R generic function `predict` for R objects of class `"glm"` produced by R function `glm`. Say

```
predict(out, type = "probs")
```

to get predicted class probabilities. The way I figured this out was to look at the source, which is found with

```
getS3method("predict", "polr")
```

(shouldn't have to do that). Maybe Venables and Ripley expect you to look in their book.

Our problem 4.4. Agresti problem 4.13. The data can be read into R as follows

```
library(CatDataAnalysis)
data(exercise_4.13)
sapply(exercise_4.13, class)

##      game      made  attempts
## "integer" "integer" "integer"
```