**FISH 458/558 - Fish Population Dynamics**

**LAB ASSIGNMENT #13 (CPUE model-based standardization)**

*Complete and return your assignment (via Canvas) in the form of a Word document (with any answers and figures requested and with the R script copied in), by the due date.*

*Guidelines:*

* *Include your Name, course, lab number, and date at the top of the document*
* *Number and label the questions and answers clearly! (We should easily be able to find your answers!)*
* *Include all of the requested output (e.g., values, data tables, and plots), not just the code for them. (We will not copy your code into R to see if it works).*
* *Include informative captions for figures and tables. See research articles for examples. [We will take points off this time if these are not included!].*
* *Submit a Word document unless directed otherwise (no r files or pdfs please).*
* *Include all your code used for the problems.*
* *Answer ALL questions using complete sentences that are clear and informative.*

**FOR 458 UNDERGRADUATES ONLY (NOT FOR GRADUATE STUDENTS) (26 pts) [extra credit]**

1. In Lab we used statistical models to calculate standardized annual indices for Atlantic Croaker based on catch-per-unit-effort (CPUE) data from a trawl survey. Now, you will be using similar methodology to develop a standardized annual index for pinfish from Tampa Bay. The data can be found in “**pinfish1992-2006.csv**”. Standardized seine pulls were conducted from 1992-2006, and data were obtained on the number of pinfish caught (*number*), the *year*, vegetation (*veg*: none vs. seagrass), bottom type (*bot*: Sand vs. Mud), *depth*, temperature (*temp*), and salinity (*sal*). Make sure that you treat *year*, *veg*, and *bot* as factors. I recommend naming the data “AC1” to match the lab script (for ease of coding). Also, you can think of the *number* column as CPUE (where the unit of effort is one standard seine pull), and you can create a column named *CPUE* for easier coding (to match the lab script). All of this data setup can be accomplished with this code:

**AC1=read.csv("pinfish1992-2006.csv")**

**AC1$CPUE = AC1$number**

**AC1$year = as.factor(AC1$year)**

**AC1$veg = as.factor(AC1$veg)**

**AC1$bot = as.factor(AC1$bot)**

Use the dataset to do the following:

* 1. Generate some exploratory plots to examine the data. For example, create a boxplot of log(CPUE+1) for each categorical variable, and a scatter plot of log(CPUE+1) for each continuous variable. What patterns do you see in the data that may be important to account for when developing an index using a statistical model? (4 pts)
  2. Modify the lab script to calculate the nominal annual mean CPUE and generate standardized indices for annual pinfish trends using these statistical models:
     1. 0) nominal means (i.e., no standardization; use this for comparison)
     2. 1) generalized linear model (GLM) with a normal distribution,
     3. 2) a lognormal GLM (with bias-corrected estimates),
     4. 3) a gamma GLM
     5. 4) a poisson GLM
     6. 5) a negative binomial GLM

Include all available explanatory variables (year, veg, bot, depth, sal, temp) in each of models 1-5 (variables are not needed for the “nominal means”). For each of the 6 methods above, generate standardized annual indices with a 95% confidence interval (CI) for each year.

Some notes/tips for modifying the lab script to accomplish this task:

* To make predictions for your models, you will need to create a different “p.data” object than what we did in lab. This will have to be a data frame that includes all of the covariates in your models (year, veg, bot, depth, sal, temp). For categorical variables, you will need to pick one of the factor levels (e.g., veg=”Seagrass”), except for year which should have all factor levels represented (e.g., year=levels(AC1$year) ). For continuous quantitative variables you will need to pick a specific value like the mean (e.g., temp = mean(AC1$temp) ).
* Be sure to bias correct your lognormal GLM estimates; do NOT use the non-bias corrected code.

For the two count models (poisson and negative binomial (NB) GLM):

* To use the rootogram() function, you will need to install the “countreg” and “topmodels” packages, and that can be done using the following code:

install.packages("countreg", repos="http://R-Forge.R-project.org")

install.packages("topmodels", repos = "https://R-Forge.R-project.org")

* You should REMOVE the “offset(..)” term in your Poisson and NB model, because we are treating all of the seine hauls as identical (i.e., we are not accounting for different amounts of area swept).
* When generating predictions using the predict() function, your “newdata=” can refer to “p.data” because we aren’t using the offset of log.area. Your predict code would look like this:
  + out<-predict(mod4, newdata=p.data, type="response", se.fit=T)

**Include all of your code at the end of your homework to receive points for this question** (3 pts).

* 1. Create a multipanel plot with all 6 indices. Include a title on each panel to easily distinguish the models, and include the 95% CI. (4 pts)
  2. Create a single plot with all indices overlaid, but standardize each index to its overall mean such that each index is centered on 1. How do the different indices compare? (4 pts)
  3. Determine which model you think is most appropriate for generating the annual index. Justify your answer based on the type of data you are modeling, model diagnostics (e.g., residual plots, rootogram), and AIC. Rootograms are only used for the count models (poisson and negative binomial). Recall that AIC can only be used to compare models that use the same response variable, so in this case, AIC can only be used to compare the normal, poisson, and negative binomial GLMs. You should also address how much Percent Deviance is explained by each model. (3 pts)
  4. Using your best model, generate a multipanel plot to visualize the effect that each covariate or explanatory factor has on the pinfish catches (use the **termplot()** or the **visreg()** function). Describe what trends and patterns you see for the pinfish data (make sure to address all modeled variables). For example, what are the patterns by year? Are catches higher for a specific bottom type? Etc. (5 pts)
  5. In your own words, describe what is meant when we say we are calculating a “standardized CPUE”, and use this exercise for reference. For example, how did the annual index from your best model compare to the nominal index (i.e., annual means with no standardization)? What does this comparison tell you about the effects of using standardized CPUE vs. nominal CPUE? (3 pts)

1. Answer these questions (+1 pt):
   1. How many hours did you spend on this assignment as a whole?
   2. I continue to encourage you to work with classmates collaboratively. Did you work with anyone else or at least consult with someone? Who? How did you communicate?
   3. Were there any particular things you struggled with in this lab and how did you overcome them?

**FOR 558 GRADUATE STUDENTS ONLY (5 pts)**

Work on your class projects this week. Include a short synopsis of what you worked on and where you are with the analysis. Address any obstacles or challenges you may be facing (analytical or conceptual). (5 pts)