**The question:**

Which types of storms are the most fatal?

**The model (need to pick one):**

1. Variables that result in the maximum number of fatalities
2. Variables that result in the maximum likelihood (MLE) of n+ fatalities
3. Variables that result in the MLE of the fatality of (a) particular demographic(s)

**Response var:**

* # Fatalities

**Explanatory vars:**

1. State
2. Month
3. Year?
4. Event Type
5. Population?
6. Fatalities / Population?
7. Fatality Location?

**Pseudo-R^2**

Package = pscl

<http://stackoverflow.com/questions/23067475/how-do-i-obtain-pseudo-r2-measures-in-stata-when-using-glm-regression>

Function = pR2(mod1)

Sample output

llh llhNull G2 McFadden r2ML r2CU

-8.536571e+03 -9.097875e+03 1.122607e+03 6.169612e-02 1.823561e-01 1.896116e-01

Interpreting different pseudo R^2 variables

<http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm>

* llh The log-likelihood from the fitted model
* llhNull The log-likelihood from the intercept-only restricted model
* G2 Minus two times the difference in the log-likelihoods 56 predict.hurdle
* McFadden McFadden’s pseudo r-squared
* **r2ML Maximum likelihood pseudo r-squared (Cox & Snell)**
* r2CU Cragg and Uhler’s pseudo r-squared

**ANOVA test w/chi-square**

Function= anova(AllModel, test="Chisq")

Sample Output:

Analysis of Deviance Table

Model: poisson, link: log

Response: FATALITIES\_COUNT

Terms added sequentially (first to last)

Df Deviance Resid. Df Resid. Dev Pr(>Chi)

NULL 5575 6015.1

EVENT\_TYPE 14 900.91 5561 5114.2 < 2.2e-16 \*\*\*

STATE 51 170.98 5510 4943.2 7.220e-15 \*\*\*

MONTH\_NAME 11 50.71 5499 4892.5 4.656e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Measuring the strength of our models:**

R^2: fraction of variance of response that is explained by explain vars

F: takes into accoumt # of variables (penalized for mult variables)

write.csv(MyData, file = "MyData.csv")