Problem Set 6

QPM II - Fall 2018

11/20/2018

Counts

Run the following code to import the `africa` dataset. If you have not already installed the `faraway` package, you will need to do so with `install.package()`.

```
#install.packages('faraway')
library(faraway)
data(africa)
# recode political liberalization as factor
africa$pollib <- as.factor(africa$pollib)
```

1. Fit a Poisson regression where military coups are predicted by years of oligarchic government, number of political parties, political liberalization, and population.

   $$\texttt{miltcoup} \sim \texttt{oligarchy} + \texttt{parties} + \texttt{pollib} + \texttt{popn}$$

   Print the summary and interpret the results. In your interpretation, discuss both the statistical significance and substantive implications for each covariate. You may find `help(africa)` useful for variable descriptions.

2. Discuss the model fit with respect to the deviance. Overall is the model fit good or not? Why?

3. Is the data overdispersed? Calculate the dispersion parameter in `R`.

4. Re-print the model summary, using your calculated dispersion parameter as the argument to `dispersion`. Your `R` code will look something like this.

```
DispersionParam <- _____
summary(MyModel, dispersion = DispersionParam)
```

Have the model results changed at all? Why or why not?

5. Re-run the model as a negative binomial regression (use `MASS::glm.nb()`. Interpret the model results as you did in Question 1. Compare these results with the Poisson model in Question 1. How have things changed statistically? Substantively?

Choices

Run the following code to import the `uncviet` dataset and the `nnet` package. If you have not already installed the `faraway` or `nnet` packages, you will need to do so with `install.package()`.

```
#install.packages('faraway')
#install.packages('nnet')
library(faraway)
library(nnet)
data(uncviet)
# reformat data observation-wise
uncdata <- do.call(rbind,
                   sapply(1:nrow(uncviet),
       function(i){
         n <- uncviet[i, 1]
         policy <- as.character(uncviet[i, 2])
         sex <- as.character(uncviet[i, 3])
         year <- as.character(uncviet[i, 4])
         cbind(rep(policy, n),
               rep(sex, n),
               rep(year, n))
       }))
colnames(uncdata) <- c('policy', 'sex', 'year')
uncdata <- data.frame(uncdata, stringsAsFactors = T)
```

6. Fit a multinomial regression (using `multinom`) where policy preference is predicted by sex and academic year.

   $$\texttt{policy} \sim \texttt{sex} + \texttt{year}$$

   Print the summary and interpret the results. In your interpretation, discuss both the statistical significance and substantive implications for each covariate. You may find `help(uncviet)` useful for variable descriptions.

7. Discuss the model fit with respect to the deviance. Overall is the model fit good or not? Why?

8. Re-run the model as an ordered logistic regression (using `polr`). Make sure the levels of the outcome are properly ordered. Interpret the model results as you did in Question 6. Compare these results with the multinomial model in Question 6. How have things changed statistically? Substantively?

9. How does the fit of the ordered choice model compare to the multinomial model? Use the deviance and AIC to justify your answer.

Spatial Lag

Run the following code to import the `columbus` dataset and the `spdep` package. If you have not already installed the `spdep` package, you will need to do so with `install.package()`.

```
#install.packages('spdep')
library(spdep)
data(columbus)
# convert neighborhood object to list
col.listw <- nb2listw(col.gal.nb)
```

10. Fit an OLS regression where crime is predicted by household income and housing values.

$$\texttt{CRIME} \sim \texttt{INC} + \texttt{HOVAL}$$

    Print the summary and interpret the results. In your interpretation, discuss both the statistical significance and substantive implications for each covariate. You may find `help(uncviet)` useful for variable descriptions.

11. Run a Moran test (`lm.morantest()`) on the model fit in Question 9. Determine the degree of spatial autocorrelation. Does the test suggest that spatial dependence modeling is appropriate?

12. Run a spatial lag model (`lagsarlm()`) with the specification above. Interpret the model results (including the spatial dependence components). Compare these results with the linear model in Question 9. How have things changed statistically? Substantively?

Ridge

Load the `USJudgeRatings` data, contained in base `R`.

```
data(USJudgeRatings)
```

13. Run a naive OLS regression where appraised worthiness of retention is predicted by all other covariates.

$$\texttt{RTEN} \sim \texttt{CONT} + \texttt{INTG} + \texttt{DMNR} + \texttt{DILG} + \texttt{CFMG} + \texttt{DECI} + \texttt{PREP} + \texttt{FAMI} + \texttt{ORAL} + \texttt{WRIT} + \texttt{PHYS}$$

14. Run a ridge regression on the above model using `glmnet::glmnet()`. Set `alpha = 0` and `lambda = 1`. Compare the results to the OLS regression results. Specifically, note the MSE and indicate which model is better fitting.

15. Run a cross-validated ridge regression on the above model using `glmnet::cv.glmnet()`. Set `alpha = 0`. What is the optimal value of lambda? Compare the results to the ridge regression in Question 13 and the OLS in Question 12. Specifically, note the MSE and indicate which model is best fitting.