STOR 455
STATISTICAL METHODS I

Jan Hannig

Project

- I have decided to assign a project. It will be part of the midterm exam score (10% of the final grade).
- The two in-class midterm score make 40% of the grade.
- Due December 2.
- There are three files (info, more info and data)

Steps in Data Analysis

- Plot and investigate the data!
  - %scatter.sas
  - proc univariate, proc means, proc print
- Run some exploratory models, look for outliers
  - Recode categorical variables, add interaction
  - Look for transformations
    residual plots & proc transreg

Steps in Data Analysis

- Do a model selection
  - proc reg, %allsubsreg.sas
- Some suggestions
  - If an interaction is in the model, make sure that the main effect is in the model.
  - If you recoded a categorical variable into several 0/1 variables, you might want to include either all or none of them.
  - Use your head, do the estimated values make sense?

Steps in Data Analysis

- Fit the final model
  - Run tests, compute predictions, answer scientific questions
- Make sure that your results make sense and write a report.

SENIC Data

- Info about 113 hospitals between 1975 and 1976
- Variables: id, length of stay, age, infection risk, routine culturing ratio, routine chest X-ray ratio, number of beds, medical school affiliation, region (1=NE, 2=NC, 3=S, 4=W), average daily census, number of nurses, available facilities
- File AppendixC01.txt in the extra data sets

Recode the variable

```
data hospital1;
set hospital;
region1=0; region2=0;region3=0;
if region=1 then region1=1;
if region=2 then region2=1;
if region=3 then region3=1;
output;
run;
proc print data=hospital1; run;
```

Run various models

- Now the exploration begins. Start with the full model.
- Find a good model. Do you recommend any
- Questions: All things being equal, is there a difference between region 1 and 2?
- Question: All things being equal is there an effect of the medical school?