Week 4, Meetings 1-2





Field notes

- Overall, everything looks good
- Perhaps too much descriptions of things going right -- that's not that interesting

Themes in the field notes so far

- General syntax issues + semantics of basic constructs
 - Functions
 - Conditionals
- "Idiomatic R"
 - Themes: group_by %>% summarize, group_by %>% filter, group_by %>% mutate, sapply, indexing with logicals
 - What are alternative approaches students took? (Why?)
 - Process
 - Debugging systematically
 - Function design recipe
- Attitudes toward the course
 - Enthusiastic?
 - Tried problem 4?
 - Motivation

Three approaches to regression

- Minimizing a cost function
- Maximum likelihood
- Maximum A-Posteriori

Minimizing a cost function

- Want to predict log(lifeExp) from log(gdpPercap)
- On example i, we are off by (log(lifeExp[i]) (a+ blog(gdpPercap[i]))
- Want to minimize by how much off we are overall
- One possibility: penalize being off by d with d^2
 - Minimize SUM_i ((log(lifeExp[i]) (a+ blog(gdpPercap[i])))^2
 - What are the implications of that?
 - Think what happens when the predictions are larger

Minimizing a cost function

(Switch to R code)

Maximum Likelihood

• Model: y ~ N(ax+b, sigma^2)

```
So: y[i] ~ N(ax[i]+b, sigma^2)
```

- Likelihood: the probability of the observations we see under a model Product(Prob(y | x, a, b))
- Maximum likelihood: find the model parameters that assign the highest possible probability to the data that we actually observe

Gaussian residuals ⇔ Quadratic cost function

(on the board)

Non-Gaussian residuals

Bayes' Rule



Bayesian inference for a location-scale model