Survival Analysis

What we have seen so far in STAD29

To summarize

- Regression including multicollinearity and predictions
- Logistic regression (basic)
- Logistic regression with ordered response
- Logistic regression with unordered (multinomial) response
- Dates and times
- Survival analysis
- Two-way ANOVA including simple effects

Regression

▶ having explanatory variables be correlated is bad (multicollinearity)
▶ predictions:
  ▶ set up explanatory variables to predict for (with `datagrid`)
  ▶ do predictions with `predictions` (or `cbind(predictions)` as necessary)

Logistic regression (basic)

- response variable categorical with two categories/levels (eg "lived/died")
- explanatory variables: anything
- modelling probability of response being in one category rather than the other
- actually: predicting log-odds of response category from explanatories
- compare one individual per row of dataframe vs. multiple individuals counted

Logistic regression with ordered response

- categorical response:
  - with *more than two* categories
  - that have a natural order
- explanatory variables: anything
- predict probability of each response category as they depend on explanatories
- understand effects by looking at trends in probabilities

Logistic regression with unordered response

- categorical response:
  - with *more than two* categories
  - that *do not* have a natural order (labels only)
- explanatory variables: anything
- predict probability of each response category as they depend on explanatories
- understand effects by looking at patterns of differences in probabilities

Dates and times

- difference between dates as text and dates as `Date`s
- underlying: days since Jan 1, 1970
- turning a text date into a `Date`:
  - `as.Date`
  - `mdy`, `dmy` and friends
- time between dates
- extracting month, day, etc
- constructing dates from month, day, etc
- times: seconds since midnight Jan 1, 1970
- making times (eg `ymd_hms`), time zones

Survival analysis

- response: time until event (eg death)
  - may not be observed ("censored")
- explanatories: anything, eg treatment, age, sex
- predictions: survival curve, prob. of surviving until a certain time given explanatory values
- usually a higher probability of surviving longer is better
- make a graph of these for different values of explanatories

ANOVA

- ▶ for us, two-way ANOVA
- ▶ response: quantitative
- ▶ explanatories: two categorical
  - ▶ might be chosen (as in experiment) or just observed
- ▶ interaction: effect of one explanatory on response depends on level of other
- ▶ test for interaction first
  - ▶ if not significant, remove then do Tukey for main effects
  - ▶ if significant, look at simple effects of one explanatory variable at a fixed level of other