QMEE: March 2nd 2021

Linear models

* All about using matrix to solve a sum of least squares
* Linear model in R=lm
* Gets it down to a matrix equation that you can solve
* Generalized additive models:

Basic theory

* Response variables: linear functions of predictor variable (something you derive from what you observe)
* Assume that once we know how the predictor variables predicts the response variables, all other variation is independent for each individual
* Common error: to assume that your response variable is normally distributed when it doesn’t actually assume this : don’t need normally distributed data to use a linear model
* Least squares fit: minimize the squared differences between predictions and observations
* One parameter variables:
  + Assumes one input variable
  + Ex: linear regression
  + T-test
* More than one parameter per variable (multi-parameter variables)
  + Want to know how much time 3 species of lizards spend basking in the sun
  + Can’t just say species (0,1,2)
    - Just 0 and 1: the model is going to say the difference is 1
    - If you do 0, 1,2 the model is going to try and put 1 between 0 and 2 but it may not belong there
  + Non-linear response: also need more than one variable
    - Can calculate these using linear models
    - Commons source for having more than one parameter per variable
  + First step: variable level P value
    - Can only get confidence intervals on parameters (not on the variables)
* Interactions:
  + Value of one predictor can affect the relationship between another predictor and the response variable
  + Interpreting main effects in the presence of interactions is tricky
* Experimental example
  + Does a drug treatment change the metabolism of some rabbits
  + No change in control group, change in metabolism of the treatment group
    - You should feel a pain when you hear: no change, no significant change
    - Not absolutely true that the control group didn’t change
    - The change may be small, but not 0: cannot clearly see the change in the metabolism
  + Testing interactions:
    - Use interaction
    - Use time and group and interaction between the two
    - Maybe all the rabbits slowed down a little bit
    - Bxt: the difference in the response between the two groups
    - Did the treatment group change differently than the control group
    - M=a+BxX+BtT+BxtXT
      * X=the group
      * T=the time period
      * XT= the product of the two
    - How to use it in a paper: lead with this and say that you measuring the interaction between the two groups
      * The interaction between time and treatment is…
      * The treatment rabbits increase their metabolism X more than the control rabbits
      * Also give the P values for the interactions
  + Statistical philosophy
    - Don’t accept the null hypothesis
    - Don’t throw out predictors you wanted to test because theyre not significant
  + Interactions example:
    - Bear road crossing
    - Predictor variables: sex, road type road length
    - Two-way interactions:
      * Sex – road length : are females more sensitive to amount of road than males
      * Sex- road type : do females vary behavior between road type more than males
      * Road type- road length : does amount of road affect crossings differently for different road types
  + Diagnostics
    - Made a bunch of assumptions
    - Don’t want to do this
      * Don’t want to say you cant reject the null hypothesis
      * Recommend instead:
      * People often normality first: should always to this first
        + Are things linear or not?
        + Is the variance the same for all aspects of your data sets
        + Are the residuals normally distributed
        + Independence: difficult to test
  + Transformations
    - Linear scales or log scales
    - 1:10 (additive of 9) and the other group goes from 10 to 19 or is the increase proportional and it goes from 10 to 100?
    - Percentages are tricky
      * 1% to 10%
      * But 10% to 100% also doesn’t make sense
    - Is your original value linear for proportional
    - Transforming data tends to help with the assumptions we need to check in the diagnostics
      * Sometimes the transformation that helps one assumption can screw another assumption up
      * Normality is the last assumption you should worry about
      * Should try to deal with these issues in order: homoscedacity, linearity, then normality
      * Log transform the y axis (response) but leave the x axis ( the predictor variable) alone
    - Don’t!: don’t do a transformation and say its better because the P value gets better
    - Can do: do the transformation and see if the diagnostic tests look any better
    - Box-Cox: tries a bunch of transformations and tells you which ones look the best for your data
    - Lambda: the power to which you are raising your data 1: not transformation, 0= log transforming, ½: scare rooting it
    - ggplot, geom\_smooth(method="lm"): to add a regression line on a ggplot