# Interdisciplinary Approaches to Geriatric Care and the Analysis of Relevant Data

Introduction to the analysis of longitudinal data

# Outline

- Two examples
- Attempts at statistical models
- Summaries of what statistical models should address

# Longitudinal Data

Longitudinal data occur when responses are collected on each subject repeatedly over time.

- Clinical trials, observational studies in humans
- Studies of the growth of tumors, decline in brain function

Such data have special features that are addressed in statistical models to ensure valid inferences on study questions

# A “famous” example

In 1964, Pothoff and Roy published a dental study on 27 children (16 boys and 11 girls).

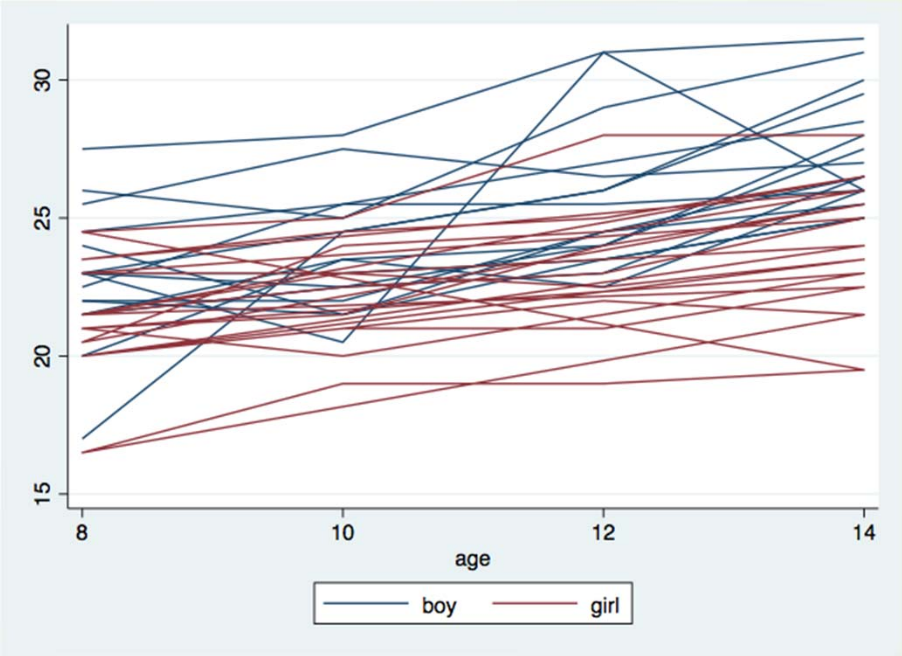
On each child, the distance in mm from the center of the pituitary to the pterygomaxillary fissure was measured at ages 8, 10, 12, and 14 years.

Does the distance change over time?

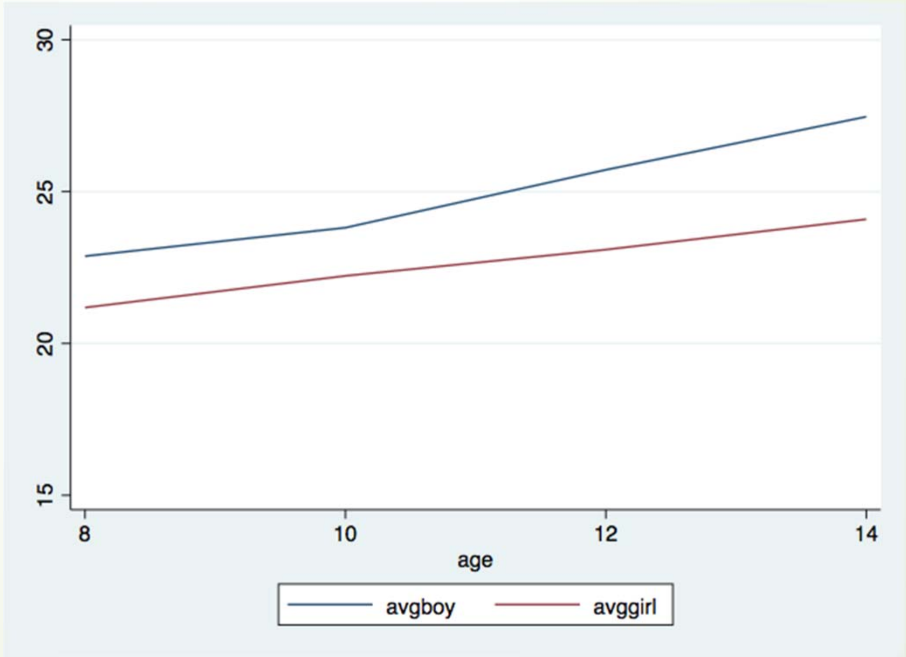
What is the pattern of the change?

How is the pattern of change different for boys and girls?

Spaghetti plot showing the growth curves for all of the boys and girls in the study



# Sex-specific means at each age



# What do we notice about these data?

- Each child is measured at the same 4 times (ages).
- The distance measure of interest seems to be greater for boys.
- Children who start with a lower distance (compared to the other children), tend to maintain their relative position over time.
- The individual pattern for most children seems to follow a line (with some variability around that).
- The sex-specific mean also seems to follow a line (with some variability around that).
- The individual and mean patterns seem to increase over time.

# A second “famous” example

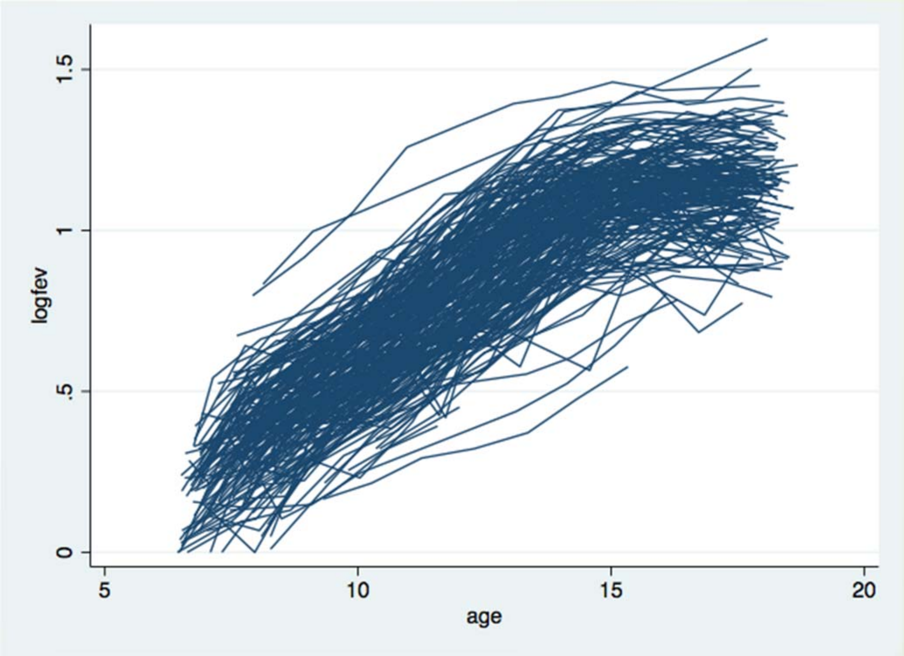
In 1983, Dockery, Berkey, Ware, Speizer, and Ferris published a study on children aged 6 to 11 years. Six cities were involved in an investigation into the effects of air pollution. A cohort of 13,379 children were enrolled.

The pulmonary function of each child was measured at doctor visits.

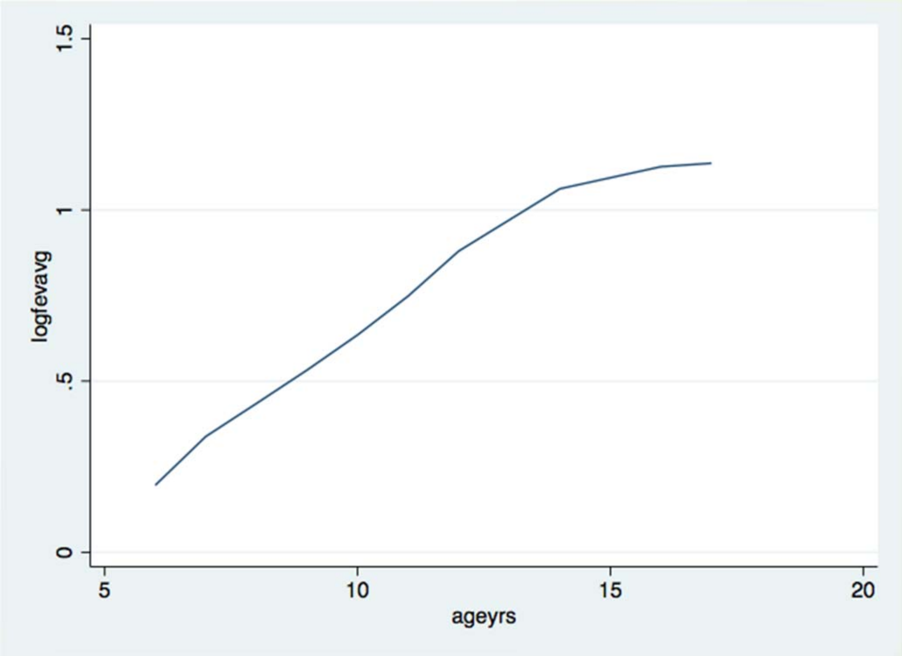
Does the pulmonary function change over time?

What is the pattern of the change?

# Spaghetti plot of log(fev) by age from 300 randomly chosen girls who lived in Topeka, Kansas



# Mean pulmonary function at ages (in years)



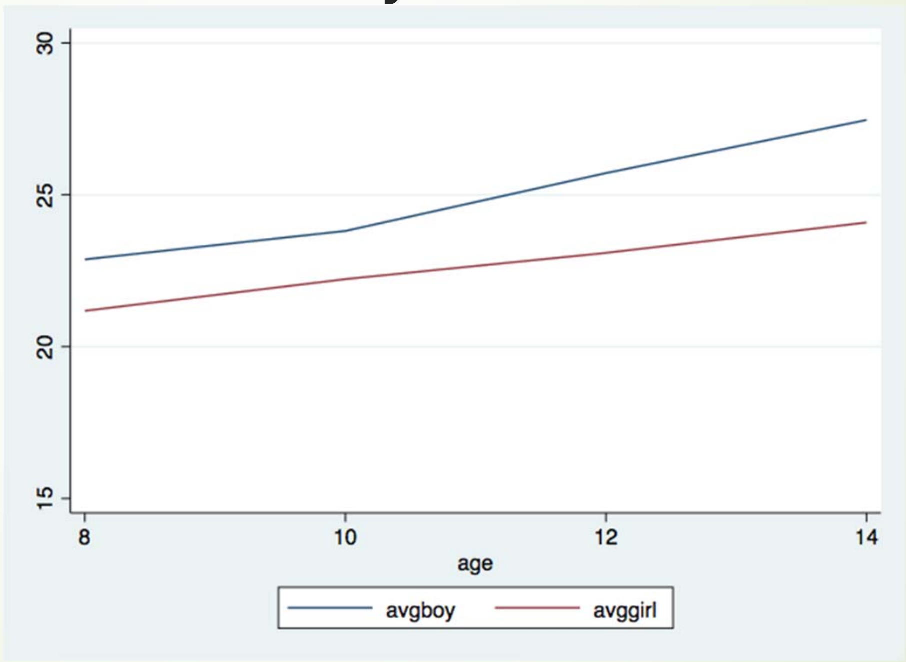
# What do we notice about these data?

- Each child is measured at different times and at a different number of times.
- The pulmonary function starts out low and grows over time.
- Once a child reaches late teen years, the growth of pulmonary function seems to level off.

- Note that in the original study of all 6 cities, the number of cigarettes smoked per day by the child’s mother, and whether the child had respiratory difficulties (infection or wheezing) were additional variables. What could we have investigated about that?

# REVISIT: Sex-specific means at each age

## Cross-sectional analysis?



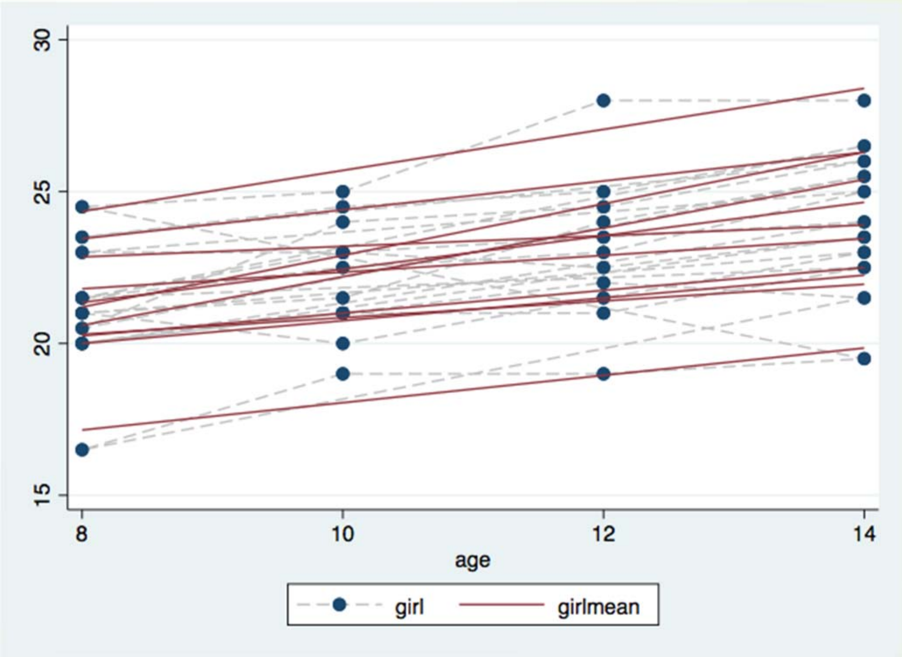
# We need a better way!

- When we look at the sex-specific trajectories, it seems reasonable to conclude that the population mean distances are linear over this range of ages.
- Formalizing our question of interest How is the pattern of change different for boys and girls? can be stated as Are the slopes of the population mean profiles different for boys and girls?
- Thus, we need a statistical model that incorporates our belief that each sex-specific trajectory is a straight line.

# First Attempt

- Let $y_{ij}$ be the distance for child $i$ at time $t_{ij} = 8{,}10{,}12{,}14$
- Let $f_i$ be an indicator of whether child $i$ is female
- The models for the two sex-specific lines are then
  - Girl: $y_{ij} = \beta_{0g} + t_{ij}\beta_{1g} + \varepsilon_{ij}$
  - Boy: $y_{ij} = \beta_{0b} + t_{ij}\beta_{1b} + \varepsilon_{ij}$
  - Combined: $y_{ij} = \beta_{0g}(f_i) + \beta_{0b}(1 - f_i) + t_{ij}\beta_{1g}(f_i) + t_{ij}\,\beta_{1b}(1 - f_i) + \varepsilon_{ij}$
- Assumption: $\varepsilon_{ij}$ is our error term and is assumed to be iid. Is that reasonable?

# Focus on individuals



# Second attempt

- Now, we assume that each child has their own underlying straight-line trajectory and we want to know Is the typical (average) slope among girls different from the typical (average) slope among boys?
- Thus, our model becomes
  - $y_{ij} = \beta_{0i} + t_{ij}\beta_{1i} + \varepsilon_{ij}$
  - where $\beta_{0i}$ is the child-specific intercept and $\beta_{1i}$ is the child-specific slope.
- To answer the research question, we could estimate a linear regression for each child's data and then do a two-sample t-test of the collection of girl slopes versus boy slopes. Is this reasonable?

# What do our models need to acknowledge?

- While observations from different subjects may be assumed to be independent, observations within subject are correlated. If we don't address that correlation, we effectively assume we have more information than is actually present and our analyses are in error.
- The two most popular types of models are population-averaged and subject-specific models.

# Subject-Specific (SS) Models

- Use these when your question of interest is in terms of typical individual behavior.

- Mixed-effects models

# Population-averaged (PA) models

- Use these when your question of interest is in terms of population-means over time.

- Generalized estimating equation (GEE) models

# Example questions about mother's smoking and the likelihood of child's respiratory illness

- The difference in the likelihood of respiratory illness for children whose mothers smoke and children whose mothers don't smoke - PA

- The difference in the likelihood of respiratory illness for a child whose mother changes from non-smoker to smoker - SS

# Questions?