Chapter 9 - Repeated Measures Data

- The data we have previously seen consist of several variables measured for each individual.
- Repeated Measures data consist of several measurements of the same variable on each individual.
- Often these repeated measurements are taken at different time points.
- If this is the case, we sometimes call such data longitudinal data.
- We generally have a response variable and one or more explanatory variables measured repeatedly in the study.
- The repeated measurements are likely to be correlated - measurements on the same individual will not be independent.
- Our model should allow for this correlation.

- We need our model to contain parameters measuring the effect of the explanatory variable(s) on the response — these are typically the parameters of main interest.
- We also need parameters measuring the correlation among within-subject measurements — these are not as interesting and are called <u>nuisance parameters</u>.
- <u>But</u> having the correct structure for these correlations will improve the inference on the parameters of interest.

Linear Mixed Effect Models

- We assume that an individual's responses depend on some unobserved random variables, which are included in the model as <u>random effects</u>.

- The correlation among responses on the same individual arises from these random effects.

Random Intercept Model

- Let Y_{ij} be the response for the j-th measurement on the i-th subject.
- Let X_j be the explanatory variable value associated with the j-th measurement.
- Then the model is:

$$Y_{ij} = \beta_0 + \beta_1 X_j + u_i + \varepsilon_{ij}$$

$$i = 1, \ldots, I \quad , \quad j = 1, \ldots, J$$

- The u_i's are <u>subject-specific random effects</u>.
- We assume the u_i's are normal with mean zero, variance σ_u^2.
- The random errors ε_{ij} are normal with mean zero, variance σ^2.
- The u_i, ε_{ij}, and X_j are all independent.

- The intercepts of each individual, $i = 1, \ldots, I$, are:

$$\beta_0 + u_1, \quad \beta_0 + u_2, \quad \ldots, \quad \beta_0 + u_I$$

- Since the u_i's are random, this is called the random intercept model.
- The parameter β_1 is a fixed effect.
- This model implies that the variance of each repeated measurement is the same, and the covariance between any pair of measurements is the same ("Compound symmetry")
- This may be unrealistic – in reality, measurements near each other in time may be more highly correlated than those far apart in time.
  - And the variances of later measurements are often greater than variances of earlier measurements.

- The <u>random slope and intercept model</u> is more flexible, having two sets of random effects, u_{i1} and u_{i2}:

$$Y_{ij} = \beta_0 + \beta_1 X_j + u_{i1} + u_{i2} X_j + \varepsilon_{ij}$$

- Note the intercepts for the individuals are: $\beta_0 + u_{11}$, $\beta_0 + u_{21}$, etc.

and the slopes are

$\beta_1 + u_{12}$, $\beta_1 + u_{22}$, etc.

- The random effects u_1 and u_2 have a bivariate normal distribution with mean vector $\underset{\sim}{0}$ and covariance matrix

$$\begin{bmatrix} \sigma^2_{u_1} & \sigma_{u_1,u_2} \\ \sigma_{u_1,u_2} & \sigma^2_{u_2} \end{bmatrix}$$

- With this model, the variance of each repeated measure depends on X_j and the correlation between a pair of measurements depends on $(X_j, X_{j'})$.

$\Rightarrow$ <u>Not</u> compound symmetry.

- We will estimate the models using <u>maximum</u> <u>likelihood</u>.

(Note that improved estimates may be found using a method called <u>restricted maximum likelihood</u> (REML).)

- A likelihood ratio test may be used to compare two different models.

- See timber data example.