Repeated Measures

SOME EXAMPLES

REPEATED MEASURES CONSIDERATIONS

ANOTHER EXAMPLE AND EXERCISE

Examples

Consider the following data sets

John Snow & Cholera: Record the location and times of cholera cases to identify probably causes of the outbreaks

The Framingham Heart Study: A randomly selected group of patients from Framingham, MA was examined and found free from cardio-vascular disease (CVD). These patients were re-examined periodically and their mortality cause was recorded

(This is known as a longitudinal cohort design)

Electronic Testing: Testing has transitioned from pencil/paper to physical calculators to computer-based calculators. The effect of these changes needed to be assessed. A random group of students take a standardized test 3 times, each with different calculation methods.

(this is known as a cross-over design)

Repeated Measures

A general repeated measure occurs when the “same” subject/location is measured multiple times

It is a particular example in which correlations exist (either deliberately or not) in the data set based on this repeated measuring

This could be over time:

-Pre and post testing for the efficacy of an educational program

Or over space:

-Measure the amount of oxygen at different depths in a core sample of ice near the south pole.

The key idea is that each subject can function as its own control

Repeated Measures

Foundations

The key feature of repeated measures is measuring the changes in the response over a relevant interval of time or space

A fundamental premise of repeated measures modeling is that observations on the same subject are likely to be correlated:

→ observations closer together in time or space are more highly correlated than observations farther apart in time or space (serial correlation)

Repeated measures tend to resemble a split design with treatment being the whole plot and time being the split plot

The primary difference is that the “within” sums of squares model requires independence for split plot, whereas the repeated measures is dependent

Two Main Ingredients

We need to select a suitable model for

- the mean and
- the covariance structure

The model for the mean is determined by the design of the experiment (ignoring the repeated measurements aspect)

The experimental unit on which the measurements will be repeated through time or space is referred to as the subject

Between subjects model: The portion of the experiment that does not include the repeated measure

Within subjects model: The portion of the experiment that involves the repeated measurements

Choosing the Covariance Structure

Choosing the Covariance Structure

We have already explored the most relevant covariance structures:

- Unstructured model: All variances/covariances are arbitrary
- Independence model: All variances are equal, and all covariances are zero.
- Compound symmetry: All variances are equal and all covariances are equal.

(This model represents the simplest nontrivial covariance structure)

- Autoregressive: All variances are equal and the covariances go to zero like $(\sigma_{AR}^2)^h$ for time/distance equal to h

(Note that this can be generalized to where unequal time/space intervals occur)

Power and Type I Error

- In general, if the selected covariance structure under-models the true correlation, the type I error rate will be inflated

  (e.g. the independence model is used when the AR model is needed)

- If the selected covariance structure over-models the true correlation, then the power suffers.

  (e.g. the unstructured model is used when an AR model would be adequate)

  It is important to identify the simplest covariance model that adequately accounts for the correlation structure in the data.

  (This maximizes power without compromising control over type I error)

In-class Example

Some Simulated Data

Suppose we measure 6 subjects at 3 time points each:

Subject	T1	T2	T3	Subject Means
1	45	50	55	***50***
2	42	42	45	***43***
3	36	41	43	***40***
4	39	35	40	***38***
5	51	55	59	***55***
6	44	49	56	***49.66666667***
Time Means	***42.833333***	***45.333333***	***49.666667***	**45.9444**

Resist the temptation to treat this as:

- A one-way ANOVA with groups given by either subject or time or
- A two-way ANOVA without replication

(Go to HW 7)