STA 674

Regression Analysis And Design Of Experiments

Advanced Concepts – Lecture 3

Pseudoreplication

STA 674, RADOE:
Pseudoreplication

- Last time, talked about analysis of split plot designs and fixed versus random effects.
- This time, we're going to discuss a pitfall to avoid in study design: pseudoreplication.

STA 674, RADOE:
Pseudoreplication

Example – aerial spraying of crop types, revisited

- An agricultural scientist wants to compare the effects of two aerially sprayed insecticides on the yield of corn using a field that is 100 m wide by 75 m deep. She can only spray the insecticides in strips that are 25 m wide, so she proceeds as follows:

1. She divides the field width-wise into 4 25-m. strips,
2. she randomly assigns an insecticide to each strip,
3. she plants corn in 3 25 m×25 m plots within each strip.

- After the corn is harvested compares the mean yield for the 2 insecticides using a two-sample t-test with 6 replicates of each insecticide.
- *Why is this not a split-plot design?* there's no second randomization of the corn...because it's all corn
- *What is the problem with the test she has used?* overstating the amount of information obtained...it's really just two observations of each insecticide (even though she 'split' each strip into threes)...pseudoreplicates

STA 674, RADOE:
Pseudoreplication

Pseudoreplication

- Pseudoreplication occurs when repeated observations are obtained from a single experimental unit and treated as if they were independent measurements.

Problem

- Pseudoreplication artificially decreases the experimental error:
  - p-values will be too small, over-expressing significance,
  - confidence intervals will be too narrow, over-stating precision.

STA 674, RADOE:
Pseudoreplication

Example – fish fat

- A fisheries researcher wants to compare the accumulation of toxins by three species of fish. He places three fish from each species in 9 separate tanks and adds the same amount of toxin to each tank. At the end of the experiment he measures the amount of toxin from:

➢ 3 samples of fat

➢ from each of 3 fish

➢ from each of the 3 species.

1. What are the experimental units? tanks
2. How many replicates are there for each species?
3. What type of variation do we need to estimate to test differences between species?

STA 674, RADOE:

Pseudoreplication

Pseudoreplication

- Extent of the Problem

  Scrutiny of 176 experimental studies published between 1960 and the present revealed that Pseudoreplication occurred in 27% of them, or 48% of all such studies that applied inferential statistics. (Hurlbert, 1983)

STA 674, RADOE:
Pseudoreplication

Analysis of Balanced Data with Pseudoreplicates

Two equivalent approaches:

1. Model both levels of variation with a mixed effects model.
2. Average over the pseudoreplicates for each experimental unit and apply a fixed effects model

Analysis of Unbalanced Data with Pseudoreplicates

Only one approach:

1. Model both levels of variation with a mixed effects model.

STA 674, RADOE:
Pseudoreplication

Analysis of Balanced Data with Pseudoreplicates

Approach #1: Model both levels of variation with a mixed effects model

```
PROC MIXED data=fish;
CLASS species fish sample;
MODEL fat=species / DDFM=satterth;
RANDOM fish(species);
LSMEANS species / CL;
RUN;
```

STA 674, RADOE:
Pseudoreplication

Analysis of Balanced Data with Pseudoreplicates

Incorrect approach:

```
TITLE 'Fish Fat -- incorrect analysis';
PROC GLM DATA=fish;
CLASS species;
MODEL fat=species;
LSMEANS species / CL;
RUN;
```