Within Cases

- A case contributes a DV value for every value of a categorical IV
- It is natural to expect data from the same case to be correlated - NOT independent
- For example, the same subject appears in several treatment conditions
- Hearing study: How does pitch affect our ability to hear faint sounds? The same subjects will hear a variety of different pitch and volume levels.

You may hear terms like

- **Longitudinal**: The same variables are measured repeatedly over time. (Usually lots of variables, including categorical ones, and large samples. If there's experimental treatment, it's usually once at the beginning, like a surgery. Basically it's *tracking* what happens over time.)
- **Repeated measures**: Usually, same subjects experience two or more experimental treatments. Usually quantitative DVs and small samples.

Advantages (If measurement of the DV does not mess things up too much)

- Convenience
- Each case serves as its own control. A huge number of extraneous variables are automatically held constant. The result can be a very sensitive analysis.
- Sometimes, ignoring a within-cases structure can inflate the Type I error rate.

Quantitative DVs: Three main approaches

- Classical Univariate
- Multivariate
- Covariance Structure

Classical Univariate

- "Case" (or Subject) is one of the factors
- Case is a *random effects* factor that is *nested* within combinations of the between-cases factors, and crosses the within-cases factors.
- Uses a mixed model ANOVA
- Only works for balanced experimental designs.
- Some other day.

Multivariate Approach

- When a case (subject) provides data under more than one set of conditions, it is natural to think of the measurements as multiple dependent variables.
- Start with the humble matched *t*-test

Why the matched *t*-test makes sense

- Let Y_1 = HS English mark
- Let Y_2 = HS Calculus mark
- Let D = Y_1 - Y_2

$$E(Y_1) = \mu_1 \text{ and } E(Y_2) = \mu_2$$

$$H_0 : \mu_1 = \mu_2$$

$$E(D) = E(Y_1 - Y_2) = E(Y_1) - E(Y_2) = \mu_1 - \mu_2$$

$$E(D) = E(Y_1 - Y_2) = E(Y_1) - E(Y_2) = \mu_1 - \mu_2$$

- Mean difference is the difference between means
- So test whether the population mean difference equals zero
- Use a one-sample t-test of $H_0 : \mu_D = 0$
- It's called a matched t-test

Multivariate extension of the *t*-test

- It's called Hotelling's T-squared
- Can be one-sample or two-sample
- Multivariate matched *t*: can test whether several mean differences are zero, simultaneously.
- Optimal for small samples if the data are multivariate normal
- Robust against normality for moderately large n.
- No equal variance assumption.

Example: Grapefruit sales

- Cases are stores
- Sales measured at three different price levels, yielding Y_1, Y_2 and Y_3
- $D_1 = Y_1 - Y_2$ and $D_2 = Y_2 - Y_3$
- Test whether mean of D_1 and mean of $D_2 = 0$, simultaneously
- Same as testing

$$H_0 : \mu_1 - \mu_2 = 0 \text{ and } \mu_2 - \mu_3 = 0$$

Mean of a contrast is a contrast of means

- To test a collection of contrasts, calculate a corresponding set of difference variables.
- Test whether the mean differences all equal zero.
- The only trick is persuading SAS to calculate Hotelling's T-squared.
- $T^2 = (n-1)\left(\frac{1}{\lambda} - 1\right)$ so just use Wilks' Lambda

One between, one within

- Grapefruit study
- Three price levels: Within stores factor
- Incentive program for produce managers (Yes-No): Between-stores factor

Monkey Study

- Train monkeys on discrimination task, at 16, 12, 8, 4 and 2 weeks prior to treatment
- Treatment is to block function of the hippocampus (with drug, not surgery) , re-tested. Get 5 scores.
- 11 randomly assigned to treatment, 7 to control
- Treatment is between, time elapsed since training is within

Form linear combinations

- Four difference variables for the five time levels
- Also, calculate average score across the 5 time levels
- Now we have 5 potential DVs

Main effect of Time

- DVs are the difference variables
- With effect coding, intercepts are grand means - Averages of the between cases cell means
- Test of the intercepts = zero is a test of whether, averaging across the between subjects factor, the marginal means of the within subjects factor are equal
- This is the main effect of time

Main effect of Treatment

- DV is average discrimination score, averaging across time
- Means of this variable for the two treatment groups are the marginal means, averaging across time
- So just do an ordinary F test for difference between means
- This is testing the main effect of Treatment.

Interaction

- DV is the set of difference variables
- IV is Treatment
- If the mean difference (trend over time) *depends* on treatment, then there is an interaction

In general, with both within and between-cases factors

- Suppose there are *k* measurement for each case
- Calculate contrast variables to represent main effects and interactions of the within-cases factors. There will be *k*-1 of them.
- Calculate one more variable: The mean or sum of the *k* measurements

Tests

- Tests of within-cases main effects and interactions are tests on the intercepts
- Tests of between-cases main effects and interactions are tests on the sum or average variable
- Interactions between between and within-cases factors: Tests of the between-cases IVs on the contrast variables
- For example, if A and B are between and C and D are within, test of AxB on the contrasts representing the CxD interaction is: the AxBxCxD interaction, because the CxD interaction depends on the AxB combination.