Within Cases

The Humble t-test

Overview

Independent Observations

- Most statistical models assume independent observations.
- Sometimes the assumption of independence is unreasonable.
- For example, times series and within cases designs.

Within Cases

- A case contributes a value of the response variable for every value of a categorical explanatory variable.
- As opposed to explanatory variables that are *Between Cases*: Explanatory variables partition the sample.
- 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? Subjects are presented with tones at a variety of different pitch and volume levels (in a random order). They press a key when they think they hear something.
- A study can have both within and between cases factors.

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 an experimental treatment, its usually once at the beginning, like a surgery. Basically its *tracking* what happens over time.
- **Repeated measures**: Usually, same subjects experience two or more experimental treatments. Usually quantitative explanatory variables and small samples.

Student's Sleep Study (*Biometrika*, 1908)

First Published Example of a t-test

- Patients take two sleeping medicines several days apart.
- Half get A first, half get B first.
- Reported hours of sleep are recorded.
- It's natural to subtract, and test whether the mean *difference* equals zero.
- That's what Gossett did.
- But some might do an independent t-test with $n_1 = n_2$.
- Is it harmful?

Conclusions from an earlier discussion

- When covariance is positive, matched t-test has better power
- Each case serves as its own control.
- A huge number of unknown influences are removed by subtraction.
- This makes the analysis more precise.

Hotelling's t^2

Multivariate Matched t-test

- $\mathbf{X}_1, \ldots, \mathbf{X}_n \stackrel{i.i.d.}{\sim} N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- $\overline{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i$ and $\mathbf{S} = \frac{1}{n-1}\sum_{i=1}^n \left(\mathbf{X}_i - \overline{\mathbf{X}}_n\right)\left(\mathbf{X}_i - \overline{\mathbf{X}}_n\right)'$
- $t^2 = n\left(\overline{\mathbf{X}}_n - \boldsymbol{\mu}\right)' \mathbf{S}^{-1}\left(\overline{\mathbf{X}}_n - \boldsymbol{\mu}\right) \sim T^2(k, n-1)$
- That is, $\frac{n-k}{k(n-1)}t^2 \sim F(k, n-k)$
- When $k=1$, reduces to the familiar $t^2 = F(1, n-1)$
- Test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$

Test *Collections* of Contrasts

$H_0 : \mathbf{L}\boldsymbol{\mu} = \mathbf{h}$, where $\mathbf{L}$ is $r \times k$

- $t^2 = n \left(\overline{\mathbf{X}}_n - \boldsymbol{\mu}\right)' \mathbf{S}^{-1} \left(\overline{\mathbf{X}}_n - \boldsymbol{\mu}\right) \sim T^2(k, n-1)$, so if H_0 is true
- $t^2 = n \left(\mathbf{L}\overline{\mathbf{X}}_n - \mathbf{h}\right)' \left(\mathbf{L}\mathbf{S}\mathbf{L}'\right)^{-1} \left(\mathbf{L}\overline{\mathbf{X}}_n - \mathbf{h}\right) \sim T^2(r, n-1)$
- Could also calculate contrast variables, like differences.
  - Expected value of the contrast is the contrast of expected values.
  - Just test (simultaneously) whether the means of the contrast variables are zero, using the first formula.
- For 2 or more within-cases factors, use contrasts to test for main effects, interactions

Compare Wald-like tests

Recall

- If $\mathbf{Y}_n = \sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \stackrel{d}{\rightarrow} \mathbf{Y} \sim N_k(\mathbf{0}, \boldsymbol{\Sigma})$, then

$$\begin{aligned} W_n &= n(\mathbf{L}\mathbf{T}_n - \mathbf{h})' \left(\mathbf{L}\widehat{\boldsymbol{\Sigma}}_n\mathbf{L}'\right)^{-1} (\mathbf{L}\mathbf{T}_n - \mathbf{h}) \stackrel{d}{\rightarrow} W \sim \chi^2(r) \\ t^2 &= n\left(\mathbf{L}\overline{\mathbf{X}}_n - \mathbf{h}\right)' \left(\mathbf{L}\mathbf{S}\mathbf{L}'\right)^{-1} \left(\mathbf{L}\overline{\mathbf{X}}_n - \mathbf{h}\right) \sim T^2(r, n-1) \end{aligned}$$

- And
  $F = \frac{n-r}{r(n-1)} t^2 \sim F(r, n-r) \Rightarrow t^2 = \frac{n-1}{n-r}\, rF \stackrel{d}{\rightarrow} Y \sim \chi^2(r)$
- So the Hotelling t-squared test is robust with respect to normality.