Within Cases

The Humble t-test

Overview

The Issue

Analysis

Simulation

Multivariate

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.

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 $n_1 = n_2$.
- It's wrong, but is it harmful?

The True Model

And the Correct Test

Independently for $i = 1, \ldots, n$, observe X_i, Y_i.

- $X_i \sim N(\mu_1, \sigma_1^2)$, $Y_i \sim N(\mu_2, \sigma_2^2)$.
- $Cov(X_i, Y_i) = \sigma_{12}$.
- Calculate Differences $D_i = X_i - Y_i$
- Matched t-test on $D_1, \ldots, D_n$
- $H_0 : \mu = 0$, where $\mu = E(D_i) = \mu_1 - \mu_2$

Test statistic is

$$t_1 = \frac{\sqrt{n}(\overline{D} - 0)}{S}$$

with $df = n - 1$.

Independent t-test

Correct if $\sigma_1^2 = \sigma_2^2$ and $\sigma_{12} = 0$

$$t_2 = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

with $df = n_1 + n_2 - 2$, where

$$S_p^2 = \frac{\sum_{i=1}^{n_1}(X_i - \overline{X})^2 + \sum_{i=1}^{n_2}(Y_i - \overline{Y})^2}{n_1 + n_2 - 2}$$

Comparing the Tests

$$t_1 = \frac{\sqrt{n}(\overline{D} - 0)}{S}, \quad df = n - 1$$

$$t_2 = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad df = 2(n - 1)$$

- The two-sample test pretends it has twice the degrees of freedom.
- Could cause worry about inflated Type I error rate
- But both critical values go to $z_{\alpha/2}$ as $n \to \infty$.
- For example, for $n = 100$, $t_{0.975}(99) = 1.98$ while $t_{0.975}(198) = 1.97$.
- So if there is a problem with df, it will be for small samples.

Comparing the Test Statistics

$$t_1 = \frac{(\overline{D} - 0)}{S/\sqrt{n}}, \quad df = n - 1$$

$$t_2 = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad df = 2(n - 1)$$

- $\overline{D} = \frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i) = \overline{X} - \overline{Y}$
- So the numerators are the same.
- Compare denominators

One-Sample (Matched) t-Test

$$
\begin{aligned}
S^2/n &= \frac{1}{n(n-1)} \sum_{i=1}^{n} (D_i - \overline{D})^2 \\
&= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(X_i - Y_i - (\overline{X} - \overline{Y})\right)^2 \\
&= \frac{1}{n(n-1)} \sum_{i=1}^{n} \left((X_i - \overline{X}) - (Y_i - \overline{Y})\right)^2 \\
&= \frac{1}{n} \left[\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1} - 2\frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{n-1} \right. \\
&\quad \left. + \frac{\sum_{i=1}^{n} (Y_i - \overline{Y})^2}{n-1} \right] \\
&= \frac{1}{n} \left[S_x^2 - 2S_{xy} + S_y^2 \right]
\end{aligned}
$$

Two-Sample (Independent) t-Test

With $n_1 = n_2 = n$

$$
\begin{aligned}
S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right) &= \frac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right) \\
&= \frac{(n-1)(S_x^2+S_y^2)}{n+n-2}\left(\frac{2}{n}\right) \\
&= \frac{(n-1)(S_x^2+S_y^2)}{2(n-1)}\left(\frac{2}{n}\right) \\
&= \frac{S_x^2+S_y^2}{n}
\end{aligned}
$$

Comparing (Squared) Denominators

Of the t Statistics

$$S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) = \frac{1}{n} \left[S_x^2 + S_y^2 \right]$$

$$S^2/n = \frac{1}{n} \left[S_x^2 - 2S_{xy} + S_y^2 \right]$$

- If covariance is zero, they are the same
- If covariance is negative
  - Denominator of two-sample t is too small
  - Value of t too large
  - Null hypothesis rejected too often
  - Not likely to be a problem in practice
- If covariance is positive (realistic)
  - Denominator of two-sample t is too large
  - Value of t too small
  - Null hypothesis *less* likely to be rejected
  - If H_0 is false, expect loss of power

Estimate Power by Simulation

- (X_i, Y_i) bivariate normal
- Equal Variances: $\sigma_1^2 = \sigma_2^2 = \sigma^2 = 1$
- $|\mu_1 - \mu_2| = \frac{\sigma}{2}$, so let $\mu_1 = 1, \mu_2 = 1.5$
- $Corr(X_i, Y_i) = +0.50$
- $n = 25$
- What is the power of the correct test and the incorrect test?

Simulate From a Multivariate Normal

```
rmvn <- function(nn,mu,sigma)
# Returns an nn by kk matrix, rows are independent
# MVN(mu,sigma)
    {
    kk <- length(mu)
    dsig <- dim(sigma)
    if(dsig[1] != dsig[2]) stop("Sigma must be square.")
    if(dsig[1] != kk)
        stop("Sizes of sigma and mu are inconsistent.")
    ev <- eigen(sigma,symmetric=T)
    sqrl <- diag(sqrt(ev$values))
    PP <- ev$vectors
    ZZ <- rnorm(nn*kk) ; dim(ZZ) <- c(kk,nn)
    rmvn <- t(PP%*%sqrl%*%ZZ+mu)
    rmvn
    }# End of function rmvn
```

Simulation Code

```
set.seed(9999)
n = 25; r = 0.5; nsim=1000
crit1 = qt(0.975,n-1); crit2 = qt(0.975,2*(n-1))
Mu = c(1,1.5); Sigma = rbind(c(1,r),
                             c(r,1))
nsig1 = nsig2 = 0
for(sim in 1:nsim)
    {
    dat = rmvn(n,Mu,Sigma); X = dat[,1]; Y = dat[,2]
    sig1 = t.test(x=X,y=Y,paired=T)$p.value<0.05
    if(sig1) nsig1=nsig1+1
    sig2 = t.test(x=X,y=Y,var.equal=T)$p.value<0.05
    if(sig2) nsig2=nsig2+1
    }
cat(" \n")
cat(" Based on ",nsim," simulations, Estimated Power \n")
cat(" Matched t-test: ",round(nsig1/nsim,3),"\n")
cat(" Two-sample t-test: ",round(nsig2/nsim,3),"\n")
cat(" \n")
```

Output

```
 Based on  1000  simulations, Estimated Power
 Matched t-test:  0.675
 Two-sample t-test:  0.385

Mu = c(1,1) # H0 is true -- estimate significance level

 Based on  1000  simulations, Estimated Power
 Matched t-test:  0.063
 T-sample t-test:  0.006

Based on  10000  simulations, Estimated Power
 Matched t-test:  0.053
 Two-sample t-test:  0.007
```

Conclusions

- 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

- $\mathbf{X}_1, \ldots, \mathbf{X}_n \overset{i.i.d.}{\sim} N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- $\overline{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i$ 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

$H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{h}$, where $\mathbf{C}$ 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
- $t^2 = n\left(\mathbf{C}\overline{\mathbf{X}}_n - \mathbf{h}\right)' \left(\mathbf{CSC}'\right)^{-1} \left(\mathbf{C}\overline{\mathbf{X}}_n - \mathbf{h}\right) \sim T^2(r, n-1)$
- 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

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

- 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.