PH7430 Statistical Methods for Correlated Data - Fall 2009

Paired Data

Kyle D. Rudser

September 24, 2009
Paired Data

- Pairing (matching) seeks to reduce or remove entirely extraneous sources of variation from comparisons of a summary measure of interest (e.g., means)
- If successful, smaller variance results in smaller sample size/more power/ability to detect smaller difference
Paired Data

- In the independent case, we have two sources of variation:
  - *between-subject* (cluster) variability: usually the larger component of variation
  - *within-subject* (cluster) variability is the other
Matched samples

- In the setting of matched samples (e.g., paired data), the data can be handled by combing the measurements for a single individual, and then performing analyses based on the measurements from those independent pairs.

- EX. Two measurements are taken on each individual, one for each of two experimental conditions. The between subject variability can be effectively eliminated for the comparison between experimental conditions (e.g., two treatments).
Model:

\[ Y_{ij} = \mu_0 + \Delta T_{ij} + b_i + \varepsilon_{ij}, \quad i = 1, \ldots, N; j = 0, 1 \]

- \( Y_{ij} \) = response on subject \( i \) at time point \( j \) (e.g., \( j = 0 \) for baseline, \( j = 1 \) after treatment)
- \( \mu_0 \) = population mean
- \( \Delta \) = treatment effect
- \( T_{ij} \) = indicator of treatment
- \( b_i \) = between-subjects component of variability
- \( \varepsilon_{ij} \) = within-subject component of variability
With paired data, we can remove the between-subjects variability:

\[ d_i = Y_{i1} - Y_{i0} = (\mu_0 + \Delta T_{i1} + b_i + \varepsilon_{i1}) - (\mu_0 + \Delta T_{i0} + b_i + \varepsilon_{i0}) \]

\[ = (\mu_0 + \Delta(1) + b_i + \varepsilon_{i1}) - (\mu_0 + \Delta(0) + b_i + \varepsilon_{i0}) \]

\[ = (\mu_0 - \mu_0) + \Delta + (b_i - b_i) + (\varepsilon_{i1} - \varepsilon_{i0}) \]

\[ = \Delta + (\varepsilon_{i1} - \varepsilon_{i0}) \]

Why can’t we remove the between-subject variability, \( b_i \), with independent data?
Difference of parameters estimated from matched samples

\[ \{X_1, \ldots, X_n\} \simiid (\mu, \sigma^2) \]
\[ \{Y_1, \ldots, Y_n\} \simiid (\nu, \tau^2) \]

where \(X_i\) and \(Y_i\) are matched \(\forall i\)

\[ D_i = X_i - Y_i \simiid (\mu - \nu, \omega^2) \]

the we have a 100(1 - \(\alpha\))% CI for \(\mu - \nu\) with

\[ \bar{D} \pm z_{1-\alpha/2} \sqrt{\frac{s_D^2}{n}} \]
Difference of parameters estimated from matched samples: Paired t-test

- When we use the t-distribution for conservatism in small samples, this is related to the “paired t-test”

\[ D_i = X_i - Y_i \]

\[ 100(1 - \alpha)\% \text{ CI for } \mu - \nu \text{ is} \]

\[ \bar{D} \pm t_{n-1,1-\alpha/2} \sqrt{\frac{s_D^2}{n}} \]
Paired t-test

- The paired t-test is perhaps the most basic (familiar) correlated data analysis
  - data are observed in matched pairs (clusters)
  - comparisons are evaluated by first taking differences within clusters and then averaging across clusters
  - independence conditional on group is assumed

- Common set-up is a random sample measured under scenario 1 and again under scenario 2 (e.g., before treatment and again after)

\[
\begin{align*}
\{ Y_{10}, \ldots, Y_{n0} \} & \quad \text{before} \\
\{ Y_{11}, \ldots, Y_{n1} \} & \quad \text{after}
\end{align*}
\]

- \( E[Y_{i0}] = \mu_0, \quad E[Y_{i1}] = \mu_0 + \Delta = \mu_1, \) and \( \text{Var}(Y_{ij}) = \sigma^2 \)
Paired t-test

Inference then is based upon $D_i = Y_{i1} - Y_{i0}$

\[
E[D_i] = E[Y_{i1} - Y_{i0}] = E[Y_{i1}] - E[Y_{i0}]
\]
\[
= \mu_1 - \mu_0
\]
\[
= \Delta
\]

\[
\text{Var}(D_i) = \text{Var}(Y_{i1} - Y_{i0})
\]
\[
= \text{Var}(Y_{i1}) + \text{Var}(Y_{i0}) - 2\text{Cov}(Y_{i1}, Y_{i0})
\]
\[
= \text{Var}(Y_{i1}) + \text{Var}(Y_{i0}) - 2\rho \sqrt{\text{Var}(Y_{i1})\text{Var}(Y_{i0})}
\]
\[
= 2\sigma^2 - 2\rho \sigma^2
\]
\[
= 2\sigma^2(1 - \rho)
\]
That was the parameter space, what about the sample space?

- Consider the same average of differences $d_i = y_{i1} - y_{i0}$
  - $\bar{d}$ estimates $\hat{\Delta}$ and $s^2_d$ estimates $2\sigma^2(1 - \rho)$
  - we will test the hypothesis $H_0: \Delta = 0$.
  - the estimate of effect is $\bar{d}$
  - our test statistic can be the t-statistic

$$t = \frac{\hat{\Delta}}{\hat{se}(\hat{\Delta})} = \frac{\bar{d}}{\sqrt{s^2_d/n}}$$

- $t$ will be approximately normally distributed for a sufficiently large value of $n$
- $t$ will be $t$-distributed if $D$ is normally distributed
Example: R paired t-test

- What is estimate of treatment effect ($\hat{\Delta}$)?
- How could you evaluate the estimated variance?

Slide aside: we could evaluate the geometric mean by first taking the log transformation of the data
  - one sample t test the mean difference is 0
  - back transform to consider geometric mean of ratios
There are other approaches to matched continuous data

- **Sign test**
  - test if the median of the differences is zero
  - if true, we would expect as many differences to be above zero as below
  - if not true, then we have an indication of differences tending to be positive (negative)

**Note:**

- median difference is not difference of medians
- E.g., \( X = c(1, 2, 5), \ Y = c(2, 4, 5) \)
  - \( mdn(\ Y) - mdn(\ X) = 2; \ mdn(\ Y - \ X) = mdn(c(1, 2, 0)) = 1 \)
- median difference is not transitive
- E.g., \( X = c(1, 2, 3), \ Y = c(2, 3, 1), \ Z = c(3, 0, 2) \)
  - \( mdn(\ X - \ Y) = 1 > 0 \Rightarrow \ Y \) “bigger” than \( X \)
  - \( mdn(\ Y - \ Z) = 1 > 0 \Rightarrow \ Z \) “bigger” than \( Y \)
  - \( mdn(\ X - \ Z) = 1 > 0 \Rightarrow \ X \) “bigger” than \( Z \)
There are other approaches to matched continuous data

- **(Wilcoxon) Signed Rank test**
  - similar to sign test, but takes into account (to some degree) magnitude
    - perhaps just as many positive and negative differences, but the magnitude of those positive differences tend to be bigger than the negative ones
  - premise is the number of positive and negative differences should tend to be the same and there should not be a tendency for the magnitude of the positive differences to be bigger than the negative ones
  - if not true, then we have an indication of differences in the distributions
    - the standard error of the test statistic is based on a permutation distribution, thus has correct type I error for testing equality of distributions
    - arbitrary differences between distributions may result in a ‘significant’ statistic
Back to means, recall formulating our ‘model’ in the form:

\[ Y = X\beta + \varepsilon \]

Writing this for pairs of data \((Y_{i0}, Y_{i1})\):

\[
\begin{pmatrix}
  y_{10} \\
  y_{11} \\
  \vdots \\
  y_{m0} \\
  y_{m1}
\end{pmatrix}
= 
\begin{pmatrix}
  1 & 0 \\
  1 & 1 \\
  \vdots & \vdots \\
  1 & 0 \\
  1 & 1
\end{pmatrix}
\begin{pmatrix}
  \beta_0 \\
  \beta_1
\end{pmatrix}
+ 
\begin{pmatrix}
  \varepsilon_{10} \\
  \varepsilon_{11} \\
  \vdots \\
  \varepsilon_{m0} \\
  \varepsilon_{m1}
\end{pmatrix}
Example: R paired data in linear model

- What is estimate of treatment effect ($\hat{\Delta}$)? (is it valid?)
- What is the result of the statistical test? (is it valid?)
- Estimate of treatment effect is valid
- Test for effect is suspect
  - Variance estimate does not take into account correlated nature of the observations
- Will the estimated variance be too small, or too large?
  - Variance estimate for independent data (of the difference of means): $\frac{2\sigma^2}{n}$
  - Variance estimate for paired data (of the mean of the differences): $\frac{2\sigma^2(1-\rho)}{n}$
Consider formulating our ‘model’ as follows:

\[
Y_i = \begin{pmatrix} Y_{i0} \\ Y_{i1} \end{pmatrix}
\]

\[
E(Y_i) = E \begin{pmatrix} Y_{i0} \\ Y_{i1} \end{pmatrix} = \begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix} = \begin{pmatrix} \mu_0 \\ \mu_0 + \Delta \end{pmatrix}
\]

\[
Var(Y_i) = Var \begin{pmatrix} Y_{i0} \\ Y_{i1} \end{pmatrix} = \begin{pmatrix} \sigma^2 & \rho \sigma^2 \\ \rho \sigma^2 & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}
\]
Since the clusters (subjects) are uncorrelated, we could estimate a cluster-level mean and variance and then use those to test.

\[
\bar{Y} = \left( \begin{array}{c} \bar{Y}_0 \\ \bar{Y}_1 \end{array} \right)
\]

\[
\text{Var}(\bar{Y}) \approx \frac{\hat{\sigma}^2}{n} \left( \begin{array}{cc} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{array} \right)
\]

where \( \hat{\sigma}^2 = (s_0^2 + s_1^2)/2 \) and

\[
\hat{\rho} = \frac{1}{n-1} \sum_{i=1}^{n}(Y_{i0} - \bar{Y}_0)(Y_{i1} - \bar{Y}_1)/\sqrt{s_0^2 s_1^2}
\]
We can now proceed with $\bar{Y}$ and $\hat{\text{Var}}(\bar{Y})$ for inference and will have accounted for the inherent correlation. This will be done by constructing a *contrast*.

- A contrast is defined as $\sum a\mu$, where $\sum a = 0$
- E.g., $(-1\mu_0 + 1\mu_1) = (-1, 1)\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}$
- Then the estimate of our contrast is $a'\hat{\mu} = (-1, 1)\begin{pmatrix} \bar{Y}_0 \\ \bar{Y}_1 \end{pmatrix} = \bar{Y}_1 - \bar{Y}_0$
- And the corresponding variance is

$$\hat{\text{Var}}(a'\hat{\mu}) = a'\hat{\text{Var}}(\hat{\mu})a \approx (-1, 1) \begin{pmatrix} \hat{\sigma}^2/n \left( \begin{array}{cc} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{array} \right) \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$= \frac{2\hat{\sigma}^2(1 - \hat{\rho})}{n}$$
We see now that performing inference on pairwise differences (paired t-test)

- creates a univariate summary measure to work with
- foregoes necessarily using a more complicated ‘model’ for the data
- corresponds to a particular contrast (...perhaps we could consider other contrasts...)

The “multivariate” model can accommodate larger data structures (and hence correlation structures). We will seek to approach data analyses in a robust fashion, with minimal use of assumptions.
Example: R paired data in linear model taking correlation into account
Suppose we have paired (matched) binary data, i.e., $Y_{ij}$ are dichotomous (e.g., disease or disease-free).

- E.g., cross-over studies
  - relief of headaches from aspirin vs. Tylenol
  - each patient receives both treatments (in random order)
- E.g., ophthalmology study
  - new treatment for conjunctivitis versus placebo
  - each subject receives each treatment with random allocation to which eye receives which treatment

Data presented in 2x2 tables:

$$
\begin{array}{c|cc|c}
+ & - & n_1. \\
\hline
+ & a & b \\
- & c & d \\
\hline
n_1 & n_2 & n
\end{array}
$$
perhaps we then would be interested in the typical proportions (probabilities) $\hat{\pi}_0 = \sum Y_{i0}/n$ and $\hat{\pi}_1 = \sum Y_{i1}/n$

\[
\frac{n_1}{n} - \frac{n'}{n} = \frac{b - c}{n}
\]

how to evaluate?
could compare proportions with response in each group in similar fashion as the paired t-test

McNemar’s test

- focus is on discordant pairs
- if no effect then it ought to be that $b \approx c$
- we will condition on the number of discordant pairs ($b + c$), test based on binominal proportion:

$$b \sim B(b + c, 0.5)$$
Note: difference between McNemar’s test and chi square

- McNemar: Is “success” with treatment 1 equally prevalent as “success” with treatment 2?
- Chi square: Does “success” with treatment 1 differ between subjects with and with out “success” with treatment 2?

Example: For a given patient, edema and ascites in liver

- McNemar: Are edema and ascites equally prevalent?
- Chi Square: Does prevalence of edema differ between patients with and with out ascites?
Sign test vs. McNemar

- McNemar’s test is really the sign test on binary data
  - We could just as well call it the sign test
  - Perhaps McNemar would not agree
ANCOVA or Pre-Post Analyses

- often times we may have a study with a baseline measurement and a follow-up measurement
- for these settings, we typically are interested in the change and if the average change is different across groups
- we may have either randomized groups or unrandomized groups
Consider the following three ‘models’:

- **Post:** \( Y_{i1} = \beta_0 + \beta_1 T_i + \varepsilon_i \)
  
  The baseline value is ignored.

- **Change:** \( (Y_{i1} - Y_{i0}) = \beta_0 + \beta_1 T_i + \varepsilon_i \)
  
  The difference between final and baseline measurements are used.

- **ANCOVA:** \( Y_{i1} = \beta_0 + \beta_1 T_i + \beta_2 Y_{i0} + \varepsilon_i \)
  
  The analysis ‘adjusts’ for the baseline value.

Here \( T_i \) will denote group 1 versus group 2 (e.g., treatment indicator).
Let’s have a look at the set-up when we have randomized groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Baseline</th>
<th>Final</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>$\mu_0$</td>
<td>$\mu_0 + \delta_t$</td>
<td>$\delta_t$</td>
</tr>
<tr>
<td>Treatment</td>
<td>$\mu_0$</td>
<td>$\mu_0 + \delta_t + \Delta_T$</td>
<td>$\delta_t + \Delta_T$</td>
</tr>
<tr>
<td>Difference</td>
<td>0</td>
<td>$\Delta_T$</td>
<td>$\Delta_T$</td>
</tr>
</tbody>
</table>

Note: by randomization the expected difference at baseline is zero.
Randomized groups

- What is the expected value of $\hat{\beta}_1$ for each of the approaches?
  - Post: $E(\hat{\beta}_1) = \Delta_T$
  - Change: $E(\hat{\beta}_1) = \Delta_T$
  - ANCOVA: $E(\hat{\beta}_1) = \Delta_T$

- All three are unbiased estimators of treatment effect.

- What about the variance (precision) of these estimators?
Randomized groups

- Variance of $\hat{\beta}_1$ for each of the approaches:
  - Post: $E(\hat{\beta}_1) = \frac{2\sigma^2}{n}$
  - Change: $E(\hat{\beta}_1) = \frac{4\sigma^2(1-\rho)}{n}$
  - ANCOVA: $E(\hat{\beta}_1) = \frac{2\sigma^2(1-\rho^2)}{n}$

- Clearly they are not necessarily the same, and will depend on the value of $\rho$.

- Some equivalences are immediate: $\rho \in \{0, 0.5, 1\}$
Randomized groups

► while there is not a unique dominating estimator, the ANCOVA ("adjusted for baseline") is uniformly "as good as"

► between the ‘Post’ and ‘Change’ approaches, ordering based on precision changes at $\rho = 0.5$

► result is increased power/smaller sample size required/ability to detect a smaller clinical benefit

► follow-up: suppose we are in the situation of limited resources, and costs are directly related to the number of measurements

  ► For what value of $\rho$ will $\text{Var}(C) \leq \text{Var}(P)$?
    $\rho \geq 0.75$
  
  ► For what value of $\rho$ will $\text{Var}(A) \leq \text{Var}(C)$?
    $\rho \geq \sqrt{1/2} \approx 0.707$
Back to our three ‘models’, all can be viewed as variations of what was previously presented as the third:

\[ Y_{i1} = \beta_0 + \beta_1 T_i + \beta_2 Y_{i0} + \varepsilon_i \]

\[ \Rightarrow Y_{i1} - \beta_2 Y_{i0} = \beta_0 + \beta_1 T_i + \varepsilon_i \]

- **Post:** \( Y_{i1} = \beta_0 + \beta_1 T_i + \varepsilon_i \)
  \( \beta_2 = 0 \), i.e., value is fixed to be 0

- **Change:** \( (Y_{i1} - Y_{i0}) = \beta_0 + \beta_1 T_i + \varepsilon_i \)
  \( \beta_2 = 1 \), i.e., value is fixed to be 1

- **ANCOVA:** \( Y_{i1} = \beta_0 + \beta_1 T_i + \beta_2 Y_{i0} + \varepsilon_i \)
  \( \beta_2 \) is not predetermined and is allowed to be estimated
Examples:

- The role of angiotensin II in fat mass homeostasis
  - randomized trial to determine effect of combined ACEI/AT II type I receptor blockade on weight and fat mass loss
  - weight and fat mass measured at baseline and at 12 months

- Promoting home-based walking among African Americans with PAD
  - randomized trial to determine effect of Patient-centered Assessment and Counseling for Exercise (PACE) in combination with motivational interviewing (i.e., counseling method) to increase home-based walking and reduce walking impairment
  - 6 min walk test measured at baseline and 6 months
What about non-randomized experiments (e.g., observational studies)

- There was a key feature in the randomized setting...

What was it?
Let’s have a look at the set-up when we do not have randomized groups

<table>
<thead>
<tr>
<th>Group</th>
<th>Baseline</th>
<th>Final</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>( \mu_0 )</td>
<td>( \mu_0 + \delta_t )</td>
<td>( \delta_t )</td>
</tr>
<tr>
<td>Treatment</td>
<td>( \mu_0 + \delta_o )</td>
<td>( \mu_0 + \delta_t + \Delta_T )</td>
<td>( \delta_t + \Delta_T - \delta_o )</td>
</tr>
<tr>
<td>Difference</td>
<td>( \delta_o )</td>
<td>( \Delta_T )</td>
<td>( \Delta_T - \delta_o )</td>
</tr>
</tbody>
</table>

Issues:

- Major issue is potential confounding of the observed treatment effect
- Post analysis may not be helpful if the two groups have a large difference at baseline
- Change and ANCOVA analyses estimate different quantities
  - suppose we are interested in evaluating weight loss
  - comparing mean follow-up weight for men and women with equal weight at baseline will likely entail a comparison of a man who is lighter than the average man and a woman who is heavier than the average woman
Summary

- for randomized studies, all three approaches have the same expectation for estimate of treatment effect
- for non-randomized studies, the three approaches answer different scientific questions:
  - What is the difference in outcome of the two groups at follow-up?
  - What is difference in the change in outcome for the two groups?
  - What is the expected difference in outcome at follow-up, adjusting for the value at baseline?