Research Directions: Mixed Two-Sample Deign and Zooarchaeology Classification

Mixed Paired and Two-Sample Designs

Coyotes.

- It was desired to compare two methods (QIAGEN DNeasy Kit, traditional chloroform isoamyl alcohol method) for extracting DNA from coyote blood samples.
- The response variable (Y) was mean concentration of DNA.
- A total of 30 coyotes were available for the study.
- Ideally, both methods would be used on each coyote
  - Reduced variability, as differences between treatments would be on the same subject
  - Fewer subject (coyotes) required

Mixed Paired and Two-Sample Designs-Coyotes

Due to constraints, however, both methods were used on only 6 coyotes (randomly selected). The kit was randomly assigned to be used for 8 of the remaining coyotes and the traditional method for the remaining 16 coyotes.

Thus,

- 6 coyotes were measured twice in a paired design (dependent samples)
- the remaining 24 were in a completely randomized design (independent samples)

Can the paired and unpaired *t*-tests be combined in order to use all of the resulting data for a single test?

Consider the statistic

$$t_{combined} = \lambda t_{paired} + (1 - \lambda) t_{unpaired}$$

Note that $t_{combined} = t_{paired}$ when $\lambda = 1$ and $t_{combined} = t_{unpaired}$ when $\lambda = 1$.

- Much of the previous research has been focused on trying to approximate the distribution of $t_{combined}$.

Previous Research: Methods based on normal distributions

- Bhoj (1978). Showed $t_{combined}$ can be approximated by a Student's t-distribution. However, the degrees of freedom must be odd, which limited its usefulness. Findings
  - Paired t-test better when $\rho \geq 0.9$ and proportion of missing pairs small
  - $t_{combined}$ better in all other cases

Previous Research: Methods based on normal distributions

- Bhoj (1984, 1989). Used transformations of t_{paired} and $t_{unpaired}$ to achieve an approximate normal distribution for $t_{combined}$
  - Better approximation resulted in more powerful tests
- None of the statistics was shown to be superior under all conditions
- It was not clear how they are affected by
  - nonnormality
  - outliers
  - **unequal variance**

Previous Research-Nonparametric approaches

- Dubnicka, et. al. (2002). Proposed weighted and unweighted combinations of Wilcoxon statistics
  - Statistic: $W_{combined} = \lambda W_{paired} + (1 - \lambda) W_{unpaired}$
  - Derived the asymptotic distribution of $W_{combined}$
- Magel & Fu (2014).
  - Dubnicka (2002) standardized $W_{combined}$
  - Magel & Fu (2014) suggested first standardizing W_{paired} and $W_{unpaired}$, and then combining
    - **Showed slightly higher power in *some* cases**

Previous Research-Johnson & Richter (2020)

- Proposed using randomization versions of Dubnicka's (2002) methods
- Performed extensive simulation study
  - Considered all previous statistics
  - Investigated a wide range of
    - Data distributions
    - Sample size combinations
    - Paired observation correlations

Previous Research-Johnson & Richter (2020)

- No single method was always best
  - Parametric methods (e.g., Bhoj) generally best with normal data
  - Nonparametric methods generally better with nonnormal data
- Rank-based methods recommended as "default"
- Choice of weight more important as the proportion of complete pairs decreased
- With high correlation and most pairs complete, W_{paired} may be adequate

Possible Research Directions

1. Adaptive test

- Idea: Choose which test to apply for a given sample based on data
- We know that certain tests work better under certain conditions
  - $t_{combined}$ tends to work better for *lighter*-tailed distributions
  - $W_{combined}$ tends to work better for *heavier*-tailed distributions
  - Suppose we had a measure of "tail weight", say tw. Then propose the procedure
    - $New_{combined} = t_{combined}, \ if \ tw = 0,$
    - $New_{combined} = W_{combined}, \ if \ tw > 0.$

1. Adaptive test

- Could also focus just on the nonparametric tests
- Since either of Dubnicka or Magel & Fu statistics have been shown to be better in certain cases
  - $W_{combined} = Dubnicka\ test,\ if\ ??,$
  - $W_{combined} = Magel\ \&\ Fu\ test,\ otherwise.$
- Would need to examine results of Magel & Fu (2014) to understand how to predict which test to use

2. Combined test

Suppose we have two or more tests that test similar hypotheses

- Compute p-value using each test
- Combine the p-value into one using a *combining function*

2. Combined test

- Could combine the "better" of the parametric and nonparametric tests, or
- Could focus just on the nonparametric tests

Combined or Adaptive test: Research Process

- Review previous simulation results
  - Einsporn & Habtzghi (2013); Magel & Fu (2014); Johnson & Richter (2020)
  - Run additional simulations?
  - How to choose the preferred method based on data?
- Research theory of combining tests
  - Mathematical derivation: combined test statistic
  - variance of combined test statistic
  - distribution of combined test statistic
- Simulation study: Investigate properties of adaptive/combined test
  - Type I error
  - Power

3. Effect of Unequal Variance

- All previous methods assumed equal population variances
  - What is the effect of unequal variance on the various tests?

3. Effect of Unequal Variance: Research Process

- Simulation study: Investigate the effect of unequal variance over a variety of scenarios
  - Power
  - Type I error
- Challenge: Designing the simulation
  - How to choose appropriate variance/covariance structures

Zooarchaeology Classification

Goal

- Create a classification algorithm from these data to predict percent of human vs. carnivore presence for a sample from a site with unknown percents

Zooarchaeology Classification: Work so far...

- Created a "proof of concept" (Ryan Parks-Master's Project (2022))

Process:

- Generated 500 bootstrap samples from each of 5 classes: 0%, 25%, 50%, 75%, 100% carnivore
- Calculated summary statistics for several features (e.g., %cut marks, average # of percussive marks, etc.)
- Combined into "superpopulation" of summary statistics
- Trained classification algorithm on this superpopulation
- Able to produce probability estimates of belonging to each class

Zooarchaeology Classification: Future directions

- Incorporate more granularity: more than 5 classes
- Investigate the effect of
  - Sample size of each bootstrap sample
  - number of bootstrap samples
  - classification algorithm used

Zooarchaeology Classification: Future directions

- Incorporate dependencies between observations
  - Some bone specimens come from the same carcass: expect these to have characterictics more similar than bones from different carcasses

Zooarchaeology Classification: Research Process

- Collaborate with zooarchaeologist to discuss issues and ideas
- Design simulation experiment to investigate aforementioned effects
- Carry out simulation

Questions?

Happy Researching!

References

- Bhoj, D.S. (1978). Testing equality of means of correlated variates with missing data on both responses. *Biometrika*, 65:225-228.
- Bhoj, D.S. (1989). On comparing correlated means in the presence of incomplete data. *Biometrical Journal*, 31:279–288.
- Dubnicka, S.R., Blair, R.C., Hettmansperger, T.P. (2002). Rank-based procedures for mixed paired and two-sample designs. *Journal of Modern Applied Statistical Methods*, 1:32-41.
- Einsporn, R.L., Habtzghi, D. (2013). Combining paired and two-sample data using a permutation test. *Journal of Data Science*, 11:767-779.
- Johnson, E.N. & Richter, S.J. (2022). Permutation tests for mixed paired and two-sample designs. *Computational Statistics*, 37:739-750.

References

- Lin, P.E. & Stivers, L.E. (1974). On the difference of means with incomplete data. *Biometrika*, 61:325–334.
- Magel, R.C., Fu, R. (2014). Proposed nonparametric test for the mixed two-sample design. *Journal of Statistical Theory and Practice*, 8:221-237.