Tests for Mixed Paired and Two-Sample Designs

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 (2018)

- 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 (2018)

- 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 proportion of complete pairs decreased
- With high correlation and most pairs complete, W_{paired} may be adequate

Possible Research Projects

1. "Combined" combined nonparametric test

2. Evaluate effect of unequal variance

1. “Combined” combined Nonparametric Test

- Since either of Dubnicka or Magel & Fu statistics could be better in certain cases, can they be further combined to give a better test (i.e., a “combined” combined test!)?

1. "Combined" combined nonparametric test

- Dubnicka, et. al. (2002).
  - Proposed weighted and unweighted combinations of Wilcoxon statistics: $W_{combined} = \lambda W_{paired} + (1 - \lambda) W_{unpaired}$
  - Combined then standardized
- Magel & Fu (2014).
  - Suggested first standardizing W_{paired} and $W_{unpaired}$, and then combining
  - **Showed slightly higher power in some cases**
  - **Dubnicka methods was more powerful in other cases**

What if the better test to use could be predicted in advance?

1. "Combined" combined nonparametric test

- Review previous simulation results
  - Magel & Fu (2014); Johnson (2018)
  - Run additional simulations?
  - Can the preferred method be predicted?
- 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 combined test statistic
  - compare to Dubnicka and M&F tests
  - compare to parametric tests

2. Effect of Unequal Variance

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

2. Effect of Unequal Variance

- Simulation study: Investigate the effect of unequal variance over a variety of scenarios
  - Power
  - Type I error