Paired Data Analysis

Jennifer Le-Rademacher, PhD
Assistant Professor of Biostatistics

Sponsored by

Clinical and Translational Science Institute (CTSI), Division of Biostatistics, The Institute for Health and Society

Speaker disclosure

In accordance with the ACCME policy on speaker disclosure, the speaker and planners who are in a position to control the educational activity of this program were asked to disclose all relevant financial relationships with any commercial interest to the audience. The speaker and program planners have no relationships to disclose.

CME Evaluations!

Please help us by filling out an evaluation even if you are not eligible for CME credit.

Outline

- What are paired data?
- How to analyze them?
  - Quantitative data (numerical measurements)
    - paired t-test
    - sign test
    - signed-rank test
  - Qualitative data (categorical values)
    - McNemar's test
- Concluding remarks

What are paired data?

Independent vs. Paired

- Data come from two independent sources
- No link
- Examples:
  - Eyes from different patients
  - Unrelated individuals

- Data come from two dependent sources
- Link – natural or artificial
- Examples:
  - Eyes from one patient
  - Twins
  - Siblings
  - Before and after measurements
  - Matched pairs

Study Setting

- **Objective:** compare the effect of two treatments
  - Treatment A vs. treatment B
  - Active treatment vs. placebo /sham
  - Intervention vs. control / standard of care
  - Patient characteristics (male vs. female)
- **Hypothesis:**

  H_0: no difference between treatments

  H_A: there is a difference between treatments

Possible Designs

Prospective randomized studies:

Treatment A vs. treatment B

Active treatment vs. placebo/sham

- Accrue 2*n* patients, randomly assign *n* to treatment A and *n* to treatment B
  - Independent samples
- Advantages:
  - Applicable to more outcomes and treatments (clinical and ethical concerns)
  - Shorter on-study time

- Accrue *n* patients, each will receive both treatments in random order
  - Paired sample
- Advantages:
  - Fewer patients required
  - Lower per-patient cost
  - Shorter accrual time
  - Reduced variability

Possible Designs

Prospective observational studies:

Intervention vs. control

- Accrue *n* to control group and *n* to intervention group
  - Independent samples
- Advantages:
  - Applicable to wider range of studies
  - Shorter on-study time

- Accrue *n* patients, assign all to intervention, compare before and after measurements
  - Paired sample
- Advantages:
  - Fewer patients required
  - Lower per-patient cost
  - Reduced variability
  - Reduced systematic bias

Possible Designs

Retrospective studies (case-control):

- Select a random sample, classify patients based on outcome, compare characteristics or exposure
  - ➢ Independent samples
- Advantages:
  - – Larger sample size
  - – Possible to estimate covariate effect

- Select *n* patients with outcome 1 and *n* matched patients with outcome 2, compare characteristics or exposure
  - ➢ Paired sample (matching)
- Advantages
  - – Smaller sample (also a disadvantage)
  - – Reduced systematic bias
  - – Reduced variability

What is Matching?

- Matched cases and controls:
  - for every case, select a control who has the same (or very similar) values of the matching variables.
- Common matching variables:
  - sex, age, ethnicity, etc.
- More than one control may be selected for each case.

Reasons for Matching

- To eliminate sources of extraneous variation by making the pairs (e.g. cases and controls) similar in variables which may be associated with outcome.
- Extraneous variables may mask the effect of the variable of interest or may be confounded with the variable of interest.

How to analyze paired data?

Example 1

Connolly B, McNamara A, Sharma S, Regillo C, and Tasman W. A Comparison of Laser Photocoagulation with Trans-scleral Cryotherapy in the Treatment of Threshold Retinopathy of Prematurity. *Ophthalmology* Vol 105:1628-31, 1998. (link)

Objective: Determine whether there was a difference between visual outcomes of eyes treated with trans-scleral cryotherapy vs. laser photocoagulation

Design: Extended follow-up of a prospective clinical trial where patients eyes were randomized to one of two treatments.

Outcomes: Best-corrected visual acuity (qualitative) and spherical equivalent (quantitative)

Example 1: Data

Patient	Gender	Cryo	Laser	SE Cryo	SE L
1	M	OD	OS	-8.50	-5.38
2	F	OD	OS	-1.63	0.38
3	M	OD	OS	-11.13	-2.75
4	M	OD	OS	-15.50	.
5	F	OD	OS	-9.00	-7.50
6	F	OD	OS	-15.88	-12.63
7	F	OD	OS	1.25	2.25
8	M	OD	OS	-2.50	-0.13
9	F	OD	OS	3.00	3.50
10	F	OD	OS	-7.50	-1.00
11	F	OD	OS	-1.38	1.63
12	F	OD	OS	-6.38	-7.5
13	M	OD	OS	-5.25	-6.25
14	M	OD	OS	0.25	-0.38
15	F	OS	OD	-6.00	-7.00
16	M	OS	OD	-13.00	-9.25
17	F	OS	OD	-9.00	-5.25
18	F	OS	OD	-2.63	-4.25
19	M	OS	OD	1.88	2.00
20	F	OS	OD	-3.13	-4.75
21	F	OS	OD	-8.38	2.38
22	M	OS	OD	-5.50	-0.88
23	M	OS	OD	-1.75	-1.75
24	F	OS	OD	-4.63	-5.75
25	M	OS	OD	.	.

Example 1: Hypothesis & Test

- Hypothesis:

  H_0: The mean difference between the SE of eyes treated with cryotherapy and eyes treated with laser photocoagulation is zero.

  H_A: The mean difference is not zero.

- Testing method: paired t-test
- Significance level α=0.05

Paired t-test

- Applications:
  - used to compare continuous measurements obtained from two dependent samples
- Assumption:
  - differences are independent and are normally distributed

Paired t-test (cont'd)

- Test statistic:

$$t = \frac{\text{mean(difference)}}{\sqrt{\text{var(difference)} / n}}$$

- Decision making: if observed test statistic is too large in magnitude, reject H_0 and conclude that the mean difference is not zero.
- Inference is based on *t* distribution with *n*-1 degrees of freedom

Example 1: Paired t-test

Patient	SE Cryo	SE Laser	SE Laser – SE Cryo
1	-8.50	-5.38	3.12
2	-1.63	0.38	2.01
3	-11.13	-2.75	8.38
4	-15.50	.	.
5	-9.00	-7.50	1.50
6	-15.88	-12.63	3.25
7	1.25	2.25	1.00
8	-2.50	-0.13	2.37
9	3.00	3.50	0.50
10	-7.50	-1.00	6.50
11	-1.38	1.63	3.01
12	-6.38	-7.50	-1.12
13	-5.25	-6.25	-1.00
14	0.25	-0.38	-0.63
15	-6.00	-7.00	-1.00
16	-13.00	-9.25	3.75
17	-9.00	-5.25	3.75
18	-2.63	-4.25	-1.62
19	1.88	2.00	0.12
20	-3.13	-4.75	-1.62
21	-8.38	2.38	10.76
22	-5.50	-0.88	4.62
23	-1.75	-1.75	0.00
24	-4.63	-5.75	-1.12
25	.	.	.

n = 23

mean(difference) = 2.02

var(difference) = 10.74

Example 1: Paired t-test

Patient	SE Cryo	SE Laser	SE Laser – SE Cryo
1	-8.50	-5.38	3.12
2	-1.63	0.38	2.01
3	-11.13	-2.75	8.38
4	-15.50	.	.
5	-9.00	-7.50	1.50
6	-15.88	-12.63	3.25
7	1.25	2.25	1.00
8	-2.50	-0.13	2.37
9	3.00	3.50	0.50
10	-7.50	-1.00	6.50
11	-1.38	1.63	3.01
12	-6.38	-7.50	-1.12
13	-5.25	-6.25	-1.00
14	0.25	-0.38	-0.63
15	-6.00	-7.00	-1.00
16	-13.00	-9.25	3.75
17	-9.00	-5.25	3.75
18	-2.63	-4.25	-1.62
19	1.88	2.00	0.12
20	-3.13	-4.75	-1.62
21	-8.38	2.38	10.76
22	-5.50	-0.88	4.62
23	-1.75	-1.75	0.00
24	-4.63	-5.75	-1.12
25	.	.	.

n = 23

mean(difference) = 2.02

var(difference) = 10.74

$$t_{22} = \frac{2.02}{\sqrt{10.74 / 23}} = 2.96$$

Example 1: Paired t-test

Patient	SE Cryo	SE Laser	SE Laser – SE Cryo
1	-8.50	-5.38	3.12
2	-1.63	0.38	2.01
3	-11.13	-2.75	8.38
4	-15.50	.	.
5	-9.00	-7.50	1.50
6	-15.88	-12.63	3.25
7	1.25	2.25	1.00
8	-2.50	-0.13	2.37
9	3.00	3.50	0.50
10	-7.50	-1.00	6.50
11	-1.38	1.63	3.01
12	-6.38	-7.50	-1.12
13	-5.25	-6.25	-1.00
14	0.25	-0.38	-0.63
15	-6.00	-7.00	-1.00
16	-13.00	-9.25	3.75
17	-9.00	-5.25	3.75
18	-2.63	-4.25	-1.62
19	1.88	2.00	0.12
20	-3.13	-4.75	-1.62
21	-8.38	2.38	10.76
22	-5.50	-0.88	4.62
23	-1.75	-1.75	0.00
24	-4.63	-5.75	-1.12
25	.	.	.

n = 23

mean(difference) = 2.02

var(difference) = 10.74

$$t_{22} = \frac{2.02}{\sqrt{10.74/23}} = 2.96$$

p-value = 0.007 < .05

Example 1: Paired t-test

Patient	SE Cryo	SE Laser	SE Laser – SE Cryo
1	-8.50	-5.38	3.12
2	-1.63	0.38	2.01
3	-11.13	-2.75	8.38
4	-15.50	.	.
5	-9.00	-7.50	1.50
6	-15.88	-12.63	3.25
7	1.25	2.25	1.00
8	-2.50	-0.13	2.37
9	3.00	3.50	0.50
10	-7.50	-1.00	6.50
11	-1.38	1.63	3.01
12	-6.38	-7.50	-1.12
13	-5.25	-6.25	-1.00
14	0.25	-0.38	-0.63
15	-6.00	-7.00	-1.00
16	-13.00	-9.25	3.75
17	-9.00	-5.25	3.75
18	-2.63	-4.25	-1.62
19	1.88	2.00	0.12
20	-3.13	-4.75	-1.62
21	-8.38	2.38	10.76
22	-5.50	-0.88	4.62
23	-1.75	-1.75	0.00
24	-4.63	-5.75	-1.12
25	.	.	.

n = 23

mean(difference) = 2.02

var(difference) = 10.74

$$t_{22} = \frac{2.02}{\sqrt{10.74 / 23}} = 2.96$$

p-value = 0.007 < .05

- Reject the null hypothesis. There is significant evidence that the mean difference between SE after laser coagulation and cryotherapy is not zero.

Example 1: Paired t-test

Patient	SE Cryo	SE Laser	SE Laser – SE Cryo
1	-8.50	-5.38	3.12
2	-1.63	0.38	2.01
3	-11.13	-2.75	8.38
4	-15.50	.	.
5	-9.00	-7.50	1.50
6	-15.88	-12.63	3.25
7	1.25	2.25	1.00
8	-2.50	-0.13	2.37
9	3.00	3.50	0.50
10	-7.50	-1.00	6.50
11	-1.38	1.63	3.01
12	-6.38	-7.50	-1.12
13	-5.25	-6.25	-1.00
14	0.25	-0.38	-0.63
15	-6.00	-7.00	-1.00
16	-13.00	-9.25	3.75
17	-9.00	-5.25	3.75
18	-2.63	-4.25	-1.62
19	1.88	2.00	0.12
20	-3.13	-4.75	-1.62
21	-8.38	2.38	10.76
22	-5.50	-0.88	4.62
23	-1.75	-1.75	0.00
24	-4.63	-5.75	-1.12
25	.	.	.

n = 23

mean(difference) = 2.02

var(difference) = 10.74

$$t_{22} = \frac{2.02}{\sqrt{10.74 / 23}} = 2.96$$

p-value = 0.007 < .05

- Reject the null hypothesis. There is significant evidence that SE are difference between laser coagulation and cryotherapy.

The 95% CI of the mean difference = (0.61, 3.44).

Example 1: Ignore Pairing

Treat two eyes from each patient as independent :

$$t_{44} = \frac{\text{mean(SE Laser)} - \text{mean(SE Cryo)}}{\sqrt{\frac{\text{var(SE Laser)}}{n\text{(SE Laser)}} + \frac{\text{var(SE Cryo)}}{n\text{(SE Cryo)}}}} = \frac{2.02}{\sqrt{\frac{18.34}{23} + \frac{23.34}{23}}} = 1.50$$

p-value = 0.14 > .05

- Do *not* reject the null hypothesis. There is *not* enough evidence to conclude that the mean SEs between laser coagulation and cryotherapy are different.

The 95% CI of the mean difference = (-.69, 4.73).

Non-normality

- Key assumption for a valid paired t-test:
  - differences are normally distributed and/or sample size is large
- Violation of the assumption:
  - non-normal differences
- Alternative:
  - nonparametric testing

Example 2

Gefffen G, Bradshaw JL, and Nettleton NC. Attention and Hemispheric Differences in Reaction Time during Simultaneous Audio-Visual Tasks. *Quart. J. of Expt. Psychol.,* 25:404-412, 1973.

Objective: To determine whether certain numbers presented in random order were perceived more rapidly in the right (RVF) or left visual fields (LVF)

Design: Present numbers to right or left visual field in random order to 12 right-handed subjects

Outcomes: Response time

Example 2: Data

LVF	RVF
564	557
521	505
495	465
564	562
560	544
481	448
545	531
478	458
580	560
484	485
539	520
467	445

Example 2: Hypothesis & Test

- Hypothesis of interest:

  H_0: median of the differences in response time between LVF and RVF is zero

  H_A: median of the differences is zero
- Note: We do not assume the differences in response times are normally distributed.
- Testing methods:
  - sign test
  - signed rank test
- Level of significance $\alpha=0.05$

Nonparametric Tests

- Applications:
  - used to compare numerical measurements from two dependent samples
  - does not assume normality of the differences
- Advantages:
  - no normality assumption
- Disadvantages:
  - wasteful of information
  - less powerful if normality holds

Sign Test

- Test statistic:
  - number of positive differences
- Decision making:
  - if H_0 is true, n/2 positive differences are expected
  - reject H_0 if number of positive differences is too large or too small
- Inference is based on binomial distribution (or its normal approximation)

Example 2: Sign Test

LVF	RVF	LVF-RVF	Sign
564	557	7	+
521	505	16	+
495	465	30	+
564	562	2	+
560	544	16	+
481	448	33	+
545	531	14	+
478	458	20	+
580	560	20	+
484	485	-1	-
539	520	19	+
467	445	22	+

n = 12

positive differences = 11

p-value = 2* P(11 or more "+")

= 2*0.003

= 0.006 < 0.05

- Reject the null hypothesis. There is significant evidence that the median of the differences in response times between LVF and RVF is not zero.
- The 95% CI of the median difference = (7, 22).

Signed-rank Test

- Test procedure:
  - rank the differences ignoring their sign
  - assign each rank the original sign of the difference
  - find the sum of the positive and negative ranks
- Test statistic: the smaller of the two sums.
- Decision making:
  - if H_0 is true, two sums should be similar in magnitude
  - reject H_0 if one of the sums is small

Example 2: Signed-Rank Test

LVF-RVF	Rank	Signed Rank
7	3	3
16	5.5	5.5
30	11	11
2	2	2
16	5.5	5.5
33	12	12
14	4	4
20	8.5	8.5
20	8.5	8.5
-1	1	-1
19	7	7
22	10	10

Sum of "-" ranks = 1

Sum of "+" ranks = 77

P-value = 2*P (one of the sums is 1 or less)

= 2×0.00049

= 0.00098 < .05

- Reject the null hypothesis. There is significant evidence that median of the differences in response times between LVF and RVF is zero.

Example 2: Ignore Pairing

- Treat two visual fields from each patient as independent

  p-value = 0.29 > .05

- Do *not* reject the null hypothesis. There is *not* enough evidence that the median response times from LVF and RVF are different.

Example 3

Coulehan JL, Lerner G, Helzlsouer K, Welty T, and McLaughlin J. Acute Myocardial Infarction among Navajo Indians, 1976-83. *American Journal of Public Health,* Vol 76:412-414, 1986. (link)

Objective: To determine risk factors associated with acute myocardial infarction (AMI) in Navajo Indians

Design: Retrospective matched case-control (age- and sex-matched)

Outcomes: presence of AMI – interest is on predictors

Example 3: Data

AMI Controls	AMI Cases		Total
	Diabetes	No Diabetes	
Diabetes	9	16	25
No Diabetes	37	82	119
Total	46	98	144

Example 3: Hypothesis & Test

- Hypothesis:

  H_0: There is no association between Diabetes and AMI.

  H_A: There is an association between Diabetes and AMI.

- Note: the outcome is a categorical variable
- Testing method: McNemar's test
- Level of significance $\alpha=0.05$

McNemar's Test

- Applications:
  - used to compare proportions from two dependent samples
- Test compares # of discordant pairs:

  r = # pairs (Present, Absent),

  s = # pairs (Absent, Present)
- If H_0 is true, r and s should be nearly equal. Reject H_0 if r & s are very different.

McNemar's Test (cont'd)

- Formally, test statistic is

$$X^2 = \frac{(r-s)^2}{(r+s)}$$

- Inference is based on chi-square distribution with 1 degree of freedom.

Example 3: McNemar's test

AMI Controls	AMI Cases		Total
	Diabetes	No Diabetes	
Diabetes	9	16	25
No Diabetes	37	82	119
Total	46	98	144

Total pairs = 144

r = # pairs (Diabetes, No Diabetes) = 16

s = # pairs (No Diabetes, Diabetes) = 37

Test statistic:

$$\chi^2 = \frac{(16-37)^2}{(16+37)} = 8.32$$

P-value = 0.0039 < 0.05

- Reject the null hypothesis. There is significant evidence of an association between AMI and Diabetes.

Example 3: Ignore Pairing

AMI Controls	AMI Cases: Diabetes	AMI Cases: No Diabetes	Total
Diabetes	9	16	25
No Diabetes	37	82	119
Total	46	98	144

AMI	Diabetes	No Diabetes	Total
Yes	46	98	144
No	25	119	144
Total	71	217	288

Pairs = 144

McNemar's test:

P-value = 0.0039

Observations = 288

Chi-square test:

P-value = 0.0041

Given the same marginal totals, the McNemar's statistics vary for different values of cell 1,1 (**9**) whereas the Chi-square test statistics remain unchanged.

Summary

- Paired analysis:
  - differences are calculated within each pair and single sample of differences is examined
- Tests for paired data:
  - paired t-test (numerical measurements; differences are assumed to be normally distributed)
  - sign or signed rank test (numerical measurements, no assumption of normality)
  - McNemar's test

    (proportions from two dependent samples)

Concluding Remarks

- The need for paired analysis is established by the study design.
- The analysis must reflect the design that generated the data.
- Overlooking pairing:
  - mean difference will be estimated properly
  - variability will be assessed incorrectly (variability related to matching variables becomes part of the unexplained variation and may obscure existing differences)

Concluding Remarks (cont'd)

- Considerations for paired studies:
  - carry-over effect when treatments must be administered in sequence
  - If matching is desired, only cases having a matched control can be used
  - matching is more useful in small studies
  - use regression methods for adjustment instead of matching for larger studies
- Use appropriate regression techniques to adjust for non-matched variables (e.g. conditional logistic regression)

Resources

- The Clinical and Translation Science Institute (CTSI) supports education, collaboration, and research in clinical and translational science: www.ctsi.mcw.edu
- The Biostatistics Consulting Service provides comprehensive statistical support www.mcw.edu/biostatistics.htm

Free drop-in consulting

- MCW: Tuesdays & Thursdays 1 – 3 PM
  - Health Research Center, H2400
- Froedtert: Mondays, Wednesdays, Fridays 1 – 3 PM
  - Froedert Pavilion (#L772A)
- VA: Every Mondays 9:30-10:30 am
  - VA Medical Center, Room 111-B-5423
    call x46494 for access
- Marquette: Tuesdays 8:30-10:30 am
  - School of Nursing, Clark Hall, Office of Research & Scholarship