Models for Matched Pairs

Jee Dong Jun

09.9.2020

Models for Paired Data

- Methods for comparing categorical responses for two samples when each observation in one sample pairs with an observation in the other sample.
- Example : longitudinal studies that observe subjects over time on the same categorical scale.
- It is not sensible to treat samples as independent.

TABLE 10.1 Rating of Performance of Prime Minister

First Survey	Second Survey: Approve	Second Survey: Disapprove	Total
Approve	794	150	944
Disapprove	86	570	656
Total	880	720	1600

Figure 1: Example

Comparing Dependent Proportion for binary

- It is not reasonable to test for independence.
- $\pi_{1+} = \pi_{+1}$ implies $\pi_{2+} = \pi_{+2}$
- This is called marginal homogeneity.
- This implies there is no change in paired sample.

Confidence intervals comparing dependent proportions

- $d = p_{+1} - p_{1+}$
- We can calculate variance of d.
- $var(\sqrt{n}d) = \pi_{1+}(1-\pi_{1+}) + \pi_{+1}(1-\pi_{+1}) - 2(\pi_{11}\pi_{22} - \pi_{12}\pi_{21})$
- Then we can form Wald type of intervals.

McNemar Test

- Under null hypothesis, estimated variance is $(p_{12} + p_{21})/n$
- Score test statistics $z = \frac{n_{21} - n_{12}}{\sqrt{n_{21} + n_{12}}}$
- Note this does not depend on diagonal but all cases contribute how much differ.

Connection Between McNemar and CMH Tests

TABLE 10.2 Representation of Four Types of Matched Pairs Contributing to Counts in Table 10.1

		Response	
Subject	Survey	Approve	Disapprove
1	First	1	0
	Second	1	0
2	First	0	1
	Second	0	1
3	First	0	1
	Second	1	0
4	First	1	0
	Second	0	1

Figure 2: Example

Connection Between McNemar and CMH Tests

- Conditional independence given subject
- This results in marginal homogeneity
- McNemar test and CMH test are algebraically identical!

Conditional Logistic Regression for Binary Matched Pairs

- Let (Y_{i1}, Y_{i2}) denote pair of observations for subject i.
- Let Y=1 is success and Y= 0 is fail
- Let $P(Y_t = 1)$ be mean of $P(Y_{it} = 1)$
- $P(Y_t = 1) = \alpha + \beta x_t$ such that $x_2 = 1$
- logit model is also possible

Subject Specific Model

- Above models are called marginal models
- The effects in marginal models are averaged.
- An alternative modeling approach focuses on the subject-specific tables.

Subject Specific Probabilities

- link $P(Y_{it} = 1) = \alpha_i + \beta x_t$
- When link is identity, then we get same as marginal model.
- When link is logistic, we get different model.
- However, this model still assumes same β

Subject-Specific Model

- $P(Y_{i1} = 1) = \frac{\exp a_i}{1+\exp a_i}$ and $P(Y_{i2} = 1) = \frac{\exp a_i+\beta}{1+\exp a_i+\beta}$
- Note $P(Y_{i1} = 1)$ and $P(Y_{i2} = 1)$ are correlated by model.
- This model has too many parameters.
- One possible model is to treat α_i as random effect from normal distribution.

Conditional Logistic Regression for Matched Case-Control Studies

TABLE 10.3 Previous Diagnoses of Diabetes for Myocardial Infarction (MI) Case–Control Pairs

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

Source: J. L. Coulehan et al., *Amer. J. Public Health* **76**: 412–414 (1986), reprinted with permission from the American Public Health Association.

TABLE 10.4 Possible Case–Control Pairs for Table 10.3

	a		b		c		d	
Diabetes	Case	Control	Case	Control	Case	Control	Case	Control
Yes	1	0	0	1	1	1	0	0
No	0	1	1	0	0	0	1	1

Figure 3: Example

- Does not necessarily same person.
- Role of X and Y are reversed.

Marginal Models for Square Contingency Tables

- Generalize to I categories. *IxI* table
- Extension of above model by $\log(p(Y_t = j)/p(Y_t = I) = a_j + \beta_j x_t$
- Marginal Homogeneity is special case where all $\beta_j = 0$
- For ordered categories, we can use cumulative logit model.

Symmetry Model

- I x I joint distribution satisfies symmetry if $\pi_{ab} = \pi_{ba}$
- Under symmetry, marginal homogeneity follows.
- They are equivalent in binary case but not necessarily in square cases.

Symmetry Model is logistic and loglinear model

- Symmetry model is logistic model where $\log(\pi_{ab}\pi_{ba}) = 0$
- It also has loglinear form $\log \mu_{ab} = \lambda + \lambda_a + \lambda_b + \lambda_{ab}$
- $\lambda_{ab} = \lambda_{ba}$
- The solution taht satisifes symmetry is $\hat{\mu}_{ab} = n_{ab} + n_{ba}/2$

Quasi-symmetry

- Permit main effect terms in the symmetry model to differ.
- $\log \mu_{ab} = \lambda + \lambda_a^x + \lambda_b^Y + \lambda_{ab}$
- Marginal homogeneity is not log linear
- But symmetry is equivalent to quasi + marginal homogeneity
- Quasi independence model $log \mu = \lambda + \lambda_a^x + \lambda_b^y + \delta_a I(a = b)$