STATISTICS 475: Applied Discrete Data Analysis

Matched Pairs Binomials

(B&L Section 1.2.6)

1 Problem to be solved

- All of the comparisons we have done so far assume that the measurements in each group are independent of measurements in the other.

- *Very often*, binary data are measured in a way that causes one observation in each group to be matched with one observation in the other group

- This usually causes there to be a positive correlation between these pairs of measurements

- We the methods we have learned so far are not appropriate for these problems, so we need to come up with new ones.

2 Paired Binary Data   *eg. test 2 body creams on same person.*

- PAIRED DATA means that there is a concept of a "subject" that consists of two parts that are different in a particular way, and the same measurements are made on each of the two parts. This creates two groups, where each member of one group has a specific "partner" in the other group.

  - Husband and wife are both asked opinions
    * "Family" is the subject concept
    * "Man" and "woman" are the two parts of the subject
  - Measurements are taken on a person's disease before and after exposure to some treatment
    * "Person" is the subject concept
    * "Before" and "after" are the two parts of the subject

1

- Fields of wheat are split in half, with one half receiving a pesticide and the other getting nothing.
    * "Field" is the subject
    * "Herbicide" and "no herbicide" are the two parts of the subject.
- Here we consider the case where the measurements are binary

    - Do people favour a new law about child custody in divorce cases?
    - Is a certain kind of insect found in traps in the fields?

- Interested in answering questions that compare the probability of success in one group to the probability of success in the other group
    - Do male/female spouses like new child custody law with different probabilities?
    - Is the insect less likely in traps in field that has been sprayed?
- Two members of pair each have two possible outcomes, so "subject" has four possible outcomes
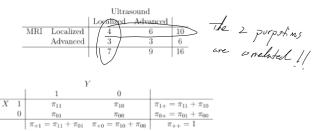   - (0,0), (0,1), (1,0), (1,1)

2.1 Notation

- X is the RV representing the binary result for the group 1 measurement
- Y is the RV representing the binary result for the group 2 measurement
- The pair X, Y has a JOINT DISTRIBUTION representing results measured simultaneously or *jointly* on the pair.
    - Joint distribution is denoted by $\pi_{ij} = P(X = i, Y = j)$ for $i = 0, 1$ and $j = 0, 1$
    - $\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11}$
    - These are called JOINT PROBABILITIES
- We are most interested in comparing $P(X = 1)$ and $P(Y = 1)$:
    - $P(X = 1) = \pi_{11} + \pi_{10} = \pi_{1+}$ '+' represents adding over the another group.
    - $P(Y = 1) = \pi_{11} + \pi_{01} = \pi_{+1}$
    - These are called MARGINAL PROBABILITIES of success → cuz at margin of the
    - Represent probabilities of success without regard to what partner's response is.   table below
- Easily depicted in a table (get used to this!!):

Table 1: Diagnoses established using MRI and ultrasound technology among individuals who truly have localized prostate cancer. Data source is Zhou and Qin (2005).

		Ultrasound		
		Localized	Advanced	
MRI	Localized	4	6	10
	Advanced	3	3	6
		7	9	16

(handwritten note: the 2 proportions are correlated !!)

		Y		
		1	0	
X	1	π_{11}	π_{10}	$\pi_{1+} = \pi_{11} + \pi_{10}$
	0	π_{01}	π_{00}	$\pi_{0+} = \pi_{01} + \pi_{00}$
		$\pi_{+1} = \pi_{11} + \pi_{01}$	$\pi_{+0} = \pi_{10} + \pi_{00}$	$\pi_{++} = 1$

- Data: Observed counts use same notation as probabilities, except with n instead of π:

		Y		
		1	0	
X	1	n_{11}	n_{10}	$n_{1+} = n_{11} + n_{10}$
	0	n_{01}	n_{00}	$n_{0+} = n_{01} + n_{00}$
		$n_{+1} = n_{11} + n_{01}$	$n_{+0} = n_{10} + n_{00}$	$n_{++} = n$

- MLEs of all probabilities can be shown to be the sample proportions

    - For example, $\hat{\pi}_{ij} = n_{ij}/n$, $\pi_{i+} = n_{1+}/n$, etc.

Example: Prostate cancer diagnosis procedures (Adapted from the book)

Table 1 shows data from a study that was used to compare the diagnostic accuracy of magnetic resonance imaging (MRI) vs. ultrasound in patients who had been established as having "localized" prostate cancer by direct examination through surgery. We are interested in comparing the probability of a correct (i.e., localized) diagnosis by an MRI to the probability of a correct diagnosis by an ultrasound. In other words, we want to see whether one of the diagnostic methods gets the right answer more often than the other.

Both procedures are performed on tumours from the same 16 patients. Therefore, the two MLEs of the probabilities of success—10/16 and 7/16, respectively—are based on the same counts, and hence are not independent. This violates the assumption of independent binomial counts used we used previously.

3

3 Inference with Matched Pairs Binary Data

Typically, we want to do two things: Test whether the probability of success is the same in each group, and a find a confidence interval for the difference (or some other measure) between the probabilities of success.

- First, consider the difference, $\pi_{1+} - \pi_{+1}$ *main parameter*

  - A funny thing happens here: $\pi_{1+} - \pi_{+1} = (\pi_{11} + \pi_{10}) - (\pi_{11} + \pi_{01}) = \pi_{10} - \pi_{01}$
  - So the probability that was causing the two MLEs to be correlated disappears!
  - Only the two cells "off the diagonal" matter.

- This fact is used in different ways to develop tests and confidence intervals

  - Also for this reason, the difference is used for comparing probabilities, and not relative risk or odds ratios

Difference between 2 "disagree" cells.
→ off diagonal cells

3.1 Hypothesis test of equal probabilities of success: McNemar's Test[1]

- The null hypothesis that the two groups have the same probability of success, $H_0 : \pi_{1+} = \pi_{+1}$, is equivalent to $H_0 : \pi_{10} = \pi_{01}$
- Since only these two cells of the table matter, ignore the rest and look only at how the counts in these two cells are distributed

  - That is, focus on the $n_{10} + n_{01}$ observations in these two cells and pretend that $(1,0)$ and $(0,1)$ are the only two outcomes
  - Call $(1,0)$ a success (doesn't matter which one is used)
  - *The number of these observations in the (1,0) cell is a binomial RV with $n_{10} + n_{01}$ trials and probability of success $\tau = \pi_{10}/(\pi_{10} + \pi_{01})$!*

(t-owl)

- A test for $H_0 : \pi_{10} = \pi_{01}$ is equivalent to what hypotheses about τ? $\underline{\tau = 0.5}$

- Do a score test for this hypothesis:

$$Z_0 = \frac{\hat{\tau} - \tau_0}{\sqrt{\frac{\tau_0(1-\tau_0)}{n_{10}+n_{01}}}}$$

where $\hat{\tau} = n_{10}/(n_{10} + n_{01})$

  - It turns out that if you square this statistic and do a little algebra, you end up with

$$M = Z_0^2 = \frac{(n_{10} - n_{01})^2}{n_{10} + n_{01}}$$

Score test
↓
McNemar's test

If H_0 is true, people in 2 off-diagonal cells have the same probability to fall in other cell.
→ Treat as 'success and failure'!!!

---

[1] The book gives a different derivation of McNemar's test, but it results in the same thing. I think this derivation is more fun!

4

* Stat students: Try proving this!
  – Compare M to χ_1^2 (It turns out that that χ_1^2 is equivalent to square of standard normal)
    * This is McNemar's Test

Example: Prostate cancer diagnosis procedures (Lecture 6 scripts.R)

Recall the example data:

		Ultrasound		
		Localized	Advanced	
MRI	Localized	4	6	10
	Advanced	3	3	6
		7	9	16

A "localized" diagnosis is correct, so we want to compare $\pi_{1+} = P(\text{MRI} = \text{Localized})$ against $\pi_{+1} = P(\text{Ultrasound} = \text{Localized})$.

- While it is true that $\hat{\pi}_{1+} = 10/16$ and $\hat{\pi}_{+1} = 7/16$, we will not need these probabilities.

- Instead, we replace $H_0 : \pi_{1+} = \pi_{+1}$ with $H_0 : \pi_{10} = \pi_{01}$, which in turn becomes $H_0 : \tau = 0.5$ where $\tau = \pi_{12}/(\pi_{12} + \pi_{21})$

  – We have $\hat{\tau} = 6/(6+3) = 0.667$

- Then
$$M = \frac{(6-3)^2}{6+3} = 1.0$$

  – Assuming $\alpha = 0.05$ we have $\chi_{1,1-\alpha}^2 = 3.84$, so we do not reject the null hypothesis (p-value=0.32)
  – There is insufficient evidence to conclude that there is a difference in the probability of correct diagnosis between MRI and Ultrasound.

3.2 Confidence interval for $\pi_{1+} - \pi_{+1}$

- The easiest thing to do is the Wald interval, but even that requires slightly deeper application of stat theory than other intervals have so far.

- Again, we rely on the fact that $\pi_{1+} - \pi_{+1} = \pi_{10} - \pi_{01}$.

  – The ML estimate is $\hat{\pi}_{10} - \hat{\pi}_{01} = (n_{10} - n_{01})/n$

5

- Its variance has a more complicated form because we have to account for the COVARIANCE between the counts n_{10} and n_{01},[2] but a little algebra simplifies it to something relatively easy:

$$\widehat{Var}(\hat{\pi}_{10} - \hat{\pi}_{01}) = \frac{\hat{\pi}_{10} + \hat{\pi}_{01} + (\hat{\pi}_{10} - \hat{\pi}_{01})^2}{n}$$

- A Wald interval is therefore

$$\hat{\pi}_{10} - \hat{\pi}_{01} \pm Z_{1-\alpha/2}\sqrt{\widehat{Var}(\hat{\pi}_{10} - \hat{\pi}_{01})}$$

- A score interval can be computed, but of course has no formula and must be found using iterative numerical methods. It is available in the `scoreci.mp()` function in the `PropCIs` package.

Example: Prostate cancer diagnosis procedures (Lecture 6 scripts.R)

Recall the example data:

		Ultrasound		
		Localized	Advanced	
MRI	Localized	4	6	10
	Advanced	3	3	6
		7	9	16

$\pi_{1+} = P(\text{MRI} = \text{Localized})$ against $\pi_{+1} = P(\text{Ultrasound} = \text{Localized})$. We find a 95% confidence interval for $\pi_{10} - \pi_{01}$:

- Wald Interval:

$$\frac{6}{16} - \frac{3}{16} \pm Z_{1-\alpha/2}\sqrt{\frac{\frac{6}{16} + \frac{3}{16} + (\frac{6}{16} - \frac{3}{16})^2}{16}} = (-0.19, 0.57).$$

So we are 95 confident that the interval from -0.19 to 0.57 covers the true difference in probability of correct diagnosis between MRI and Ultrasound. In particular, note that the interval contains 0, so that we cannot rule out the possibility that there is no difference in probability of correct diagnosis between the two diagnostic methods.

- The score interval is shown in the program and yields in interval from -0.18 to 0.51 with the same conclusion.

---

[2] Covariance is related to correlation. It comes into the problem here because we are fixing n and distributing the observations among four possible outcomes. Every time a count goes into one outcome, there is one less possible count to go into the other outcomes, so there is a negative correlation between the counts of the outcomes. This affects the variance when sums or differences are taken of the counts. Simply, if two RVs X_1 and X_2 are correlated, and c_1 and c_2 are two constants $Var(c_1X_1 + c_2X_2) = c_1^2 Var(X_1) + c_2^2 Var(X_2) + 2c_1c_2 Cov(X_1, X_2)$, where $Cov(,)$ is the covariance between the two listed RVs. In

6

4 Notes

1. In general, when we have more information about the structure of data, we can try to exploit that knowledge to make statistical models more accurate

   (a) The closer a model is to the correct structure, the better the inferences are likely to be

5 Conclusions: What to learn from this

1. Measuring multiple variables in the same "subject" is common

2. Comparison of proportions from paired data are done differently from those on independent groups

   (a) Comparison depends only on those subjects where the two measurements gave different results

   (b) Treat these subjects as Bernoulli trials, with success representing either of the two outcomes (0,1) or (1,0)

   (c) Another score test!

3. Cross-tabulation and notation representing counts and probabilities will show up a few more times

   (a) So will "marginal" and "joint" probability—learn it!

7

6 Exercises (due at time to be announced)

Compete the following exercises from B&L, Chapter 1:

1. Exercise 23. In (b), use the score confidence interval instead of the Agresti-Min interval requested.

8