Comparing groups – categorical data

Dr. Myitzu Tin Oung
M.B.,B.S.; MA; MPH; PhD
Deputy Director
Medical Statistics Division
DMR (POLB)

Data Management Workshop, July 2019

Bivariate analysis

Choosing the correct test in the assessment of association between two variables depends on

1. the type of risk factor and outcome variable you have
2. whether the data are unpaired or paired/matched

Categorical variables (Unpaired)

Type of variables : Dichotomous, Nominal & Ordinal

Outcome (Dependent variable)				
		Dichotomous	**Nominal**	**Ordinal**
Risk factor (Independent variable)	**Dichotomous**	Chi-squared, Fisher's exact, risk ratio, odds ratio	Chi-squared	Chi-squared for trend
	Nominal	Chi-squared	Chi-squared	
	Ordinal	Chi-squared for trend		

Categorical variables (Paired)

Scale of measurement: Dichotomous

	Paired observations (2 observations)	Repeated observations (≥ 3 observations)
Dichotomous	McNemar's test	Cochran's Q

Outline

1. Chi-squared test for association
2. Fisher's exact test
3. Chi-squared test for trend
4. McNemar's test
5. Cochran’s Q

Categorical variables (Unpaired)

Chi-squared test for association

- Types of variables are nominal or dichotomous
- Used to determine whether a relationship exists in a 2 x 2 contingency table or R x C table

Two by two contingency table

	Death		
Diabetes	**Yes**	**No**	**Total**
Yes	**47**	**679**	**726**
No	**158**	**4048**	**4206**
Total	**205**	**4727**	**4932**

R by C (row by column) table

Ethnicity	Poor glycemic control (HbA_{1c} > 10%)		Total
	Yes	No	
African-American	2379	6117	8496
Asian	1679	5953	7632
Latino	1695	4584	6279
Caucasians	7205	32820	40025
Total	12958	49474	62432

Assumptions for Chi-squared

1. Random sampling
2. The values must be number of subjects actually observed not percentages or rates.
3. The categories must be mutually exclusive.
4. No more than 20% of cell should have the expected frequencies less than 5 (for 2 x 2 table, none of the cells should have expected frequencies less than 5)
5. Independent samples
6. Independent observations

Chi-squared test

E.g. Association between diabetes and death

Null hypothesis (Ho):

Presence of diabetes is not associated with death.

Alternative hypothesis (Ha):

Presence of diabetes is associated with death.

Chi-squared test

The test is based on the relative ratio of observed frequencies (O) and expected frequencies (E).

Chi-squared test

Expected frequency can be calculated for each cell of the table if H_o were true

$$\text{Expected Frequency} = \frac{\text{Column Total} \times \text{Row Total}}{\text{Grand Total}}$$

Chi-squared test

$$\text{Expected Frequency} = \frac{\text{Column Total x Row Total}}{\text{Grand Total}}$$

Diabetes	Death					
	Yes		No		Total	
Yes	47	14.7%	679	14.7%	726	14.7%
No	158	85.3%	4048	85.3%	4206	85.3%
Total	205		4727		4932	

Chi-squared test

$$\text{Expected Frequency} = \frac{\text{Column Total x Row Total}}{\text{Grand Total}}$$

	Death		
Diabetes	**Yes**	**No**	**Total**
Yes	**47 (30)**	**679 (696)**	**726**
No	**158 (175)**	**4048 (4031)**	**4206**
Total	**205**	**4727**	**4932**

Chi-squared test

$$\chi^2_{(r-1)(c-1)} = \frac{\Sigma (O_i - E_i)^2}{E_i}$$

$$= \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \dots + \frac{(O_k - E_k)^2}{E_k}$$

Degree of freedom = df = (r – 1) (c– 1)

When calculated χ^2 > critical value of χ^2 from χ^2 Table with df = (r – 1) (c – 1) at 95% CI (i.e., $\alpha = 0.05$), reject H_o and will conclude that there is relationship between the two factors at 95% CI ($p < 0.05$)

DIY

Chi-squared test

E.g. Association between ethnicity and poor glycemic control (HbA1c > 10%) among persons with diabetes

Null hypothesis (Ho):

Ethnicity is not associated with poor glycemic control.

Alternative hypothesis (Ha):

Ethnicity is associated with poor glycemic control.

Chi-squared test

Nominal variable

Dichotomous variable

Table 5.11. Association between ethnicity and poor glycemic control

	Poor glycemic control ($HbA_{1c} > 10\%$)	
	Yes	No
African-American	2379 (28)	6117 (72)
Asian	1679 (22)	5953 (78)
Latino	1695 (27)	4584 (73)
Caucasians	7205 (18)	32,820 (82)

Values are represented as *n* (%).

$\chi^2 = 612; P < 0.0001.$

Data from Karter, A.J., et al. Ethnic disparities in diabetic complications in an insured population. *J. Am. Med. Assoc.* 2002; 287: 2519–27.

Chi-squared test

Table 5.11. Association between ethnicity and poor glycemic control

	Poor glycemic control ($HbA_{1c} > 10\%$)	
	Yes	No
African-American	2379 (28)	6117 (72)
Asian	1679 (22)	5953 (78)
Latino	1695 (27)	4584 (73)
Caucasians	7205 (18)	32,820 (82)

Values are represented as *n* (%).

$\chi^2 = 612$; $P < 0.0001$.

Data from Karter, A.J., et al. Ethnic disparities in diabetic complications in an insured population. *J. Am. Med. Assoc.* 2002; 287: 2519–27.

- the differences in glycemic control across ethnicities are unlikely to have occurred by chance.

- does *not* tell which groups are significantly different from one another.

Chi-squared test

A group of patients (480) with bronchopneumonia were treated with either amoxicillin or erythromycin.

Null hypothesis (Ho):
The treatment effect of amoxicillin and erythromycin on bronchopneumonia is not different.

Alternative hypothesis (Ha):
The treatment effect of amoxicillin and erythromycin on bronchopneumonia is different.

Table 3 Comparison of effect of treatment of bronchopneumonia with amoxicillin or erythromycin

Antibiotics given	Effect of treatment: Improvement at 5 days	Effect of treatment: No improvement at 5 days	Total
Amoxicillin	144 (60)	96 (40)	240 (100)
Erythromycin	160 (67)	80 (33)	240 (100)
Total	304	176	480

Values are represented as n (%)

$\chi^2 = 2.3$; $P = 0.13$

A chi-squared statistics was used to examine the effect of two different antibiotics on bronchopneumonia. The treatment effect of amoxicillin and erythromycin on bronchopneumonia was not different, $\chi^2 = 2.3$, $df = 1$, $N = 480$, $p > 0.05$.

Chi-squared test

Do male and female differ on whether they take biostatistics course or not?

Null hypothesis (Ho):
Taking biostatistics course between male and female is not different.

Alternative hypothesis (Ha):
Taking biostatistics course between male and female is different.

Chi-squared test

Gender	Biostatistics course		
	Taken	Not taken	Total
Male	24 (71)	10 (29)	34 (100)
Female	12 (29)	29 (71)	41 (100)
Total	36	39	75

Values are represented as n (%)

$\chi^2 = 12.714$; $P = 0.000$

A chi-squared statistics was used to investigate whether males and females differ on whether they take biostatistics or not. The Chi-squared result indicates that male and females are significantly different on whether they have taken or have not taken biostatistics, $\chi^2 = 12.71$, $df = 1$, $N = 75$, $p < .001$. Males are more likely to take biostatistics than females.

Chi-squared test

- Chi-squared statistic is **not an index of the strength** of the association.
- The chi-squared test is **valid if at least 80%** of the expected frequencies **exceed 5** and **all** the expected frequencies **exceed 1**.
    - ➢ If the criterion is not satisfied, we can usually **combine or delete rows and columns** to give bigger expected values.
    - ➢ This **cannot be done for 2 by 2 tables**

Exercise

Fisher's exact test

- **When the sample is not large and expected values are less than 5, we can turn to an exact distribution called Fisher's exact test.**
- **However, Fisher's exact test can be done for any 2×2 table**

Fisher's exact test

1. Calculations are messy

2. Shouldn't try to do them manually

3. Use a computer program

Assumptions for Fisher's exact test

1. Random sampling
2. The values must be number of subjects actually observed not percentages or rates.
3. Both variables are dichotomous (i.e. 2 x 2 table).
4. The categories must be mutually exclusive.
5. Fisher's test can be used for any number of subjects
6. Independent observations
7. Independent samples

Fisher's exact test

Villar and colleagues investigated the cause of a botulism outbreak among bus drivers in Argentina. One of the foods they investigated is *matambre*, a traditional meat dish of Argentina that is cooked at temperatures too low to kill *Clostridium botulinum* spores.

Is eating *matambre* associated with botulism?

Fisher's exact test

Data for consumption of matambre and botulism

	Botulism		
Ate *matambre*	Yes	No	Total
Yes	9	2	11
No	0	10	10
Total	9	12	21

Ate matambre – Botulism (yes) =

Ate matambre – Botulism (no) =

Don't eat matambre – Botulism (yes) =

Don't eat matambre – Botulism (no) =

Expected frequency

Fisher’s exact test

Table 5.8. Actual data showing association between consumption of matambre and botulism

Ate *matambre*	Botulism		
	Yes	No	Total
Yes	9 (82)	2 (18)	11 (52)
No	0 (0)	10 (100)	10 (48)
Total	9	12	21 (100)

Values are represented as *n* (%).
Fisher’s exact test (two-tailed) = 0.0002.
Data from Villar, R.G., et al. Outbreak of Type A botulism and development of a botulism surveillance and antitoxin release system in Argentina. *J. Am. Med. Assoc.* 1999; 281: 1334–8, 1340.

There is a relationship between eating *matambre* and developing botulism. It was significant at $p<.001$.

Limitations

- A limitation of both the chi-squared and Fisher's exact test - they ***do not measure the strength of the association*** between the risk factor and the outcome.

- Only tell whether the ***relationship is statistically significant*** (not likely to be due to chance)

Choosing between Chi-squared and Fisher's exact

- 2 x 2 contingency table
- Depends on the size of samples

Choosing between Chi-squared and Fisher's exact

Size of sample	Suggested test
Small samples	• Use Fisher's test
Medium-sized samples	• Use Fisher's test if you use computer program (exact P value) • Use Chi-square if you can't access computer
Large samples	• Use Chi-squared (give almost identical P value)

Chi-squared test for trend

- **Assesses whether there is an increasing (or decreasing) linear trend in the proportion of subjects at each level of the ordinal variable.**

Chi-squared test for trend

- The standard calculations for chi-square do not take into account any order among the rows or columns.
- If your experiment was designed so that the rows/columns have a distinct order (i.e., time points, ages, or doses), this information is not used in the standard chi-square calculations.

Chi-squared test for trend

Null hypothesis (Ho):

- There is no linear trend between two variables

Alternative hypothesis (Ha):

- There is a linear trend between two variables

Chi-squared test for trend

Cough during the day or at night and cig smoking by 12-year-old boys

	Boy's smoking		
	Non-smoker	**Occasional**	**Regular**
Cough	266 (20.4%)	395 (28.8%)	80 (46.5%)
No cough	1037 (79.6%)	977 (71.2%)	92 (53.5%)
Total	1303 (100.0%)	1372 (100.0%)	172 (100.0%)

Chi-squared test for trend = 59.47; $p < 0.001$

Chi-squared test for trend

- Persons over the age of 70 years were randomized to an intervention designed to increase the independence or to usual care.

- Main outcome was "change in patients' ability to perform basic activities of daily living from admission to discharge" (much worse, worse, unchanged, better, much better)

Assess the efficacy of the intervention in increasing the independence of elderly persons.

Chi-squared test for trend

Table 5.20. Functional changes in elders' ability to perform basic activities of daily living

Change from admission to discharge	Intervention group	Usual care
Much worse	26 (9)	25 (8)
Worse	22 (7)	39 (13)
Unchanged	151 (50)	163 (54)
Better	39 (13)	33 (11)
Much better	65 (21)	40 (13)

Values are represented as *n* (%).

P-value for chi-squared for trend = 0.009.

Data from Landefeld, C.S., et al. A randomized trial of care in a hospital medical unit especially designed to improve the functional outcomes of acutely ill older patients. *New Engl. J. Med.* 1995; 332: 1338–44.

In comparing the effects of two groups on functional changes in elders' ability to perform basic activities of daily living by using chi-squared test for trend, the intervention group is more likely to be better or much better than the usual-care group. The linear trend is significant with $p = 0.009$.

Chi-squared test for trend

1. The trend may be significant even if the contingency table chi-squared is not. (The test for trend has greater power for detecting trends than has the ordinary chi-squared test.)
2. The trend test would not detect if there had an association where those who were occasional smokers had far more symptoms than either non-smokers or regular smokers

If the hypothesis we wish to test involves the order of the categories, we should use the chi-squared test for trend.

Chi-squared test for trend

- The test should not be used with tables with small cell counts.
- As long as there are at least 30 observations the approximation should be valid.

Categorical variables (Paired)

McNemar's test

- **Paired observations (i.e., before and after) of a dichotomous variable**

Assumptions for McNemar's test

- Observations are dichotomous, categories are mutually exclusive
- Dichotomous measures are paired observations of the same subjects or matched pairs (Data contained in the cells of table are number of pairs)

	Severe colds at age 14		
Severe colds at age 12	**Yes**	**No**	**Total**
Yes	212	144	356
No	256	707	963
Total	468	851	1319

McNemar's test

Null hypothesis:

- There is no change from pre test condition to post test condition

  (*the numbers of discordant cells are equal*)

Alternative hypothesis:

- There is a change from pre test condition to post test condition

  (*the numbers of discordant cells are not equal*)

McNemar's test

If the resultant $\chi2$ is significant,

- there was a change from pre test condition to post test condition

or

- there was an association between treatment and the effect

McNemar's test

E.g. To determine the likelihood that long-term acid suppression would result in gastritis (Kuipers and colleagues, 1996)

- 59 patients with gastroesophageal reflux disease and Heliobacter pylori infection
- All patients were treated - an acid suppresser (omeprazole)
- Average duration of treatment - 5 years
- Follow up

McNemar's test

Table 5.23. Presence of gastritis among patients with gastroesophageal reflux and *H. pylori* infection

	Follow-up		
Baseline	Normal	Gastritis	Total
Normal			24 (41%)
Gastritis			35 (59%)
Total	11 (19%)	48 (81%)	59 (100%)

P = 0.007 by McNemar's test.

Data from Kuipers, E.J., et al. Atrophic gastritis and *Helicobacter pylori* in patients with reflux esophagitis treated with omeprazole or fundoplication. *New Engl. J. Med.* 1996; 334: 1018–22.

21

10.5

10.5

P-value associated with a two-tailed Fisher's exact test is 0.10

McNemar's test

E.g. Holland *et al.* (1978) obtained respiratory symptom questionnaires for 1319 Kent schoolchildren at ages 12 and 14. At age 12, **356 (27%)** children were reported to have had severe colds in the past 12 months compared to **468 (35%)** at age 14.

Investigate whether the prevalence of reported symptoms was different at the two ages

McNemar’s test

Table 13.13. Severe colds reported at two ages for Kent schoolchildren (Holland et al. 1978)

Severe colds at age 12	Severe colds at age 14		
	Yes	No	Total
Yes	212	144	356
No	256	707	963
Total	468	851	1319

p= 0.000 by McNemar’s test

There was a significant difference in reported severe colds in the past 12 months between the two ages, 12 and 14 ($p<0.001$).

Cochran's Q

- **Multiple observations (≥3) of a dichotomous variable on the same subjects**

Cochran's Q

- Use the test when a group of people perform a series of tasks or getting a set of treatments where the outcome is dichotomous, a "Success" or "Failure" or "Yes" or "No".

Null hypothesis:

- The proportion of "successes" is equal for all groups.

Alternative hypothesis:

- The proportion of "successes" is different for at least one group.

Assumptions in Cochran's Q test

- Random sampling
- Dichotomous dependent variable
- Independent variable with ≥ 3 groups or levels
- The groups/levels must be mutually exclusive
- The sample size is sufficiently "large".

(As a rule of thumb, $n \geq 4$; $nk \geq 24$ (the number of subjects, n, multiplied by the number of levels in the independent variable, k)

Cochran's Q

E.g. Assessed **changes in behaviors associated with HIV transmission**

- 1256 HIV-negative men who have sex with men
- Subjects were assessed **at baseline, at 6, 12, and 18 months.**
- Calculate Cochran's Q-test for each of the four behaviors.

Cochran's Q

Table 5.24. Changes in behaviors associated with HIV transmission*

	Baseline (%)	6 months (%)	12 months (%)	18 months (%)	*P*-value
HIV-positive sex partner	20	18	18	10	<0.001
Unprotected receptive anal sex	32	27	28	29	0.02
Condom failure	19	13	10	12	<0.001
Urethritis	9	3	2	2	<0.001

* Behaviors are coded as yes or no.

Data from Buchbinder, S.P., et al. Feasibility of human immunodeficiency virus vaccine trials in homosexual men in the United States: risk behavior, seroincidence, and willingness to participate. *J. Infect. Dis.* 1996; 174: 954–61.

There were significant differences in the percentages of HIV-negative men who have sex with men engaging in the four behaviors associated with HIV transmission over time.

Cochran's Q

- The test will tell if there is a difference.
- But it won't tell where those differences are.
- If the null hypothesis is rejected (i.e. the test identifies differences), perform a follow-up pairwise Cochran's Q tests to identify the areas which have differences.

Summary

		Outcome (Dependent variable)		
Risk factor (Independent variable)		**Dichotomous**	**Nominal**	**Ordinal**
	Dichotomous	Chi-squared, Fisher's exact, risk ratio, odds ratio	Chi-squared	Chi-squared for trend
	Nominal	Chi-squared	Chi-squared	
	Ordinal	Chi-squared for trend		

Summary

	Paired observations (2 observations)	Repeated observations (≥ 3 observations)
Dichotomous	McNemar's test	Cochran's Q

References

- Mitchell H. Katz (2011). Multivariable analysis: a practical guide for clinicians and public health researchers (3rd Edition), the United States of America by Cambridge University Press, New York.
- Bland & Martin (2000). Introduction to Medical Statistics (3rd Edition), Oxford University Press.
- HarveyM otulsk (1995). Intuitive Biostatistics, Oxford University Press.

THANK YOU!

Photos credit: Dr. Ei Ei Swe

Group work: Exercises