University of New Mexico
Hypothesis Testing-5 (Fall 2016)
PH 538: Public Health Biostatistical Methods I (by Fares Qeadan)

---

hypothesis testing for association between two categorical variables with small sample sizes: (Fisher-Exact Test of Independence)

The assumptions of this test are:
The Fisher's exact test is used when you want to conduct a chi-square test, but more than 20% of the expected cell frequencies < 5. Remember that the chi-square test assumes that each cell has an expected frequency of five or more, but the Fisher's exact test has no such assumption and can be used regardless of how small the expected frequency is. The name of this test implies that the chi-square test is an approximate method. If the number of rows, number of columns, or total sample size become too large, the program you're using may not be able to perform the calculations for Fisher's exact test in a reasonable length of time, or it may fail entirely.

The null and alternative hypotheses are:

H_0 : In the population, the two categorical variables are independent.
H_1 : In the population, two categorical variables are dependent.

(1) **Problem 1:** Smith, Delgado and Rutledge (1976) report data on ovarian carcinoma. Individuals had different numbers of courses of chemotherapy. The 5-year survival data for those with 1-4 and 10 or more courses of chemotherapy are provided in the following table.

		Five Year Status		
		Dead	Alive	Total
Courses	$1-4$	21	2	23
	≥ 10	2	8	10
	Total	23	10	33

In this case our hypotheses can be stated as follows:

H_0 : Survival status among patients with ovarian carcinoma is independent from the numbers of courses of chemotherapy.
H_1 : Survival status among patients with ovarian carcinoma is dependent on the numbers of courses of chemotherapy.

Please test the null hypothesis, at the 5% significance level.

(a) The significance level α is:

(b) The null and alternative hypotheses are:

(c) The decision rule (about H_0) is:

(d) Conduct the test in STATA:

```
tabi 21 2 \ 2 8, row all exact expected
```

(e) Get the p-value from STATA's output:

(f) Decision:

(g) Conclusion:

(h) Were the assumptions of this test satisfied? Explain.

hypothesis testing for Comparing Proportions with Paired data (McNemar Test):
[Test for Two Correlated Proportions]

The McNemar test is a test on a 2x2 classification table when you want to test the difference between paired proportions. This test is generally used for

- Pair-Matched data : Pair-Matched data can come from Case-control studies where each case has a matching control (matched on age, gender, race, etc.)
- Before - After data: the outcome is presence (+) or absence (-) of some characteristic measured on the same individual at two time points.
- Comparison of sensitivity and specificity between two diagnostic tests, each measured on the same patient.

Notes:
1. The rule-of-thumb for the McNemar test version of the chi-square test is that when the discordant pairs are less than 10, the exact form of the test should be used.
2. Unfortunately, in STATA's output, the variables are labeled cases and controls, which is rather confusing.

(2) **problem 2 (Before - After data):** Fifty-three study participants assessed twice for plaque index (PI), at baseline and 4 weeks later. We wish to assess whether the proportion of patients with high PI changes. Conduct the test at the 5% significance level?

		PI at 4 weeks		
		low	high	Total
PI at baseline	low	29	1	30
	high	13	10	23
	Total	42	11	53

(a) The significance level α is:

(b) The null and alternative hypotheses are:

H0: P(PI high at baseline) = P(PI high at 4 week)
H1: P(PI high at baseline) $\neq$ P(PI high at 4 week)

(c) The decision rule (about H_0) is:

(d) Conduct the test in STATA:
```
mcci 29 1 13 10
```

(e) Get the p-value from STATA's output:

(f) Decision:

(g) Conclusion:

(3) **problem 3 (Matched Case-Control Study):** A study was carried out on post-menopausal women in City A. Cases of women with endometrial cancer were identified from this city. A control group was selected matched to the case on age and length of residence in city A. The medical question was whether endometrial cancer was related to estrogen use.

		Control		
		Estrogen	No estrogen	Total
Cases	Estrogen	27	29	56
	No estrogen	3	4	7
	Total	30	33	63

(a) The significance level α is:

(b) The null and alternative hypotheses are:

H0: There is no association between estrogen use and endometrial cancer
H1: There is an association between estrogen use and endometrial cancer

OR

H0: P(Estrogen among cases) = P(Estrogen among controls)
H1: P(Estrogen among cases) $\neq$ P(Estrogen among controls)

(c) The decision rule (about H_0) is:

(d) Conduct the test in STATA:

```
mcci 27 29 3 4
```

(e) Get the p-value from STATA's output:

(f) Decision:

(g) Conclusion:

(h) In this problem, STATA gives the following measures in the output:
Cases: the estimated risk of being exposed among cases.
Controls: the estimated risk of being exposed among controls.
Rel. diff: the estimated risk difference of being exposed relative to the risk of being unexposed among controls.
odds ratio: the odds of endometrial cancer is approximately 10 times greater for women who were on estrogen therapy compared to those who were not.