Module 10: Confounding and MH Method

With categorical data, we are classifying data instead of measuring it. As a review:

A Review of Statistical Procedures:
When to do What?

Study Goal:	Type of Outcome Data: Measurement Data	Dichotomous Data	Survival Data	Messy Measurement Data (non-normal)
Describe a population	CI for mean One sample t-test	CI for proportion One sample test proportion/chi-square goodness of fit	Survival Curve (Kaplan-Meier)	Median
Compare two independent groups	Two sample t-test CI for difference in means	z-test for proportion Chi-square test CI for RD, RR, OR Fisher's Exact test	Log rank test	Wilcoxon rank sum test (Mann-Whitney test)
Compare two dependent groups	Paired sample t-test CI for difference in means	McNemar's paired sample chi-square		Wilcoxon signed rank test
Compare more than two groups	Analysis of variance	Chi-square test	Log rank test	Kruskal-Wallis test
Association between variables	Correlation Linear regression	Chi-square test CI for RR, OR	Cox's proportional hazards regression	Spearman's rank correlation
Multivariable analyses (multiple predictors, adjust for confounding)	Multiple linear regression (Analysis of Covariance)	Multiple logistic regression	Cox proportional hazards regression	

Notice we never use a z test, a t test is almost always more appropriate even for large samples. Likewise, for dichotomous outcomes could use a z-test but a chi-square test is usually used in practice. Chi-square reflects categorical outcomes.

chi-square = sum((obs-exp)^2 / exp), df=n-1; where n is the number of random variables or categories.

Mantel-Hansel Method

Cochran-Mantel-Haenszel method is a technique that generates an estimate of an association between an exposure and an outcome after adjusting for or taking into account confounding. We stratify the data into two or more levels of the confounding factor (as we did in the example above). In essence, we create a series of two-by-two tables showing the association between the risk factor and outcome at two or more levels of the confounding factor, and we then compute a weighted average of the risk ratios or odds ratios across the strata

	Disease?		
Exposure?	Yes	No	Total
Yes	a	b	r_1
No	c	d	r_2
Total	c_1	c_2	n

Given the above table, we have the below MH Equations:

For Cell(1,1) frequency has

$$E_0(a) = r_1 c_1 / n$$

and

$$Var_0(a) = r_1 r_2 c_1 c_2 / n^2(n-1)$$

For a given table, and the MH test statistic is:

$$\chi^2_{MH} = \frac{\left(\sum a - \sum(r_1 c_1 / n)\right)^2}{\sum \frac{r_1 r_2 c_1 c_2}{n^2(n-1)}}, \text{1df}$$

It follows a Chi-Square distribution with 1 degree of freedom.

We can also derive the following:

Cochran-Mantel-Haenszel Estimate for a Risk Ratio

$$\widehat{RR}_{cmh} = \frac{\sum \frac{a_i(c_i+d_i)}{n_i}}{\sum \frac{c_i(a_i+b_i)}{n_i}}$$

Cochran-Mantel-Haenszel Estimate for an Odds Ratio

$$\widehat{OR}_{cmh} = \frac{\sum \frac{a_i d_i}{n_i}}{\frac{b_i c_i}{n_i}}$$

Though, in practice we just have the computer solve for MH estimates.

Continuity Corrections

Chi-Squared distributions are continuous whereas some variables are categorical, therefore we must correct by adding or subtracting .5. The "corrected" result is more accurate. This is often used when there are small samples or 0's in some cells.

$$X_c^2 = \sum_{i,j} \frac{(|x_{ij} - e_{ij}| - 0.5)^2}{e_{ij}}$$

Fisher's Exact Test

When a single epected value is small (n<5) a continuity correction only helps so much. In such cases we use Fisher's Exact test. Prepare your 2x2 table with the cell containing the smallest number in the upper left corner. Then identify all possible tables with that cell the same as or more extreme than observed and sum the probabilities. This gives the probability of that specific table.

Pr(a,b,c,d) = [(a+b)!(c+d)!(a+c)!(b+d)!] / N!a!b!c!d!

Again, this is usually performed by a computer. In R: fisher.test(Count)

Revision #5
Created 26 August 2022 14:17:51 by Elkip
Updated 27 August 2022 19:32:20 by Elkip