Biostatisics

Lecture14

Analysis Techniques for Epidemiologic Studies and Survival Analysis

Odds Ratios

- Odds ratios used extensively in epidemiology to express the odds of developing a condition due to the presence of a factor compared to the absence of the factor
  - Odds for success are the ratio of the probability of success to the probability of failure
  - Odds ratio is the ratio of the odds of a disease with a specific factor to the odds of a disease without the factor
- Thus, an odds ratio of 2.0 indicates that a person with the specific factor has twice the odds of contracting the disease than a person without that factor

Relative Risk

- From a prospective study
- Risk of developing disease among those with risk factor is a / (a+b) and among those without risk factor c / (c+d)

Risk Factor	Disease Present	Disease Absent	Total at risk
Present	a	b	a+b
Absent	c	d	c + d
Total	a + c	b + d	n

Relative Risk

- Relative risk is defined as

$$RR = \frac{a/(a+b)}{c/(c+d)}$$

  - If RR = 1, no difference in risk of developing disease
  - If RR > 1, greater risk among those with risk factor
  - If RR < 1, greater risk among those without risk factor
  - Can construct CI and, thus, can test RR for difference from 1.0

Odds Ratio

- From a retrospective study – case-control
- Odds of being a case to being a control among those with risk factor is a/b and odds of being a case to being a control among those with out the risk factor is c/d

Risk Factor	Cases (disease)	Controls (no disease)	Total
Present	a	b	a + b
Absent	c	d	c + d
Total	a + c	b + d	n

Odds Ratio

- Odds ratio is defined as
  - OR = ad / bc
  - Interpretation similar to RR
  - Can construct CI and, thus, can test OR for difference from 1.0

Mantel-Haenszel Statistic

- Used to combine data across i strata to provide an overall test
- Uses the observed and expected frequency for the cell for cases with the risk factor within each stratum and then sums across strata to test for any differences

Risk Factor	Cases (disease)	Controls (no disease)	Total
Present	a(i)	b(i)	a(i) + b(i)
Absent	c(i)	d(i)	c(i) + d(i)
Total	a(i) + c(i)	b(i) + d(i)	n(i)

Mantel-Haenszel Statistic

- Different than homogeneity of effect because gives an overall test of frequencies
- Used to analyze both retrospective and prospective studies
- Can also get an estimate of the common odds ratio

Survival Analysis

- Used for time to ‘death’ analyses with predictor factors
  - e.g., compare time to death for bladder cancer patients with single vs multiple tumors
  - For long-term conditions, time to event analyses are more useful than straight event analyses
  - Other areas: time to second MI, time to obesity, time to diabetes, time to hypertension
- Time can be measured in hours, days, weeks, months, years

Survival Analysis

- Estimates hazard rate, conditional probabilities of survival taking into account ‘censored’ data
- Use the Mantel-Haenszel chi-square (log rank) and Wilcoxon statistic to test for differences between the two survival curves
- All subjects start at time 0 and are followed from that point forward
- Two major life table procedures
  - Kaplan-Meier and Cutler-Ederer

Life Tables

- Cutler-Ederer
  - Grouped data
  - Useful if only know an interval for ‘death’
    - Interval between visits
  - Also useful if there are large numbers of people with large numbers of ‘deaths’
- Kaplan-Meier product limit method
  - Each event is its own ‘interval’
  - Non-parametric procedure
  - At each event time, calculate the probability of surviving given that one person just ‘died’: n-1 / n
  - Multiply the new probability by the cumulative survival probability

Cox Proportional Hazards Model

- Essentially the multivariate approach to life table analysis
- Models the hazard function (probability of an event at a time just larger than t, conditional on surviving to time t) with predictor variables
- $H(t_i) = h0(t_i) \exp(\beta_1 z_{i1} + \beta_2 z_{i2} + \ldots + \beta_k z_{ki})$
- Interpretation of regression coefficients is same as of linear regression coefficients
  - Coefficients represent effect of that covariate on the hazard function

Linear Contrasts

- While ANOVA F-tests are used to compare all groups simultaneously, frequently we want to look at a set of groups vs. another set or a placebo
  - In NOPAIN, one analysis could be the combined active treatment groups compared to the placebo
  - In a factorial study, one analysis could be the groups who were given treatment x either alone or in conjunction with treatment y compared to those who did not receive treatment x

Linear Contrasts

- Linear contrasts are essentially weighted combinations of group means

$$L = \sum_{i=1}^{k} c_i \bar{y}_i \quad where \quad \sum_{i=1}^{k} c_i = 0$$

- Comparison of two means is a special case of the linear contrast

$$Var(L) = s^2 \sum_{i=1}^{k} c_i^2 / n_i$$

- where s^2 is the Within MS from one-way ANOVA

Testing Linear Contrasts

- Using these results, the testing procedure for linear contrasts is a t-test with the test statistic

$$t = \frac{L}{\sqrt{s^2 \sum_{i=1}^{k} \frac{c_i^2}{n_i}}}$$

- Which is distributed as a t-statistic at 1-α/2 level and n-k degrees of freedom