Biostatistics

Lecture 9

Hypothesis Testing: Categorical Data

Two Sample Test of Proportions

- Normal Theory Method
  - Extension of test of means

$$z = \frac{\left|\hat{p}_1 - \hat{p}_2\right| - \left[\frac{1}{2n_1} + \frac{1}{2n_2}\right]}{\sqrt{\hat{p}\hat{q}\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

  - where the new feature is the continuity correction for the normal approximation to the binomial

Analysis of frequencies with contingency tables

- Analysis of n-way tables
  - Ex.: is age distribution similar in black and white girls in NGHS
  - Ex.: does income differ by race and does # parents effect this relationship
- Depends on the chi-square distribution
  - Tests how close the observed frequencies in each cell are to the expected frequencies

Chi-Square Distribution

- Distribution of the sum of the differences between (observed and expected frequencies)^2 divided by the expected
- Equivalent to the square of the z-statistic
  - i.e., $\chi^2_{[1]} = ((y - \mu) / \sigma)^2 = z^2$

Types of Chi-square Tests

- Tests of independence
  - e.g., is there a relationship between treatment and outcome
- Tests of homogeneity
  - e.g., is the relationship between treatment and outcome the same across gender
- Tests of goodness of fit
  - e.g., does the frequency of education follow a normal distribution

Type of frequencies

- Observed frequencies
  - Frequencies of each combinations of data values in a sample
    - e.g., number of girls who are black and from a middle income household
  - Frequencies tabulated and presented in a contingency table
- Expected frequencies
  - Frequencies that we would expect for each combination of data values in a sample
    - Calculated by multiplying the two marginals and dividing by the total

Chi-square statistic

- $\chi^2 = \sum \left[\frac{(O_i - E_i)^2}{E_i} \right]$ for each of i cells

- Distribution tabled in Table 6
- Reject for high values of chi-square only
- Degrees of freedom determined by (r-1)*(c-1)

Fisher's Exact Test

- Used when there are small sample sizes in at least one cell
- Test for independence in a 2x2 table (extended to rxc tables)
- Gives the exact p-value for the result (or more extreme) where the chi-square test is an approximation
- Today, can be used in virtually any situation, not just for small sample sizes
- Limitations on the chi-square test: not good when $n < 20$ or when $20 <= n <= 40$ and one cell size $<= 5$

Fisher's Exact Test

- Computationally, Fisher's Exact Test is:

Status	Factor	No factor	Total
Alive	a	b	a+b
Dead	c	d	c+d
Total	a+c	b+d	n

$$\frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

Fisher’s Exact Test

- Gives us the probability for only the observed table.
  - We need the probability of that table and all tables more extreme to be consistent with the approach to hypothesis testing
  - Use the hypergeometric distribution to test this

McNemar's Test for Matched-Pair Data

- Standard chi-square test assumes that the frequencies are from independent samples
- If data are not independent but have a correlation structure, then use McNemar's test
- Typical use is in a matched-pair study design where subjects are matched on a variety of factors so that the members of each pair are very similar with the exception of the factor under study

McNemar's Test for Matched-Pair Data

- Form a contingency table of the pairs to represent the concordant and discordant pairs
  - Concordant pairs have the same outcome
  - Discordant pairs have different outcomes
  - Test statistic (1 df) is

$$\chi^2 = \left(\left|n_A - n_B\right| - 1\right)^2 / (n_A + n_B)$$

  - Reject for high values of chi-square

Sample Size Estimates for Binomial Proportions

- Sample size required to to compare two binomial proportions using a two-sided test with specified alpha and power

$$n_1 = \left[\sqrt{\bar{p}\bar{q}(1+1/k)}\, z_{1-\alpha/2} + \sqrt{p_1 q_1 + \frac{p_2 q_2}{k}}\, z_{1-\beta} \right]^2 / \Delta^2$$

- and $n_2 = kn_1$
- and $\Delta = abs(p_2 - p_1)$

Sample Size Estimates for Matched-Pair Data

- Sample size required to to compare two correlated binomial proportions using a two-sided test with specified alpha and power

$$n = \frac{\left(z_{1-\alpha/2} + 2z_{1-\beta}\sqrt{p_A q_A}\right)^2}{4\left(p_A - 0.5\right)^2 p_D}$$

- for n matched pairs

Sample Size Estimates In Clinical Trials

- Compliance with study treatment or regimen is critical to successful conduct of study
- Problem with compliance is both drop-out in ‘active’ treatment and drop-in in ‘standard’ treatment
- If we estimate λ_1 as the drop-out rate and λ_2 as the drop-in rate, then we can adjust the sample size estimates as originally calculated as
  - $n / (1-\lambda_1 - \lambda_2)^2$

RxC Contingency Tables

- Similar approach as for the 2x2 table
  - Calculate the expected values in the same way
  - Calculate the chi-square by summing over all cells
  - Degrees of freedom are calculated as (r-1)*(c-1)

Chi-Square Test for Trend

- Can be used to look for trends in the data over ordered intervals
  - e.g., does obesity (BMI > 85th %ile) decrease with increasing income where income is a three level categorical variable
- The hypothesis of trends can be of great interest in studies
  - A ‘half-way’ point between contingency table analysis and regression
- It is closely related to the Wilcoxon Rank-Sum Test

Goodness of Fit Chi-Square Test

- Not really used much in research
- Procedure: compare observed frequencies to the frequencies expected from a distribution
  - Only a one sample test
- e.g., number of clustered risk factors
- To some extent, however, all chi-square tests are goodness of fit tests since always testing the fit of the observed frequencies to the expected frequencies

Goodness of Fit Chi-Square Test

- Once the expected frequencies are known, apply the usual chi-square test
- However, generating the expected frequencies can be challenging
- For the normal, standardize the values
  - After dividing the raw data into intervals, calculate the expected values from the standard normal distribution

Tests of Independence

- Probably the most frequent use of the chi-square test
- Compare the observed and expected frequencies in a contingency table
- We are testing that the two classifications in the table are independent; that is, that the chi-square statistic for the table is significantly different from zero
  - Significance indicates that there is at least one observed-expected pair that is non-zero and that the two classifications are related

Tests of Homogeneity

- Used primarily in situations involving a third classification variable
- Testing if the association between the two main variables varies according to the value of a third variable
  - e.g., does the association between income and TV watching vary by race

Kappa Statistic

- Used primarily to assess reproducibility
  - Depends on concordant pairs
  - e.g., the same questionnaire administered twice to determine whether people will answer similarly
    - $k > 0.75$ = excellent
    - $0.4 \leq k \leq 0.75$ = good
    - $k < 0.4$ = poor
- Produced by virtually all packages that perform a chi-square test