

HANDY REFERENCE SHEET – HRP/STATS 261, Discrete Data

2x2 Contingency Tables

| | Disease (D) | No Disease (~D) |
|----------------|-------------|-----------------|
| Exposed (E) | a | b |
| Unexposed (~E) | c | d |

Measures of Association

$$\text{Risk Ratio} = \frac{a/(a+b)}{c/(c+d)}$$

$$95\% \text{ CI: } RR * \exp\left[-1.96\sqrt{\frac{1-a/(a+b)}{a} + \frac{1-c/(c+d)}{c}}\right], RR * \exp\left[+1.96\sqrt{\frac{1-a/(a+b)}{a} + \frac{1-c/(c+d)}{c}}\right]$$

$$\text{Odds Ratio} = \frac{ad}{bc}$$

$$95\% \text{ CI: } OR * \exp\left[-1.96\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\right], OR * \exp\left[+1.96\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\right]$$

Difference in Proportions:

$$H_0: p_{D/E} - p_{D/\sim E} = 0 \quad [\text{for case-control study: } p_{e/d} - p_{e/\sim d} = 0]$$

Test Statistic:

$$\frac{\hat{p}_{d/e} - \hat{p}_{d/\sim e}}{\sqrt{\frac{(\bar{p}_{d/e})(1 - \bar{p}_{d/e})}{n_e} + \frac{(\bar{p}_{d/\sim e})(1 - \bar{p}_{d/\sim e})}{n_{\sim e}}}} \sim Z$$

$$95\% \text{ CI: } (\hat{p}_{d/e} - \hat{p}_{d/\sim e}) \pm 1.96 * \sqrt{\frac{(\hat{p}_{d/e})(1 - \hat{p}_{d/e})}{n_e} + \frac{(\hat{p}_{d/\sim e})(1 - \hat{p}_{d/\sim e})}{n_{\sim e}}}$$

SAS CODE:

```
proc freq data=yourdata;
  tables yourExposure*yourOutcome /measures cl;
  weight counts; *if you have grouped data;
run;
```

2x2xK Contingency Tables

Notation: outcome= d ; predictor= e ; categorical covariate= k

Steps

1. Calculate crude OR_{d-e} (or RR_{d-e})
2. Calculate stratum-specific OR's: $OR_{d-e/k=K}$
3. If crude OR and stratum-specific OR's are all similar \rightarrow **STOP**. k is unlikely to be a confounder or an effect modifier, and you may use usual methods for 2x2 table (see page i), ignoring k .
4. If crude OR and stratum-specific OR's differ \rightarrow proceed to (a) or (b) below:
 - a) If stratum-specific OR's are similar to each other \rightarrow suspect confounding
 - (i) Apply Cochran-Mantel-Haenszel test of conditional independence:
 - H_0 : e and d are conditionally independent.
 - Test Statistic:

| | | |
|-----------|---------|------------|
| | Disease | No Disease |
| Exposed | a | b |
| Unexposed | c | d |

$$\frac{[\sum_{i=1}^k (a_k - E(a_k))]^2}{\sum_{i=1}^k Var(a_k)} \sim \chi_1^2$$

$$E(a_k) = \frac{(a+b)_k (a+c)_k}{n_k}; \text{ where } n_k = \text{total in stratum } k$$

$$Var(a_k) = \frac{(a+b)_k (a+c)_k (b+d)_k (c+d)_k}{n_k^2 (n_k - 1)}$$

- (ii) Calculate Mantel-Haenszel (MH) summary OR (adjusted for confounding by k). If you rejected the null in (i), then MH OR should be $\neq 1.0$. If you did not reject null in (i), then MH OR should be $\cong 1.0$.

$$OR_{d-e/k} = \frac{\sum_{i=1}^k \frac{a_i d_i}{T_i}}{\sum_{i=1}^k \frac{b_i c_i}{T_i}}$$

$$\text{For RR: } RR_{d-e/k} = \frac{\sum_{i=1}^k \frac{a_i (c_i + d_i)}{T_i}}{\sum_{i=1}^k \frac{c_i (a_i + b_i)}{T_i}}$$

- b) If stratum-specific OR's differ from each other \rightarrow suspect interaction (effect modification)
 - (i) Apply Breslow-Day test of homogeneity of the OR's:
 - H_0 : stratum-specific OR's are equal (homogenous)
 - Test Statistic: complex \rightarrow use computer software
 - If you reject null \rightarrow effect modification is present. Report stratum-specific OR's
 - If you fail to reject null \rightarrow insufficient evidence of effect modification; calculate Mantel-Haenszel summary OR, as above.

SAS CODE:

```
proc freq data=yourdata;
  tables yourCovariate*yourExposure*yourOutcome /cmh;
  weight counts; *if you have grouped data;
run;
```

Multi-way Contingency Tables:

Notation: x , y , and z are categorical variables

Chi-square test of independence ($R \times C$ table):

H_0 : x , y , and z are independent

Test Statistic:

$$\sum_{\text{cells}} \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}} \sim \chi^2_{(rows-1)(columns-1)}$$

Log-Linear Model

$\log(\text{counts}) = \alpha + \beta_x x + \beta_y y + \beta_z z + \text{interactions if appropriate}$

Odds ratio interpretation (if 2×2 table):

$$\log(OR_{d-e}) = \log\left(\frac{ad}{bc}\right) = \log a + \log d - \log b - \log c$$

$$\log\left(\frac{ad}{bc}\right) = (\alpha + \beta_d + \beta_e + \beta_{d*e}) + (\alpha) - (\alpha + \beta_d) - (\alpha + \beta_e) = \beta_{d*e}$$

$$\therefore OR = e^{\beta_{d*e}}$$

SAS CODE:

Chi-square test

```
proc freq data=yourdata;
  tables yourX*yourY*yourK / chisq;
  weight counts; *if you have grouped data;
run;
```

Log-linear models

```
proc genmod data=yourdata;
  model total = yourX yourY yourK YourInteractions /
    dist=poisson link=log pred;
run;
```

Logistic Regression

Model:

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots$$

Tests of Model Fit:

- Wald Test

$$Z = \frac{\hat{\beta} - 0}{\text{asymptotic standard error } (\hat{\beta})}$$

- Likelihood Ratio Test: where full model has n parameters and reduced model has $n-p$

$$\begin{aligned} -2 \ln \frac{L(\text{reduced})}{L(\text{full})} &= \\ -2 \ln(L(\text{reduced})) - [-2 \ln(L(\text{full}))] &\sim \chi_p^2 \end{aligned}$$

Simplest form of the logistic likelihood (from 2x2):

$$l(\beta) = \left(\frac{e^{\alpha+\beta_e}}{1+e^{\alpha+\beta_e}}\right)^a x \left(\frac{1}{1+e^{\alpha+\beta_e}}\right)^b x \left(\frac{e^{\alpha}}{1+e^{\alpha}}\right)^c x \left(\frac{1}{1+e^{\alpha}}\right)^d$$

| | Disease | No Disease |
|-----------|---------|------------|
| Exposed | a | b |
| Unexposed | c | d |

Interpretation of Estimated Coefficients in Odds and Probabilities:

OR interpretation: $OR_{\text{exposure}} = e^{\beta_{\text{exp}}}$

OR interpretation in the presence of interaction (two binary predictors):

$$OR_{\text{exposure/interacting factor present}} = e^{\beta_e + \beta_k + \beta_{e \times k}}$$

$$OR_{\text{exposure/interacting factor absent}} = e^{\beta_e}$$

$$\text{Probability interpretation: } P(D/E) = \frac{e^{\alpha+\beta_e}}{1+e^{\alpha+\beta_e}}$$

SAS CODE:

```
proc logistic data = analysis;
  class YourPredictor_c (ref= "YourBaseline");
    * automatic dummy coding to get >1 OR against the reference group;
  model YourOutcome_c (event = "Yes") = YourPredictor / lackfit;
    *lackfit gives Hosmer and Lemeshow goodness of fit chi-square test;
  output out = OutDataSet p = Predicted;
    *outputs predicted probabilities to new dataset;
run;
```

Pair-Matched Data

| <u>Case</u> | <u>Matched-control</u> | |
|-------------|------------------------|-----------|
| | Exposed | Unexposed |
| Exposed | a | b |
| Unexposed | c | d |

Odds Ratio

$$OR = \frac{b}{c}$$

McNemar's Test

- H_0 : exposure and disease are independent
- Test Statistic:

$$\frac{(b - c)^2}{b + c} \sim \chi_1^2$$

1:M Matched Data

Matched Data
Unrestricted Predictors and Covariates
Binary Outcome

Conditional Logistic Regression:

The simplest likelihood (from 2x2):

$$l(\beta) = \left(\frac{e^{\beta_e}}{1 + e^{\beta_e}} \right)^b x \left(\frac{1}{1 + e^{\beta_e}} \right)^c$$

OR interpretation: $OR_{\text{exposure}} = e^{\beta_{\text{exp}}}$

SAS CODE:** SAS V9 ONLY

```
proc logistic data = yourData ;
  model YourOutcpme_c (event= "Yes") = YourPredictor1
    YourPredictor2 YourPredictor3...;
  strata MatchingVariable1 MatchingVariable2 / info;
  output out = OutDataset    p = Predicted;
run;
```