Grant MacEwan University

Stat 371 – Categorical Data Analysis

Formula Sheet for Final Exam

- **Probability Theory**
  - Binomial distribution, with parameters n and π

$$P(X=x)=\frac{n!}{x!(n-x)!}\,\pi^x(1-\pi)^{n-x},\ \mu=n\cdot\pi,\sigma=\sqrt{n\pi(1-\pi)}$$

  - Multinomial distribution

$$P(X_1=x_1,X_2=x_2,\ldots,X_c=x_c)=\frac{n!}{x_1!x_2!,\ldots x_c!}\,\pi_1^{x_1}\pi_2^{x_2}\ldots\pi_c^{n_c}$$

  - Poisson distribution

$$P(X=x)=\frac{e^{-\mu}\mu^x}{x!},\ \sigma=\sqrt{\mu}$$

  - Clinical tests: Sensitivity = P(Test+|Disease+), Specificity = P(Test-|Disease-)

- **Inferential statistics for one probability, π**

Wald - Test Statistic and Confidence Interval:

$$z_0=\frac{\hat\pi-\pi_0}{\sqrt{\hat\pi(1-\hat\pi)/n}}\qquad \hat\pi\pm z_{\alpha/2}\sqrt{\hat\pi(1-\hat\pi)/n}$$

Score - Test Statistic:

$$z_0=\frac{\hat\pi-\pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$$

Likelihood-ratio - Test Statistic:
l_0 = max of the likelihood function under the null hypothesis, l_1 = likelihood function at ML-estimator

$$\chi_0^2=-2\ln(l_0/l_1),\quad df=1$$

- **Inferential statistics for Contingency Tables**

Confidence Interval for $\pi_1-\pi_2$.

$$(\hat\pi_1-\hat\pi_2)\pm z_{\alpha/2}\sqrt{\frac{\hat\pi_1(1-\hat\pi_1)}{n_1}+\frac{\hat\pi_2(1-\hat\pi_2)}{n_2}}$$

Test statistic for $\pi_1-\pi_2$.

$$z_0=\frac{(\hat\pi_1-\hat\pi_2)}{\sqrt{\frac{\hat\pi_1(1-\hat\pi_1)}{n_1}+\frac{\hat\pi_2(1-\hat\pi_2)}{n_2}}}$$

Confidence Interval for $\ln(\pi_1/\pi_2)$.

$$\ln(\hat\pi_1/\hat\pi_2)\pm z_{\alpha/2}\sqrt{\frac{1-\hat\pi_1}{n_1\hat\pi_1}+\frac{1-\hat\pi_2}{n_2\hat\pi_2}}$$

Confidence Interval for $\ln(\theta)$ (ln(odds ratio)).

$$\ln(\hat{\theta}) \pm z_{\alpha/2}\sqrt{\frac{1}{n_{11}}+\frac{1}{n_{12}}+\frac{1}{n_{21}}+\frac{1}{n_{22}}}$$

Pearson's χ^2 statistic for test of independence

$$\text{Expected cell count for cell } (i,j) = \hat{\mu}_{ij} = \frac{\text{row}_i\text{total} \times \text{column}_j\text{total}}{\text{Grand total}} = \frac{n_{i+}n_{+j}}{n}$$

then

$$\chi_0^2 = \sum \frac{(n_{ij}-\hat{\mu}_{ij})^2}{\hat{\mu}_{ij}},\ df = (I-1)(J-1)$$

Likelihood-ratio χ^2 statistic for test of independence

$$\chi_0^2 = 2\ \sum n_{ij}\ln\left(\frac{n_{ij}}{\hat{\mu}_{ij}}\right),\ df = (I-1)(J-1)$$

Standardized cell residuals

$$z_{ij} = \frac{n_{ij}-\hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}(1-\hat{\pi}_{i+})(1-\hat{\pi}_{+j})}}$$

- **Linear Association between Ordinal Variables**
  Pearson's Correlation Coefficient

$$r = \frac{\sum(u_i-\bar{u})(v_i-\bar{v})p_{ij}}{\sqrt{\left[\sum(u_i-\bar{u})^2p_{i+}\right]\left[\sum(v_j-\bar{v})^2p_{+j}\right]}}$$

  Spearman's Correlation Coefficient is Pearson's Correlation Coefficient for midranks

  Test Statistic for testing for a linear association

$$\chi_0^2 = (n-1)r^2, df = 1$$

  Kendall's τ-b ($n_c(n_d)$ is the number of concordant (discordant) pairs of observations)

$$\hat{\tau} = 2\ \frac{n_c-n_d}{n(n-1)},$$

- **Model Comparison for Nested Models**
  Likelihood-ratio - Test Statistic:
  Let L_0 denote the maximal value of the log likelihood function for Model M_0 (null hypothesis), and L_1 is the value of the log likelihood function for Model M_1, then

$$\chi_0^2 = -2(L_0-L_1),\quad df = \text{number of parameters in } M_1 - \text{number of parameters in } M_0$$

  Let D_0 denote the deviance Model M_0 (null hypothesis), and D_1 the deviance for Model M_1, then also

$$\chi_0^2 = D_0 - D_1,\quad df = \text{number of parameters in } M_1 - \text{number of parameters in } M_0$$

- **Model Comparison** *Akaike's Information Criterion, AIC*: L=log likelihood value for the model

$$AIC = -2(L - \#\text{parameters})$$

- **Logistic Regression Model**
  Median effective level:

$$EL_{50} = -\alpha/\beta$$

- **McNemar Test**

$$\chi_0^2 = \frac{(n_{12}-n_{21})^2}{n_{12}+n_{21}},\quad df = 1$$