Stat 565 Formula Sheet

Case-control studies
Cohort studies
Randomized clinical trials
Observational studies
Experiments

Relative risk:

$$RR = \frac{Pr\{\text{disease} \mid \text{exposure}\}}{Pr\{\text{disease} \mid \text{no exposure}\}} = \frac{P_1}{P_0} \qquad S^2_{\widehat{log(RR)}} = \frac{1-\hat{P}_1}{n_1\hat{P}_1} + \frac{1-\hat{P}_0}{n_0\hat{P}_0}$$

Odds ratio: $OR = \frac{P_1/(1-P_1)}{P_0/(1-P_0)} \qquad S^2_{\widehat{log(OR)}} = \sum_i \sum_j \frac{1}{Y_{ij}}$

Bayes theorem: $P(A \mid B) = Pr\{B \mid A\}\ Pr\{A\}/Pr\{B\}$

Chi-square test of independence for a two-way contingency table:

$$X^2 = \sum_i \sum_j \frac{(Y_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}} \qquad \text{where } \hat{m}_{ij} = \frac{Y_{i+}Y_{+j}}{Y_{++}}$$

McNemar's test:

$$X^2_C = \frac{(\mid n_{12} - n_{21} \mid -1)^2}{n_{12} + n_{21}}$$

Sign test (exact binomial test):

$$\sum_{j=x}^{n} \binom{n}{j} \left(\frac{1}{2}\right)^n$$

Marginal odds ratio:

$$\hat{\phi} = \frac{P_{+1}P_{2+}}{P_{+2}P_{1+}} \qquad S^2_{\widehat{1og(\phi)}} = Y_{++}\left(\frac{1}{Y_{1+}Y_{2+}} + \frac{1}{Y_{+1}Y_{+2}} - 2\frac{Y_{11}Y_{22} - Y_{12}Y_{21}}{Y_{1+}Y_{2+}Y_{+1}Y_{+2}}\right)$$

Logistic regression:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} = \mathbf{X}_i^T\beta$$

$$y_i \sim \text{Bin}(n_i, \pi_i),\ i = 1, \cdots, m \qquad \pi_i = P_r\{\text{success} \mid \mathbf{X}_i\} = \frac{\exp(\mathbf{X}_i^T\beta)}{1 + \exp(\mathbf{X}_i^T\beta)}$$

Conditional logistic regression:

$$\log\left(\frac{Pr\{Y_{ij}=1\}}{Pr\{Y_{ij}=0\}}\right)=\alpha_i+\beta Z_j \qquad \text{where} \qquad Z_j=\begin{cases}1 & j=1\\ 0 & j=2\end{cases}$$

$$\hat{\beta}=\log\left(\frac{n_{21}}{n_{12}}\right) \qquad S^2_{\hat{\beta}}=\tfrac{1}{n_{12}}+\tfrac{1}{n_{21}}$$

Randomization test:

Delta Method: $\hat{\theta}=g(\hat{\alpha},\hat{\beta})$

$$V(\hat{\theta})\approx\left[\frac{\partial g(\alpha,\beta)}{\partial\alpha}\ \frac{\partial g(\alpha,\beta)}{\partial\beta}\right]\left[\begin{array}{cc}Var(\hat{\alpha}) & Cov(\hat{\alpha},\hat{\beta})\\ Cov(\hat{\alpha},\hat{\beta}) & Var(\hat{\alpha})\end{array}\right]\left[\frac{\partial g(\alpha,\beta)}{\partial\alpha}\ \frac{\partial g(\alpha,\beta)}{\partial\beta}\right]^T$$

Survivor function: $S(t)=Pr\{T\geq t\}$

Density: $f(t)=\frac{-\partial S(t)}{\partial t}$

Hazard function:

$$h(t) \ = \ \overset{lim}{\Delta t\rightarrow 0}\ \ \frac{S(t)-S(t+\Delta t)}{(\Delta t)\ S(t)} \ = \ \frac{f(t)}{S(t)} \ = \ \frac{-\partial log(S(t))}{\partial t}$$

Cummulative hazard: $H(t) \ = \ \int_0^t h(u)du \ = \ -\log(S(t))$

Proportional hazards: $h(t) \ = \ h_0(t)\ e^{(X^T\beta)}$

Accelerated failure time model: $S(t) \ = \ S_0(te^{X^T\beta})$

Maximum Likelihood Estimation:

Weibull distribution: $S(t)=exp(-\theta t^{\alpha})$ $\qquad t(p)=\left[\frac{1}{\theta}log\left(\frac{1}{1-p}\right)\right]^{1/\alpha}$

Log-normal distribution: $S(t)=1-\Phi\left(\frac{log(t)-\mu}{\sigma}\right)$ $\qquad t(p)=exp\left(\sigma\Phi^{-1}(p)+\mu\right)$

Kaplan-Meier estimation:

$$\hat{S}(t) = \prod_{j=1}^{K} \frac{n_j - d_j}{n_j} \qquad \widehat{Var}\left(\hat{S}(t)\right) \approx [\hat{S}(t)]^2 \sum_{j=1}^{K} \frac{d_j}{n_j(n_j - d_j)}$$

$$\widehat{Var}\left(\ log(\hat{S}(t))\right) \approx \sum_j \frac{d_j}{n_j(n_j - d_j)}$$

$$\widehat{Var}\left(\log[-\log(\hat{S}(t))]\right) \approx \left[\log(\hat{S}(t))\right]^{-2} \sum_j \frac{d_j}{n_j(n_j - d_j)}$$

Life table estimator: $\hat{S}(t) = \prod_{j=1}^{K} \frac{n_j - d_j - .5c_j}{n_j - .5c_j}$

Nelson-Aalen estimator: $\widehat{H}(t) = \sum_{t_{(j)} < t} d_j / n_j$

Generalized log-rank test:

	Group		
	1	2	Totals
die	d_{1j}	d_{2j}	d_j
survive	$n_{1j} - d_{1j}$	$n_{2j} - d_{2j}$	$n_j - d_j$
Totals	n_{1j}	n_{2j}	

$$\hat{e}_{1j} = \frac{n_{ij} d_j}{n_j} \qquad \hat{V}_{1j} \ = \ n_{1j} n_{2j} d_j (n_j - d_j) / \left[n_j^2 (n_j - 1)\right]$$

$$LR = \sum_{j=1}^{m} w_j (d_{1j} - \hat{e}_{1j})^2 / \left(\sum_{j=1}^{m} w_j^2 \hat{V}_{1i}\right)$$

Power calculations and sample size determination:

Cox proportional hazards model:

$$h(t) = h_0(t)\ e^{X^T B} \qquad\qquad S(t) = [S_0(t)]^{\exp(\mathbf{X}^T B)}$$

Partial likelihood:

$$L(\beta) = \prod_{j=1}^{n} \left[\frac{\exp(X_j^T B)}{\sum_{k \in R(t_j)} \exp(X_k^T B)} \right]^{\delta_j} \qquad \delta_j = \begin{cases} 1 & \text{if } t_j \text{ is a failure time} \\ 0 & \text{if } t_j \text{ is a censoring time.} \end{cases}$$

Ties:

Breslow $\left(\frac{r_1}{r_1 + r_2 + r_3 + r_4 + r_5} \right) \left(\frac{r_2}{r_1 + r_2 + r_3 + r_4 + r_5} \right)$

Efron $\left(\frac{r_1}{r_1 + r_2 + r_3 + r_4 + r_5} \right) \left(\frac{r_2}{.5(r_1 + r_2) + r_3 + r_4 + r_5} \right)$

Average Partial Likelihood

Breslow(Nelson-Aalen) Estimator:

$$\widehat{H}(t) = -log(\hat{S}_0(t)) = \sum_{j=1}^{k} \left(\frac{-d_j}{\sum_{i \in R(t_{(j)})} exp(x_i^T \hat{\beta})} \right)$$

Time dependent covariates:

Stratification:

Diagnostics:

Schoenfeld residuals
Deviance residuals
Martingale residuals
Score (*df* beta) residuals

Frailty models:

Limited failure (cure) models:

Cross-sectional studies:

Longitudinal studies:
Let $\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{i,n_i})^T$ denote a set of n_i repeated measurements on the i-th subject (or unit), $i = 1, \ldots, m$

Linear models:

$$E(Y_{ij}) = X_{ij}^T \beta \qquad j = 1, \cdots, n_i, \qquad i = 1, \cdots, m_i$$

$$V_i = Var(\mathbf{Y}_i) \qquad i = 1, \cdots, m$$

Generalized least square estimation:

$$0 = \sum_{i=1}^{m} X_i^T V_i^{-1} (\mathbf{Y}_i - X_i \beta)$$

$$\hat{B} = \left(\sum_{i=1}^{m} X_i^T V_i^{-1} X_i \right)^{-1} \sum_{i=1}^{m} X_i^T V_i^{-1} \mathbf{Y}_i \sim N \left(B, \left(\sum_{i=1}^{m} X_i^T V_i^{-1} X_i \right)^{-1} \right)$$

REML estimation:

$$\mathbf{r} = M(I - P_X)\mathbf{Y} \quad \text{where} \quad P_X = X(X^T X)^{-1} X^T$$

$$\text{and} \quad \text{rank } (M) = n - \text{rank}(X)$$

Linear Models with Random Effects:

$$\mathbf{Y}_i \mid \mathbf{U}_i \sim N(X_i \beta + Z_i \mathbf{U}, R_i) \quad \text{with} \quad \mathbf{U}_i \sim N(0, V)$$

$$\mathbf{Y}_i \sim N(X_i \beta, Z_i V Z_i^T + R_i)$$

Generalized linear models:
$\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{i,n_i})^T$ is the vector of n_i repeated observations on the i-th subject or unit.

$$E(Y_{ij}) = \mu_{ij} = h^{-1}(X_{ij}^T \beta) \quad \text{or} \quad h(\mu_{ij}) = X_{ij}^T \beta$$

where h is the link function. Also specify a distribution for $\mathbf{Y}_i$.

Logistic regression:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_0 + \beta_1 X_{1ij} + \cdots + \beta_K X_{Kij} = X_{ij}^T \beta$$

$$E(Y_{ij}) = \pi_{ij} = \frac{\exp(\mathbf{X}_{ij}^T\beta)}{1+\exp(X_{ij}^T\beta)} \quad j = 1, ..., n_i \quad i = 1, ..., m$$

$V_i = Var(\mathbf{Y}_i) =$

$$\begin{bmatrix} \sqrt{\pi_{i1}(1-\pi_{i1})} & & \\ & \ddots & \\ & & \sqrt{\pi_{in_1}(1-\pi_{in_i})} \end{bmatrix} R_i \begin{bmatrix} \sqrt{\pi_{i1}(1-\pi_{i1})} & & \\ & \ddots & \\ & & \sqrt{\pi_{i,n_1}(1-\pi_{i,n_i})} \end{bmatrix}$$

Poisson regression (log-linear models):

$$\log(\mu_{ij}) = \beta_0 + \beta_1 X_{1ij} + \cdots + \beta_k X_{Kij}$$

$$Y_{ij} \sim \text{Poisson } (\mu_{ij}) \qquad E(Y_{ij}) = \mu_{ij} = Var(Y_{ij})$$

$$V_i = Var(\mathbf{Y}_i) \;\; = \begin{bmatrix} \sqrt{\mu_{i1}} & & \\ & \ddots & \\ & & \sqrt{\mu_{i,n_i}} \end{bmatrix} R_i \begin{bmatrix} \sqrt{\mu_{i1}} & & \\ & \ddots & \\ & & \sqrt{\mu_{i,n_i}} \end{bmatrix}$$

Generalized Estimating Equations (GEE):

$$0 = \sum_{i=1}^{m} D_i^T V_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) \quad \text{where} \quad D_i = \begin{bmatrix} \frac{0\mu_{i1}}{0\beta_0} & \cdots & \frac{0\mu_{i1}}{0\beta_K} \\ \vdots & & \\ \frac{0\mu_{i,n_i}}{0\beta_0} & \cdots & \frac{0\mu_{i,n_i}}{0\beta_K} \end{bmatrix}$$

$$\hat{\beta}_{GEE} \;\dot{\sim}\; N\left(\beta, \left[\sum_{i=1}^{m} D_i V_i^{-1} D_i\right]^{-1}\right)$$

Robust covariance estimation.

$$\left[\sum_{i=1}^{m}\hat{D}_i^T\hat{V}_i^{-1}\hat{D}_i\right]^{-1}\left[\sum_{i=1}^{m}\hat{D}_i^T\hat{V}_i^{-1}[\mathbf{Y}_i-\hat{\boldsymbol{\mu}}_i]\,[\mathbf{Y}_i-\hat{\boldsymbol{\mu}}_i]^T\hat{V}_i^{-1}\hat{D}_i\right]\left[\sum_{i=1}^{m}\hat{D}_i^T\hat{V}_i^{-1}\hat{D}_i\right]^{-1}$$

Negative Binomial Distribution:

Random Effects:

Logistic Regression:

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)=X_{ij}^T\beta+Z_{ij}^T\mathbf{U}\quad\text{where}\quad\pi_{ij}=E(Y_{ij}\mid\mathbf{U})=\frac{\exp(\mathbf{X}_{ij}^T\beta+\mathbf{Z}_{ij}^T\mathbf{U})}{1+\exp(X_{ij}^T\beta+\mathbf{Z}_{ij}^T\mathbf{U})}$$

$$Var(Y_{ij}\mid U_i)=\pi_{ij}(1-\pi_{ij})$$

$$\mathbf{U}\sim N(0,V)$$

Poisson Regression:

$$\log(\mu_{ij})=\mathbf{X}_{ij}^T\beta+\mathbf{Z}_{ij}^T\mathbf{U}\quad\text{where}\quad\mu_{ij}=E(Y_{ij}\mid\mathbf{U})$$

$$V_{ij}=Var(Y_{ij}\mid\mathbf{U})=\mu_{ij}$$

$$\mathbf{U}\sim N(0,V)$$