Formula sheet 1

Distributions

Binomial:

$$p(y) = \binom{n}{y}\pi^y(1-\pi)^{n-y}, \quad y = 0, 1, \ldots, n$$

where $\binom{n}{y} = \frac{n!}{y!(n-y)!}$

$$E(Y) = n\pi$$
$$V(Y) = n\pi(1-\pi)$$

Note that ! is the "factorial" operator.

Hypergeometric:

$$p(n_{11}) = \frac{\binom{n_{1+}}{n_{11}}\binom{n_{2+}}{n_{+1}-n_{11}}}{\binom{n}{n_{+1}}}$$

for $n_{11} \in \{\max(0, n_{+1}+n_{1+}-n), \ldots, \min(n_{+1}, n_{1+})\}$

Poisson:

$$p(y) = \frac{\mu^y e^{-\mu}}{y!}, \quad y = 0, 1, \ldots \;\; \mu > 0$$

with $E(Y) = V(Y) = \mu$.

Inference on π

Wald:

- $TS = \frac{p-\pi_0}{\sqrt{p(1-p)/n}} \overset{\text{approx.}}{\sim} N(0,1)$
- $p \mp z_{1-\alpha/2}\sqrt{p(1-p)/n}$

Score/Wilson:

- $TS = \frac{p-\pi_0}{\sqrt{\pi_0(1-\pi_0)/n}} \overset{\text{approx.}}{\sim} N(0,1)$
- Solving for π_0, $\left|\frac{p-\pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}\right| < z_{1-\alpha/2}$

Inference on π_1, π_2

Difference: $\pi_1 - \pi_2$

$$p_1 - p_2 \mp z_{1-\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_{1+}} + \frac{p_2(1-p_2)}{n_{2+}}}$$

Relative Risk: $R.R. = \frac{\pi_1}{\pi_2}$

$$\log\left(\frac{p_1}{p_2}\right) \mp z_{1-\alpha/2}\sqrt{\frac{1-p_1}{n_{1+}p_1} + \frac{1-p_2}{n_{2+}p_2}} \rightarrow (L, U)$$

and then (e^L, e^U).

Odd Ratio: $\theta = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)} = \frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}$

$$\log\left(\hat{\theta}\right) \mp z_{1-\alpha/2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} \rightarrow (L, U)$$

and then (e^L, e^U).

Independence

Pearson:

$$X^2 = \sum_{ij}\frac{(n_{ij}-\hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \underset{\text{approx.}}{\overset{H_0}{\sim}} \chi^2_{(I-1)(J-1)}$$

where $\hat{\mu}_{ij} = \frac{n_{i+}}{n}\frac{n_{+j}}{n}$.

LRT:

$$G^2 = 2\sum_{ij} n_{ij}\log\left(\frac{n_{ij}}{\hat{\mu}_{ij}}\right) \underset{\text{approx.}}{\overset{H_0}{\sim}} \chi^2_{(I-1)(J-1)}$$

Standardized/Adjusted Residuals:

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

Pearson residuals:

$$e_{ij} = \frac{n_{ij}-\hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}}}$$

CMH:

$$CMH = \frac{\left[\sum_{k=1}^{K}(n_{11k} - E(n_{11k}))\right]^2}{\sum_{k=1}^{K}V(n_{11k})} \underset{\text{approx.}}{\overset{H_0}{\sim}} \chi^2_1$$

where

$$E(n_{11k}) = \frac{n_{1+k}n_{+1k}}{n}$$
$$V(n_{11k}) = \frac{n_{1+k}n_{2+k}n_{+1k}n_{+2k}}{{n_{++k}}^2(n_{++k}-1)}$$

and

$$\hat{\theta}_{MH} = \frac{\sum_{k=1}^{K}(n_{11k}n_{22k}/n_{++k})}{\sum_{k=1}^{K}(n_{12k}n_{21k}/n_{++k})}$$

GLM

Logistic

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \alpha + \beta x \Rightarrow \pi(x) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$$

Poisson log-linear

$$\log(\mu) = \alpha + \beta x \quad \Rightarrow \quad \mu = e^{(\alpha+\beta x)} = e^{\alpha}(e^{\beta})^x$$

Poisson rates:

$$\log\left(\frac{\mu}{t}\right) = \log(\mu) - \log(t) = \alpha + \beta x$$
$$\Rightarrow \log(\mu) = \alpha + \beta x + \underbrace{\beta_2}_{=1}\underbrace{x_2}_{\log(t)}$$

Inference

Wald:

- $TS = \frac{\hat{\beta}-\beta_0}{s_{\hat{\beta}}} \overset{H_0}{\sim} N(0,1)$
- $\hat{\beta} \mp z_{1-\alpha/2}\left(s_{\hat{\beta}}\right)$

LRT-Goodness of fit:

$$D(y;\hat{\mu}) := G^2 = -2[L(\hat{\mu};y) - L(y;y)] \xrightarrow[H_0]{d} \chi^2_{df}$$

LRT:

$$G^2 = D(y;\hat{\mu}_0) - D(y;\hat{\mu}_1) \xrightarrow[H_0]{d} \chi^2_{df}$$

For binomial and Poisson models

$$D(y;\hat{\mu}) = 2\sum_{i=1}^{n} y_i \log(y_i/\hat{\mu}_i)$$

Overdispersion

Check (ideally) $X^2 \gg df$, or $X^2/df \gg 1$

- count data: Negative Binomial, with θ

$$V(Y) = \mu + \left(\frac{1}{\theta}\right)\mu^2$$

- binomial data: Beta-Binomial, with ρ

$$V(Y) = n\pi(1-\pi)[1+(n-1)\rho]$$