60

STAT 8004, Homework 6

Group # 2
Members: Yinghui Lu, Du Ye, Scott Bruce

April 10, 2014

This homework is due Thu., 2014/04/10, 5:30pm.

Instructions: Generate a PDF file from it and submit the PDF file to blackboard. Each group should submit one file with file names **hw[number]-[groupnumber].pdf**. For example, "hw01-1.pdf" for homework 1 and group 1. Please also include your R code in the appendix.

Problem 1. (50 points) A medical study is conducted to compare the success rates of two treatments for kidney stones. The table shows the success rates and numbers of treatments for treatments involving both small and large kidney stones, where Treatment A includes all open procedures and Treatment B is percutaneous nephrolithotomy:

	Treatment A	Treatment B
Small Stones	93%(81/87)	87%(234/270)
Large Stones	73%(192/263)	69%(55/80)
Both	78%(273/350)	83%(289/350)

a) (10 points) Perform within-strata analysis and marginal unadjusted analysis on the data. Display your results as Table 3 in the notes of Lecture 8.

First let's look at the contingency tables for the stratified and marginal data:

Patients with Small Stones			
Treatment	Success	Partial Success	
A	81	6	87
B	234	36	270
	315	42	357

Table 1: Success by treatment for Small Stones

Patients with Large Stones			
Treatment	Success	Partial Success	
A	192	71	263
B	55	25	80
	247	96	343

Table 2: Success by treatment for Large Stones

All Patients			
Treatment	Success	Partial Success	
A	273	77	350
B	289	61	350
	562	138	700

Table 3: Success by treatment for All Patients

Now we can calculate the point estimates and confidence intervals for RD, RR, and OR using the distributions and formulas derived in the previous homework:

Small Stones

$$\begin{aligned}
\hat{\pi}_1 &= 81/87 = 0.931 \\
\hat{\pi}_2 &= 234/270 = 0.867 \\
\widehat{RD} &= \hat{\pi}_1 - \hat{\pi}_2 = 0.931 - 0.867 = 0.064 \\
\widehat{RR} &= \hat{\pi}_1/\hat{\pi}_2 = 0.931/0.867 = 1.074 \\
\widehat{OR} &= \frac{\hat{\pi}_1/(1-\hat{\pi}_1)}{\hat{\pi}_2/(1-\hat{\pi}_2)} = \frac{0.931/0.069}{0.867/0.133} = 13.5/6.5 = 2.077
\end{aligned}$$

$$\widehat{RD} \pm z_{\alpha/2}\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1}+\frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$$

$$95\% \text{ CI } 0.064 \pm 0.067 = (-0.0026, 0.1313)$$

$$log(\widehat{RR}) \pm z_{\alpha/2}\sqrt{\frac{(1-\hat{\pi}_1)}{\hat{\pi}_1 n_1}+\frac{(1-\hat{\pi}_2)}{\hat{\pi}_2 n_2}}$$

$$95\% \text{ CI } \exp(0.0716 \pm 0.074) = (\exp(-0.00225), \exp(0.14553)) = (0.9978, 1.1567)$$

$$log(\widehat{OR}) \pm z_{\alpha/2}\sqrt{\frac{1}{\hat{\pi}_1(1-\hat{\pi}_1)n_1}+\frac{1}{\hat{\pi}_2(1-\hat{\pi}_2)n_2}}$$

$$95\% \text{ CI } \exp(0.731 \pm 0.9005) = (\exp(-0.1695), \exp(1.6314)) = (0.8441, 5.1109)$$

Large Stones

$$\begin{aligned}
\hat{\pi}_1 &= 192/263 = 0.730 \\
\hat{\pi}_2 &= 55/80 = 0.6875 \\
\widehat{RD} &= \hat{\pi}_1 - \hat{\pi}_2 = 0.730 - 0.6875 = 0.0425 \\
\widehat{RR} &= \hat{\pi}_1/\hat{\pi}_2 = 0.730/0.6875 = 1.062 \\
\widehat{OR} &= \frac{\hat{\pi}_1/(1-\hat{\pi}_1)}{\hat{\pi}_2/(1-\hat{\pi}_2)} = \frac{0.730/0.270}{0.6875/0.3125} = 2.704/2.2 = 1.229
\end{aligned}$$

$$\widehat{RD} \pm z_{\alpha/2}\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1}+\frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$$

$$95\% \text{ CI } 0.0425 \pm 0.1149 = (-0.0724, 0.1574)$$

$$log(\widehat{RR}) \pm z_{\alpha/2}\sqrt{\frac{(1-\hat{\pi}_1)}{\hat{\pi}_1 n_1}+\frac{(1-\hat{\pi}_2)}{\hat{\pi}_2 n_2}}$$

$$95\% \text{ CI } \exp(0.0602 \pm 0.165) = (\exp(-0.1048), \exp(0.2252)) = (0.9005, 1.2525)$$

$$log(\widehat{OR}) \pm z_{\alpha/2}\sqrt{\frac{1}{\hat{\pi}_1(1-\hat{\pi}_1)n_1}+\frac{1}{\hat{\pi}_2(1-\hat{\pi}_2)n_2}}$$

$$95\% \text{ CI } \exp(0.2062 \pm 0.5455) = (\exp(-0.3393), \exp(0.7517)) = (0.7122, 2.1207)$$

All Patients

$$
\begin{aligned}
\hat{\pi}_1 &= 273/350 = 0.780 \\
\hat{\pi}_2 &= 289/350 = 0.826 \\
\widehat{RD} &= \hat{\pi}_1 - \hat{\pi}_2 = 0.78 - 0.826 = -0.0457 \\
\widehat{RR} &= \hat{\pi}_1/\hat{\pi}_2 = 0.780/0.826 = 0.9446 \\
\widehat{OR} &= \frac{\hat{\pi}_1/(1-\hat{\pi}_1)}{\hat{\pi}_2/(1-\hat{\pi}_2)} = \frac{0.780/0.220}{0.826/0.0.174} = 3.545/4.738 = 0.7483
\end{aligned}
$$

$$
\begin{gathered}
\widehat{RD} \pm z_{\alpha/2}\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}} \\
95\% \text{ CI } -0.0457 \pm 0.0588 = (-0.1045, 0.0131) \\
log(\widehat{RR}) \pm z_{\alpha/2}\sqrt{\frac{(1-\hat{\pi}_1)}{\hat{\pi}_1 n_1} + \frac{(1-\hat{\pi}_2)}{\hat{\pi}_2 n_2}} \\
95\% \text{ CI } \exp(-0.057 \pm 0.0735) = (\exp(-0.1305), \exp(0.01654)) = (0.8776, 1.0167) \\
log(\widehat{OR}) \pm z_{\alpha/2}\sqrt{\frac{1}{\hat{\pi}_1(1-\hat{\pi}_1)n_1} + \frac{1}{\hat{\pi}_2(1-\hat{\pi}_2)n_2}} \\
95\% \text{ CI } \exp(-0,28995 \pm 0.3746) = (\exp(-0.6646), \exp(0.0847)) = (0.5145, 1.0883)
\end{gathered}
$$

Here is the summary table of the above calculations

Measure	Stone Size Small	Large	Marginal
Risk difference $\widehat{RD}$	0.064	0.043	-0.046
95% CI for RD	-0.003,0.131	-0.072,0.157	-0.105,0.013
Relative risk $\widehat{RR}$	1.07	1.06	0.94
95% CI for RR	0.998,1.157	0.901,1.253	0.878,1.017
Odds ratio $\widehat{OR}$	2.08	1.23	0.75
95% CI for OR	0.844,5.11	0.712,2.121	0.515,1.088

Table 4: Measure of risk within each stratum and marginal unadjusted

b) (10 points) Perform stratified analysis and display your results as Table 5 in the notes of Lecture 8.

First we need the MH estimate of the odds ratio:

$$\begin{aligned}
\widehat{OR}_{MH} &= \frac{\sum_j a_j d_j/N_j}{\sum_j b_j c_j/N_j} \\
\text{numerator} &= \frac{81*36}{357} + \frac{192*25}{343} = 22.1622 \\
\text{denominator} &= \frac{6*234}{357} + \frac{71*55}{343} = 15.3176 \\
\widehat{OR}_{MH} &= 22.1622/15.3176 = 1.4468
\end{aligned}$$

Next we can find estimates of v_j for j = 1,2:

$$\begin{aligned}
\hat{v}_j &= \frac{b_j c_j/N_j}{\sum_l b_l c_l/N_l} \\
\hat{v}_1 &= \frac{(6*234)/357}{(6*234)/357 + (71*55)/343} = 3.9328/15.3176 = 0.2567 \\
\hat{v}_2 &= \frac{(71*55)/343}{(6*234)/357 + (71*55)/343} = 11.3848/15.3176 = 0.7433
\end{aligned}$$

Now we can find the confidence interval for the Mantel-Haenszel Estimate of the odds ratio as:

$$\begin{aligned}
&(\exp(\hat{\theta}_l, \exp(\hat{\theta}_u) \\
\hat{\theta}_l &= \log(\widehat{OR}_{MH}) - z_{1-\alpha/2}\sqrt{\frac{\sum_j \hat{v}_j^2 * (\widehat{OR}_j)^2 * \mathrm{Var}(\log(\widehat{OR}_j))}{(\widehat{OR}_{MH})^2}} \\
\hat{\theta}_u &= \log(\widehat{OR}_{MH}) + z_{1-\alpha/2}\sqrt{\frac{\sum_j \hat{v}_j^2 * (\widehat{OR}_j)^2 * \mathrm{Var}(\log(\widehat{OR}_j))}{(\widehat{OR}_{MH})^2}} \\
\hat{\theta}_l &= 0.3695 - 1.96 * \sqrt{\frac{0.257^2 * 2.077^2 * 0.211 + 0.743^2 * 1.229^2 * 0.07746}{1.447^2}} \\
&= 0.3695 - 1.96 * \sqrt{0.0596} = -0.1088 \\
\hat{\theta}_u &= 0.3695 + 1.96 * \sqrt{\frac{0.257^2 * 2.077^2 * 0.211 + 0.743^2 * 1.229^2 * 0.07746}{1.447^2}} \\
&= 0.3695 + 1.96 * \sqrt{0.0596} = 0.8478 \\
&(\exp(\hat{\theta}_l, \exp(\hat{\theta}_u) = (\exp(-0.1088), \exp(0.8478)) = (0.8969, 2.3346)
\end{aligned}$$

	Stone Size		Mantel-Haenszel	
Measure	Small	Large	Estimate	95% C.I.
Odds Ratio	2.077	1.229	1.447	0.897,2.335
$\hat{v}_j$	0.257	0.743		
$\widehat{Var}(\widehat{OR}_{MH})$			0.125	

Table 5: Odds Ratios within strata and the Mantel-Haenszel Adjusted Estimates

c) (10 points) Test the global hypothesis

$$H_0 : OR_j = 1, j = 1, 2$$

using conditional Mantel-Haenszel test.

$$E_1 = E(a_1) = \frac{n_1 m_{11}}{N_1}$$
$$E_2 = E(a_2) = \frac{n_{12} m_{12}}{N_2}$$
$$\hat{V}_{c1} = \widehat{Var(a_1)} = \frac{n_{11}n_{21}m_{11}m_{21}}{N_1^2(N_1 - 1)}$$
$$\hat{V}_{c2} = \widehat{Var(a_2)} = \frac{n_{12}n_{22}m_{12}m_{22}}{N_2^2(N_2 - 1)}$$

Under H_0 asympototically for large N_j,

$$X^2_{c(MH)} = \frac{[\sum_j (a_j - E_j)]^2}{\sum_j \hat{V}_{cj}} \overset{d}{\approx} \chi^2(1)$$

Then we found out

$$X^2_{c(MH)} = 2.43 < \chi^2_{0.95}(1) = 3.841$$

Thus we cannot reject null hypthoesis

d) (10 points) Form a logistic model without interaction for the data to study the effect of the treatment. Interpret each parameter in the model.

For the ith patient ($i = 1, \cdots, n$), let

$$Y_i = \begin{cases} 1 & \text{if the ith patient's treatment is successful} \\ 0 & \text{otherwise} \end{cases}$$

Let

$$x_i = \begin{cases} 1 & \text{if the ith patient's had treatment B} \\ 0 & \text{otherwise} \end{cases}$$

and

$$z_i = \begin{cases} 1 & \text{if the ith patient's had small stone} \\ 0 & \text{otherwise} \end{cases}$$

Let $\pi_i = E(Y_i)$. Then we can set up the logistic regression as:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta_1 x_i + \beta_2 z_i$$

Interpretation of the parameter: α: log reference odds, where the reference is patients with large stone having treatment A.
β_1:log OR for treatmentB vs treatment A.
β_2:log OR for patient with small stone vs large stone.

e) (10 points) Fit the model in R. What are the estimates of each parameter? What is the estimated odds ratio for the treatment effect in each stone group?

Fit the model in R as follows:

```
kidney <- data.frame(Treatment=c("A","A","B","B"),
Stone=c("S","L","S","L"),
Success=c(81,192,234,55),
PartialSuccess=c(6,71,36,25));

logistic.fit <- glm(cbind(kidney$Success,
kidney$PartialSuccess) ~
kidney$Stone +
kidney$Treatment,
data = kidney, family = binomial);

summary(logistic.fit);

Call:
glm(formula = cbind(kidney$Success, kidney$PartialSuccess) ~
    kidney$Stone + kidney$Treatment, family = binomial, data = kidney)
```

```
Deviance Residuals:
      1        2        3        4
 0.7636  -0.2756  -0.3588   0.4695

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)         1.0332     0.1345   7.684 1.55e-14 ***
kidney$StoneS       1.2606     0.2390   5.274 1.33e-07 ***
kidney$TreatmentB  -0.3572     0.2291  -1.559    0.119
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 33.1239  on 3  degrees of freedom
Residual deviance:  1.0082  on 1  degrees of freedom
AIC: 26.355

Number of Fisher Scoring iterations: 3
```

The estimate for each parameter is:

$$\hat{\alpha} = 1.0332$$

$$\hat{\beta}_1 = -0.3572$$

$$\hat{\beta}_2 = 1.2606$$

The estimated odds ratio for the treatment B vs. treatment A is:

$$OR_{treat} = \exp(\hat{\beta}_1) = 0.69968$$

Note: Treatment B odds is in the numerator, so this is consistent with the previous analysis.

Problem 2. (10 points) Consider the exponential risk model using the log link:

$$\log \pi_i = \alpha + \mathbf{x}_i^T \boldsymbol{\beta}$$

Derive the likelihood, log-likelihood, score estimating equations, and the elements of the Hessian. The likelihood function and log likelihood function are

$$
\begin{aligned}
L(\pi) &= \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi)^{1-y_i} \\
l(\pi) &= \sum_{i=1}^{n} y_i log(\pi) + \sum_{i=1}^{n}(1-y_i)log(1-\pi)
\end{aligned}
$$

since we know exponential risk model:

$$\pi_i = \exp(\alpha + x_i^T\beta)$$

Let $\theta = (\alpha, \beta)^T$ Then that in the corresponding model parameters are:

$$L(\theta) = \prod_{i=1}^{n}(\exp(\alpha + x_i^T\beta))^{y_i}(1-\exp(\alpha + x_i^T\beta))^{1-y_i}$$

$$l(\theta) = \sum_{i=1}^{n} y_i(\alpha + x_i^T\beta) + \sum_{i=1}^{n}(1-y_i)log(1-\exp(\alpha + x_i^T\beta))$$

The score function for intercept is:

$$
\begin{aligned}
U(\theta)_\alpha = \frac{\partial l}{\partial \alpha} &= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}(1-y_i)(\frac{\exp(\alpha + x_i^T\beta)}{1-\exp(\alpha + x_i^T\beta)}) \\
&= \sum_{i=1}^{n}[y_i - (1-y_i)\frac{\pi_i}{1-\pi_i}] \\
&= \sum_{i=1}^{n} \frac{y_i - \pi_i}{1-\pi_i}
\end{aligned}
$$

The score function for the j-th coefficient is

$$
\begin{aligned}
U(\theta)_{\beta_j} = \frac{\partial l}{\partial \beta_j} &= \sum_{i=1}^{n}(x_{ij}y_i - (1-y_i)x_{ij}\frac{\pi_i}{1-\pi_i}) \\
&= \sum_{i=1}^{n} x_{ij}\frac{y_i - \pi_i}{1-\pi_i}
\end{aligned}
$$

The elements of the Hessian matrix:

$$H(\theta)_\alpha = \frac{\partial U(\theta)_\alpha}{\partial \alpha} = \sum_{i=1}^{n} \frac{(y_i - 1)\pi_i}{(1 - \pi_1)^2}$$

$$H(\theta)_{\beta_j} = \frac{\partial U(\theta)_{\beta_j}}{\partial \beta_j} = \sum_{i=1}^{n} \frac{x_{ij}^2 (y_i - 1)\pi_i}{(1 - \pi_i)^2}$$

$$H(\theta)_{\beta_j,\beta_k} = \frac{\partial U(\theta)_{\beta_j}}{\partial \beta_k} = \sum_{i=1}^{n} \frac{x_{ij} x_{ik} (y_i - 1)\pi_i}{(1 - \pi_i)^2}$$

$$H(\theta)_{\alpha,\beta_j} = \frac{\partial U(\theta)_\alpha}{\partial \beta_j} = \sum_{i=1}^{n} \frac{x_{ij} (y_i - 1)\pi_i}{(1 - \pi_i)^2}$$

Problem 3. You don't need to turn in the solution to this part. Please program the iterated reweighted least squares in R. Try the example in Problem 1 and compare the estimated value with the results yielded by the R function glm.