HW2_answer

Guojing Wu / UNI: gw2383

2/18/2019

Problem 1

i) Fill out the table and give comments.

```
data.bioassay = data.frame(dose = c(0, 1, 2, 3, 4),
                  n = c(2, 8, 15, 23, 27))
resp=cbind(data.bioassay$n, 30 - data.bioassay$n)
pred=data.bioassay$dose
fit_result = tibble(Model = c("logit", "probit", "c-log-log"),
                    Estimate = rep(0, 3),
                    CI = rep("0", 3),
                    Deviance = rep(0, 3),
                    phat = rep(0, 3))

fit_logit = glm(resp ~ pred, family = binomial(link = 'logit'))
fit_probit = glm(resp ~ pred, family = binomial(link = 'probit'))
fit_cloglog = glm(resp ~ pred, family = binomial(link = 'cloglog'))
fit_all = list(fit_logit, fit_probit, fit_cloglog)

for (i in 1:length(fit_all)) {
  beta = fit_all[[i]]$coefficients[2] # estimation for beta
  fit_result$Estimate[i] = round(beta, 3)

  se = sqrt(vcov(fit_all[[i]])[2,2]) # standard error for beta
  fit_result$CI[i] = paste("(", round(beta + qnorm(0.025) * se, 3), ", ", round(beta - qnorm(0.025) * s

  fit_result$Deviance[i] = round(sum(residuals(fit_all[[i]], type = 'deviance')^2), 3) # deviance

  predi = predict(fit_all[[i]], data.frame(pred = 0.01), se.fit = TRUE, type = 'response')$fit # predic
  fit_result$phat[i] = round(predi, 3)
}

fit_result %>% knitr::kable()
```

Model	Estimate	CI	Deviance	phat
logit	1.162	(0.806, 1.517)	0.379	0.090
probit	0.686	(0.497, 0.876)	0.314	0.085
c-log-log	0.747	(0.532, 0.961)	2.230	0.128

This is grouped data and each $m_i \geq 10$, so we could use the Deviance for goodness-of-fit test $D \sim \chi^2_3$:

- For logit: p-value = 0.9445456
- For probit: p-value = 0.9573743
- For c-log-log: p-value = 0.5260635

All failed to reject the null hypothesis, which suggests that every model fits the data well.

For logit link, we could interpret that: the log odds ratio of probability of dying is 1.162 given one unit increase in dose

ii) Suppose that the dose level is in natural logarithm scale, estimate LD50 with 90% confidence interval based on the three models.

$$\hat{x_0} = f(\hat{\alpha}, \hat{\beta}) = \frac{g(0.5) - \hat{\alpha}}{\hat{\beta}}$$
$$\frac{\partial f}{\partial \hat{\alpha}} = -\frac{1}{\hat{\beta}}$$
$$\frac{\partial f}{\partial \hat{\beta}} = \frac{g(0.5) - \hat{\alpha}}{\hat{\beta}^2}$$

```
pred_result = tibble(Model = c("logit", "probit", "c-log-log"),
                     g = c(log(0.5/(1-0.5)), qnorm(0.5), log(-log(1-0.5))),
                     Estimation = rep(0, 3),
                     Variance = rep(0, 3),
                     CI = rep("0", 3))

for (i in 1:length(fit_all)) {
  beta0 = fit_all[[i]]$coefficients[1]
  beta1 = fit_all[[i]]$coefficients[2]
  betacov = vcov(fit_all[[i]]) # inverse fisher information
  x0fit = (pred_result$g[i] - beta0)/beta1 # point estimate of LD50
  pred_result$Estimation[i] = round(exp(x0fit) , 3)

  varx0 = betacov[1,1]/(beta1^2) + betacov[2,2]*((beta0 - pred_result$g[i])^2)/(beta1^4) - 2*betacov[1,2
  pred_result$Variance[i] = round(varx0, 3)

  pred_result$CI[i] = paste("(", round(exp(x0fit + qnorm(0.05) * sqrt(varx0)), 3), ", ", round(exp(x0fi
}

pred_result %>% knitr::kable()
```

Model	g	Estimation	Variance	CI
logit	0.0000000	7.389	0.032	(5.51, 9.91)
probit	0.0000000	7.436	0.030	(5.583, 9.904)
c-log-log	-0.3665129	8.841	0.034	(6.526, 11.977)

Problem 2

i) How does the model fit the data?

This is grouped data, but it's sparse. So we use Hosmer-Lemeshow staistic $\chi^2_{HL} \sim \chi^2_8$:

```
data.mph = data.frame(amount = seq(from = 10, to = 90, by = 5),
                      offers = c(4, 6, 10, 12, 39, 36, 22, 14, 10, 12, 8, 9, 3, 1, 5, 2, 1),
                      enrolls = c(0, 2, 4, 2, 12, 14, 10, 7, 5, 5, 3, 5, 2, 0, 4, 2, 1))
resp = cbind(data.mph$enrolls, data.mph$offers - data.mph$enrolls)
pred = data.mph$amount
```

```r
fit_mph = glm(resp~pred,family=binomial(link='logit'))

hi = hoslem.test(fit_mph$y, fitted(fit_mph), g = 10)
```

The test result showed p-value = 0.991, we failed to reject H_0, which suggests that the model fits the data well.

ii) How do you interpret the relationship between the scholarship amount and the enrollment rate? What is 95% CI?

```r
beta = fit_mph$coefficients[2]
se = sqrt(vcov(fit_mph)[2,2]) # standard error
```

Let p stand for enrollment rate, $logit(p) = \beta_0 + \beta_1 x$.

We calculate $\hat{\beta}_1 = 0.031$, so for per \$1,000 increase in scholarship, the log odds ratio of enrollment rate is 0.031.

$CI_{\hat{\beta}_1} = (0.012, 0.05)$

iii) How much scholarship should we provide to get 40% yield rate? What is the 95% CI?

p = 0.4

```r
beta0 = fit_mph$coefficients[1]
beta1 = fit_mph$coefficients[2]
betacov = vcov(fit_mph) # inverse fisher information
x0fit = (log(0.4/0.6) - beta0)/beta1
varx0 = betacov[1,1]/(beta1^2) + betacov[2,2]*((beta0 - log(0.4/0.6))^2)/(beta1^4)-2*betacov[1,2]*(beta
```

So we need 40.134 thousands scholarship to get 40% yield rate.

$CI_{\hat{x_0}} = (30.583, 49.686)$