Question 1

Joseph Froelicher

3/12/2021

Part A

$$f(y|p) = p^y(1-p)^{1-y}, \quad p = 1 - e^{-k2^{t/\theta}}$$

$$f(y|\theta) = (1 - e^{-k2^{t/\theta}})^y(-e^{-k2^{t/\theta}})^{1-y}$$

$$L(\theta|\mathbf{y}) = \prod_{i=1}^{n}(1 - e^{-k2^{t/\theta}})^y(-e^{-k2^{t/\theta}})^{1-y}$$

$$L(\theta|\mathbf{y}) = (1 - e^{-k2^{t/\theta}})^{\sum_{i=1}^{n} y_i}(-e^{-k2^{t/\theta}})^{n-\sum_{i=1}^{n} y_i}$$

$$l(\theta|\mathbf{y}) = \sum_{i=1}^{n} y_i \; ln(1 - e^{-k2^{t/\theta}}) + (n - \sum_{i=1}^{n} y_i) \; ln(-e^{-k2^{t/\theta}})$$

$$l(\theta|\mathbf{y}) = \sum_{i=1}^{n} y_i \; (ln(1) - (-k2^{t/\theta})) + (n - \sum_{i=1}^{n} y_i) \; (-k2^{t/\theta})$$

$$l(\theta|\mathbf{y}) = \sum_{i=1}^{n} y_i \; (k2^{t/\theta}) + (n - \sum_{i=1}^{n} y_i) \; (-k2^{t/\theta})$$

$$\begin{aligned} l(\theta|\mathbf{y}, t = \{8, 12, 16\}) = &\sum_{i=1}^{n} y_i \; (k2^{8/\theta}) + (n - \sum_{i=1}^{n} y_i) \; (-k2^{8/\theta}) \\ &+ \sum_{i=1}^{n} y_i \; (k2^{12/\theta}) + (n - \sum_{i=1}^{n} y_i) \; (-k2^{12/\theta}) \\ &+ \sum_{i=1}^{n} y_i \; (k2^{16/\theta}) + (n - \sum_{i=1}^{n} y_i) \; (-k2^{16/\theta}) \end{aligned}$$

Part B

$$g(e[Y_i|X_i]) = \mathbf{X}_i^T\beta, \quad g(p_i) = ln(-ln(1-p_i)), \quad \mathbf{X}_i^T\beta = \beta_1 t_i + ln(k)$$

From our model:

$$p_i = 1 - e^{-e^{\beta_1 t_i + ln(k)}}$$

$$p_i = 1 - e^{-ke^{\beta_1 t_i}}$$

Given in the problem:

$$p_i = 1 - e^{-k2^{t_i/\theta}}$$

We know that:

$$\theta = f(\beta_1)$$

And thus by the invariance property of MLE's:

$$\hat{\theta} = f(\hat{\beta}_1)$$

Therefore, Set them equal and solve:

$$1 - e^{-ke^{\beta_1 t_i}} = 1 - e^{-k2^{t_i/\theta}}$$

$$-ke^{\beta_1 t_i} = -k2^{t_i/\theta}$$

$$e^{\beta_1 t_i} = 2^{t_i/\theta}$$

$$\beta_1 t_i = \frac{t_i}{\theta} ln(2)$$

$$\beta_1 = \theta \ ln(2)$$

$$\hat{\theta} = \frac{ln(2)}{\hat{\beta}_1}$$

Q.E.D.

Part C

```r
# Part C
model1 = glm(
  pos ~ hours + offset(log(k)) - 1, family = binomial(link = "cloglog"), data = data.glm
)
summary(model1)
```

```
##
## Call:
## glm(formula = pos ~ hours + offset(log(k)) - 1, family = binomial(link = "cloglog"),
##     data = data.glm)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5035  -0.5950   0.2984   0.5076   1.9072
##
## Coefficients:
##       Estimate Std. Error z value Pr(>|z|)
## hours  0.35921    0.01593   22.55   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 314.011  on 60  degrees of freedom
## Residual deviance:  59.234  on 59  degrees of freedom
## AIC: 61.234
##
## Number of Fisher Scoring iterations: 5
```

Part D

$\hat{\theta}_{MLE} = 1.9296389$ [1.7753414, 2.1133096]

Part E

We reject the null hypothesis that the full model containing the time group variable is not a better fit that the saturated model containing no time group variable ($p < 0.001$).

Appendix

```r
n = 20
k = rep(1 / 100, n * 3)

data = data.frame(
  "hours" = c(8, 12, 16),
  "pos" = c(5, 10, 19)
)

data.glm = data.frame(
  "hours" = c(rep(data$hours[1], n), rep(data$hours[2], n), rep(data$hours[3], n)),
  "pos" = c(
    rep(1, data$pos[1]),
    rep(0, n - data$pos[1]),
    rep(1, data$pos[2]),
    rep(0, n - data$pos[2]),
    rep(1, data$pos[3]),
    rep(0, n - data$pos[3])
  )
)

# Part C
model1 = glm(
  pos ~ hours + offset(log(k)) - 1, family = binomial(link = "cloglog"), data = data.glm
)
summary(model1)

# Part D
b1 = summary(model1)$coefficients[1, 1]
se = summary(model1)$coefficients[1, 2]
alpha = 0.05
z = qnorm(1 - (alpha / 2))
b1_ci = c(b1 - (z * se), b1 + (z * se))
theta = log(2) / b1
theta_ci = c(log(2) / b1_ci[1], log(2) / b1_ci[2])

# part E
model0 = glm(
  pos ~ offset(log(k)) - 1, family = binomial(link = "cloglog"), data = data.glm
)
summary(model0)
anova(model1, model0, test = "LRT")
```