

# HW4 solutions

*Eric Scott*

*2/11/2020*

## load packages

```
library(tidyverse)
library(here)
library(lmtest) #for lrtest()
library(bbmle) #for Ictab()
```

## Load data

```
mice <- tibble(site = 1:6,
              status = c("burn", "unburn", "burn", "unburn", "unburn", "unburn"),
              fleas = c(7, 1, 9, 7, 8, 3),
              nofleas = c(16, 3, 22, 3, 1, 2))

galls <- read_csv(here("data", "goldenrod_galls.csv"))

plant <-
  tibble(year = 1988:2007,
         dead = c(4, 10, 7, 28, 2, 9, 7, 22, 13, 9, 10, 58, 7, 5, 8, 11, 2, 17, 13, 15),
         alive = c(96, 90, 93, 72, 98, 91, 93, 78, 87, 91, 90, 42, 93, 95, 92, 89, 98, 83, 87, 85))
```

## 1. Mice and fleas

### a. Fit 3 models

Model 1: intercept only

```
mice1 <- glm(cbind(fleas, nofleas) ~ 1, family = binomial(), data = mice)
coef(mice1) %>% plogis()
```

```
## (Intercept)
## 0.4268293
```

42.7% of mice had fleas

Model 2: effect of burning treatment

```
mice2 <- glm(cbind(fleas, nofleas) ~ -1 + status, family = binomial(), data = mice)
coef(mice2) %>% plogis()
```

```
##      statusburn statusunburn
##      0.2962963    0.6785714
```

29.6% of mice at burned sites and 67.9% of mice at unburned sites had fleas

### Model 3: differences among sites

```
# -1+ is important! otherwise, glm will fit a line
mice3 <- glm(cbind(fleas, nofleas) ~ -1 + as.factor(site),
             family = binomial(), data = mice)
coef(mice3) %>% plogis()
```

```
## as.factor(site)1 as.factor(site)2 as.factor(site)3 as.factor(site)4
##      0.3043478      0.2500000      0.2903226      0.7000000
## as.factor(site)5 as.factor(site)6
##      0.8888889      0.6000000
```

### b. AIC

```
AIC(mice1, mice2, mice3)
```

```
##      df      AIC
## mice1  1 34.10879
## mice2  2 24.99069
## mice3  6 27.53858
```

Model 2 wins because it has the lowest AIC

### c. LRT

```
lrtest(mice1, mice3)
```

```
## Likelihood ratio test
##
## Model 1: cbind(fleas, nofleas) ~ 1
## Model 2: cbind(fleas, nofleas) ~ -1 + as.factor(site)
##   #Df   LogLik Df Chisq Pr(>Chisq)
## 1    1 -16.0544
## 2    6 -7.7693  5 16.57  0.005391 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 3 is significantly better than model 1

```
lrtest(mice1, mice2)
```

```
## Likelihood ratio test
##
## Model 1: cbind(fleas, nofleas) ~ 1
## Model 2: cbind(fleas, nofleas) ~ -1 + status
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    1 -16.054
## 2    2 -10.495  1 11.118  0.0008549 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 2 is significantly better than model 1

```
lrtest(mice2, mice3)
```

```
## Likelihood ratio test
##
## Model 1: cbind(fleas, nofleas) ~ -1 + status
## Model 2: cbind(fleas, nofleas) ~ -1 + as.factor(site)
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    2 -10.4953
## 2    6 -7.7693  4 5.4521      0.244
```

model 3 is not significantly different from model 2 (yes, they are nested. Coefficients for each site *could* turn out to perfectly match burn status). Note: you **cannot** compare p-values between likelihood ratio tests and interpret that as relative support for one model over another.

## 2. Goldenrod ball galls

### a. MLE proportion of successful galling events for the entire dataset

```
m1 <- glm(cbind(galls, punctures - galls) ~ 1, family = binomial(), data = galls)
coef(m1) %>% plogis()
```

```
## (Intercept)
## 0.1809524
```

MLE of galling success rate is 0.18

### b. Compare three models:

1. the model you fit above in part a,
2. separate galling rates for fertilized and unfertilized plants
3. model that estimates different galling rates by clone.

```
m2 <- glm(cbind(galls, punctures - galls) ~ fertilized, family = binomial(), data = galls)
m3 <- glm(cbind(galls, punctures - galls) ~ as.factor(clone), family = binomial(), data = galls)

AIC(m1, m2, m3)
```

```
##      df      AIC
## m1   1 97.60337
## m2   2 98.90307
## m3   6 64.39147
```

Model 3, differences by clone, wins!

### c. Confidence intervals

For the third model, the one that estimates different rates for each clone, get the MLE galling success on each clone along with confidence intervals. Give an explanation for the 95% confidence interval for clone #5

```
m3 <- glm(cbind(galls, punctures - galls) ~ -1 + as.factor(clone),
          family = binomial(link = "logit"), data = galls)
coef(m3) %>% plogis()
```

```
## as.factor(clone)1 as.factor(clone)2 as.factor(clone)3 as.factor(clone)4
##      3.088235e-01      2.835821e-01      1.384615e-01      1.052632e-01
## as.factor(clone)5 as.factor(clone)6
##      5.705164e-10      2.258065e-01
```

```
confint(m3) %>% plogis()
```

```
## Waiting for profiling to be done...
```

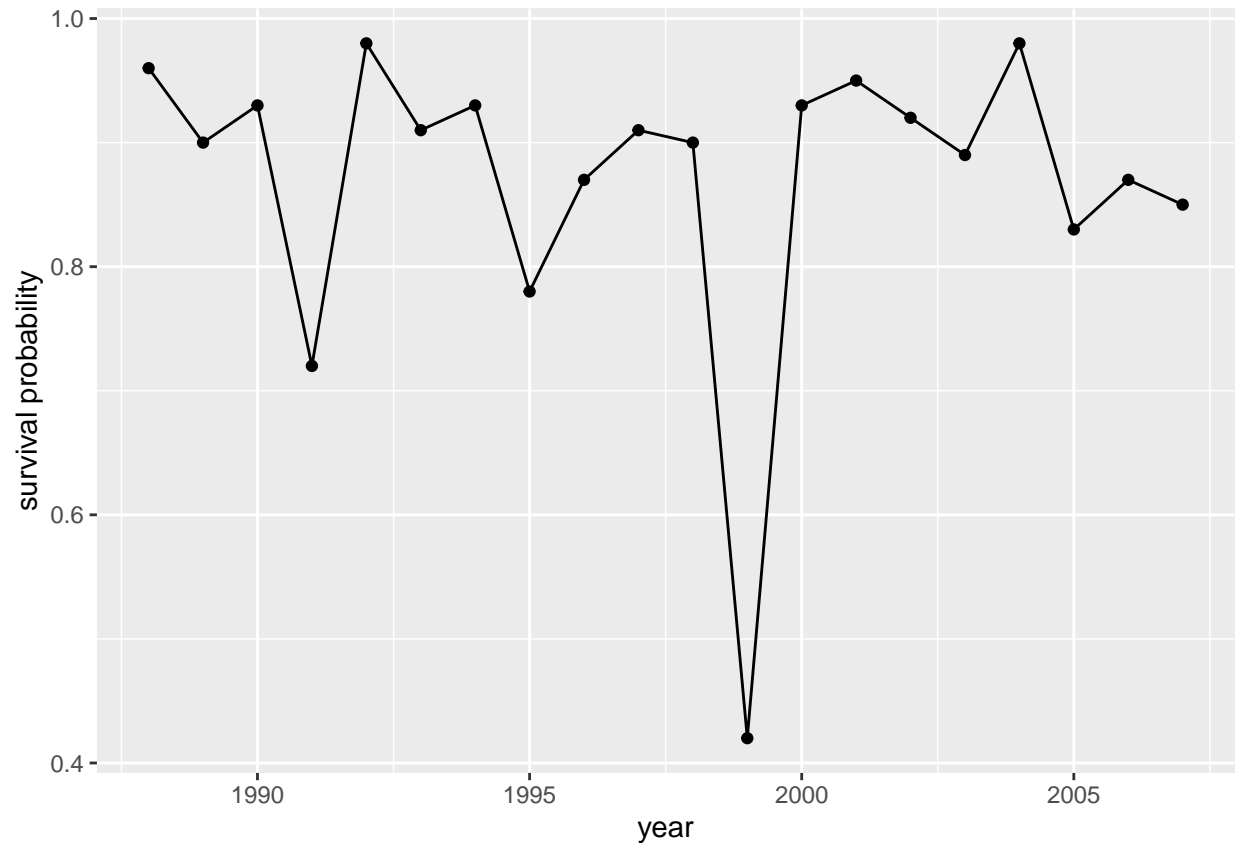
```
##              2.5 %    97.5 %
## as.factor(clone)1 0.20750241 0.4242916
## as.factor(clone)2 0.18515124 0.3982622
## as.factor(clone)3 0.06907135 0.2357409
## as.factor(clone)4 0.04323093 0.2018644
## as.factor(clone)5 0.00000000 1.0000000
## as.factor(clone)6 0.14898445 0.3175298
```

Because there were no successful galls for clone 5, it's impossible to estimate a probability of galling. This is why the 95% CI spans the entire range of possibilities from 0 to 1.

## 3. Bitterroot milkvetch.

### a. Make a plot of survival probability over time.

```
p1 <- ggplot(plant, aes(x = year, y = alive/(alive + dead))) +
  geom_point() +
  geom_line() +
  labs(y = "survival probability")
p1
```



## b. GLM for linear change in survival probability over time.

Use the coefficients of that model to write the equation for the line.

```
m1 <- glm(cbind(alive, dead) ~ year, family = binomial(link = "identity"), data = plant)
coef(m1)
```

```
## (Intercept)      year
## 4.508933798 -0.001820983
```

$$\frac{k}{N} = 4.5089 - 0.00182 \times year$$

Alternatively, if you used the logit link:

```
m1a <- glm(cbind(alive, dead) ~ year, family = binomial(link = "logit"), data = plant)
coef(m1a)
```

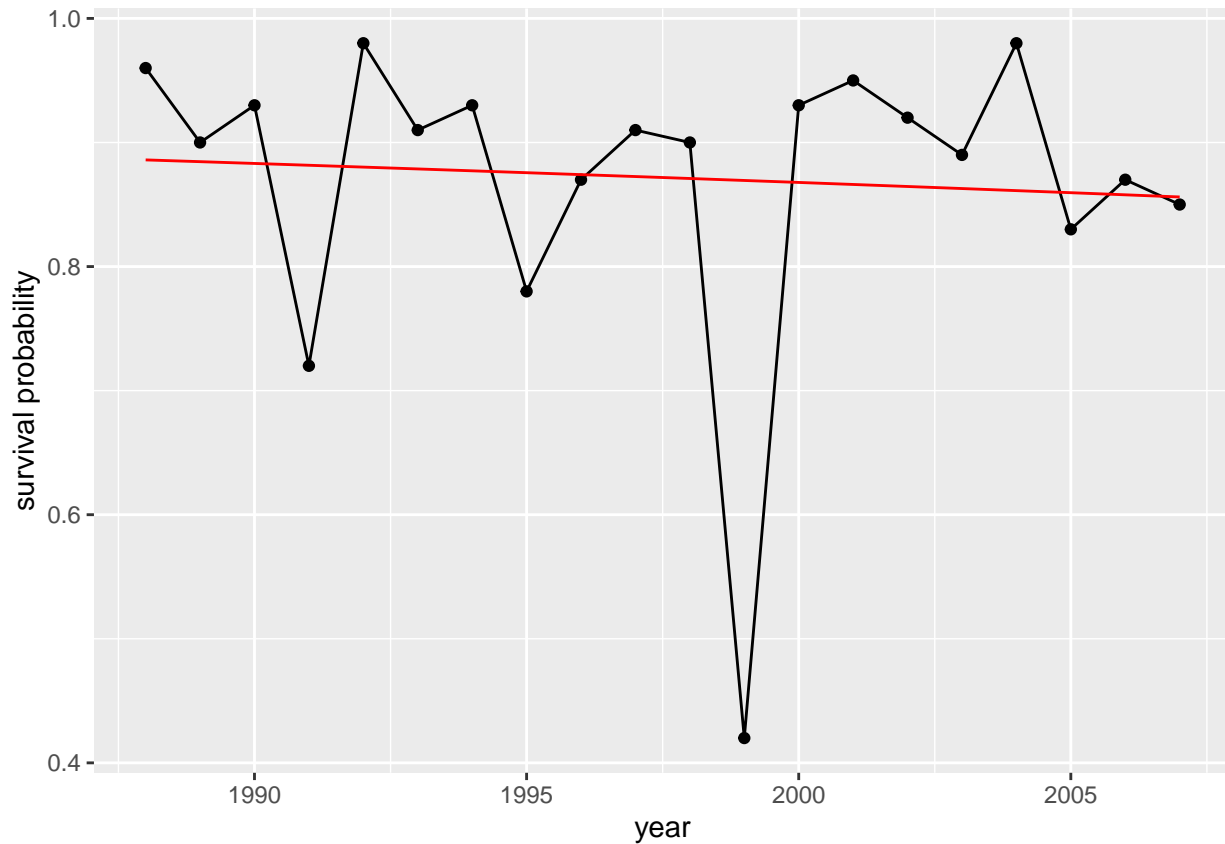
```
## (Intercept)      year
## 29.97864870 -0.01404852
```

$$\text{logit}\left(\frac{k}{N}\right) = 29.98 - 0.014 \times year$$

### c. Add the fit line

Add the fit line to the plot using the `predict()` function

```
p1 + geom_line(aes(y = predict(m1a, type = "response")), color = "red")
```



### d. Significance test

Use model comparison (LRT or AIC, your choice) to determine if there is a significant change in survival probability through time over this time period.

```
m0 <- glm(cbind(alive, dead) ~ 1, family = binomial(link = "logit"), data = plant)
lrtest(m0, m1)
```

```
## Likelihood ratio test
##
## Model 1: cbind(alive, dead) ~ 1
## Model 2: cbind(alive, dead) ~ year
##   #Df LogLik Df  Chisq Pr(>Chisq)
## 1    1 -141.04
## 2    2 -140.19  1  1.6987    0.1925
```

No significant effect of year on survival probability ( $p = 0.193$ )