QSCI 482 Lab 5

Simon Hans Edasi

2023-11-03

What is a reasonable value for fish survival?

In an experimental lake in Minnesota, 3,000 age-1 paddlefish are stocked every year (added to the lake). You have been told that 60% survive every year in this lake, so you set up an experiment where you randomly sample 200 fish from the lake, and find their ages, resulting in the following data.

1.) In R, store the observed counts in a vector called Nobserved using the c() function.

```
age <- c(1,2,3,4,5,6,7,8,9,10)
Nobserved <- c(47,41,26,27,21,10,12,7,4,5)

musn <- 3000

preds <- vector()
preds[1] <- musn
for (i in 2:10)
  preds[i] <-(0.6 * preds[i-1])
Npredicted <- (preds)

fobs <- Nobserved / 200
fpred <- Npredicted / musn
```

What is the predicted count in each category?

```
cbind(age, Nobserved,Npredicted)
```

```
##       age Nobserved Npredicted
## [1,]    1       47 3000.00000
## [2,]    2       41 1800.00000
## [3,]    3       26 1080.00000
## [4,]    4       27  648.00000
## [5,]    5       21  388.80000
## [6,]    6       10  233.28000
## [7,]    7       12  139.96800
## [8,]    8        7   83.98080
## [9,]    9        4   50.38848
## [10,]  10        5   30.23309
```

Divide by the sum of all the fish to find the proportion of fish at each age in the lake, and multiply by 200 to find the predicted number of fish at each age in a sample of 200 fish. Write the code to do this, and save the results in a vector called Npredicted. The sum of the values in Npredicted should be 200.

```
pred_freq <- Npredicted / sum(Npredicted) * 200
obs_freq <- Nobserved / sum(Nobserved) * 200
cbind(obs_freq,pred_freq)
```

```
##        obs_freq  pred_freq
##  [1,]       47 80.4866721
##  [2,]       41 48.2920033
##  [3,]       26 28.9752020
##  [4,]       27 17.3851212
##  [5,]       21 10.4310727
##  [6,]       10  6.2586436
##  [7,]       12  3.7551862
##  [8,]        7  2.2531117
##  [9,]        4  1.3518670
## [10,]        5  0.8111202
```

Calculate the χ^2-statistic based on the values in Nobserved and Npredicted.

```
chisq <- sum(  (obs_freq - pred_freq)^2 / pred_freq  )
print(chisq)
```

```
## [1] 88.52447
```

Find the p-value for the χ^2-statistic using the χ^2 distribution.

```
p_value <- pchisq(chisq, df = 9,lower.tail = F)
print(p_value)
```

```
## [1] 3.217612e-15
```

What do you conclude about whether the survival is 60% in the lake?

$H_0 : \hat{f}_i = 0.60$

$H_A : \hat{f}_i \neq 0.60$

With a p-value of 3.217612e-15, we reject the null hypothesis that $\hat{f}_i = 0.60$

Repeat the analysis, but use the built-in function chisq.test() which requires a vector x of observed numbers, and a vector p of predicted probabilities

```
chisq.test(x = Nobserved, p = pred_freq / 200)
```

```
## Warning in chisq.test(x = Nobserved, p = pred_freq/200): Chi-squared
## approximation may be incorrect
```

```
##
##  Chi-squared test for given probabilities
##
## data:  Nobserved
## X-squared = 88.524, df = 9, p-value = 3.218e-15
```

Now, repeat the above analysis with survival values of 50%, 70%, 80%, 90%, and 100%. Are there any values of survival that are very unlikely? Are there any values of survival that are likely? Which value of survival is the most likely (among that set of 5 values)? What is the best value of survival that you can find?

```
library(glue)
chisq_calculator <- function(sr){

  age <- c(1,2,3,4,5,6,7,8,9,10)
  Nobserved <- c(47,41,26,27,21,10,12,7,4,5)

  musn <- 3000
```

```r
  preds <- vector()
  preds[1] <- musn
  for (i in 2:10)
    preds[i] <-(sr * preds[i-1])
  Npredicted <- (preds)

  fobs <- Nobserved / 200
  fpred <- Npredicted / musn

  pred_freq <- Npredicted / sum(Npredicted) * 200
  obs_freq <- Nobserved / sum(Nobserved) * 200
  chisq <- sum(  (obs_freq - pred_freq)^2 / pred_freq  )
  return(chisq)



}

survival_rates <- c(0.5, 0.7, 0.8,0.9, 1)
for (i in survival_rates){
  chisq <- chisq_calculator(i)
  p_value <- pchisq(chisq, df = 9,lower.tail = F)
  cat(i,p_value,'// ')
}
```

```
## 0.5 1.549859e-73 // 0.7 0.09633789 // 0.8 0.7274143 // 0.9 5.092414e-05 // 1 3.076736e-18 //
```

A believable survival rate could be 0.7 or 0.8.