Assignment-2
Course: Data Visualization and Modelling

Professor: Pere Puig
Student: Jamia Begum
NIU: 1676891

Keating (2002) studied the annual numbers of females with cubs of-the-year in the Yellowstone grizzly bear population, from 1986 to 2001. The next table shows the number of unique females with cubs-of-the-year that was seen exactly j times during the year 1998.

Table 1: Sights of unique females with cubs-of-the-year (1998).

Sights	0	1	2	3	4	5	6	7
No. of bears (frequency, f_k)	?	11	13	5	1	1	0	2

This is multiple capture-recapture problems where the main objective is to know the total number of females.

1. Use Chao's estimator to estimate the total number of females.

Answer:

Let N be the total number of females, $\mathbf{x} = (x_1, x_2, \dots, x_n)$ a sample of size n, and let f_k denote the number (frequency) of x_i equal to k, k = 1,2,…,7. It is evident that
$f_1 + f_2 + \_\_\_ + f_7 = n$

Therefore, n=11+13+5+1+1+0+2=33

Let f_0 denote the number of non-observed zeros, to be estimated. The size of the complete sample (counting the zeros) would be

$$N = f_0 + n$$

From the well-known Chao's lower bound estimator, the lower bound estimates of f_0 is the following:

$f_{0,c}= f_1^2/2f_2=11^2/2*13 =121/26 = 4.65 \sim 5$

Therefore, The total number of females using **Chao's estimator is N**= 33+5 =38

2. A random variable X follows a zero-truncated Poisson distribution when its probability function is,
$P(X = k) =(e^{-\lambda} *\lambda^k)/(1 - e^{-\lambda})*k!$
, k = 1, 2,...
Assuming that the counts shown in Table 1 follow a zero-truncated Poisson distribution, estimate the parameter λ of this distribution (for instance, by maximum likelihood or by the method of the moments). Then, estimate the total number of females using the Horvitz-Thompson estimator, $N\text{^} =n/(1 - e^{-\hat{\lambda}})$
Where n is the total number of observed females. Use parametric bootstrap to compute a confidence interval of the total number of females N. Compare the results with those obtained with Chao's estimator.

Answer:
Since the random variable x_i follows Poisson distribution, the likelihood of $(x_1,x_2,\dots,x_{33})$ is
$L(\lambda|x) = \prod_{1}^{33} ((e^{-\lambda} *\lambda^k)/(1 - e^{-\lambda})*k!$

So, the log-likelihood is

$$\ell(\lambda \mid x) = -33\log(e^{\lambda} - 1) + 33 * \bar{x} * \log \lambda$$

Here, $\bar{x}$ is the sample mean given by $= (1 * 11 + 2 * 13 + 3 * 5 + 4 * 1 + 5 * 1 + 6 * 0 + 7 * 2)/33$

$$= 2.2727$$

and, its derivative with respect to λ is

$$\partial\ell/\partial\lambda = e^{\lambda}/(1 - e^{\lambda}) + \bar{x}/\lambda$$

Now, MLE(the maximum likelihood estimator λ) must satisfy,

$$\partial\ell/\partial\lambda = e^{\lambda}/(1 - e^{\lambda}) + \bar{x}/\lambda = 0$$

Multiplying both sides by$(1 - e^{\lambda})$ and λ,

$$\lambda * e^{\lambda} + (1 - e^{\lambda}) * \bar{x} = 0$$

$$\text{or, } e^{\lambda} \cdot (\lambda - \bar{x}) + \bar{x} = 0$$

$$\text{or, } e^{\lambda} \cdot (\lambda - \bar{x}) = -\bar{x}$$

$$\text{or, } e^{\lambda} \cdot (\lambda - 2.2727) = -2.2727$$

Therefore,soving this equaltion we get the MLE estimated value of λ is 1.95

Now, estimating the total number of females using
the **Horvitz-Thompson estimator**, $\hat{N} = n/(1 - e^{-\hat{\lambda}}) = \hat{N} = 33/(1-e^{-1.95}) =$ **38.5**

Using R to estimate the parameter and the total number of female bear:

Source Visual

```r
#to calculate MLE of the parameter λ in the zero truncated Poisson distribution
#input the data for the random variable
#rep(x,n) gives n times the r.v x has been repeated

sights  <-  c(rep(1,11),rep(2,13),rep(3,5),4,5,rep(7,2))

#get the size of the sample
n<-length(sights)
#define the log-likelihood function as follows
loglike<-function(lamda,sights){
  -n*log(exp(lamda)-1)+n*mean(sights)*log(lamda)
}
# Find maximum likelihood estimate(MLE) of lambda
lambda_MLE <- optimize(loglike, c(0, 33), sights, maximum = TRUE)$maximum
lambda_MLE
```

22:47 Chunk 2

Console Terminal Background Jobs

R 4.2.1 · F:/INTERMATH/intermath 2021-2023/spain/DV/Resampling/ex-2/DVM_Assig2Sim_PedroPaixao/

```
> loglike<-function(lamda,sights){
+   -n*log(exp(lamda)-1)+n*mean(sights)*log(lamda)
+ }
> # Find maximum likelihood estimate(MLE) of lambda
> lambda_MLE <- optimize(loglike, c(0, 33), sights, maximum = TRUE)$maximum
> lambda_MLE
[1] 1.949087
> #estimating the total number of females using
> #the Horvitz-Thompson estimator
> N<-n/(1-exp(-lambda_MLE))
> N
[1] 38.47966
```

We get the same result from R.

Now, Using simulation of 1000 parametric bootstrap in R to compute a confidence interval of the total number of females N,

```
nsim<-1000 #setting the number of simulation
#we will do simulation for computing the parameter lambda
lambdab<-numeric( nsim) #create a vector to store data

for (i in 1:nsim){

  sightsb <- sample (sights , n, replace =T)

  # Obtaining parameter estimates via bootstrap
  lambdab[i] <- optimize(loglike, c(0, 33), sightsb, maximum = TRUE)$maximum

}

# Using Horvitz-Thompson estimator to obtain total number of females
count <-n/(1-exp(-lambdab))
# Bootstrap estimate of total number of females(N)
Nb<-mean(count)
Nb
hist(count)
```

```
R 4.2.1 · F:/INTERMATH/intermath 2021-2023/spain/DV/Resampling/ex-2/DVM_Assig2Sim_PedroPaixao/
> # Bootstrap estimate of total number of females(N)
> Nb<-mean(count)
> Nb
[1] 39.06166
> hist(count)
> # Calculating 95% confidence interval for N
> quantile(count, probs = c(0.025, 0.975))
    2.5%    97.5%
35.51811 44.66032
```

Results Comparison:

Methods	**Chao's Estimator**	**Horvitz-Thompson Estimator**	**Parametric Bootstrap**
Counted Number of Female Bear	38	38.5	39 with confidence interval [35.5,44.7]