Mini-project

YOUNG JAE LEE

2018.12.25

Personal motive and purpose:

My grandfather had suffered from dementia for several years before he died. When he was suffering from dementia, my parents had difficulty supporting him. Alzheimer's disease caught my attention in appling bayesian statistical methods.

The purpose : to predict future population exposed to higher risk of Alzheimer's disease because of APOE e4.

What is Alzheimer's disease?

- The disease is named after German psychiatrist Dr. Alois Alzheimer.
- It is characterized by unusual behavioral symptoms: including short-term memory loss in a middle-aged patient.
- As the disease progress, more symptoms appear and patients cannot carry out ordinary tasks. For example, some patients wash their hands in toilet
- Different nerve cells, plaques, and tangles are seen in people who suffer from this disorder.
- Health officials predict that the U.S. population may result in 3 times as many cases by 2050.
- Various kinds of efforts are being made to heve better knowledge of Alzheimer's disease.
- Reference: The webpage of Natoinal Institutes of Health ->link

Related paper :

'The Distribution of Apolipoprotein E Genotype Over the Adult Lifespan and in Relation to Country of Birth'
American Journal of Epidemiology, Volume 181, Issue 3, 1 February 2015, Pages 214-217

According to the paper:

- The dna APOE e4 is related to onset of Alzheimer's disease.
- APOE dna has different forms of e2, e3, and e4.
- Frequency of different forms is reported to vary depending on ethnicity

The data:

- The data is about the frequency of APOE dna of 4,579 persons
- 4,210 persons are from northern countries (latitude >= 55 degree)

Methods used to analyze the data:

- Let y be each person with e4
- Set the probability of APOE dna e4 passing down to offspring as p
- Donors : from age 60-99. Receiver : from age 25-59.
- 659 donors have e4 dna, and 322 receivers have e4 dna
- $y\ Bernoulli(p) -> p(y1, ..., y659|p) \propto p^{322} \times (1-p)^{659-322}$
- Set non-informative prior of p as Beta(0.5,0.5)

Findings and implications(prediction):

$posterior \propto Beta(322.5, 337.5)$

To predict the future population of Denmark with e4 dna:

- Set the probability of APOE dna e4 passing down to offspring as p
- DNA : a group of pairs - e4 is inherited from pairs of e2/e4, e3/e4, and e4/e4
- Calculate the expected number of persons with e4 as below
- 1. calculate the number of persons with e2/e4, e3/e4, and e4/e4, respectively.
- 2. Set the number as a,b,c, respectively.
- 3. ((0 or 1) sampling from a with probability p)+((0 or 1)sampleing from b with probability p)+(c) is the expected number of offspring with e4 dna.
- Considering that the current population of Denmark is 5,750,000 : a = 1,437,500 b=c=143,750

```
a<-1437500;b<-143750;c<-143750

post_a<-rbeta(a,322.5,303.5)
prediction_a<-numeric(a)
for(k in 1:a){
  prediction_a[k]<-sample(0:1,size = 1, prob = c(1-post_a[k],post_a[k]))
}
pred_a<-length(prediction_a[which(prediction_a==1)])

post_b<-rbeta(b,322.5,303.5)
prediction_b<-numeric(b)
for(k in 1:b){
  prediction_b[k]<-sample(0:1,size = 1, prob = c(1-post_b[k],post_b[k]))
}
pred_b<-length(prediction_b[which(prediction_b==1)])

pred_a+pred_b+c
```

```
## [1] 957792
```

- If a couple has one child, about 958,000 chlidren would have e4 dna.
- If a couple has more than two chlidren, more than 1,900,000 would have e4 dna.

Conclusions as limitations:

- First of all, I must study more to have solid understanding of prior probabilities and statistical knowledge.
- It was a truly precious experience because I could learn from my mistake.
- Besides, I should study further to know about factors influencing the probability p.
- It is because Alzheimer's disease depends on not only genetic factors, but also environment.
- Varying proportion of APOE e4 depending on ethnicity should also be studied better.