26 BE 7082 + 26 BE 7028 + 20 BME 7082

Introduction to Data Science

Autumn 2020

MB Rao

Homework 8: Due Date: November 12, 2020 Maximum points: 30

Theme: Disease Progression Models

Use the data ‘aneur’ for this exercise from the ‘msm’ package. When you activate the package, the data automatically gets into the console. A Markov process with 4 states is used to model the data.

1. What is the dimension of the data? What is aneurysm? Describe the data and its variables. How many subjects are involved in the study? (Here state 4 is not death.)

1+2+3+1 points

**Dimensions of the data:**

> dim(aneur)

[1] 4337 4

**What is aneurysm?**

Aneurysm is the dilation or expansion of an artery. The data in particular is concerned with aortic aneurysm which is the ballooning of the abdominal aortic artery.

**Description of data:**

?aneur

Description

This dataset contains longitudinal measurements of grades of aortic aneurysms, measured by ultrasound examination of the diameter of the aorta.

Usage

aneur

Format

A data frame containing 4337 rows, with each row corresponding to an ultrasound scan from one of **838 men over 65 years of age**.

**Description of variables:**

**ptnum (numeric) Patient identification number**

**age (numeric) Recipient age at examination (years)**

**diam (numeric) Aortic diameter**

**state (numeric) State of aneurysm.**

The states represent successive degrees of aneurysm severity, as indicated by the aortic diameter.

**State 1 Aneurysm-free < 30 cm**

**State 2 Mild aneurysm 30-44 cm**

**State 3 Moderate aneurysm 45-54 cm**

**State 4 Severe aneurysm > 55 cm**

683 of these men were aneurysm-free at age 65 and were re-screened every two years. The remaining men were aneurysmal at entry and had successive screens with frequency depending on the state of the aneurysm. Severe aneurysms are repaired by surgery

**Subjects:**

There were a total of 838 men (over age 65) involved in the study.

2. Trace the progress of ptnum 1. 2 points

**Using the dplyr package, I was able to extract rows containing specific patients’ data as following:**

> ptnum1<-df%>%filter(ptnum==1)

> ptnum1

ptnum age diam state

1 1 60.00000 29 1

2 1 65.47671 29 1

3 1 67.50411 29 1

4 1 70.04384 29 1

5 1 72.07671 29 1

6 1 74.08767 29 1

7 1 76.03836 29 1

3. Trace the progress of ptnum 101. 2 points

**Using the dplyr package, I was able to extract rows containing specific patients’ data as following:**

> ptnum101<-df%>%filter(ptnum==101)

> ptnum101

ptnum age diam state

1 101 60.0000 29 1

2 101 65.2548 29 1

4. Estimate the crude transition matrix from the data. Interpret the matrix.

2+3 points

**Initial generator for iteration purposes:**

**Q <- rbind(c(0.25, 0.25, 0.25, 0.25), c(0.166, 0, 0.166, 0.166), c(0.25, 0.25, 0, 0.25),**

**+ c(0.1, 0.1, 0.1, 0))**

**Crude transition matrix:**

**> crude<-crudeinits.msm(state~years,ptnum, data=df, qmatrix=Q)**

**> crude**

[,1] [,2] [,3] [,4]

[1,] -0.02478157 0.02026629 0.00189012 0.002625166

[2,] 0.14163529 -0.28622131 0.14163529 0.002950735

[3,] 0.03397722 0.30579497 -0.88340768 0.543635498

[4,] 0.00000000 0.00000000 0.78102491 -0.781024913

**Interpretation:**

1. From state 1, there is a 2% chance of progressing to state 2 and a 0.2% chance of progressing directly to state 3 or state 4.

2. From state 2, there is a 14% chance of improving and going back to state 1 or worsening and progressing to state 3, and there is a 0.3% chance of progressing directly to state 4.

3. From state 3 there a 4% chance of improving and going directly back to state 1, a 30% chance of improving and going back to state 2 while a 54% chance of worsening and progressing to state 4.

4. From state 4 there is a 78% chance of improving and going back to state 3 while there is no chance of improving directly to either state 2 or state 1.

5. Estimate the infinitesimal generator. Use a reasonable initial generator.

2 points

**Initial Generator:**

Q <- rbind(c(0.25, 0.25, 0.25, 0.25), c(0.166, 0, 0.166, 0.166), c(0.25, 0.25, 0, 0.25),

+ c(0.1, 0.1, 0.1, 0))

**Infinitesimal generator:**

> aneur.msm<-msm(state~years, subject=ptnum, data=df, qmatrix=Q)

> aneur.msm

Call:

msm(formula = state ~ years, subject = ptnum, data = df, qmatrix = Q)

Maximum likelihood estimates

Transition intensities

Baseline

State 1 - State 1 -3.240e-02 (-3.711e-02,-2.829e-02)

State 1 - State 2 3.183e-02 ( 2.756e-02, 3.677e-02)

State 1 - State 3 5.662e-04 ( 4.537e-05, 7.066e-03)

State 1 - State 4 1.811e-06 ( 1.382e-20, 2.373e+08)

State 2 - State 1 8.433e-02 ( 6.178e-02, 1.151e-01)

State 2 - State 2 -2.511e-01 (-3.121e-01,-2.020e-01)

State 2 - State 3 1.668e-01 ( 1.262e-01, 2.204e-01)

State 2 - State 4 1.312e-06 ( 2.052e-61, 8.393e+48)

State 3 - State 1 1.264e-02 ( 1.734e-03, 9.215e-02)

State 3 - State 2 4.762e-01 ( 3.212e-01, 7.060e-01)

State 3 - State 3 -1.229e+00 (-1.588e+00,-9.511e-01)

State 3 - State 4 7.402e-01 ( 5.258e-01, 1.042e+00)

State 4 - State 1 1.571e-12 ( 0.000e+00, Inf)

State 4 - State 2 3.284e-08 ( 0.000e+00, Inf)

State 4 - State 3 8.316e-01 ( 5.084e-01, 1.360e+00)

State 4 - State 4 -8.316e-01 (-1.360e+00,-5.084e-01)

-2 \* log-likelihood: 2797.756

[Note, to obtain old print format, use "printold.msm"]

6. Estimate the transition matrix when t = 1. Comment. 2+2 points

> pmatrix.msm(aneur.msm, t=1)

State 1 State 2 State 3 State 4

State 1 0.969346629 0.02806761 0.002087062 0.0004987025

State 2 0.074804484 0.80309349 0.091373798 0.0307282249

State 3 0.020958883 0.26098679 0.419694687 0.2983596417

State 4 0.006104747 0.09858593 0.335182508 0.5601268125

**Comment:**

**In t=1 years time,**

1. Starting from state 1, there are 97% chances of the patient to stay in state 1, 3% chance of progressing to state 2, 0.2% chance of progressing directly to state 3 and negligible (<0.1%) chance of progressing directly to state 4.

2. Starting from state 2, there is a 7.5% chance of improving and going back to state 1, 80% chance of staying at state 2, 9% chance of progressing to state 3 and 3% chance of progressing directly to state 4.

3. Starting from state 3, there is a 2% chance of improving and going directly to state 1, a 26% chance of improving and going back to state 2, a 42% chance of staying at state 3, and a 30% chance of worsening and progressing to state 4.

4. Starting from state 4, there is a 0.6% chance of improving and going directly back to state 1, 10% chance of improving and going directly to state 2 and a 34% chance of improving and going to state 3 and a 56% chance of staying at state 4.

7. Estimate the transition matrix when t = 5. Comment. 2+2 points

> pmatrix.msm(aneur.msm, t=5)

State 1 State 2 State 3 State 4

State 1 0.8732705 0.09683571 0.01811604 0.01177775

State 2 0.2610256 0.47156450 0.14565384 0.12175601

State 3 0.1502222 0.41674586 0.21628910 0.21674282

State 4 0.1111870 0.39117543 0.24349190 0.25414567

**Comment:**

**In t=5 years time,**

1. Starting from state 1, there are 87% chances of the patient to stay in state 1, 9.6% chance of progressing to state 2, 2% chance of progressing directly to state 3 and 0.1% chance of progressing directly to state 4.

2. Starting from state 2, there is a 26.1% chance of improving and going back to state 1, 41.67% chance of staying at state 2, 21.63% chance of progressing to state 3 and 21.67% chance of progressing directly to state 4.

3. Starting from state 3, there is a 15% chance of improving and going directly to state 1, a 41.67% chance of improving and going back to state 2, a 21.63% chance of staying at state 3, and a 21.67% chance of worsening and progressing to state 4.

4. Starting from state 4, there is a 11.1% chance of improving and going directly back to state 1, 39.11% chance of improving and going directly to state 2 and a 24.35% chance of improving and going to state 3 and a 25.41% chance of staying at state 4.

8. Estimate the transition matrix when t = 10. Comment. 2+2 points

> pmatrix.msm(aneur.msm, t=10)

State 1 State 2 State 3 State 4

State 1 0.7919089 0.1423850 0.03671079 0.02899528

State 2 0.3864546 0.3559783 0.13456388 0.12300330

State 3 0.2965564 0.3859915 0.16297816 0.15447395

State 4 0.2640387 0.3961211 0.17353753 0.16630264

**Comment:**

**In t=5 years time,**

1. Starting from state 1, there are 79.2% chances of the patient to stay in state 1, 14.24% chance of progressing to state 2, 3.67% chance of progressing directly to state 3 and 2.9% chance of progressing directly to state 4.

2. Starting from state 2, there is a 38.65% chance of improving and going back to state 1, 35.6% chance of staying at state 2, 13.45% chance of progressing to state 3 and 12.30% chance of progressing directly to state 4.

3. Starting from state 3, there is a 29.65% chance of improving and going directly to state 1, a 38.6% chance of improving and going back to state 2, a 16.3% chance of staying at state 3, and a 15.44% chance of worsening and progressing to state 4.

4. Starting from state 4, there is a 26.4% chance of improving and going directly back to state 1, 39.61% chance of improving and going directly to state 2 and a 17.35% chance of improving and going to state 3 and a 16.63% chance of staying at state 4.