Homework 3

Data Science
March 11, 2020

- Collaboration: Homeworks are individual work. Visit here for more info on academic integrity.
- Formatting: You are supposed to upload a file `hw3_writeup.pdf` . You can produce the file however you like (e.g. LaTeX, Microsoft Word, scanner), as long as it is readable.

Create a pdf file `hw3_writeup.pdf`. Answer the problems in the pdf file only.

Problem 1 (Markov Transition Matrix)

Assume the status of an automobile is one of the following four: (Excellent, Good, OK, Poor) Assume an annual Markov transition matrix equals to

$$\begin{bmatrix} 1/2 & 1/4 & 1/8 & 1/8 \\ 0 & 5/8 & 1/4 & 1/8 \\ 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Calculate:

(a)[2pt] If a car is in excellent condition, the probability it is in good condition in 2 years.

(b)[2pt] If a car is in good condition, the probability it is still in good condition in half a year.

(c)[4pt] For a new car, the probability that it is in excellent condition over time (a graph)

(d)[4pt] For a new car, the probability that it is in poor condition over time (a graph)

Problem 2 (Statistical Analysis)

An insurance company undertook a study in one of its health care plans to assess whether customers' reports of plan satisfaction were related to their leaving the plan within 1 year of joining. In a random sample of 130 customers who reported that they were satisfied with the plan, 35 left that plan within 1 year. In another random sample of 140 customers who reported that they were not satisfied with the plan, 58 left within 1 year.

(a)[4pt] Treating "satisfaction" as the exposure and "leaving the plan within one year" as the outcome, construct a 2×2 table for these data, compute and interpret the relative risk (RR) and odds ratio (OR) for leaving the plan.

(b)[3pt] Compute the 95% confidence interval for the OR, do you think it is different from 1?

(c)[4pt] Confirm your observations of last question using both χ^2 and Z test with the hypothesis below:

$$H_0 : \text{OR} = 1 \text{ vs } .H_1 : \text{OR} \neq 1$$

(d)[2pt] What other options/tests could have been used in order to see if "satisfaction" is related to "leaving the plan within one year"?

$$\begin{bmatrix} 1/2 & 1/4 & 1/8 & 1/8 \\ 0 & 5/8 & 1/4 & 1/8 \\ 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

E G O P

$$\begin{bmatrix} 1/2 & 1/4 & 1/8 & 1/8 \\ 0 & 5/8 & 1/4 & 1/8 \\ 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

let $A = \begin{bmatrix} 1/2 & 1/4 & 1/8 & 1/8 \\ 0 & 5/8 & 1/4 & 1/8 \\ 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

(a)[2pt] If a car is in excellent condition, the probability it is in good condition in 2 years.

by python, we know $A^2 = \begin{bmatrix} 0.25 & 0.2812 & \cdots \\ \vdots & \ddots & \\ & & \end{bmatrix}$

wanted = 0.2812

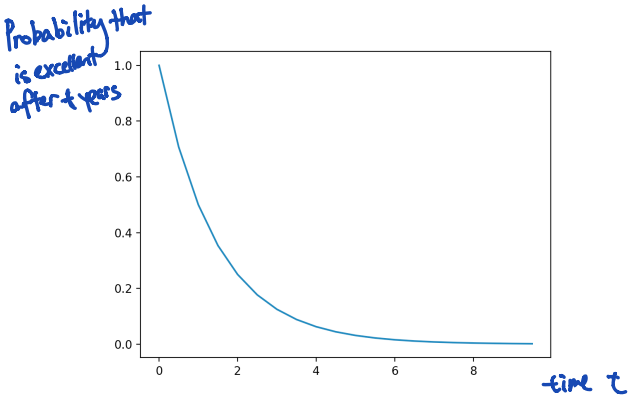
(b)[2pt] If a car is in good condition, the probability it is still in good condition in half a year.

By python, we know that $A^{\frac{1}{2}} = \begin{bmatrix} 0.70710678 & \cdots & \\ \vdots & \ddots & \\ & & \end{bmatrix}$

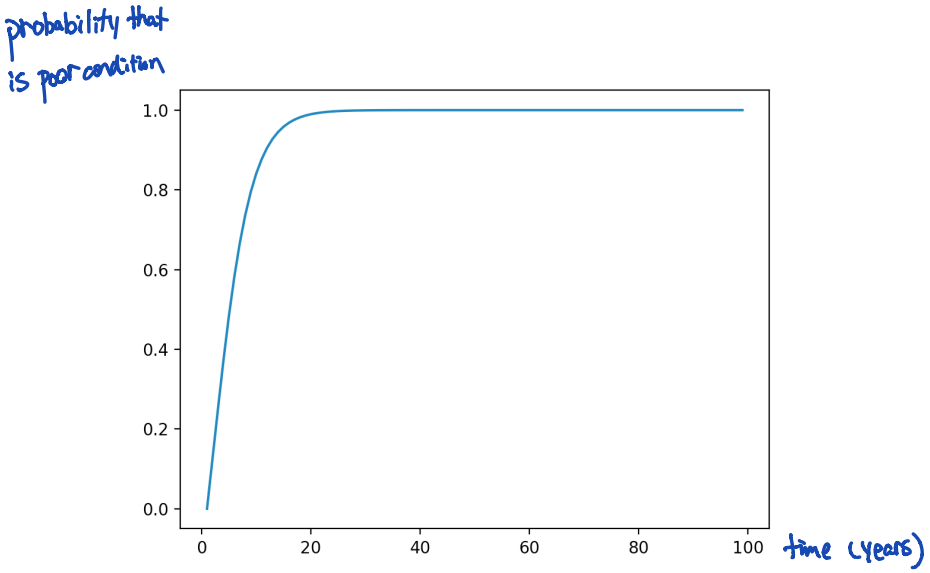
wanted = 0.7071

(c)[4pt] For a new car, the probability that it is in excellent condition over time (a graph)

P(Excellent after t years) = $\frac{1}{2}^t$



(d)[4pt] For a new car, the probability that it is in poor condition over time (a graph)



(a)[4pt] Treating "satisfaction" as the exposure and "leaving the plan within one year" as the outcome, construct a 2×2 table for these data, compute and interpret the relative risk (RR) and odds ratio (OR) for leaving the plan.

$$RR = \frac{\frac{95}{130}}{\frac{82}{140}} = \frac{95 \times 140}{130 \times 82} = 1.248$$

	Stay	leave
satisfy	95	35
not satisfy	82	58

$$OR = \frac{\frac{95}{35}}{\frac{82}{58}} = \frac{95 \times 58}{35 \times 82} = 1.920$$

(b)[3pt] Compute the 95% confidence interval for the OR, do you think it is different from 1?

$$CI = \exp\left(\log(OR) \pm \frac{1.96}{2} \cdot \sqrt{\frac{1}{95} + \frac{1}{35} + \frac{1}{82} + \frac{1}{58}}\right) = (1.15, 3.21)$$

$1 \notin (1.15, 3.21)$. Yes. 1 is different.

(c)[**4pt**] Confirm your observations of last question using both χ^2 and Z test with the hypothesis below:

$$H_0 : \text{OR} = 1 \text{ vs } .H_1 : \text{OR} \neq 1$$

$$\chi^2 = \frac{270(95 \cdot 58 - 35 \cdot 82)^2}{(95+35)(95+82)(35+58)(82+58)} = 6.281$$

For $\alpha = 0.05$. $\chi^{2(2)}_1$ critical value is 3.84. since $\chi^2 > 3.84$, we reject null hypothesis.

$$\chi^2_{\text{corrected}} = \frac{270(|95 \cdot 58 - 35 \cdot 82| - 0.5 \cdot 270)^2}{(95+35)(95+82)(35+58)(82+58)} = 5.655$$

For $\alpha = 0.05$. $\chi^{2(2)}_1$ critical value is 3.84. since $\chi^2 > 3.84$, we reject null hypothesis.

(d)[**2pt**] What other options/tests could have been used in order to see if "satisfaction" is related to "leaving the plan within one year"?

fisher's exact test