

Divorce Analysis Fall 2023

Background

Using divorce data, determine the factors that influence divorce.

Dataset

Use the dataset divorce.xlsx spreadsheet. This data is from the University of California Irvine machine learning dataset repository.

Divorce dataset

Reference: <https://archive.ics.uci.edu/ml/datasets/Divorce+Predictors+data+set>

Articles which discuss the dataset:

<https://medium.com/analytics-vidhya/a-logistic-model-for-predicting-divorce-rates-among-couples-implemented-using-the-statsmodel-api-763e03d29d68>

<https://dergipark.org.tr/en/pub/nevsosbilen/issue/46568/549416>

The divorce dataset from the Machine Learning Repository at University of California Irvine contains 170 responses to a survey. Of the 170 respondents, 84 (49%) were divorced, 86 (51%) remained married. They were asked 54 questions and the answers to the questions were collected on a 5-point scale (0-4):

| | |
|---|------------|
| 0 | Never |
| 1 | Seldom |
| 2 | Average |
| 3 | Frequently |
| 4 | Always |

The questions were:

1. If one of us apologizes when our discussion deteriorates, the discussion ends.
2. I know we can ignore our differences, even if things get hard sometimes.
3. When we need it, we can take our discussions with my spouse from the beginning and correct it.
4. When I discuss with my spouse, to contact him will eventually work.
5. The time I spent with my wife is special for us.
6. We don't have time at home as partners.
7. We are like two strangers who share the same environment at home rather than family.
8. I enjoy our holidays with my wife.
9. I enjoy traveling with my wife.
10. Most of our goals are common to my spouse.
11. I think that one day in the future, when I look back, I see that my spouse and I have been in harmony with each other.
12. My spouse and I have similar values in terms of personal freedom.
13. My spouse and I have similar sense of entertainment.
14. Most of our goals for people (children, friends, etc.) are the same.
15. Our dreams with my spouse are similar and harmonious.
16. We're compatible with my spouse about what love should be.
17. We share the same views about being happy in our life with my spouse
18. My spouse and I have similar ideas about how marriage should be

BUA 751: Machine Learning for Business

19. My spouse and I have similar ideas about how roles should be in marriage
20. My spouse and I have similar values in trust.
21. I know exactly what my wife likes.
22. I know how my spouse wants to be taken care of when she/he sick.
23. I know my spouse's favorite food.
24. I can tell you what kind of stress my spouse is facing in her/his life.
25. I have knowledge of my spouse's inner world.
26. I know my spouse's basic anxieties.
27. I know what my spouse's current sources of stress are.
28. I know my spouse's hopes and wishes.
29. I know my spouse very well.
30. I know my spouse's friends and their social relationships.
31. I feel aggressive when I argue with my spouse.
32. When discussing with my spouse, I usually use expressions such as 'you always' or 'you never'.
33. I can use negative statements about my spouse's personality during our discussions.
34. I can use offensive expressions during our discussions.
35. I can insult my spouse during our discussions.
36. I can be humiliating when we discussions.
37. My discussion with my spouse is not calm.
38. I hate my spouse's way of open a subject.
39. Our discussions often occur suddenly.
40. We're just starting a discussion before I know what's going on.
41. When I talk to my spouse about something, my calm suddenly breaks.
42. When I argue with my spouse, I only go out and I don't say a word.
43. I mostly stay silent to calm the environment a little bit.
44. Sometimes I think it's good for me to leave home for a while.
45. I'd rather stay silent than discuss with my spouse.
46. Even if I'm right in the discussion, I stay silent to hurt my spouse.
47. When I discuss with my spouse, I stay silent because I am afraid of not being able to control my anger.
48. I feel right in our discussions.
49. I have nothing to do with what I've been accused of.
50. I'm not actually the one who's guilty about what I'm accused of.
51. I'm not the one who's wrong about problems at home.
52. I wouldn't hesitate to tell my spouse about her/his inadequacy.
53. When I discuss, I remind my spouse of her/his inadequacy.
54. I'm not afraid to tell my spouse about her/his incompetence.

Assignment

What's due:

PowerPoint presentation due before class on Monday, October 23, 2023. The expected length of presentation is 15-20 minutes, approximately 20 slides. Please send me the slides at least one hour before class. You can describe the slides from your seat; I'll advance the slides.

Outline

Using the divorce dataset, perform a thorough analysis of the following aspects of the data.

1. Visualization (20 points)
 - a. Develop an overall view of relationship of dependent variable (divorce) with all continuous variables (see iris example with scatterplot matrix in Rcmdr). Since there are 54 questions, break this down into groups of 9 questions and Class, with six groupings (10 points)
 - b. Highlight at least two graphs where there are NO relationships between pairs of questions (see iris example with scatterplots in Rcmdr) (5 points)
 - c. Show at least two graphs with pairs of continuous variables and plot divorce events (see iris example in RStudio) (5 points)
2. Variance Inflation Factor (VIF) (20 points)
 - a. Perform a VIF analysis (see session 6 notes) (5 points)
 - b. Identify all questions (Atr) with a VIF less than 8 (5 points)
 - c. What are these questions (list the complete questions) (5 points)
 - d. Do these questions appear distinct from each other? Justify your answer (5 points)
3. Neural network – binary output (50 points)
 - a. Develop neural networks with binary output Class using the nine Attributes (Atr) from the VIF analysis and all data points using 70% training data, randomly selected. Vary the number of hidden nodes from 1 to 5 (15 points)
 - b. Test the neural networks with the remaining 30% testing data (5 points)
 - c. Measure the accuracy of the neural networks (5 points)
 - d. Develop a second neural network with binary output Class using the Attributes (Atr) Atr7, Atr43, Atr46, Atr48, Atr52 and 70% training data, randomly selected. Vary the number of hidden nodes from 1 to 5 (15 points)
 - e. Test these neural networks with the remaining 30% testing data (5 points)
 - f. Measure the accuracy of these neural networks (5 points)
4. Identify a list of lessons learned (10 points)
 - a. When do visualizations help? When do they not help? (5 points)
 - b. Which neural network was the most accurate? (5 points)