Data description

The dataset “Risk Factor Prevalence Study, 1980” was obtained from the Australian Data Archive (ADA). The survey documentation is public, available for anyone to read, but to download the data, one needs permission from the ADA. Citing requirements, as well as a copyright and disclaimer can be found below.

After requesting and gaining access to the dataset, I could download 7 files: the documentation and data of the survey from the website:
https://dataverse.ada.edu.au/dataset.xhtml?persistentId=doi:10.26193/BYE1RE

The data type is a survey; clinical data which was collected from Australian residents in capital cities between the ages 25–64 in 1980.
The primary investigator was the above-mentioned Australia, N. H. F. O., and the survey was partially funded by the Commonwealth Department of Health.

Citation Requirement: “Australia, N. H. F. O. Risk Factor Prevalence Study, 1980 [computer file]. Canberra: Australian Data Archive, The Australian National University”.
Rights & Disclaimer:
“Use of the material is solely at the user's risk. The depositor, The Australian National University and the Australian Data Archive shall not be held responsible for the accuracy and completeness of the material supplied.”
Copyright:

The dataset has 169 variables with 5,603 cases, the sas7bdat data file is 7.7 MB. The variables of the dataset follow the questions of the survey. From the 169 variables, I used 34 from the original dataset to predict a heart attack. I added a calculated variable: BMI, since it is a better indicator of obesity than weight and height separately. I used the formula:

$$\text{BMI} = \text{weight (kg)} / (\text{height (m)})^2$$

taken from the CDC website:
https://www.cdc.gov/nccdphp/dnpao/growthcharts/training/bmiage/page5_1.html

I renamed the variables from the question numbers to more specified names. Variable 8, “heart_attack”, from the list below is the target data, and all remaining variables are the feature data.

Variables description:

	Variable name	Description	Values	Data type
0	sex	sex of the patient	1: male 2: female	discrete
1	age	age of the patient	1: 24 – 29 2: 30 – 34 3: 35 – 39 4: 40 – 44 5: 45 – 49 6: 50 – 54 7: 55 – 59 8: 60 – 64	discrete
2	BMI	calculated from weight and height	numeric	continuous
3	chestpain	responder experienced chest pain	1: no 2: yes	discrete
4	chestpressure	esponder experienced chest pressure	1: no 2: yes	discrete
5	diabetes	responder has diabetes	1: no 2: yes	discrete
6	HBP	responder has HBP	1: no 2: yes	discrete
7	angina	responder has angina pectoris	1: no 2: yes	discrete
8	heart_attack	responder had heart_attack	1: no 2: yes	discrete
9	stroke	responder had stroke	1: no 2: yes	discrete

10	highcholesterol	responder has high cholesterol	1: no 2: yes	discrete
11	hightriglyc	responder has high triglyceride levels	1: no 2: yes	discrete
12	sleep_aid	frequency in which the responder takes sleeping aids	1: every day 2: a few days a week 3: once a week 4: occasionally 5: rarely 6: never	discrete
13	sedatives	frequency in which the responder takes sedatives or tranquilizers	1: every day 2: a few days a week 3: once a week 4: occasionally 5: rarely 6: never	discrete
14	vitamins	how often the responder takes vitamins or mineral supplements	1: every day 2: a few days a week 3: once a week 4: occasionally 5: rarely 6: never	discrete
15	sleep_hours	the number of hours the responder usually sleeps in a night	1: 5 hours or less 2: 6 hours 3: 7 hours 4: 8 hours 5: 9 hours or more	discrete
16	sleep_quality	how the responder describes their normal sleep	1: poor 2: fair 3: good	discrete
17	add_salt	how often the responder adds salt to food	1: not at all 2: sometimes 3: only after fasting 4: always	discrete
18	meat	how often the responder eats meat	1: every day 2: most days 3: at least once a week 4: infrequently 5: never	discrete

19	fat	how often the responder eats the fat on meat	1: every day 2: most days 3: at least once a week 4: infrequently 5: never	discrete
20	eggs	how many eggs the responder eats per week	numeric	numeric
21	fat_type	which type of fat the responder eats most often	1: butter 2: polyunsaturated margarine 3: other table margarines 4: I rarely eat any of these 5: I don't eat any of these	discrete
22	alcohol	how often the responder drinks alcohol	1: don't drink alcohol 2: less than once a week 3: on 1 or 2 days a week 4: on 3 or 4 days a week 5: on 5 or 6 days a week 6: every day	discrete
23	num_drinks	Number of drinks the responder consumes per occasion	1: don't drink alcohol 2: 1 or 2 drinks 3: 3 or 4 drinks 4: 5 to 8 drinks 5: 9 to 12 drinks 6: 13 to 20 drinks 7: more than 20 drinks	discrete

24	walking	how often the responder walks for exercise	1: three times or more a week 2: once or twice a week 3: once a month 4: rarely 5: never	discrete
25	cardio	how often the responder exercises vigorously	1: three times or more a week 2: once or twice a week 3: once a month 4: rarely 5: never	discrete
26	sport	how often the responder engages in other sports	1: three times or more a week 2: once or twice a week 3: once a month 4: rarely 5: never	discrete
27	hobby	other physical activities, like hobbies	1: three times or more a week 2: once or twice a week 3: once a month 4: rarely 5: never	discrete
28	work	how much time the responder spends walking while working	1: practically all 2: more than half 3: about half 4: less than half 5: almost none	discrete
29	SYS1	systolic blood pressure first measure	numeric	continuous
30	SYS2	systolic blood pressure second measure	numeric	continuous

31	DIA1	diastolic blood pressure first measure	numeric	continuous
32	DIA2	diastolic blood pressure second measure	numeric	continuous
33	SERCHOL	serum cholesterol level	numeric	continuous
34	HDLCHOL	HDL cholesterol level	numeric	continuous
35	TRIGLYC	triglyceride levels	numeric	continuous