2023 STAT 528: Homework 1

Due Wednesday, January 18 at 11:59pm

1 Background

In this exercise you will investigate the association of mortality with exercise variables and a number of other potential risk factors. The data arise from the Cardiovascular Health Study (CHS) which is described in detail in Fried et al. (1991).

The CHS study is a population-based, longitudinal study of coronary heart disease and stroke in adults aged 65 years and older. Study participants were recruited in 1989–1990 from four communities: Forsyth County, North Carolina; Sacramento County, California; Washington County, Maryland; and Pittsburgh, Pennsylvania. Eligible participants were sampled from Medicare eligibility lists in each area. It is thought that these lists are 98–99% complete for U.S. citizens aged 65 years and older. Eligible participants included all persons who were living in the household of each individual sampled from the lists and who were: 1) 65 years or older; 2) non-institutionalized; 3) expected to remain in the area for 3 years; and 4) able to give informed consent and did not require a proxy respondent. Participants were eligible for CHS whether or not they had a history of cardiovascular disease. However, persons were excluded if they were wheelchair-bound or if they were receiving hospice treatment or treatment for cancer.

The data for this exercise consist of the subset of participants who were recruited in the first wave of recruitment and were "healthy", that is had no history of heart or circulation disease, no restriction of daily activities by illness and no medications that would indicate heart disease. A large number of variables were determined for each study participant at *baseline*, that is, at the time of recruitment. A number of these measurements were also determined approximately three years after recruitment (it may be assumed that the season was the same for both baseline and 3 year examinations). The baseline examination consisted of a home interview and a clinic examination. During the home interview, information was collected on

prior medical history, medical usage and physical activity. Information was also obtained regarding the presence of impairments of physical functioning. The clinic examination included a fasting blood draw and seated blood pressure measurements.

In the data provided, mortality refers to the status of study participants in 1998.

With regard to the exercise measurement, participants were asked if they engaged in any of 15 leisure-time activities over the prior 2 weeks. For each activity, information was also obtained on the frequency and duration of participation (time spent) in the activity; this information was used to estimate the kilocalories of energy expended and the time spent per week in leisure-time physical activity. Participants were also asked the usual pace of walking outside the home.

Based on the highest intensity leisure-time activity reported over the prior 2 weeks, an *exercise intensity* variable was derived. Participants were categorized as having engaged in high-intensity, moderate-intensity, or low-intensity, where high-intensity activities were estimated to require a work metabolic rate/resting metabolic rate of > 6 METs. MET is short for metabolic equivalent and relates the intensity relative to resting so that, for example, a 6 MET activity requires a metabolic rate 6 times that when resting.

Participants who engaged in one or more of six high-intensity activities including swimming, hiking, aerobics, tennis, jogging, or raquetball, or who walked for exercise at a brisk (> 4 mph) pace were categorized as having engaged in high-intensity activity (> 6 METs). Participants who engaged in one or more of nine light- or moderate-intensity activities (< 6 METs) including gardening, mowing, raking, golf, bowling, biking, dancing, calisthenics, or exercise cycle, or who walked for exercise at an average pace (2–3 mph) or fairly brisk (3–4 mph) were categorized as having engaged in moderate-intensity activity. Participants who did not report participating in any of the 15 leisure-time activities and who walked for exercise at a casual or strolling pace (< 2 mph) were categorized as having engaged in low-intensity activity. If an individual was engaged in no activity and no walking they were categorized as "no exercise". The same questionnaire was used for the baseline and 3 year measurements.

The *blocks walked* variable was given by the number of blocks walked in the previous 2 weeks.

Further information on the exercise variables may be found in Siscovick et al. (1997).

It is known that exercise reduces weight, blood pressure, and risk of diabetes, and that smokers are at higher risk of death than non-smokers.

2 Data

Data Documentation

Table 1 contains descriptions of the data set. The outcome variable of interest is mortality. The other variables are baseline measurements, 3 year exercise measurements, myocardial infarction, and mortality status at 3 years.

Data access

A file containing the data can be obtained from the class website.

3 Questions

1. What are the different components of the scientific question for this problem, i.e. the population, response (outcome), and the specific aims?

2. The goal of the analysis is to determine whether there is an association between mortality at the end of the study (1998) and exercise. Your analysis should revolve around the following:

- How are the exercise variables related to each other, and how do they change over time?
- Describe the relationship between the exercise variables and baseline variables, not including mortality and myocardial infarction status.
- Investigate the association between mortality at the end of the study and baseline variables, not including the exercise variables.

- Do these data suggest that exercise is beneficial with respect to mortality in healthy individuals over 65 years of age? Explain your answer.

4 Report Details

Using .tex file or word document, write a report that answers each of the 4 questions in Section 3. Be sure that for each question you provide a brief summary of your findings and/or conclusions in addition to more detailed writing.

Your report should consist of a main body that includes no more than 10 pages of text and key figures/tables. Make sure you cite all the references you use in your analysis. You may include a short Appendix of supplementary material, which will not be read closely by the instructors. Your final upload should include both your final .tex or word document file and the compiled report (either .pdf or .html).

References

Fried, L.P. et al. (1991). The Cardiovascular Health Study: Design and Rationale. *Annals of Epidemiology*, **1**, 263–276.

Siscovick, D.S. et al. (1997). Exercise intensity and subclinical cardiovascular disease in the elderly. *American Journal of Epidemiology*, **145**, 977–986.

VARIABLE	DESCRIPTION
MORTALITY	0=alive, 1=dead (status at 1998)
CLINIC	1=Sacramento, 2=Forsyth, 3=Washington, 4=Pittsburgh
INITDATE	Date of recruitment (in days, measured from a day zero of no significance)
MORT3	Indicator of mortality, 0=alive, 1=dead, after 3 years
INCMI	1=heart attack (myocardial infarction), 0=no MI (status at 1998)
SEASON	Season at baseline, 1=summer, 2=fall, 3=winter, 4=spring
GENDER	0=female, 1=male
AGE	Age at baseline (in years)
WEIGHT	Weight at baseline (in KGs)
WEIGHT50	Recalled weight at 50 years old (in KGs)
GRADE	Years of education
ARTH	1=arthritis, 0=none (at baseline)
SBP	Systolic blood pressure (in mm of HG)
PKYRS	Pack years of smoking history. Number of years as smoker $\times$ number of packs/day $\times$ 365
DIAB	Diabetes, 1=none, 2=borderline diabetes, 3=diabetes
INCOME	Household income (in 1000s of dollars)
	1 $\leq$ 5, 2=5–8, 3=8–12, 4=12–16, 5=16–24, 6=24–35, 7=35–50, 8$\geq$50
EXINT0	Baseline measure of exercise intensity,
	0=no exercise, 1=low intensity, 2=moderate intensity, 3=high intensity
BLOCK0	Baseline measure of blocks walked in last 2 weeks (at about 12 blocks/mile)
KCAL0	Baseline measure of estimated kilocalories expended in exercise activity in past 2 weeks
EXINT3	3 year measure of exercise intensity,
	0=no exercise, 1=low intensity, 2=moderate intensity, 3=high intensity
BLOCK3	3 year measure of blocks walked in last 2 weeks (at about 12 blocks/mile)
KCAL3	3 year measure of estimated kilocalories expended in exercise activity in past 2 weeks

Table 1: Variable names and description.