Homework 01

Spencer Pease

10/07/2019

Population Overview

This study aimed to observe the population of elderly (65+ year) and generally healthy adults in order to better access the risk factors associated with cardiovascular and cerebrovascular disease. To do this, the study randomly selected US adults enrolled in Medicare who were at least 65 years old and generally healthy. Due to the high enrollment of of individuals in the study, the cohort presents a good representation of older adults in the US.

In total, this study included **735** adults aged **65** to **99**. Of these individuals **369** were reported as female, and **366** as male.

Variable Description

In this report we are examining a subset of variable within the data. These variables are listed below, along with the type of variable they represent:

Variable	Type
age	Quantitative continuous (ratio, measured discretely)
male	Binary
weight	Quantitative continuous (ratio)
height	Quantitative continuous (ratio)
atrophy	Quantitative continuous (ratio, measured discretely)
obstime	Quantitative continuous (censored)
death	Binary (indicator of censoring time)

Summary Statistics

We can also generate summary statistics for the subset of variables we are interested in. Not every variable benefits form the same summary statistics, so only relevant statistics for each variable are displayed in the tables below, along with the reasoning for why some variables include fewer summary statistics.

- *male* is a binary variable, so all we need to understand the distribution is the mean to understand the proportion of data in each category, as well as the min and max to verify that the variable is indeed binary.
- *obstime* is a censored variable, so it doesn't make sense to include information about the percentiles or mean, since the data has inherent missing-ness.
- *death* is an indicator of the censoring time, so again it doesn't make sense to include summary statistics due to the inherent missing-ness of the variable. The mean is included however to get an idea of the proportions of the uncensored observations, along with the min and max to verify the binary nature of the uncensored observations.

Table 2: All observations

variable	obs	ms	mean	sd	min	q25	median	q75	max
age	735	0	74.57	5.45	65	71.0	74.0	78.00	99.0
male	735	0	0.50	-	0	-	-	-	1.0
weight	735	0	159.95	30.74	74	138.5	158.0	179.00	264.0
height	735	0	165.78	9.71	139	158.0	165.9	173.25	190.5
atrophy	735	0	35.98	12.92	5	27.0	35.0	44.00	84.0
obstime	735	0	-	-	-	-	-	-	-
death	735	0	0.18	-	0	-	-	-	1.0

In addition to looking at the complete set of observations, we can split the data to look at trends within different groups of the cohort. Here we look at the summary of the female and male members of the sample separately. Note that we don't display any statistics for *male* since we have filtered the data to include only one category.

Table 3: Female observations (male = 0)

variable	obs	ms	mean	sd	min	q25	median	q75	max
age	369	0	74.41	5.26	65	71	73.0	78.0	91.0
male	369	0	-	-	-	-	-	-	-
weight	369	0	146.54	28.60	74	127	143.0	162.0	257.0
height	369	0	158.52	6.35	139	155	158.1	162.4	177.5
atrophy	369	0	32.91	12.22	5	24	32.0	40.0	82.0
obstime	369	0	-	-	-	-	-	-	-
death	369	0	0.13	-	0	-	-	-	1.0

Table 4: Male observations (male = 1)

variable	obs	ms	mean	sd	min	q25	median	q75	max
age	366	0	74.73	5.64	66	71.00	74	78.0	99.0
male	366	0	-	-	-	-	-	-	-
weight	366	0	173.47	26.66	110	154.62	171	190.0	264.0
height	366	0	173.10	6.47	157	168.70	173	177.3	190.5
atrophy	366	0	39.09	12.88	10	30.00	39	48.0	84.0
obstime	366	0	-	-	-	-	-	-	-
death	366	0	0.23	-	0	-	-	-	1.0

Minimum Follow-up

In order to get a better understanding of how long members of the cohort were observed, one thing we can do is look at the minimum amount of years a person who was living at the end of the study was observed. To do this, we must filter the data to include only members who were alive at the end of the study (*death = 0*), convert observation time from days to years (*obstime_yr = obstime / 365*), then find the minimum value (*min(obstime_yr)*).

Doing this, we arrive at a minimum follow-up time of **5.01 years**.