

Unit 4 – Introduction to R Studio (Part 2)

Homework

Due: Wednesday October 21, 2020

Last Submission Date for Credit: Wednesday October 28, 2020

Before you begin: Download from the course website
[hersdata.csv](#)

Description of [hersdata.csv](#)

These data are a simple random sample of $n=2763$ observations from the HERS study. The HERS study was a randomized clinical trial of hormone therapy (estrogen plus progestin) for the reduction of cardiovascular disease risk in post-menopausal women with established coronary disease. Study participants were $n=2,763$ women who were: (1) post-menopausal (2) with coronary disease; and (3) with an intact uterus. *We will be using just a few of the variables in this dataset in this assignment.*

Following is the codebook produced by Stata

. codebook, compact

Variable	Obs	Unique	Mean	Min	Max	Label
HT	2763	2	.4994571	0	1	random assignment to hormone therapy
age	2763	36	66.64857	44	79	age in years
raceth	2763	3	1.146942	1	3	race/ethnicity
nonwhite	2763	2	.1129207	0	1	nonwhite race/ethnicity
smoking	2763	2	.1302932	0	1	current smoker
drinkany	2761	2	.3915248	0	1	any current alcohol consumption
exercise	2763	2	.3865364	0	1	exercise at least 3 times per week
physact	2763	5	3.200145	1	5	comparative physical activity
globrat	2760	5	3.063406	1	5	self-reported health
poorfair	2760	2	.240942	0	1	poor/fair self-reported health
medcond	2763	2	.3720594	0	1	other serious conditions by self-report
htnmeds	2763	2	.8215708	0	1	anti-hypertensive use
statins	2763	2	.3633731	0	1	statin use
diabetes	2763	2	.2645675	0	1	diabetes
dmpills	2763	2	.0962722	0	1	oral DM medication by self-report
insulin	2763	2	.0988056	0	1	insulin use by self-report
weight	2761	549	72.73484	37.5	132	weight (kg)
BMI	2758	1465	28.57925	15.21	54.13	BMI (kg/m^2)
waist	2761	496	91.73676	56.9	170	waist (cm)
WHR	2760	385	.869979	.624	1.218	waist/hip ratio
glucose	2763	202	112.1516	29	298	fasting glucose (mg/dl)
weight1	2613	540	72.04141	37.7	142	year 1 weight (kg)
BMI1	2610	1448	28.36211	14.73	54.04	year 1 BMI (kg/m^2)

codebook, compact

```
. codebook, compact
```

Variable	Obs	Unique	Mean	Min	Max	Label
waist1	2612	490	91.12136	59	142	year 1 waist (cm)
WHR1	2612	366	.8667994	.606	1.15	year 1 waist/hip ratio
glucose1	2613	230	114.4826	42	440	year 1 fasting glucose (mg/dl)
tchol	2759	225	228.5799	110	465	total cholesterol (mg/dl)
LDL	2752	758	145.0385	36.8	393.4	LDL cholesterol (mg/dl)
HDL	2752	85	50.26199	14	130	HDL cholesterol (mg/dl)
TG	2759	287	166.1493	31	476	triglycerides (mg/dl)
tchol1	2613	229	219.2346	92	535	year 1 total cholesterol (mg/dl)
LDL1	2608	752	132.3915	-20	450.2	year 1 LDL cholesterol (mg/dl)
HDL1	2608	89	51.78183	14	124	year 1 HDL cholesterol (mg/dl)
TG1	2613	367	175.7976	31	1016	year 1 triglycerides (mg/dl)
SBP	2763	110	135.0695	83	224	systolic blood pressure
DBP	2762	58	73.1517	45	102	diastolic blood pressure
age10	2763	36	6.664857	4.4	7.9	age (per 10 years)

1. Launch R Studio. Import **hersdata.csv**
2. Issue the command **class(hersdata\$drinkany)**. You should see that it is a character variable. Convert this variable to factor.
3. Issue the command **library(summarytools)**. Next, issue the command **freq(something)** with the correct something to produce a one-way frequency table of **drinkany**. How many missing values are there?
4. Create a numeric variable **drinkany01** that is coded as follows:


```
= 0 if drinkany == "no"
  1 if drinkany == "yes" and
  NA if drinkany == NA
```
5. Label the variable **BMI** with the label "Body Mass Index"
6. Create a subset of **hersdata** called **mytiny** as follows:


```
Include only the following variables: HT, LDL, and SBP
Include only the observations with: weight > 125
```
7. Print **mytiny**
8. Output/save **mytiny** to an R dataset named **mytiny.Rdata**