Homework 2 (50 points)
Due 1/23/24

Goals
• Facility with data exploration
• Facility with use of the "reshape" and "gen" commands
• Construction of a "Table 1"

Open the Framingham data file in STATA, and explore it a bit. Perhaps list out the first 10 records ("list in 1/10"). It is a "long" data file, with a separate record for each of three time periods (note that not all individuals have 3 observations).

*(Stata tip: When pasting stata results into a word file, use **new courier font, size 8**.)*

You can change it to a wide file, so there is only one line of data per person, preserving all of the variables (you do not need to type them all in the command line – click on each of them in the variables list which you can view in a separate window). Look back to HW1 to remember how to switch between long and wide formats, if necessary. List the first 10 records again, to see how the data file has changed. One important purpose of this data exploration is to know which variables are time-dependent (i.e. change over time) and which variables are time-constant (i.e. constant across the 3 records within each subject).

Task 1 (10 points): Describe the time differences between study periods (or examinations).
Task 2 (10 points): List 3 variables that are time constant and 3 variables that are time-varying, and describe how they changed over study period.

(Stata tip: When confused, I often use the "help" command in STATA to refresh my memory about a command. The help files are very detailed and you can find illustrations of the command's use towards the end of each of these help files. You can also read the PDF manual for further usage.)

Task 3 (30 points): Construct a nice Table 1 for a prospective cohort study. The study questions, which will guide how you think about how the Table 1 should look, are described in detail below, together with some helpful categorization schemes.

The primary study questions for the Framingham cohort analysis are:

(1) Do categories of body mass index measured in period 1 (underweight, normal, overweight and obese) predict mortality through the end of the observation period?
(2) Are the mortality effects of obesity accounted for, or independent of, cholesterol level, hypertension and diabetes? (In other words – if there are mortality consequences to obesity, are they (a) explained by a higher rate of these co-morbidities or (b) independent of these conditions?)

Remember that for this exercise you are not trying to do the statistical analyses to answer these questions. Rather, you are trying to construct a descriptive, informative Table 1 that will shed some light on the larger research questions.

There are many ways to construct such descriptive tables and categorize the data. Sometimes people are reluctant to combine means and percentages in one table – but some variables are best described by one and some by the other.

For a cohort study, Table 1 may just present descriptive characteristics for the whole cohort, but more often it describes the population in terms of some stratification variables (such as sex) or in terms of the primary exposure if there is one (for example, BMI categories would be a sensible way to stratify the descriptive data here). By contrast, in a case-control study, Table 1 divides the population by their case status (cases vs controls). For your Table 1, you will need to decide which variables to include and how to present them. You will without question need to create new variables to do this.

Don't forget to include a very clear Table legend and footnotes to describe the nature of the summarized variables (where necessary).

Write a "do file" for any variable manipulation you need (reshaping the data, generating new variables or new categorizations of variables).

a. **Create a descriptive table for the cohort analysis and interpret the table.**
b. **Turn in the do and log file together with your Table 1.**

Clinical definitions you may need at some point (not necessarily an exhaustive list)

Body weight categories
<18.5 kg/m^2 underweight
18.5 <25 kg/m^2 normal
25 < 30 kg/m^2 overweight
30+ kg/m^2 obese (35+ may be separated as morbidly obese)

Desirable Cholesterol levels
<200 total mg/dl
< 100 mg/dl LDL
>60 mg/dl HDL

Desirable Blood Pressure
<140 mmHg systolic
<90 mmHg diastolic

Casual (non-fasting) glucose
200+ mg/dl is considered evidence of diabetes, lacking better measurements

Heart rate
Normal adult heart rate is 60-100 bpm
Well-conditioned athletes may be 40-60 bpm

Epidemiologic Methods (PBHS 31001, Winter 2024)

Alexandra Chang
01/23/2024

Homework #2

Problem 1.

Solution. STATA: 'smm' command for average time between exam periods.

- For the Framingham Heart Study, the participant clinic data was collected during **3 exam periods**, approximately **6 years apart**, which was from around 1956 to 1968.
- During the study, each participant was followed for a total of 24 years.
- *Note*: not all individuals had 3 observations, which depended on the number of exams the subject attended.

Problem 2.

Solution.

- **Time-dependent** variables:
  - ***age*** (Age at exam in years)
    - Participants were initially of a certain age in 1956, then aged approximately 6 years to the next exam period, and similarly aged another 6 years to the third exam period in 1968.
    - The changes in the variable "age" would represent the aging process of the participants over these specific time intervals.

```
. list randid period age in 4/6
```

	randid	period	age
4.	6238	1	46
5.	6238	2	52
6.	6238	3	58

Figure 1: Example of aging across 3 study periods for participant No.6238

- ***heartrte*** (Ventricular heart rate in beats/min)
  - Participants exhibited diverse increases or decreases in heart rate. There was no general overall trend common to all participants.
  - Factors influencing heart rate differed widely among the study population, highlighting the complexity and heterogeneity of these physiological responses within the study cohort.

```
list randid period heartrte in 4/6
```

	randid	period	heartrte
4.	6238	1	95
5.	6238	2	80
6.	6238	3	80

Figure 2: Example of heart rate fluctuations across 3 study periods for participant No.6238

- ***bmi*** (Body Mass Index, weight in kg/height meters squared)
  - Participants exhibited fluctuations across 3 exam periods, but the overall numbers do not vary significantly across participants.
  - Since major changes in height are considered less unlikely in adulthood, factors influencing BMI are more likely a result of changes in weight, indicating that multiple and intricate reasons contribute to the weight changes for BMI over the 3 periods.

```
. list randid period bmi in 4/6
```

	randid	period	bmi
4.	6238	1	28.73
5.	6238	2	29.43
6.	6238	3	28.5

Figure 3: Example of BMI fluctuations across 3 study periods for participant No.6238

- **Time-constant** variables:
  - ***sex*** (Participant sex)
    - Biological sex did not change within each study participant for the cohort.
  - ***educ*** (Attained education)
    - Cohort maintained a constant educational profile over the years of study.

- Event data such as ***mi-fchd, timemifc*** (Hospitalized Myocardial Infarction or Fatal Coronary Heart Disease)
  - ∗ Covered the entire follow-up period and did not change according to exam.
  - ∗ Event data for each participant was added without regard for prevalent disease status or when exam data was collected.

Problem 3.

Solution.

(Refer to the STATA .do and .log file for complete steps)

```
table1 if period==1, by(bmiCat) vars(death cat\ totchol contn\ totcholCat cat\ hyperten contn\ sysbp contn\ diabp contn\ diabetes cat) format(%2.1f) missing
```

Factor	Level	underweight	normal	overweight	obese	p-value
N		57	1936	1848	574	
0=No death during followup, 1=Death during followup	0	39 (68.4%)	1365 (70.5%)	1157 (62.6%)	317 (55.2%)	<0.001
	1	18 (31.6%)	571 (29.5%)	691 (37.4%)	257 (44.8%)	
serum cholesterol mg/dl, mean (SD)		215.5 (38.4)	231.2 (44.4)	242.0 (44.5)	242.8 (43.6)	<0.001
RECODE of totchol (serum cholesterol mg/dl)	desirable	19 (33.3%)	498 (25.7%)	312 (16.9%)	96 (16.7%)	<0.001
	Undesirable	37 (64.9%)	1420 (73.3%)	1515 (82.0%)	467 (81.4%)	
		1 (1.8%)	18 (0.9%)	21 (1.1%)	11 (1.9%)	
incident hypertension, mean (SD)		0.5 (0.5)	0.6 (0.5)	0.8 (0.4)	0.9 (0.3)	<0.001
systolic bp mmhg, mean (SD)		119.8 (17.8)	126.8 (20.4)	135.9 (21.5)	145.1 (24.6)	<0.001
diastolic bp mmhg, mean (SD)		76.5 (10.1)	79.2 (10.9)	84.9 (11.3)	90.9 (13.0)	<0.001
0=Not diabetic, 1=Diabetic	0	54 (94.7%)	1906 (98.5%)	1794 (97.1%)	542 (94.4%)	<0.001
	1	3 (5.3%)	30 (1.5%)	54 (2.9%)	32 (5.6%)	

Figure 4: Descriptive Table for the Prospective Cohort Analysis in Period 1.

- Do **categories of body mass index** measured in **period 1** (underweight, normal, overweight and obese) predict **mortality** through the end of the observation period?
  - There is a general trend indicating that as BMI increases, the mortality rate also rises consistently. The overall mortality status exhibits a statistically significant difference across BMI categories. ($p < 0.001$)
  - In period 1, 44.8% of obese patients died during follow-up while normal and overweight patients only accounted for 29.5% and 37.4% respectively.
  - Notably, the underweight category stands out as an exception, demonstrating a higher mortality rate compared to the normal category. This exception aligns with evidence suggesting that being underweight can also be linked to an increased risk of underlying heart problems, subsequently elevating the likelihood of mortality, or a possibility of other underlying causes.

- Are the **mortality effects of obesity** accounted for, or independent of, **cholesterol** level, **hypertension** and **diabetes**?
  - Regarding the mortality effects of obesity, as indicated by Table 1, each variable demonstrates a statistically significant difference when stratified across BMI categories ($p < 0.001$).
  - A distinct trend is observable for **serum total cholesterol, hypertension incidence, systolic and diastolic blood pressure, and diabetic status**, revealing that, as BMI increases, these rates also consistently rise.
  - Given that mortality similarly increases with higher BMI values, the mortality consequences associated with obesity can be elucidated by a higher incidence of these co-morbidities. In essence, there appears to be an interconnected relationship between obesity, the presented health indicators, and mortality, suggesting that the **observed mortality effects are influenced by an elevated prevalence of these associated conditions.**

Reference:

In the context of STATA, **Table 1** typically refers to a summary table that provides descriptive statistics for various variables in a dataset. Researchers often create Table 1 to present baseline characteristics of study participants or other relevant information for scientific research, particularly in medical and epidemiological studies.