Problem Set 2 Solutions

BMI and CHD prevalence. The following table uses data from the NHLBI teaching data set and displays categories of body mass index for 4,415 participants in the Framingham Heart Study attending an examination in 1956 with non-missing values for body mass index. For each body mass index category, the table displays the number of subjects with existing Coronary Heart Disease (CHD) at that exam (**prevchd=1**)

Body Mass Index Category		Number of Subjects at 1956 Exam	Cases of CHD Diagnosed Prior to 1956
Under Weight	BMI < 18.5	57	0
Normal Weight	$18.5 \leq$ BMI < 25	1936	66
Overweight	$25 \leq$ BMI < 30	1848	90
Obese	BMI $\geq$ 30	574	38
Total		4415	194

1. What is the prevalence of obesity among the 4415 participants at the 1956 exam?

   574/4415 = 0.1300

2. What is the prevalence of CHD at the 1956 exam among the 4415 participants at the 1956 exam?

   194/4415 = 0.0439

3. What is the prevalence of CHD at the 1956 exam for each of the body mass index classes?

   Under Weight Participants **0/57 = 0.0**

   Normal Weight Participants **66/1936 = 0.0341**

   Overweight Participants **90/1848 = 0.0487**

   Obese Participants **38/574 = 0.0662**

Diabetes prevalence. Use Stata and the NHLBI data set to calculate the prevalence of diabetes among participants who attended and had non-missing data on diabetes at all three examinations. (**Hint: There were 3,206 such participants.**)

1. What is the prevalence of diabetes at the first exam (**diabetes1=1**)?

   58/3206 = 0.0181

2. What is the prevalence of diabetes at the second exam (**diabetes2=1**)?

   105/3206 = 0.0328

3. What is the prevalence of diabetes at the third exam (**diabetes3=1**)?

   251/3206 = 0.0783

BMI and hypertension prevalence. Use Stata and the BMI1 variable in the NHLBI data set to create the four categories of body mass index as defined in the first question.

1. What is the prevalence of hypertension (**prevhyp1=1**) at the 1956 exam for each of the body mass index classes?

   Under Weight Participants   **6/57 = 0.1053**

   Normal Weight Participants   **398/1936 = 0.2056**

   Overweight Participants   **683/1848 = 0.3696**

   Obese Participants   **336/574 = 0.5854**

Hypertension and high blood pressure. Use Stata to create a binary variable (**highbp1**) to represent the presence/absence of high blood pressure at the 1956 examination

```
gen highbp1=.
replace highbp1=1 if (sysbp1>=140 | diabp1 >= 90)
replace highbp1=0 if (sysbp1<140 & diabp1 < 90)
```

(**Note: There are no missing data on sysbp1 and diabp1. If data were missing on both sysbp1 and diabp1 then they should also be missing for highbp1. If data were missing on diabp1 only and sysbp1 $\geq$ 140 then highbp1 =1, otherwise highbp1 should be missing. Similarly, if data were missing on sysbp1 only and diabp1 $\geq$ 90 then highbp1 =1, otherwise highbp1 should be missing.)**

1. What is the prevalence of CHD (prevchd1=1) at the 1956 exam for participants with high blood pressure at the 1956 exam (highbp1=1)?

   106/1619 = 0.0655

2. What is the prevalence of CHD (prevchd1=1) at the 1956 exam for participants without high blood pressure at the 1956 exam (highbp1=0)?

   88/2815 = 0.0313

Hypothetical life table. The table below lists the number of individuals alive at age *x*, for a hypothetical population in 1950-1952 and 1990-19992.

Number of Survivors out of 100,000 Live Births

Age	1950-1952	1990-1992
0	100,000	100,000
20	73,412	96,902
40	56,884	92,638
70	31,744	79,873

1. What is the probability of surviving from birth to age 20 in 1950-1952?

    0.73412

    73412/10000 = 0.73412

2. What is the probability of surviving from age 40 to age 70 in 1990-1992?

    0.86221

    79873/92638 = 0.86221

3. Define the absolute survival increase over the 40 year span as p_1-p_2, where p_1 is the chance of surviving from age x to age x+n in 1990-1992 and p_2 is the chance of surviving from age x to age x+n in 1950-1952. Which age group has the greatest absolute survival increase?

    (a) 0 – 20 , (b) 20 – 40, **(c) 40 – 70**

    Survival: 1950 – 1952

    0 – 20: 73412/100000 = 0.73412
    20 – 40: 56884/73412 = 0.7748597
    40 – 70: 31744/56884 = 0.558048

    Survival: 1990 – 1992

    0 – 20: 96902/100000 = 0.9559968
    20 – 40: 92638/96902 = 0.9559968
    40 – 70: 79873/92638 = 0.8622056

    Absolute Survival Increase:

    0 – 20: (0.9559968 - 0.73412) = 0.239
    20 – 40: (0.9559968 - 0.7748597) = 0.181
    40 – 70: (0.8622056 - 0.558048) = 0.304

4. Define the relative survival increase over the 40 year span as $(p_1 - p_2)/p_2$, where p_1 is the chance of surviving from age x to age x+n in 1990-1992 and p_2 is the chance of surviving from age x to age x+n in 1950-1952. Which age group has the greatest relative survival increase?

 (a) 0 – 20 , (b) 20 – 40, **(c) 40 – 70**

Relative Survival Increase:

0 – 20: (0.9559968 - 0.73412)/0.73412 = 0.3022351
20 – 40: (0.9559968 - 0.7748597)/0.7748597 = 0.2337676
40 – 70: (0.8622056 - 0.558048)/0.558048 = 0.5450384