Homework 04

Spencer Pease

10/28/2019

(Q1) Creatinine Levels and Survivorship Status

One way we can examine data from the MRI data set is by dichotomizing variable to see the size of different category pairings. In this question we will look at two such variables: *5-year vital status* (did the patient live longer than five years after their MRI date) and *high creatinine level* (is the patient's creatinine level **above** $1.2\frac{mg}{dl}$).

(Q1.a)

To better compare the different categories created by the mixing of these two variables, we create a contingency table:

Table 1: Contingency table for each pairing of the dichotomous variables 'Survival Status' and 'CRT Level'.

	Non-high CRT	High CRT
Died w/in 5 years	77	44
Surv. 5+ years	493	119

(Q1.b-e)

If we want to look closer at the relationship between survival status and high creatinine level, we can use inferential statistics to make estimates about different groups within our population. For each grouping condition below, we look at the point estimate of the mean, standard error, and 95% confidence interval.

Table 2: Summary of the point estimate, standard error, and 95% confidence interval of the population probability, for various conditions.

Condition	Point est.	Std. error	95% CI (lower)	95% CI (upper)
Survive 5+ years	0.83	0.01	0.81	0.86
Low CRT level	0.78	0.02	0.75	0.81
Low CRT level, given survival of <5 years	0.64	0.02	0.60	0.67
Survive 5+ years, given high CRT level	0.73	0.02	0.70	0.76

(Q2) The ELISA Test

(Q2.a,b)

For a test with 99% sensitivity and 96% specificity, in a population of people donating blood with a prevalence of HIV-infection at 1%, the predictive values of the test are:

- Positive predictive value: **0.2**
- Negative predictive value: $1 - (1 \times 10^{-4})$

(Q2.c)

Even though the test may have a specificity of *"only* 96%*"*, it is important to consider what this means in the context of the population in which the test is being used. In this case, the prevalence of HIV is 1%, which means the test will only be wrong 4% of the time for this small fraction of the population. Overall, this means very few individuals with HIV will test negative when donating blood.

(Q2.d)

For a test used to keep HIV out of the blood supply of blood banks, it is more important to have a high specificity. In the scheme of things, rejecting someone from donating because they falsely test positive for HIV is far less consequential than accepting someone who falsely tests negative. For this reason, it is more important to have confidence in the test when it produces negative results, which is captured in the specificity of the test.

(Q2.e)

For the same test used in a clinic with a high number of intravenous drug users where the prevalence of HIV-infection is now 16%, the predictive values of the test are:

- Positive predictive value: **0.82**
- Negative predictive value: **0.998**

(Q2.f)

Finally, for the same test in a population of students entering college where the HIV-infection prevalence is .01%, the predictive values of the test are:

- Positive predictive value: **0.002**
- Negative predictive value: $1 - (1 \times 10^{-5})$

(Q3) A Biased Coin

a) $X = 1$ and $Y = 1$: 9.75×10^{-3}
b) $Y \geq 3$: **0.765**
c) $X = 1$ given $Y = 1$: **0.2**
d) $E(Y)$: **3.25**