Homework2

Jeffrey Liang

10/4/2020

Problem 1

A new test is being developed for the detection of carcinoma of the prostate (Foti et al. N Engl. JMed. 1977). When it is tested in a group of 113 patients with prostatic cancer, 79 have a positive diagnosis; in a group of 217 individuals without prostatic cancer, 10 have a positive diagnosis.

1) Calculate the specificity and sensitivity of the test. (4p)

$sensitivity = \frac{P(positive|carcinoma)}{P(carcinoma)}$
$= \frac{79/113}{79/113+10/217}$
$= 0.9381601$

$specificity = \frac{P(negative|not\ carcinoma)}{P(not\ carcinoma)}$
$= \frac{207/217}{34/113+207/217}$
$= 0.7602132$

2) In another hypothetical scenario, it is planned to use the test to screen a large sample of subjects for prostatic cancer where the test results will be the only data available. Is this information enough to assess the test characteristics, i.e., sensitivity & specificity? Yes, no, what is missing (if the case)? (2p)

NO, the sensitivity and specificity are computed base in comparision to the test result of gold standard test's result. And thus, given the data of this test only is not sufficient to calculate sensitivity and specificity of this test.

3) In patients with palpable prostatic nodules, the pretest likelihood of prostatic cancer is 50%. The test under these conditions has a sensitivity of 80% and a specificity of 95%.

a) Calculate the probability of a patient with a palpable prostatic nodule and a positive test result having prostatic cancer. What is the epidemiological term of this probability? (3.5p)

$$P(prostatic\ cancer \mid Positive \cap Palpable\ Prostatic\ Nodule)$$

$$= \frac{P(Positive \cap Palpable\ Prostatic\ Nodule\ \cap prostatic\ cancer)}{P(Positive \cap Palpable\ Prostatic\ Nodule)}$$

$= P(Positive \cap Palpable\ Prostatic\ Nodule\ \cap prostatic\ cancer) \div$
$(P(Positive \cap Palpable\ Prostatic\ Nodule \mid prostatic\ cancer) \times P(prostatic\ cancer)$
$+ P(Positive \cap Palpable\ Prostatic\ Nodule \mid prostatic\ cancer^c) \times P(prostatic\ cancer^c))$

$$= \frac{80\% \times 50\%}{80\% \times 50\% + 5\% \times 50\%} = 0.9411765$$

This is the ????

b) Re-calculate the probability in 2-a), using a pretest likelihood of prostatic cancer of 10%. Compare the two values and comment on their differences. (3.5p)

$$P(prostatic\ cancer' \mid Positive \cap Palpable\ Prostatic\ Nodule)$$
$$= \frac{P(Positive \cap Palpable\ Prostatic\ Nodule\ \cap prostatic\ cancer')}{P(Positive \cap Palpable\ Prostatic\ Nodule)}$$

$= P(Positive \cap Palpable\ Prostatic\ Nodule\ \cap prostatic\ cancer') \div$
$(P(Positive \cap Palpable\ Prostatic\ Nodule \mid prostatic\ cancer') \times P(prostatic\ cancer')$
$+ P(Positive \cap Palpable\ Prostatic\ Nodule \mid prostatic\ cancer'^c) \times P(prostatic\ cancer'^c))$

$$= \frac{80\% \times 10\%}{80\% \times 10\% + 5\% \times 90\%} = 0.64$$

Noticing that lower disease prevalence/ prior produce a lower positive predictive value. This is similar to the example given on class regarding the low positive predictive value base on the low prevalence of COVID-19 in some communities. ■

Problem 2

According to the Center for Disease Control (CDC), about 34.5% of the adult US population are prediabetic (National Diabetes Statistics Report, 2020). Suppose we randomly select a group of 50 patients seen at Columbia University Medical Center. Calculate the following exact probabilities based on the national statistics:

1) Probability that none of these patients are prediabetic. (2.5p)
2) Probability that less than 10 patients are prediabetic. (2.5p)
3) Probability that 34.5% of these patients are prediabetic (round to the nearest integer). (2.5p)
4) Could you use an approximation method to calculate the probabilities above? If yes, calculate the probabilities using approximations and compare to the exact values; otherwise, explain why approximations methods are not appropriate. (2.5p) **Proof**

1. With the information given, any sample from the population has probability of $p = 34.5$ suffer from prediabetic. Define $\mathcal{X}$ as the random variable(R.V.) , then the probability of number of samples from population suffer from prediabetic follows the distribution of Binomial of $Bin\ (50, .345)$. Therefore we have the distribution of:
$$P(\mathcal{X} = k) \ = \binom{n}{k} p^k (1-p)^{n-k}$$
And the probability of none of these patients have prediabetics is:
$$P(\mathcal{X} = 0) \ = \binom{n}{0} p^0 (1-p)^{n-0} \ = 6.4873152 \times 10^{-10}$$
2. Use the CDF of binomial distribution:
$$P(\mathcal{X} < 10) \ = \sum_{k=0}^{9} \binom{n}{k} p^k (1-p)^{n-k} \ = 0.0081532$$
3. The probability 34.5% proportion of the patients are prediabetic is: $P(\mathcal{X} = 50 \times 34.5\% \approx 17) \ = \binom{n}{17} p^{17} (1-p)^{n-17} \ = 0.1180887$
4. the sample size exceed 30, it can be approximate by Normal distribution of $N \sim (np, np(1-p))$ by CLT
    1. $P(\mathcal{X} = 0) \ \approx \phi(0 - 1/2 < \mathcal{X} < 0 + 1/2)$
    $= 2.4848654 \times 10^{-7}$
    2. $P(\mathcal{X} < 10) \ \approx \phi(\mathcal{X} < 9 + 1/2)$
    $= 0.0105661$
    3. $P(\mathcal{X} = 17) \ \approx \phi(17 - 1/2 < \mathcal{X} < 17 + 1/2)$
    $= 0.1179243$
5. ■

Problem 3

The incidence of uveal melanoma in the US is approximately 5 per million individuals per year, with a significantly higher incidence in non-Hispanic Whites (6.02 per million), when compared to Blacks and Asians: 0.31 and 0.39 per million, respectively. a) What is the probability that in NYC (population of 8.3 million reported in 2020), exactly 30 cases occur in a given year? (4p) b) Compute the same probability in a) by the mentioned racial/ethnic groups and comment on the findings. Demographic data of NYC in 2020: 14.0% Asians, 42.8% non-Hispanic Whites, 24.3% Black. (6p)