IT3030 Biostatistics Quiz#3 2019.03.26 ID:____________________ Name:_____________________

1. (30%) The Health Department reports 10% rate of the HIV infection for the “at-risk” population. A preliminary screening test for the HIV is correct 95% of the time. One person is randomly selected from the at-risk population.

(a) (15%) What is the probability that the selected person tests positive in the initial screening if this person has the virus?

(b) (15%) What is the probability that the selected person has the virus if this person tested positive in initial screening?

The general form for Bayes’ theorem is given below, in which A_1, A_2,…., and A_n are n mutually exclusive and exhaustive events.

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + \ldots + P(A_n)P(B \mid A_n)}$$

[Keep your answers with 4 decimal points]

Answer:

V+ : people having HIV virus (regardless he is tested positive or negative)

V- : people free from HIV virus (regardless he is tested positive or negative)

T+ : people tested positive (regardless he is carrying virus or not)

T- : people tested negative (regardless he is carrying virus or not)

(a) We want to know P(T+|V+). This is given as “a preliminary screening test for the HIV virus is correct 95% of the time”. Thus the answer is 0.9500.

(b) We want to know P(V+|T+):

$$P(V+ \mid T+) = \frac{P(V+)P(T+ \mid V+)}{P(V+)P(T+ \mid V+) + P(V-)P(T+ \mid V-)}$$

We already know that P(T+|V+) = 0.95. The same statement also says P(T-|V-) = 0.95 (95% accurate for not having the virus). Thus P(T+|V-) = 1 – 0.95 = 0.05. P(V+) = 0.1 and P(V-) = 0.9. Thus we have

$$P(V+ \mid T+) = \frac{0.1 \times 0.95}{0.1 \times 0.95 + 0.9 \times 0.05} = 0.6786$$

2. (40%) Given 16% of the adults in US who are smokers. Among 100 US adults, we'd like to estimate the probability for a given number of people who smoke. (a) What theoretical probability distribution is used to describe the random variable here? (b) Determine the probability that none of them are smokers. (c) The probability for exactly 20 smokers? (d) The probability for at most 16 are smokers.

Answer:

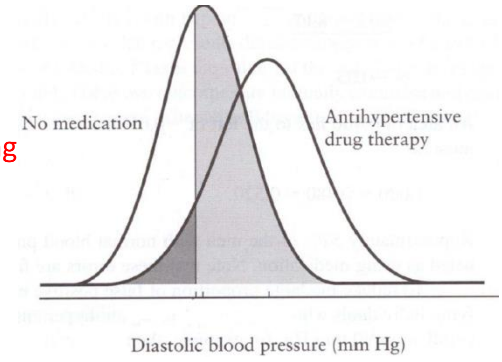
(a) Binomial Probability Distribution

(a) binopdf(0, 100, 0.16) = 2.6787e-08

(c) binopdf(20, 100, 0.16) = 0.0567

(d) binocdf(16,100,0.16) = 0.5662

3. (30%) Given two normal distributions below. Group 1 represents men of the blood pressure distribution that are normal or within acceptable range (μ=80.7, σ=9.2). Group 2 represents men of blood pressure distribution that are controlled by taking medications (μ=94.9, σ=11.5).
(a) What is the lower cut-off pressure to identify 90% of the individuals taking medication? That is, the dark gray area occupies 10% of the area under group 2 curve. (compute to 4D)
(b) Under this cut-off value, what percentage of the normal individuals will be **falsely identified** as the medication-taking individuals? That is, to find the light gray area under group 1 curve. (compute to 4D)



Ans:

(a) First determine the Z value for group 2 yielding 0.1 for the dark gray area. With F defining the standard normal distribution, we tried and found the following:

>> **int(F, -inf, -1.2815)**

ans =

0.10000904996608381492569387780556

>>

This gives Z=-1.2815. Then X = sigma*Z + mu = 11.5*(-1.2815) + 94.9 = **80.1628**

(b) Determine the Z value for group 1 curve when X = 80.1628.

Z = (X - mu) / sigma = (80.1628 – 80.7) / 9.2 = -0.0584

The following MATLAB command then gives the answer:

>> **int(F, -0.0584, inf)**

ans = **0.5233**

>>

Later when you learned and mastered normpdf, normcdf and norminv, you will know the above answers can be easily obtained with the following two commands:

>> **X = norminv(0.1, 94.9, 11.5) = 80.1622**

>> **1-normcdf(X, 80.7, 9.2) = 0.5233**

>>