IT3030 Biostatistics Quiz#1   5/8/2018   ID:____________________ Name:____________

【Work on as many problems as you can. Write down detailed steps or MATLAB commands leading to your answers.】

1. Given 16% of the adults in US who are smokers. Among 100 US adults, determine the probability that (a) none of them are smokers, (b) exactly 50 are smokers, (c) at most 16 are smokers.

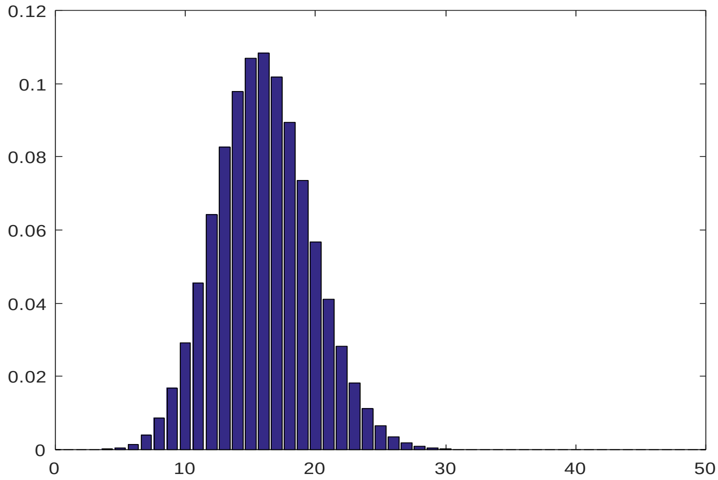
(a) binopdf(0, 100, 0.16) = 2.6787e-08

(b) binopdf(50, 100, 0.16) = 2.6535e-15

(c) binocdf(16,100,0.16)

ans =   0.5662

>>



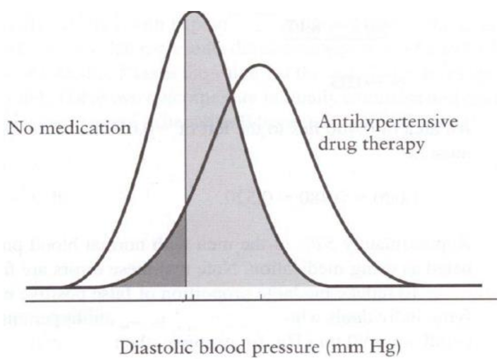
2. Assume the probability that a particular individual involved in an airplane accident each year is p=0.000009. (a) For a population of n=100,000 people, determine the value of λ assuming that we have a Poisson distribution. (b) Also with n=100,000, determine the probability that at least 5 people are involved in such accident.

λ = np = 0.9;

>> 1-poisscdf(4,lambda)

ans =   0.0023

3. Given the same distribution as in class. (a) Determine the blood pressure x for the "antihypertensive drug therapy" group that only 5% of the group have a blood pressure of x or smaller. (b) Determine the percentage of people in the "no medication" group having their blood pressure greater or equal to x.



>> x=norminv(0.05,94.9,11.5)

x =   **75.9842**

>> 1-normcdf(x,80.7,9.2)

ans =   **0.6959**

4. A rapid test for Zikavirus has been developed to screen whether a patient is infected or not. A study

shows that out of 25 people confirmed to have the virus, 19 were tested positive and others were tested negative. For 36 people who were free from this virus, 29 were tested negative, and remaining were tested positive. Determine the TP, FP, TN, FN, sensitivity, specificity, PPV and NPV of this test. (Keep 2D)

Ans:

	Virus	No virus	Total
Test positive	19 (TP)	7 (FP)	26
Test negative	6 (FN)	29 (TN)	35
Total	25	36	61

Sensitivity is 19/(19+6)=0.76

Specificity is 29/(29+7)=0.81

PPV = 19/26 = 0.73

NPV = 29/35=0.83

5. Given a test with sensitivity 0.6 and specificity 0.8 for having a disease. The prevalence for this disease is 0.02, i.e., for every 100 people, 2 are actually having the disease.

(a) If you were tested positive for this diagnosis, what is the probability you actually had the disease?

(b) Following a first positive test, if you were tested for a second time and still got positive result, what is the probability you actually had the disease? (Keep 4D.)

Ans:

(a) Let "D+" be the sample for diseased people, "D–" be non-diseased. Let "+" represents people having the test positive, and "–" be the ones tested negative. From the above statements, we have:

- (1) P(+|D+)=0.6 (sensitivity), then P(–|D+)=1-0.6 = 0.4.
- (2) P(–|D–)=0.8 (specificity), then P(+|D–)=1-0.8 = 0.2
- (3) P(D+)=0.02, then P(D–)=1-0.02=0.98.

According to Bayes' theorem, we wish to know P(D+|+), which is given by the formula (note that D– and D+ are mutually exclusive)

$$P(D+|+)=\frac{P(+|D+)P(D+)}{P(+|D+)P(D+)+P(+|D-)P(D-)}$$

$$=\frac{0.6\times0.02}{0.6\times0.02+0.2\times0.98}=\frac{1.2}{1.2+19.6}=\frac{1.2}{20.8}=0.0577$$

(b) Let "++" be people having tested positive on their second test provided their first test is also positive. We wish to know P(D+|++), which is given by the formula (note that D– and D+ are still mutually exclusive)

$$P(D+|++)=\frac{P(++|D+)P(D+)}{P(++|D+)P(D+)+P(++|D-)P(D-)}$$

$$=\frac{(0.6)^2\times0.02}{0.6^2\times0.02+(0.2)^2\times0.98}=\frac{6^2\times2}{6^2\times2+4\times98}=\frac{72}{72+392}=0.1552$$

6. Assuming p=0.35 is the percent of people smoked. Among 20 randomly selected people:

(a) Determine each of the probability that you will find 0, 5, 10, 15 and 20 people who actually smoked.
(b) Determine the mean, and standard deviation of the number of people who smoked.

Answer:

(a)

>> binopdf(0,20,0.35) = 1.8125e-004;

>> binopdf(5,20,0.35) = 0.1272;

>> binopdf(10,20,0.35) = 0.0686;

>> binopdf(15,20,0.35) = 2.6063e-004

>> binopdf(20,20,0.35) = 7.6096e-010

(b)

Mean = np = 20*0.35 = 7.0

STD = sqrt(np(1-p)) = sqrt(20*0.35*0.65) = 2.1331.

7. Considering the BMI (Body-Mass Index) statistics for all college male students. The mean value and standard deviation are 22.31 and 4.66, respectively. If we are going to sample 28 male college students:
(a) What are the mean and standard deviation of the BMI means of these samples?
(b) What proportion of these samples to give a mean BMI of 18.00 or less?
(c) What proportion of these samples to give a mean BMI of 24.00 or higher?

Answer:

(a) Mean = 22.31 according to central limit theorem. STD = 4.66/sqrt(n=28) = 0.8807.

(b,c) Given the normal distribution of mean=22.31 and STD=0.8807, the left-tail up to 18.00 would be normcdf(18, 22.31, 0.8807) = 4.9385e-007 and up to 24.00 would be normcdf(24, 22.31, 0.8807)= 0.9725. So the answers would be 4.9385e-007 and 0.0275.

8. Following the same question above:
(a) Given one particular sample of n=28 with mean BMI of 22.50. When looking for the confidence interval (CI) for this mean BMI 22.50, would you use one-tail or two-tail approach? If one-tail, would it be a left-tail-to-exclude or right-tail-to-exclude approach? Clearly explain your answer.
(b) Determine the 75% CI, 85% CI and 95% CI of this particular sample.

Answer:

SEM = 0.8807 as obtained from problem #2. 75%, 85% and 95% CI for standard normal distribution would cut-off left tail of 12.5%, 7.5% and 2.5%, which mark at z = norminv(0.125) = -1.1503, norminv(0.075) = -1.4395, and norminv(0.025) = -1.9600, respectively.

The half-CI would be abs(z)*SEM = 1.0131, 1.2677 and 1.7261, respectively.

Finally the 75% CI would be from 21.4869 to 23.5131.

The 85% CI would be from 21.2323 to 23.7677.

The 95% CI would be from 20.7739 to 24.2261.