IT3030 Biostatistics midterm exam#1 2019.04.14

Write down important intermediate steps and results, including MATLAB or Excel commands used to illustrate your solution strategies. **DO NOT** just show the final result.

1. (10%) Given the total cholesterol level for 1,067 men in the following table.

(a) Draw a histogram with relative frequencies.

(b) Determine the mean cholesterol level (2D). (Consider using the formula shown below.)

Cholesterol range	Number of people
80-119	20
120-159	143
160-199	299
200-239	342
240-279	215
280-319	24
320-359	19
360-399	5

$$\bar{x} = \frac{\sum_{i=1}^{k} m_i f_i}{\sum_{i=1}^{k} f_i},$$

```
>> m=0.5*[80+119 120+159 160+199 200+239 240+279 280+319 320+359 360+399] =
    99.5000    139.5000    179.5000    219.5000    259.5000    299.5000    339.5000    379.5000
>> f=[20 143 299 342 215 24 19 5];
>> mean = sum(m.*f)/sum(f) = 208.0661 or 208.07 (2D)
```

2. (20%) Sort and draw a box plot (with whiskers) for the data set {3, 7, 8, 5, 12, 14, 21, 13, 18}. Clearly mark the value for minimum, Q1, Q2, Q3 and maximum on your graph.

The sorted array is {3, 5, 7, 8, 12, 13, 14, 18, 21}

Min = 3, Q1=7, Q2=12, Q3=14, Max=21. So 7 and 14 are the box boundaries, 3 and 21 are the two whiskers.

3. (20%) The average number of motor accidents reported each month in a given university is 0.27. Assume that the number of accidents follows a Poisson distribution.

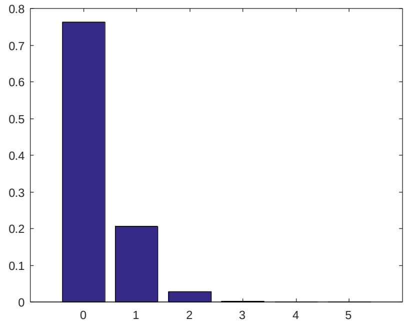
(a) What is the probability that no accident be reported in a given month?

(b) What is the probability that at most 1 accident be reported?

(c) What is the probability that 2 or more accidents be reported?

(d) Graph a probability distribution ranging from 0 to 5 accidents (a bar chart).

```
>> lambda=0.27;
(a) >> poisspdf(0,lambda) =      0.7634
(b) >> poisscdf(1,lambda) =      0.9695
(c) >> 1-poisscdf(1,lambda) =      0.0305
(d) >> x = [0:5];
    >> y = poisspdf(x,lambda);
    >> bar(x,y)
```



4. (30%) Among a population of 1,280 people being examined for the new virus infection, 30 were actually infected and the others were not. Blood tests were administered to all 1,280 people. Among those 30 having the infection, 12 showed positive results and the others were negative. Among those healthy ones, 950 showed negative results and the others showed positive.

(a) (10%) Determine the sensitivity, specificity, PPV and NPV of the test.

(b) (10%) What is the probability that a randomly selected individual has the virus that his or her blood test is positive?

(c) (10%) What is the probability that a randomly selected individual has the virus that his or her blood test is positive for <u>two consecutive tests</u>?

Formula that are useful:

Sensitivity = TP/(TP+FN); Specificity = TN/(TN+FP); PPV = TP/(TP+FP); NPV = TN/(TN+FN)

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

Answer:

TP=12, FN=18 (adding up to 30 infected people), TN=950, FP=300 (adding up to 1250 healthy ones)

(a) Sensitivity = 12/30 = **0.400**

Specificity = 950/1250 = 0.76

PPV = 12/(12+300) = 12/312= 0.0385

NPV = 950/(950+18) = 0.9814

(b) We need to know P(D+/+) = P(+/D+)P(D+) / [P(+/D+)P(D+) + P(+/D-)P(D-)]

Here P(D+) = 30/1280 = 0.0234, P(D-) = 1 - 0.0234 = 0.9766

P(+/D+) = sensitivity = 0.4, P(+/D-) = 1 - specificity = 0.24

Therefore P(D+/+) = 0.4*0.0234 / [0.4*0.0234 + 0.24*0.9766]

= 0.0094 / (0.0094 + 0.2344) = 0.0094/ 0.2438 = **0.0386** (This is actually the same as PPV=0.0385 above, with minor round-off difference.)

(c) We need to know P(D+/++) = P(++/D+)P(D+) / [P(++/D+)P(D+) + P(++/D-)P(D-)]

= 0.4^2*0.0234 / [0.4^2*0.0234 + 0.24^2*0.9766]

= 0.0037 / (0.0037 + 0.0563) = 0.0037 / 0.06 = **0.0617**

5. (20%) Given BMI statistics for university male students as μ=21.3 and σ=2.8. Assume these BMI statistics follow a normal distribution. (a) Determine the probability of male students having BMI less than 18. (b) Determine the probability for male students having BMI greater than 24.

(Formula for normal distribution is shown below.)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2}(\frac{x-\mu}{\sigma})^2)$$

```
>> F(x)='1/(sqrt(2*pi)*2.8)*exp(-0.5*((x-21.3)/2.8)^2)';
>> int(F,-inf,18) =    0.11928443711103525995497063226197
>> int(F,24,inf) =    0.16745135054973664632843483286331
or
>> normcdf(18, 21.3, 2.8) = 0.1193
>> 1-normcdf(24, 21.3, 2.8) = 0.1675
```