IT3030 Biostatistics midterm exam#1 2019.04.02 ID:________________ Name:_________________

1. (30%) Given the score distribution for students participating in a course in the following table.

(a) Determine the number of students and draw a histogram with absolute frequencies.

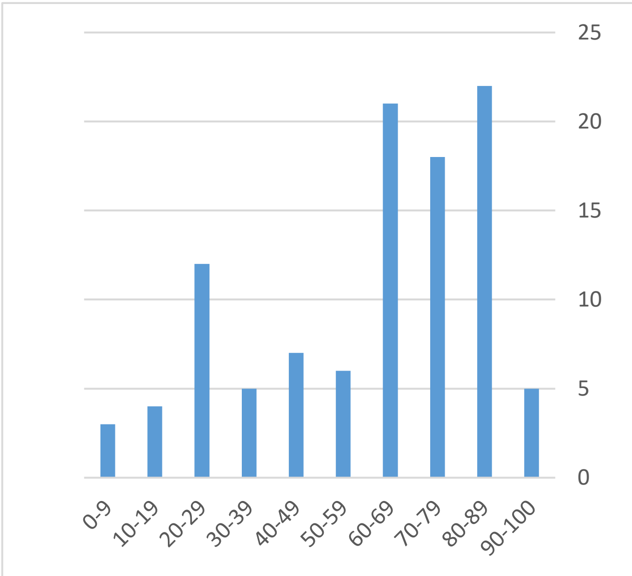
(b) Determine the mean score (2D).

(c) Determine the standard deviation for these scores (2D).

Score range	Number of students
90-100	5
80-89	22
70-79	18
60-69	21
50-59	6
40-49	7
30-39	5
20-29	12
10-19	4
0-9	3

Answer (using Excel):

	f(i)	m(i)	f(i)*m(i)	f(i)*[(m(i)-mean)^2]	
90-100	5	95	475	6037.8125	
80-89	22	84.5	1859	12937.375	
70-79	18	74.5	1341	3655.125	
60-69	21	64.5	1354.5	379.3125	
50-59	6	54.5	327	198.375	
40-49	7	44.5	311.5	1736.4375	
30-39	5	34.5	172.5	3315.3125	
20-29	12	24.5	294	15336.75	
10-19	4	14.5	58	8372.25	
0-9	3	4.5	13.5	9324.1875	
N=	103		6206	61292.9375	
		mean=	60.25	VAR=	600.91
				STD=	24.51



Using MATLAB:

```
>> f=[3 4 12 5 7 6 21 18 22 5];
>> m=[4.5 14.5 24.5 34.5 44.5 54.5 64.5 74.5 84.5 95];
>> N=sum(f)
N =     103
>> mu=sum(f.*m)/N
mu =     60.2524
>> VAR=(sum(((m-mu).^2).*f))/(N-1)
VAR =    600.9111
>> STD=sqrt(VAR)
STD =     24.5135
>>
```

2. (20%) The average number of suicides reported each month in a given city is 0.72. Assume that the number of suicides follows a Poisson distribution.

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

(b) What is the probability that 2 or more suicides be reported?

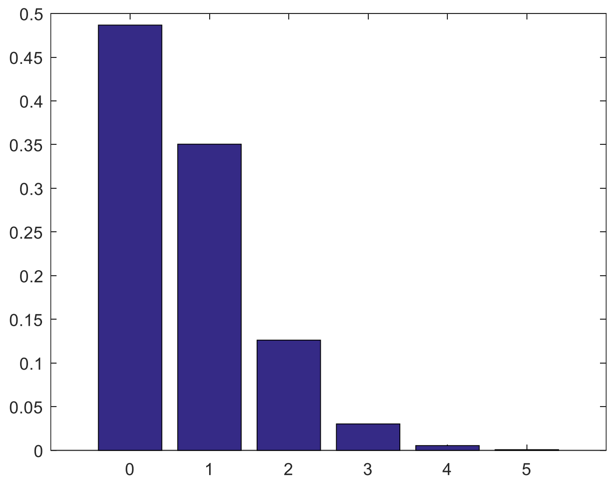
(c) What is the probability that 4 or more suicides be reported?

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

[Hint: Recall that, for an incident (like suicide mentioned in this problem) occurring at a probability of *p*, the average number of people involved among a population of *n* people would be *np*, which is the same as the parameter λ used in a Poisson distribution with the formula shown below. Useful MATLAB functions include poisspdf, poisscdf and poissinv.]

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

```
>> lambda=0.72;
(a) >> poisspdf(0,lambda) =        0.4868
(b) >> 1-poisscdf(1,lambda) =        0.1628
(c) >> 1-poisscdf(3,lambda) =        0.0063
(d) >> x = [0:5];
    >> y = poisspdf(x,lambda);
    >> bar(x,y)
```



3. (30%) Among a population of 1,820 people, 30 suffered from tuberculosis (肺結核) and the other 1790 did not. Chest X-rays were administered to all these people. Among those 30 having the disease, 22 showed positive results and the others were negative. Among those healthy individuals, 1739 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 tuberculosis that his or her X-ray is positive?

(c) (10%) What is the probability that a randomly selected individual has tuberculosis that his or her X-ray 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=22, FN=8, TN=1739, FP=51

(a)

Sensitivity = 22/30 = **0.7333**

Specificity = 1739/1790 = **0.9715**

PPV = 22/(22+51) = 22/73 = **0.3014**

NPV = 1739/(1739+8) = **0.9954**

(b)

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

Here P(D+) = 30/1820 = 0.0165, P(D-) = 1 - 0.0165 = 0.9835

P(+/D+) = sensitivity = 0.7333, P(+/D-) = 1 - specificity = 0.0285

Therefore P(D+/+) = 0.7333*0.0165 / [0.7333*0.0165 + 0.0285*0.9835]

= 0.0121 / (0.0121 + 0.0280) = 0.0121 / 0.0401 = **0.3017** (This is actually the same as PPV=0.3014, with minor round-off difference.)

(c)

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

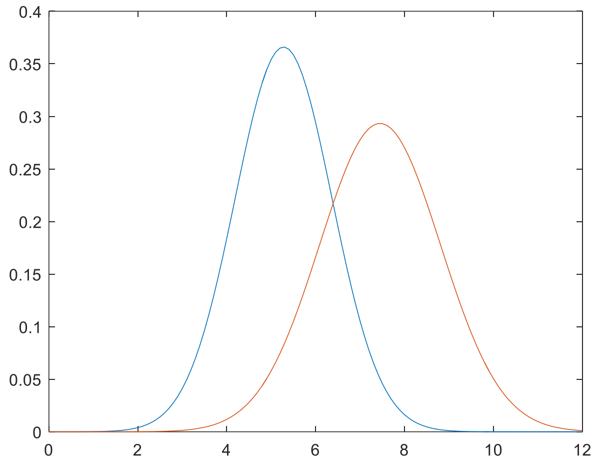
= 0.7333^2*0.0165 / [0.7333^2*0.0165 + 0.0285^2*0.9835]

= 0.0089 / (0.0089 + 0.0008) = 0.0089 / 0.0097 = **0.9175**

4. (20%) Given two normal distributions defined by their mean values and standard deviations. $\mu_1 = 5.28$, $\sigma_1 = 1.09$, $\mu_2 = 7.45$, $\sigma_2 = 1.36$. The probability density distribution for the two samples is shown below.

(a) Determine the probability for sample 2 to have its value smaller or equal to μ_1.

(b) Determine the probability for sample 1 to have its value greater or equal to μ_2.



Answer:

(a)

>> normcdf(5.28,7.45,1.36)

ans = **0.0553**

(b)

>> 1-normcdf(7.45,5.28,1.09)

ans = **0.0233**

>>