Statistics Notes

Given below are some questions. I give hints when I think that you may need help.

1. I need a sample of 20 students from a list of 805. Label those students from 1 to 805 and give me one sample of 20 using random sampling (using a table or a calculator and indicating how you did it). Also, give me a sample of 20, using systematic random sampling, indicating the procedure. (page 18).
2. Suppose that 100 of the 805 students of problem 1 are graduating seniors and that I want them well represented. What kind of sampling would you use? Approximately how many graduating seniors will be in a sample of 20? (page 22).
3. What is the difference between an observational study and a designed experiment? Look up at least one example of each.
4. Label the following variables qualitative or quantitative as appropriate:

Height,
Weight
Place of birth
Color of hair
High school class rank
Amount of water in a beaker
Telephone number of a person
Number of courses

5. Given the data respond to the requests that follow: 83, 58, 77, 57, 87, 93, 70, 85, 92, 92, 79, 68, 82, 72, 68, 76, 88, 90, 81, 76, 77, 55, 74, 73, 82, 86, 92, 84, 89, 64, 35, 73.

(i). Prepare a stem and leaf diagram of the data and comment on the shape of the distribution.

(ii). Complete the following frequency and relative frequency table.

Class	Tally	Frequency	Relative Frequency
30-39			
40-49			
50-59			
60-69			
70-79			
80-89			
90-99			

(iii). Use the above table to construct a frequency histogram of the data.

(iv). Give the "five number summary" and construct the box plot of the data. Decide if the data contains any outliers

6. Given the data: 5,14, 7, 10, 6, 8, 12, 9, 10, 9, 7, 11.

(i) Find the mean_____ , median _____, and mode _____, of the data.

(ii) Test the data for normality using the normal probability plot. (You are supposed to use Table III for the z-scores.)

(iii) Find the variance of the data using the fact that $\sum (x-\bar{x})^2 = 74$ and assuming that the above data is a sample.

(iv) Find the standard deviation of the above data assuming that it is the data of a population.

7. The following table provides 8 observations of y against x.

Observation	1	2	3	4	5	6	7	8
x	2.4	3.4	4.6	3.7	2.2	3.3	4	2.1
y	1.33	2.12	1.80	1.65	2.00	1.76	2.11	1.63

$\sum x = 25.7$, $\sum y = 14.40$, $\sum x^2 = 88.31$, $\sum y^2 = 26.4324$ and $\sum xy = 46.856$,

(I) Using your calculator, or the information above, find the (linear) coefficient of determination of the data and comment on the strength of relationship between the x's and the y's.

(II) Using your calculator, or the information above, find the slope b_1 and the intercept b_0 of the line of regression taking x as predictor and y as response.

8. The following table provides observations of y against x.

x	6	6	2	5	4	5	1	4
y	205	195	299	230	270	243	340	240

$\sum x = 33$, $\sum y = 2022$, $\sum x^2 = 159$, $\sum y^2 = 527500$ and $\sum xy = 7743$,

(III) Using your calculator, or the information above, find the liner correlation coefficient of the data and comment on the strength of relationship between the x's and the y's.

(IV) In general what can you tell about the slope from the linear correlation coefficient?

(V) Using your calculator, or the information above, find the slope b_1 and the intercept b_0 of the line of regression taking x as predictor and y as response.

(VI) Write the regression equation $\hat{y} = b_0 + b_1 x$ and use it to predict the value of *y* for x=3.

9. The following table provides the frequency distribution for the number of rooms in US housing units. The frequencies are in thousands.

Rooms	1	2	3	4	5	6	7	8+
Freq.	471	1,470	11,715	23,468	24,476	21,327	13,782	15,647

Total number of housing units: 112,356

For a US housing unit selected at random, let

A = event the unit has at most 4 rooms.

B = event the unit has at least 2 rooms.

C = event the unit has between 5 and 7 rooms inclusive.

D = event the unit has more than 7 rooms

(i) Describe each of the above events (e.g. A = {1,2,3,4}= the unit has 1 rm or 2 rms or 3 rms, or 4 rms.).

(ii) Find P(A), P(B) and P(C). (e.g. $P(A) = \frac{f}{N} = \frac{471+1470+11715+23468}{112356}$)

(iii) Describe each of the events and find their probabilities: (not A), (A &B), (C or D).

10. In Problem 9 above, X= "number of rooms" is a random variable. Using this notion the event A can be written as A = $\{X \le 4\}$. Rewrite the events B, C and D using the random variable notation.
11. The following table displays a frequency distribution for the number of crew-members on each shuttle mission from April 1981 to July 2000.

Crew Size	2	3	4	5	6	7	8
Frequency	2	1	2	36	18	33	2

Let X denote the crew size of the randomly selected shuttle mission between April 1981 and July 2000.

(i) What are the possible values of the random variable X? (Ans. X=2,3,4,5,6,7,8)

(ii) Use random variable notation to represent the event that the space shuttle mission obtained has a crew of size 6. (Ans. {X=6})

(iii) Find P(X=6). (Ans. $P(X = x) = \frac{Freq.(\{X = x\})}{sum\ of\ \ all\ frequencies} = \frac{18}{94}$).

(iv) Find P(X=4) and interpret it in terms of percentages.

(v) Obtain the probability distribution of X, by writing the value of P(X) under each possible value of X in the following table

X	2	3	4	5	6	7	8
P(X)					.1915		

12. Given in the table below are two investment options.

Option A			Option B		
Profit =x	P(X =x)	$xP(X = x)$	Profit = x	P(X =x)	
-2000	.2	-400	-3000	.2	
100	.1	10	-1000	.1	
1000	.3	300	2000	.2	
2000	.3	600	3000	.3	
4000	.1	400	4000	.2	
		910			
For option A $E(X) = \sum xP(X = x)$ =910					

For the given probability distribution of the random variable of option A, the expected value (or the mean m) = $\sum xP(X = x)$ = 910. Find the expected value of profit in option B and decide which gives higher expected value. (Note. The negative profit is loss, as you already know.)

12. A fair die is rolled. Count "six showing up" as success denoted as *s*.

(i) Find P(s) = (probability of success). (Ans. P(s) = 1/6 why?)

(ii) Denote failure by *f*, indicate what you mean by it in this context, and find P(*f*).

(iii) Give all the necessary reasons to establish that: Rolling a (fair) die *looking for six* is a Bernoulli trial.

(iv) Suppose that a fair die is rolled 4 times and let X denote the number of times six shows up. What are the possible values of X? (Ans. X = 0,1,2,3,4)

(v) Use the binomial probability formula to find P(X =2). (Note: Here *n = the number of trials =4, the probability p of success is 1/6 and the formula is:*

$$P(X = x) = \binom{n}{x} p^x (1-p)^{(n-x)}.$$

(vi) Use the formula in part (v) to verify the following table and use the following table to compute the expected value of X.

X	0	1	2	3	4
P(X)	.48225	.3858	.11574	.01543	.00077

(a) Find the expected value of X.
(b) Assess the normality of the distribution of P(X).
(c) Find P(X<2) from the above table.

More to come in the next instalment.