Middle East Technical University **Spring 2010-2011**
Department of Economics
ECON 205

Instructor: Dr. Ozan Eruygur
Research Assistant: Damla Hacıbrahimoğlu

PROBLEM SET 01

PROBLEM 1

Acquired immunodeficiency syndrome (AIDS) has become one of the most devastating diseases in modern society. The numbers of cases of AIDS (in thousands) reported in 25 major cities in the United States during 1992 are as follows:

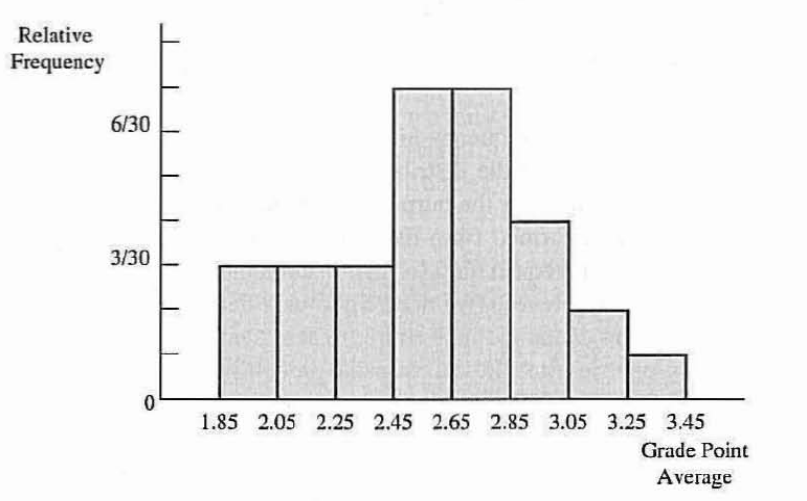
38.3	6.2	3.7	2.6	2.1
14.6	5.6	3.7	2.3	2
11.9	5.5	3.4	2.2	2
6.6	4.6	3.1	2.2	1.9
6.3	4.5	2.7	2.1	1.8

a) Construct a relative frequency histogram to describe these data.
b) What proportion of these cities reported more than 10,000 cases of AIDS in 1992?
c) If one of the cities is selected at random from the 25 for which the preceding data were taken, what is the probability that it will have reported fewer than 3000 cases of AIDS in 1992?

PROBLEM 2

Given here is the relative frequency histogram associated with grade point averages (GPAs) of a sample of 30 students.

a) Which of the GPA categories identified on the horizontal axis are associated with the largest proportion of students?
b) What proportion of students had GPAs in each of the categories that you identified?
c) What proportion of the students had GPAs less than 2.65?



PROBLEM 3

Resting breathing rates for college-age students are approximately normally distributed with mean 12 and standard deviation 2.3 breaths per minute. What fraction of all college-age students have breathing rates in the following intervals?

a) 9.7 to 14.3 breaths per minute
b) 7.4 to 16.6 breaths per minute
c) 9.7 to 16.6 breaths per minute
d) less than 5.1 or more than 18.9 breaths per minute

PROBLEM 4

It has been projected that the average and standard deviation of the amount of time spent on-line using the Internet are, respectively, 14 and 17 hours per person per year.

a) What value is exactly one standard deviation below the mean?
b) If the amount of time spent on-line using the Internet is approximately normally distributed, what proportion of the users spend an amount of time on-line that is less than the value you found in part (a)?
c) Is the amount of time spent on-line using the Internet approximately normally distributed? Why?

PROBLEM 5

The following results on summations will help us in calculating the sample variance s^2. For any constant c,

a) $\sum_{i=1}^{n} c = nc$

b) $\sum_{i=1}^{n} cy_i = c\sum_{i=1}^{n} y_i$

c) $\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

Use (a), (b) and (c) to show that $s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right]$

PROBLEM 6

Refer to Problem 1 answer the following,

a) Calculate y and s for the data given.

b) Calculate the interval $y \pm ks$ for k= 1, 2, and 3. Count the number of measurements that fall within each interval, and compare this result with the number that you would expect according to the empirical rule.

PROBLEM 7

In Problem 1, there is one extremely large value (38.3). Eliminate this value and calculate y and s for the remaining 24 observations. Also, calculate the intervals $y \pm ks$ for k= 1, 2, and 3; count the number of measurements in each; and compare these results with those predicted by the empirical rule. Compare the answers here to those found in Problem 6. Note the effect of a single large observation on y and s.

PROBLEM 8

The *range* of a set of measurements is the difference between the largest and the smallest values. The empirical rule suggests that the standard deviation of a set of measurements may be roughly approximated by one fourth of the range (i.e., range/4). Calculate this approximation to s for the data sets given below. Compare the result in each case to the actual, calculated value of s.

Data Set 1

964	3168	1256	997	1477	951	1514	1713
1040	1080	2325	1159	1151	1102	1233	1120
1358	1013	1261	1317	1615	1185	1588	948
968	1007	1103	1312	565	1500	1347	1567
1165	1189	1056	1216	1037	1088	1251	1104
786	870	923	1051	1207	1090	1592	1292
1416	1385						

Data Set 2

0.74	6.47	1.9	2.69	0.75
0.32	9.99	1.77	2.41	1.96
1.66	0.7	2.42	0.54	3.36
3.59	0.37	1.09	8.32	4.06
4.55	0.76	2.03	5.7	12.48

Data Set 3

38.3	6.2	3.7	2.6	2.1
14.6	5.6	3.7	2.3	2
11.9	5.5	3.4	2.2	2
6.6	4.6	3.1	2.2	1.9
6.3	4.5	2.7	2.1	1.8

PROBLEM 9

The manufacturer of a new food additive for beef cattle claims that 80% of the animals fed a diet including this additive should have monthly weight gains in excess of 20 pounds. A large sample of measurements on weight gains for cattle fed this diet exhibits an approximately normal distribution with mean 22 pounds and standard

deviation 2 pounds. Do you think the sample information contradicts the manufacturer's claim? (Calculate the probability of a weight gain exceeding 20 pounds.)

PROBLEM 10

Prove that the sum of the deviations of a set of measurements about their mean is equal to zero; that is,

$$\sum_{i=1}^{m}(y_i - \overline{y}) = 0$$

PROBLEM 11

There were 1521 NCAA tournament basketball games played between 1939 and 1992. If the total number of points is computed for each game, the resultant totals are approximately normally distributed with mean 143 and standard deviation 26.

a) What percentage of the games ended with a total score in excess of 169 points?
b) What percentage of the games ended with a total score between 117 and 195?
c) Would you expect the total score to exceed 225? Why or why not?