Topics in Probability and Statistics

by Robert J. Rapalje

SEMINOLE COMMUNITY COLLEGE

An Equal Access/Equal Opportunity Community College

INTRODUCTION

Statistics is the mathematics of the collection, organization, and interpretation of numerical data. The applications of statistics today to politics, athletics, higher education, and business in general seem limitless. Almost any undergraduate or graduate major involves at least one course (or more) in statistics.

Public opinion polls on various products provide statistics that indicate which products are preferred and how products may be improved. In an election year, these polls become especially significant in the political arena, from predicting voter turnout to anticipating the outcomes of the elections themselves. More will be said about this later.

In athletics, individual and team statistics are maintained to compare players and teams, to keep track of records, and to recognize and to generate improvement.

In the last few years, the application of statistics to the world of business seems to have made statistics a requirement for nearly every field of study. The reason is quite simple. Statistics is the collection and interpretation of numerical data. Numerical data is the language of computers. Thus statistics and computers go hand in hand. As our society becomes more and more dependent upon computers and data processing, the need for statistics to analyze the data will become more and more important.

This chapter is actually just an introduction to statistics. Such topics as histograms (or bar graphs), sampling, measures of central tendency, variance and the normal distribution will be covered. These are really just the tools for statistics which involves hypothesis testing, statistical inference, and many more advanced topics.

HISTOGRAMS

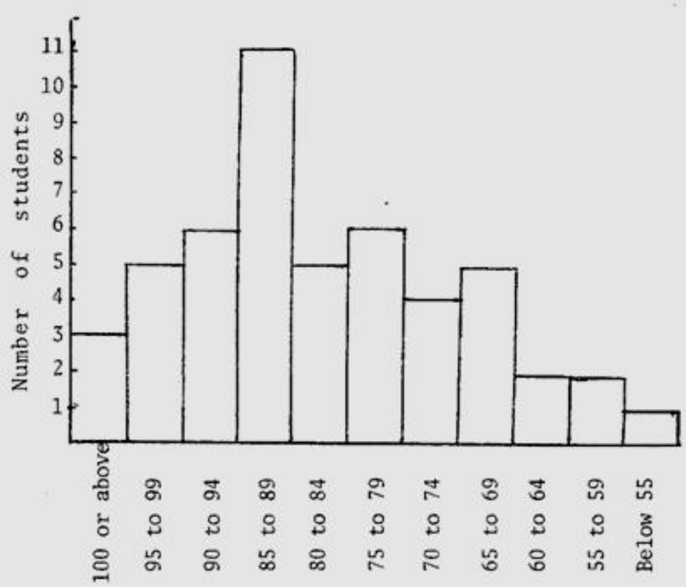
The following set of test grades were scored by a recent class:

89	91	85	81	75
92	85	68	103	70
88	63	83	76	95
93	65	85	59	78
84	89	89	98	100
60	67	75	82	65
75	83	70	101	87
90	73	89	43	97
70	56	65	85	95
89	90	96	90	77

Generally, a table of numbers such as the above does little in providing information about the numbers, such as how well or how badly did the student do? It might be helpful to arrange a frequency distribution for the grades. This is done by separating the grades into "classes", such as "100 or above," "95 to 99," "90 to 94," "85 to 89," etc. Then a tally of the grades is made as follows:

Class	Tally	Frequency
100 or above	|||	3
95 to 99	𝍸	5
90 to 94	𝍸 |	6
85 to 89	𝍸 𝍸 |	11
80 to 84	𝍸	5
75 to 79	𝍸 |	6
70 to 74	||||	4
65 to 69	𝍸	5
60 to 64	||	2
55 to 59	||	2
Below 55	|	1
		50

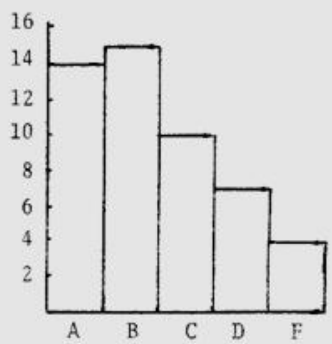
This information can then be shown in what is known as a histogram or bar graph. The horizontal axis represents the classes we have listed, and the vertical axis represents the frequency that was tallied for each class. Notice that the frequency must go up to eleven, since that was the largest.



The selection of the classes was arbitrary. However, the "width" of each class should be the same. That is, a class from "95 to 99" used in the same frequency distribution as a class from "80 to 89" would give a distorted histogram. An exception to this rule would be to separate the grades into "letter grades" where:

Class	Tally	Frequency
A = 90 or above	卌 卌 IIII	14
B = 80 to 89	卌 卌 卌	15
C = 70 to 79	卌 卌	10
D = 60 to 69	卌 II	7
F = Below 60	IIII	4

The resulting bar graph would be:



In this distribution notice that the class of "B", "C", and "D" each have equal widths (from 80-89, 70-79, and 60-69) but the "F" grade has a range of from 0 to 55. Also the "A" grade ranges from 90 to some number above 100. This may account in part for the large number of "A's".

PROBLEMS:

1. The following are free-throw averages of the members of the 1979-80 Seminole Community College basketball team.*

.660	.721	.750
.619	.778	.559
.875	.745	.538
.821	.600	.641

Using classes of .500 to .599, .600 to .699, .700 to .799, etc. and construct a frequency distribution and a histogram for this information.

2. The following are the "field shot" percentages for members of the Seminole Community College basketball team. Using classes as in #1, construct a frequency distribution and histogram.

.491	.447	.527
.400	.504	.498
.500	.505	.383
.547	.556	.496
.538		

3. The following numbers represent the number of points scored per game by the 1979-80 S.C.C. basketball team. Using class intervals of 60 to 69, 70 to 79, 80 to 89, 90 to 99, and 100 to 109, construct a frequency distribution and histogram:

81	80	77	65	89	79
89	102	107	72	99	99
76	92	71	60	66	80
84	86	75	77	85	74
68	83	74	88	69	60
86					

*This and all other basketball statistics courtesy of Coach Joseph A. Sterling.

4. Members of the 1978-79 baseball team** had the following numbers of base hits:

4	25	10	9
40	9	60	52
44	19	49	16
10	37	31	2

Construct a frequency distribution and histogram using classes 0 to 9, 10 to 19, 20 to 29, etc.

5. The members of the same baseball team scored the following numbers of runs during the year:

6	17	22	4
21	6	50	33
26	12	23	2
4	22	3	6

Construct a frequency distribution and histogram using classes as in problem #4.

6. The "getting on base" averages for the baseball players are as follows:

.304	.330	.305	.467	.278
.413	.314	.536	.288	.250
.446	.299	.167	.452	
.300	.435	.383	.378	

Construct a frequency distribution and histogram as in #1 and #2.

**This and all other baseball statistics courtesy of Coach Jack Pantelias.

7. In the fall semester of 1978, the 28 Florida Community Colleges*** had the following enrollments (by headcount):

9,773	1,643	39,562	1,457
15,385	3,551	915	14,445
2,336	11,478	3,492	6,806
1,054	3,942	8,637	3,882 SCC
5,778	2,617	2,639	849
4,009	1,926	8,029	3,192
13,833	4,644	4,474	8,957

Construct a frequency distribution and histogram using class intervals of width 2000. (That is, 0 to 2000, 2001 to 4000, ..., 12,001 to 14,000, 14001 to 16,000, above 16,000.)

MEASURES OF CENTRAL TENDENCY

There are actually three very common methods of finding an average of a set of numbers. The most common meaning of the word "average" refers to the process of adding the numbers and dividing that sum by the numbers of numbers. This is called the arithmetic mean or the mean of the numbers.

If a set of numbers is arranged in order from smallest to largest or from largest to smallest, the middle grade is said to be the median. If there are an even number of numbers, then the median is obtained by averaging the two middle numbers.

The third measure of central tendency is called the mode. The mode is simply the most frequently occurring number. It is possible to have more than one mode. However, if no number occurs more than once, there is no mode for that set of numbers.

***Community College information courtesy of Mr. A. Norris Miner, Director of Management Systems.

EXAMPLE: Suppose a set of test grades is as follows:

72, 86, 56, 92, 100, 28, 100

Find: a. the mean, b. the median, and c. the mode

SOLUTION; a. $\text{mean} = \frac{72 + 86 + 56 + 92 + 100 + 28 + 100}{7}$

$= \frac{534}{7} = 76.3$

b. median - arranging in order:

28, 56, 72, 86, 92, 100, 100

The middle grade is 86.

c. mode - the most frequent score is 100.

If a letter grade were to be assigned, notice the difference that the three measures produce: the mean is a "C", the median is a "B", and the mode is an "A+".

Some instructors use letter grades to average, instead of number grades, and use a 4-point scale.

90-100	= A	4 points
80- 89	= B	3 points
70- 79	= C	2 points
60- 69	= D	1 point
Below 60	=F	0 points

In the previous example, this translates:

72 = C = 2 points
86 = B = 3 points
56 = F = 0 points
92 = A = 4 points
100 = A = 4 points
28 = F = 0 points
100 = A = 4 points

The mean of these numbers figured in this way is:

$\text{Mean} = \frac{2 + 3 + 0 + 4 + 4 + 0 + 4}{7}$

$= \frac{17}{7} = 2.4$ or C

In this case, the letter grade average would be the same by either method. However, this is not always the case.

EXAMPLE: Given the grades 90, 60, 82, 71, 20,

a. Find the mean.

b. Convert to letter grades and find the mean of the letter grades on a 4 point scale.

SOLUTION: a. Mean $= \frac{90 + 60 + 82 + 71 + 20}{5}$

$= \frac{323}{5} = 64.6 = D$

b. 90 = A = 4

60 = D = 1

82 = B = 3

71 = C = 2

20 = F = 0

10

Mean $= \frac{10}{5} = 2.0 = C$

Sometimes it does not matter how the grade is computed, but sometimes it does. The problem here is that the numerical grade of 20 is a very low "F", and it really pulls down the average. It is therefore important to know how you are being graded.

EXAMPLE: Find the mean, median and mode for:
60, 80, 70, 100, 100, 70

SOLUTION: Mean $= \frac{60 + 80 + 70 + 100 + 100 + 70}{6}$

$= \frac{480}{6} = 80$

Median: 60, 70, 70, 80, 100, 100
Since there is an even number, we must average the middle two grades.
Median $= \frac{70 + 80}{2} = \frac{150}{2} = 75$

Mode: There are two modes = 70 and 100

If a frequency distribution is used, the mean becomes a little like expected value (see page P-17). In this situation we multiply the frequency of each grade times the grade, sum these, and then divide by the total number of grades. This is the weighted mean. In a formula, where $x_1, x_2, x_3, \ldots$ is the set of grades which occur with frequency $f_1, f_2, f_3, \ldots,$

$$\text{Weighted Mean} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + x_3 \cdot f_3 + \ldots}{f_1 + f_2 + f_3 + \ldots}$$

It really isn't as bad as it looks:

EXAMPLE: Suppose 100 students took a test and their scores are arranged with the following frequency:

Grade		Frequency
100		8
90		7
85		10
50		5
75		14
70		24
65		10
60		8
55		5
50		6
40		3
	Total	100

Find the weighted mean or average.

SOLUTION:

Grade	Frequency	(Grade) x (Frequency)
100	8	800
90	7	630
85	10	850
80	5	400
75	14	1050
70	24	1680
65	10	650
60	8	480
55	5	275
50	6	300
40	3	120
Total	100	Total 7235 (weighted total)

$$\text{Weighted Mean} = \frac{\text{weighted total}}{\text{total number of grades}}$$

$$= \frac{7235}{100} = 72.35$$

Observe that the mean of the grades without taking the frequency into account would be meaningless.

EXAMPLE: A student earns the following grades: 6 hours of A, 4 hours of B, 3 hours of C. To compute his grade point average (GPA), quality points are assigned with A = 4 points, B = 3 points, C = 2 points, D = 1 point, F = 0 points, (W or Audit do not count at all in the GPA.) Find the GPA.

SOLUTION:

Grade	Hours	Quality Pts.		Total Pts.
A	6 hrs. x	4 points	=	24 points
B	4 hrs. x	3 points	=	12 points
C	3 hrs. x	2 points	=	6 points
	13 TOTAL			42 TOTAL

$$\text{GPA} = \frac{42}{13} = 3.23$$

PROBLEMS:

1. A student in Essential Math earns the grades:

   75, 85, 95, 65, and 75.

   a. Find the mean.

   b. Find the median

   c. Find the mode.

2. A student in Essential Math earns the grades:

   75, 85, 95, 65, and 75. If the instructor allows the student to drop the lowest grade,

   a. Find the mean.

   b. Find the median

   c. Find the mode.

3. A student in Essential Math earns the grades:

   100, 72, 86, 100, 86 and 90.

   a. Find the mean.

   b. Find the median.

   c. Find the mode.

4. A student in Essential Math earns the grades:

   100, 72, 86, 100, 86 and 90. If the instructor allows the student to drop the lowest grade,

   a. Find the mean.

   b. Find the median.

   c. Find the mode.

5. A student in Essential Math earns the grades:

32, 65, 24, 76, 68 and 65.

a. Find the mean.

b. Find the median

c. Find the mode

d. If a grade of 60 is passing, did the student pass most of his tests?

e. If a mean (average) grade of 60 is required to pass the course, did the student pass the course?

6. The average monthly temperatures for the state of Florida for a certain year were as follows:

J	58°	M	70°	S	74°
F	60°	J	74°	O	70°
M	65°	J	79°	N	65°
A	68°	A	80°	D	60°

Find the annual average temperature. Is this method valid? Why or why not?

A student's grade point average (GPA) is computed as a weighted mean, as illustrated by the following example.

A student earns the following grades:

4 hours	A	(4 quality points)	16
5 hours	B	(3 quality points)	15
3 hours	C	(2 quality points)	6
12 hours total			37

$$\text{GPA} = \frac{37}{12} = 3.08$$

Notice also that an "F" is 0 quality points, but that a W or Audit does not affect the GPA.

7. Compute the GPA for a student who has earned the following grades: 3 hours A, 6 hours B, 3 hours C, and 3 hours D.

8. Compute the GPA for a student who has earned the following grades: 3 hours B, 9 hours C, 3 hours F.

9. If the student in the previous problem had dropped the course he was failing, his grades would have been 3 hours B, 9 hours C, 3 hours W. Find his GPA.

10. A student needs a 1.5 GPA to be eligible for sports and must also complete 10 semester hours. He has 3 hours B, 6 hours D, and is failing 3 hours.
    a. What is his GPA if he drops the course he is failing?
    b. What is his GPA if he takes the F?
    c. What must the student do in order to stay eligible?

11. A student needs a GPA of 2.0 and 12 semester hours to be eligible for Campus Governance Association.(CGA) The president of CGA is taking 15 semester hours, and has a B in 3 hours, C in 6 hours, D in 3 hours, and is failing 3 hours. By hard work, he is able to pull his failing grade up to a D, the day before the deadline to withdraw from a course. What should he do? Why?

Nearing graduation time, a student has completed all of his required courses, with the following grades. (He needs a 2.0 average with 60 hours of credit plus 4 hours of P.E.) Find GPA and determine if he is eligible for graduation.

12. 3 hours A, 16 hours B, 26 hours C, 19 hours D, 6 hours F, 12 hours W.

13. 32 hours A, 19 hours B, 6 hours C, 9 hours W.

14. 32 hours A, 19 hours B, 15 hours C.

15. 16 hours A, 20 hours B, 28 hours C, 2 hours D, 6 hours W.

16. 42 hours A, 24 hours B, 3 hours W.

17. On a certain job, 10 people earned \$400, 6 people earned \$200, and 24 people earned \$100. Find the weighted mean (that is, the average amount earned per person).

18. In a certain tribe (Seminoles, of course!) the big chief earns \$54,000 per year. There are 20 assistants to the big chief who earn \$32,000. Next there are 100 braves who earn an average of \$22,000, and finally there are 80 squaws, who earn only \$8500 per year.

    a. Find the unweighted mean of these salaries.

    b. Find the weighted mean of these salaries.

    c. Which if the more representative average?

19. If a basketball team has shooting averages of .491, .400, .500, .547, and .447, does this mean that their team average is .477? Why or why not?

20. To get the team average in #19, more information is needed.

Shots attempted	Shots made	Percent
224	110	.491
50	20	.400
32	16	.500
236	129	.547
367	164	.447

Now find the team average.

MEASURES OF DISPERSION

The measures of central tendency in the previous section frequently do not provide enough information. For example, consider a city in Texas whose seasonal temperatures might be 45°, 80°, 95°, and 80°, and a city in Hawaii whose seasonal temperatures might be 73°, 75°, 77°, and 75°. Both have an average temperature of 75°, but their climates are certainly not similar.

As another example, consider a baseball team that can average 12 runs per game. (Note: In the 1978-79 year, S.C.C. opponents scored more than 12 runs only twice during the year.) If the team were consistent, scoring between 10 and 14 runs per game, they would win nearly all of their games. However, if they wavered, half of the time booming out 20 runs per game but the other half of the time struggling to get 4 runs (still an average of 12) they will lose a lot of games, since the opponents usually score four or more runs. It would be a "winning" season but not a championship team.

In the classroom, a class of ten students may average 70 (passing) with only one failure in the class: 80, 80, 75, 75, 70, 70, 70, 70, 60, 50. Or it can average 70 with only three students passing: 100, 100, 100, 58, 58, 58, 57, 57, 56, 56. Certainly different teaching techniques are necessary.

From the previous examples, it is obvious that we need to <u>measure dispersion</u> (or spread) as well as to measure the central tendency (or average). A <u>deviation</u> is the difference between the individual measure and the mean. (If the measure is more than the mean, then the deviation is positive; if less, then it is negative.)

The <u>standard deviation</u> is like an average of the deviations. It is the square root of the average of the squares of the deviations. We represent the standard deviation by the Greek letter, σ, (read sigma). If x_1, x_2, x_3 ... are scores and if M is the average or mean of "n" scores, then

$$\sigma = \sqrt{\frac{(x_1-M)^2 + (x_2-M)^2 + (x_3-M)^2 + \ldots}{n}}$$

Again it is not as bad as it looks. Consider the following example:

EXAMPLE: Find the standard deviation of:

a. 8, 10, 12, 14, 16, 18, 20

b. 1, 8, 14, 21, 26

SOLUTION: a. <u>Step 1</u>: Find mean $= \frac{8 + 10 + 12 + 14 + 16 + 18 + 20}{7}$

$$= \frac{98}{7} = 14$$

<u>Step 2</u>: Find deviations:	<u>Step 3</u>: Square deviations:
$8 - 14 = -6$	$(-6)^2 = 36$
$10 - 14 = -4$	$(-4)^2 = 16$
$12 - 14 = -2$	$(-2)^2 = 4$
$14 - 14 = 0$	$0^2 = 0$
$16 - 14 = 2$	$2^2 = 4$
$18 - 14 = 4$	$4^2 = 16$
$20 - 14 = 6$	$6^2 = 36$

<u>Step 4</u>: Sum the squares ⟶ 112

<u>Step 5</u>: Divide by "n" $\frac{112}{7} = 16$

<u>Step 6</u>: Take square root: $\sigma = \sqrt{16} = 4$

b. Step 1: Find mean $= \frac{1 + 8 + 14 + 21 + 26}{5}$

$= \frac{70}{5} = 14$

Step 2: Find deviations:

$1 - 14 = -13$
$8 - 14 = -6$
$14 - 14 = 0$
$21 - 14 = 7$
$26 - 14 = 12$

Step 3: Square deviations:

$(-13)^2 = 169$
$(-6)^2 = 36$
$0^2 = 0$
$7^2 = 49$
$12^2 = 144$

Step 4: Sum the squares ⟶ 398

Step 5: Divide by "n" $\frac{398}{5} = 79.6$

Step 6: Take square root: $\sigma = \sqrt{79.6} = 8.92$ approximately

Comparing results of the previous example, it could be said of the second set of numbers that they are much more scattered. The "average deviation" from the mean is almost 9, whereas in the first set, the numbers tend to be closer to the mean with an "average deviation" of only 4.

Another measure of dispersion is known as variance. Variance is simply the square of the standard deviation, or more simply the variance is the quantity obtained in the fifth step of the standard deviation problem. In the previous examples, the variance was 16 and 79.6 respectively.

PROBLEMS:

Find the mean, the variance, and the standard deviation of each of the following sets of numbers:

1. 10, 12, 14, 16, 18
2. 1, 8, 14, 21, 26
3. 50, 58, 52, 60, 55

(cont'd)

4. 150, 158, 152, 160, 155
5. 18, 25, 25, 14, 20, 24
6. 8, 8, 8, 8, 8, 8
7. 100, 85, 60, 30, 95
8. 80, 86, 85, 83, 90, 92, 93
9. 80, 96, 42, 84, 58, 90
10. 18, 96, 30, 84, 92, 45, 63, 100

SAMPLING AND THE NORMAL POPULATION

The problems we have done thus far have been selected for their simplicity (believe it or not!) Since real life problems can be quite enormous or expensive, some short cuts must be developed. The "short cut" to be discussed in this section is "sampling." First, some terminology.

A "population" is generally whatever it is that is being studied. In a political campaign, the population is probably the set of all voters. In a baseball coach's mind, the population is probably the members of his team. To a teacher, the population is the set of all students in his class. To someone in the General Electric plant, the population may be a set of light bulbs or flash cubes. Notice that the word "population" refers to data, and not necessarily to people, as we normally think.

It is often not possible to obtain statistics on an entire population. Obviously the <u>entire</u> population may be too large. But even worse than that, how does one test a population of flash cubes or fire crackers? By using them all up in order to decide that they were all good? Certainly not! He does not test the entire population, but rather he selects a sample for testing, usually at random. Then the results of the sample test may be used to predict the results of the population.

Care must be taken to ensure that the entire population is accurately represented by the sample. For example in 1948, a random telephone poll showed that Dewey would certainly defeat Truman in the Presidential election. It was certainly a shock to Mr. Dewey when suddenly it was realized that many of Truman's supporters did not have telephones and thus they had not been counted in the sample.

Next we consider the idea of a normal population or a normal distribution. A normal population assumes that most of the data will be approximately average (or normal) with a lesser amount of the data at either extreme. For example, if the heights of a large number of men were measured and a frequency distribution made, an average height of the men would be found. Most of the men would be "close" to that average height. A few men would be extremely tall and a few extremely short.

Life expectancies, diameters of trees, weights of people, intelligence of people, and many other measures are normally distributed about some mean or average. In fact, the normal distribution is so natural that some instructors use it to determine the grades of their students. Assuming an average class or a normal class, it could be reasoned that most of the students are normal or average, and they should earn a grade of "C". A few will be above average (B), and a few will be below average (D). Then a couple of students (certainly no more than that!) will be excellent, outstanding students (A), and likewise a couple of students will fail. (F)

The validity of this method of grading hinges upon two factors. First, is the class really normal? Second, what is the goal of the course? If subject mastery is the goal of the course, then the students should have scored high grades on the tests and should receive high grades in the course. Ideally, no one should fail. However, if the goal of the course is to determine who is average, to eliminate a few that are below average, and to distinguish a few who are above average, then the normal grading system is appropriate, if the class is correctly assumed to be "normal."

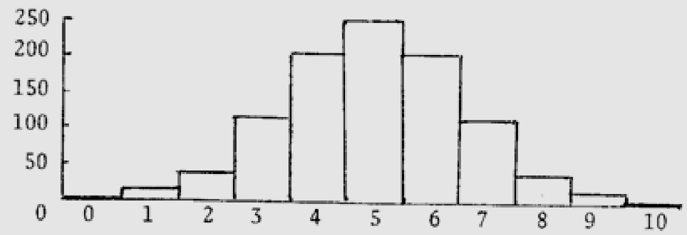
It has been my experience in the community college to have had quite "abnormal" classes. In my classes, there are usually a large number of highly motivated and well-equipped students who perform very well, and also a rather large number of students who, for a number of reasons, (from lack of background to lack of study time), are not able to pass the course and hence withdraw or change to "audit." Therefore, most of my classes are not normally distributed.

Returning to the normal distribution, consider the following:

EXAMPLE: If ten coins are tossed a large number of times, a tally of the number of heads obtained on each toss would give a normal distribution. <u>Theoretically</u>, if the coins are tossed 1024 times, the following frequency results. Draw a histogram to show the distribution.

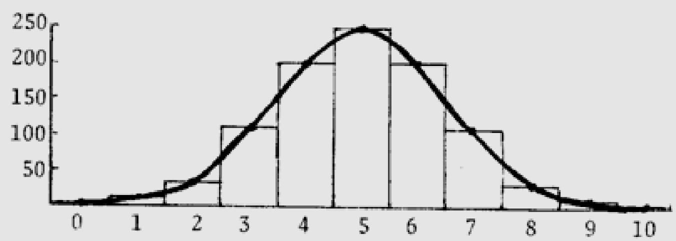
Number of heads	0	1	2	3	4	5	6	7	8	9	10
Frequency	1	10	45	120	210	252	210	120	45	10	1

ANSWER:



Notice that the most frequent outcome is "5 heads." Also an actual tossing of coins will closely approximate this distribution.

This example is a normal frequency distribution. From this, a normal curve may be drawn by connecting the mid-points of the tops of the rectangles with a smooth curve as shown:

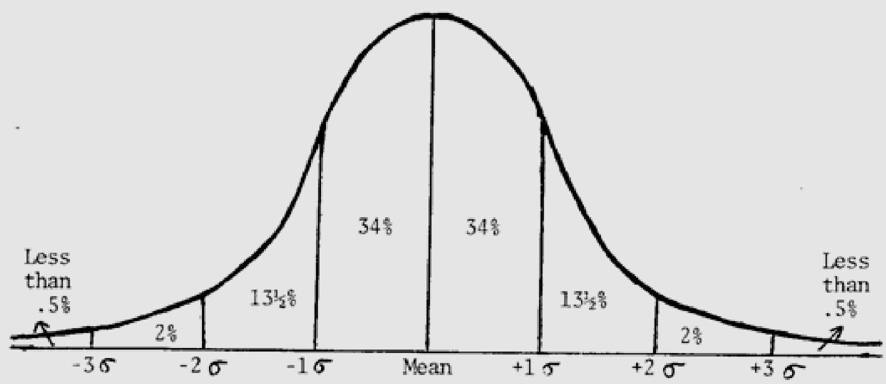


Now that we have discussed the concepts of the standard deviation and the normal curve, we combine them together in the Empirical Rule.

EMPIRICAL RULE: For a normal distribution,

1. approximately 2/3 (68%) of the measures will be within one standard deviation of the mean.
2. approximately 95% of the measures will be within two standard deviations of the mean.
3. almost all of the measures (99.7%) will be within three standard deviations of the mean.

The empirical rule can also be stated graphically. A normal distribution will break down approximately as follows:



Notice that there is 50% on each side of the mean.

Between -1σ and $+1\sigma$ is 34% + 34% = 68% (approximately 2/3)

Between -2σ and $+2\sigma$ is 13½% + 34% + 34% + 13½% = 95%

Between -3σ and $+3\sigma$ is 95% + 2% + 2% = 99%

Instructors often use the normal distribution in their grading systems. This is called "curving the grades" or "grading on the bell curve." All grades within one standard deviation of the mean are C grades, from +1 standard deviation to +2 standard deviations the grades are B, above +2 standard deviations are A grades. Similarly, those grades between -1 and -2 standard deviations are D, and those below -2 standard deviations are F. The following example demonstrates this.

EXAMPLE: A set of grades has a mean of 76 with a standard deviation of 6. Determine the grading scale if the grades are normally distributed and grading is on the bell curve.

ANSWER: C grades are from 76 + 6 = 82 to 76 - 6 = 70

B grades are from 82 to 82 + 6 = 88

D grades are from 70 to 70 - 6 = 64

In summary: A = 88 or above
B = 82 to 88
C = 70 to 82
D = 64 to 70
F = 64 or less

(Borderline grades of 88, 70, and 64 would probably be determined by the clustering of the grades.)

EXAMPLE: In measuring the weights of 1000 men, it is found that the weights are normally distributed, with mean of 160 pounds and standard deviation of 10 pounds.

a. How many are between 150 and 170? (10)
b. How many are between 140 and 180? (20)
c. How many are between 130 and 190? (30)
d. How many weigh more than 190 pounds?

ANSWER:

a. 68% of 1000 = (.68) • (1000) = 680 men
b. 95% of 1000 = (.95) • (1000) = 950 men
c. 99% of 1000 = (.99) • (1000) = 990 men
d. Less than ½% of 1000 = (.005)•(1000) = 5 men

EXAMPLE: A light bulb manufacturer selects a random sample of 100 light bulbs (from a large population) and finds that the average life expectancy of the bulbs tested is 1000 hours, with a standard deviation of 50 hours. If the manufacturer guarantees his light bulbs for 900 hours, approximately how many bulbs out of 1000 will not meet the guarantee?

ANSWER: 900 hours is 2 standard deviations less than the mean. Approximately 2½% of the 1000 will not measure up to the guarantee.

2½% of 1000 = (.025) • (1000) = 25

EXAMPLE: A battery dealer knows that a particular type of battery has an average life expectancy of 3½ years with standard deviation of six months. They advertise the battery as a 4-year battery. If the battery goes bad within 2 years, the customer is given a new battery free. If it goes bad within 4 years, the value of the battery is "pro-rated," and the customer is allowed to buy a new battery at a discounted rate depending upon how much time remains under warranty. 1000 such batteries are sold at a retail price of $56 each, but cost the dealer $35.

a. How many of the batteries will be replaced?

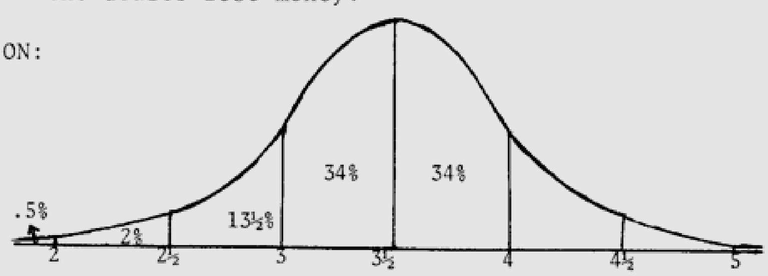
b. How many will last less than the guaranteed 4 years?

c. If the battery fails at 3½ years, how much is the customer allowed on his new battery if the price is still $56?

d. How many of the 1000 batteries will likely fail in each of the six month intervals from 2 years to 4 years?

e. On how many batteries, including those replaced, will the dealer lose money?

SOLUTION:



a. .5% of 1000 = (.005) 1000 = 5 batteries

b. 50% + 34% = 84% or 840 batteries

c. One half year equals 1/8 of life of battery. 1/8 of $56 = $7. The replacement battery costs the customer $49.

d. 34% fail between 3½ and 4 years or 340 batteries

34% fail between 3 and 3½ years or 340 batteries

13½% fail between 2½ and 3 years or 135 batteries

2% fail between 2 and 2½ years or 20 batteries

.5% fail before 2 years or 5 batteries

e.

If battery fails:	Customer is allowed:	Replacement battery $$	Dealer's Prof/loss	No.of Batt.
3½-4 years	$ 7	$56-7=$49	$49-35=$14	340
3-3½ years	$14	$56-14=42	$42-35=$ 7	340
2½-3 years	$21	$56-21-35	$35-35=$ 0	135
2-2½ years	$28	$56-28=28	$28-35=$-7	20
Before 2 yrs.	New Battery	$0	$-35	5

The dealer "loses" on only 25 of 1000 batteries. In fact, his total loss is: 20 x $7 = 140

+ 5 x 35 = 165

$305

But his profit on sales generated by the other "defective" batteries is: 340 x $14 = 4760

340 x $ 7 = 2380

$7140

which more than makes up for it!

PROBLEMS:

1. On a certain test the average grade was 70 with a standard deviation of 10. If 200 students took the test and the results were normally distributed, then how many scored?
   a. between 70 and 80?
   b. between 80 and 90?
   c. between 90 and 100?
   d. between 60 and 80?
   e. between 70 and 100?

2. On a certain test the average grade was 85 with a standard deviation of 5. If 200 students took the test and the results were normally distributed, then how many scored:

   a. between 80 and 85?
   b. between 90 and 95?
   c. between 95 and 100?
   d. between 70 and 80?
   e. less than 70?
   f. more than 90?

In 3-6, the grading system is established by the "bell curve" method of the example. Determine the range of grades for A, B, C, D, and F.

3. The mean is 75 with standard deviation 5.
4. The mean is 70 with standard deviation 10.
5. The mean is 78 with standard deviation 4.
6. The mean is 80 with standard deviation 5.
7. A certain brand of tires is found to have an average life of 20,000 miles with standard deviation 4000 miles.
   a. It may be estimated that 68% of the tires will wear out between ________ and ________miles.
   b. 95% of the tires will wear from ________ to ________ miles.
   c. Nearly all of the tires will last from _____ to _____miles.

8. A brand of batteries has an average life expectancy of 3 years with standard deviation of six months.
   a. The batteries may be guaranteed with 95% certainty to last between ______ and ______ years.
   b. Two-thirds of the batteries will last between ______ and ______ years.
   c. Nearly all of the batteries will last at least ________.
   d. Nearly all of the batteries will be expired within ________ years.

9. A set of test grades is normally distributed with mean 85 and standard deviation 5. If the instructor grades "on the curve," what grade would be required for:
   a. an A? b. a B? c. a C? d. a D? e. an F?

10. A set of test grades is normally distributed with mean 75 and standard deviation 10. If the instructor grades "on the curve," what grade would be required for:
    a. an A? b. a B? c. a C? d. a D? e. an F?

11. A set of test grades is normally distributed with mean 75 and standard deviation 10. The instructor can grade "on the curve," or use a straight grading system of 90-100 A, 80-89 B, 70-79 C, 60-69 D, 59 or less F.
    a. In which system is it harder to score an A?
    b. In which system are there more C's?
    c. In which system are there more F's?

12. Brand x of light bulbs has a mean life of 950 hours with standard deviation 200. Brand y has a mean life of 800 with standard deviation 50.
    a. Which brand is most likely to last at least 600 hours?
    b. Which brand is most likely to last at least 750 hours?
    c. Which brand is most likely to last at least 950 hours?
    d. Which brand has the most "duds"?
    e. Which brand has the best quality control?
    f. If the bulbs are the same price, which is the best buy?

Statistics Practice Test

Show work as necessary. Turn in all work sheets. CALCULATORS - YES!!

1. Find the mean, median, mode:
   a. 9, 2, 7, 11, 14, 6, 14
   b. 100, 85, 30, 85, 70, 75
   c. 79, 82, 53, 93, 77, 54, 75, 100

2. Find the G.P.A. for each of the following: (to the nearest hundredth)

   a. 6 hrs. of A
   3 hrs. of B
   3 hrs. of C

   b. 4 hrs. of A
   13 hrs. of B
   24 hrs. of C
   9 hrs. of D
   3 hrs. of F
   9 hrs. of W

3. In a company, the employees earn the following salaries. Find the average salary (weighted mean).

No. of employees	Salary
6	$40,000
30	$32,000
84	$20,000
50	$15,000
10	$ 9,000

4. Find the variance and standard deviation "σ".

a.	b.	c.
7	8	18
2	2	96
10	16	30
9	4	84
12	26	92
	34	45
		63
		100

5. A brand of light bulbs is found to have an average life of 750 hours with standard deviation of 50. Assuming a normal distribution:
   a. about 2/3 of the bulbs will last between _____ and _____ hours.
   b. almost all of the bulbs will last between _____ and _____ hours.
   c. about 95% will last between _____ and _____ hours.

6. The mean score on an IQ test (normally distribution) is 100 with a standard deviation of 15.

If 800 children took the test,

a. how many scored between 85 and 115? ________

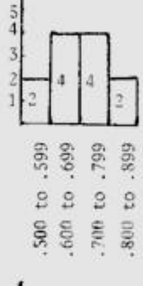
b. how many scored between 115 and 130? ________

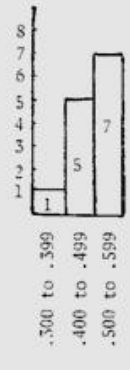
c. how many scored over 145? ________

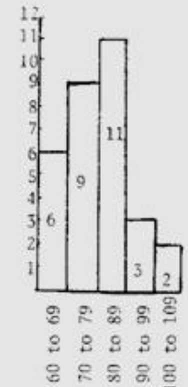
d. how many scored between 55 and 145? ________

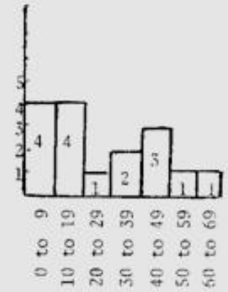
e. how many scored below 85? ________

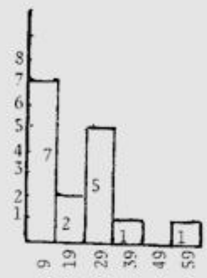
S-5 to S-7

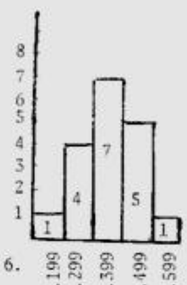
1. 

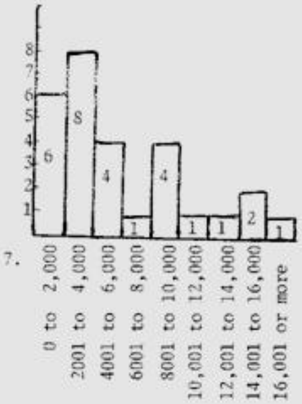
2. 

3. 

4. 

5. 

6. 

7. 

S-12 to S-15

1.a 79
b 75
c 75

2.a 82.5
b 80
c 75

3.a 89
b 88
c 86 & 100

4.a 92.4
b 90
c 86 & 100

5.a 55
b 65
c 65
d Yes
e No

6. 68.58°
Not completely.
The months are not all of equal length.

7. 2.60
8. 1.80
9. 2.25
10.a 1.67
b 1.25
c Pull failing grade up to D for 1.50 average for 12 hrs.
11. Drop course - 2.00
Take D - 1.80
To remain eligible, drop the course.
12. 1.87, Not eligible, GPA too low.
13. 3.46; not eligible (only 57 hrs.)
14. 3.26; eligible
15. 2.76; eligible
16. 3.64; eligible
17. $190.
18.a $29,125.
b. $17,781.09
c. weighted mean.
19. The average of these numbers is .477. This is the team average only if each member has the same number of shots. (See problem #20.)
20. $\frac{439}{909} = .483$

S-18
S-19

1. Mean = 14, Var = 8, σ = 2.83
2. Mean = 14, Var = 79.6, σ = 8.92
3. Mean = 55, Var = 13.6, σ = 3.69
4. Mean = 155, Var = 13.6, σ = 3.69
5. Mean = 21, Var = 16.67, σ = 4.08
6. Mean = 8, Var = 0, σ = 0
7. Mean = 74, Var = 674, σ = 25.96
8. Mean = 87, Var = 20, σ = 4.47
9. Mean = 75, Var = 358.33, σ = 18.93
10. Mean = 66, Var = 888.25, σ = 29.80

S-27

1.a 68
b 27
c 4
d 136
e 99

2.a 68
b 27
c 4
d 31
e 1
f 32

3. A 85 or above
B 80 to 85
C 70 to 80
D 65 to 70
F 65 or less

4. A 90 or above
B 80 to 90
C 60 to 80
D 50 to 60
F 50 or less

ANSWERS

S-27

5. A 86 or above
   B 82 to 86
   C 74 to 82
   D 70 to 74
   F 70 or less

6. A 90 or above
   B 85 to 90
   C 75 to 85
   D 70 to 75
   F 70 or less

7. a 16,000 & 24,000
   b 12,000 to 28,000
   c 8,000 to 32,000

8. a 2 and 4
   b $2\frac{1}{2}$ and $3\frac{1}{2}$
   c $1\frac{1}{2}$
   d $4\frac{1}{2}$

9. a 95
   b 90
   c 80
   d 75
   e less than 75

10. a 95
    b 85
    c 65
    d 55
    e less than 55

11. a curve
    b curve
    c straight

12. a Brand y-100% (Brand x-less than 98%)
    b either Brand-84%
    c Brand x-50% (Brand y-none)
    d Brand x
    e Brand y
    f Brand x

S-30

1. a. Mean = 9
   Median = 9
   Mode =14

   b. Mean = 74.167
   Median = 80
   Mode = 85

   c. Mean = 76.63
   Median = 78
   Mode = None

2. a 3.25
   b 2.11

3. $20,666.67

4. a. Mean = 8
   Var =11.6
   = 3.41

   b. Mean = 15
   Var = 137
   = 11.70

   c. Mean = 66
   Var = 888.25
   = 29.80

S-31

5. a. 700 to 800
   b. 600 to 900
   c. 650 to 850

6. a. 544
   b. 108
   c. 4
   d. 792
   e. 128