STATISTICS

Instructor: Prof. Dr. Doğan Nadi LEBLEBİCİ

Source: Kaplan, Robert M. <u>Basic Statistics for the Behavioral Sciences</u>, Allyn and Bacon, Inc., Boston, 1987. <u>SENTENCES IN THIS POWER</u> POINT PRESENTATION ARE USUALLY BORROWED FROM KAPLAN'S BOOK. © Copyright 2005, Doğan N. LEBLEBİCİ

1

DESCRIPTIVE STATISTICS

In statistics there are at least three different ways to determine if a score is typical. The typical score usually occurs in the center of distribution, and indexes of typicalness are usually called measures of central tendency.

The mean is the arithmetic average. The median is the point representing the 50th percentile in the distribution and the mode is the most common score.

The mean is the most commonly used in statistics. It is simply the arithmetic average. To give an example of how it is calculated, we need a variable which is a quantity that may take on different values.

Symbols used in calculation:

```
X_{(i=1,2,3,4...N)} = Variable
N = Number of Cases
\Sigma = Sign of summation
= Mean
Formula = \Sigma X/N
K
```

Student Name	Variable X	Score
Hasan		60
Deniz	X ₂	70
Mesut	X ₃	60
Emrah	X	80
Derya	X ₅	80
Havva	X ₆	60
Cem	X ₇	70
Salih	X ₈	90
		$\Sigma X = 570$

$$= \overline{X}/N$$

$$\Sigma X = 570$$

 $N = 8$
 $\overline{\mathbf{X}} = 570/8 = 71.25$

The mean has many properties. The most important one is that sum of deviations around the mean always equals zero.

Student Name	Variable	Score	\overline{X}	$x = X - \overline{X}$
Hasan	X ₁	60	71.25	-11.25
Deniz	X_2	70	71.25	-1.25
Mesut	X_3	60	71.25	-11.25
Emrah	X_4	80	71.25	8.75
Derya	X_5	80	71.25	8.75
Havva	X ₆	60	71.25	-11.25
Cem	X_7	70	71.25	-1.25
Salih	X ₈	90	71.25	18.75
				$\Sigma(X-\overline{X})=0$

The median is the point in distribution where the half of cases fall above and the other half fall below. For odd number of cases it is quite easy to calculate median. The formula for the depth of median is $Md_m = (N+1)/2$

Student Name	Score
Hasan	60
Deniz	70
Mesut	60
Emrah	80
Derya	80
Havva	60
Cem	70
Salih	90
Ruhi	85

Student Name	Score
Salih	90
Ruhi	85
Emrah	80
Derya	80
Cem	70
Deniz	70
Hasan	60
Mesut	60
Havva	60

Student Name	Score
Hasan	60
Deniz	70
Mesut	60
Emrah	80
Derya	80
Havva	60
Cem	70
Salih	90

Student Name	Score
Salih	90
Emrah	80
Derya	80
Cem	70
MEDIAN	70
Deniz	70
Hasan	60
Mesut	60
Havva	60

The mode is the score that occurs most frequently in the distribution. The mode in our example is 60. In some cases, no score occured more frequently than other score, each score occured just once. Thus, there is no mode in this example. The mode is useful especially when a <u>nominal scale</u> is used. The mode is rarely used in statistical analysis.

SIMILARITIES AND DIFFERENCES AMONG THE MEAN, MEDIAN, AND MODE

Three measures of central tendency are the same when the distribution of scores is normal. The normal distribution is a theoratical distribution of scores that is symmetrical.

SIMILARITIES AND DIFFERENCES AMONG THE MEAN, MEDIAN, AND MODE

The major consideration with central tendency is how weight to give to extreme scores. While the mean takes into account each score in the distribution, the median finds only the halfway point. Thus, for most of real data set, they differ.

SIMILARITIES AND DIFFERENCES AMONG THE MEAN, MEDIAN, AND MODE

Name	Weight	Name	Weight
Mertcan	49	Mertcan	49
Şule	51	Şule	51
Figen	60	Figen	60
Doğan	65	Doğan	65
Nafiz	75	Çağlar	105
$\overline{\mathbf{X}} = \mathbf{M}\mathbf{d}$	$\overline{\mathbf{X}} = 60$	$\overline{\mathbf{X}} \neq \mathbf{M}\mathbf{d}$	$\overline{\mathbf{X}} = 66$
	$\mathbf{Md} = 60$		Md = 60

SIMILARITIES AND DIFFERENCES AMONG THE MEAN, MEDIAN, AND MODE

The mean considers all scores in the distribution. The mean is often easier to calculate. The mean is used in formal statistical analysis.

SIMILARITIES AND DIFFERENCES AMONG THE MEAN, MEDIAN, AND MODE

The median is useful when we do not want scores at the extreme of a distribution to have a strong impact. Consider about following example of 5 people with their annual income.

SIMILARITIES AND DIFFERENCES AMONG THE MEAN, MEDIAN, AND MODE

Names	Annual income (YTL)
Şeyda	5.000
Hasan	6.000
Mustafa	6.500
Leyla	7.500
Yeşim	1.000.000
	$\overline{X} = 205.000$ Md = 6.500

NORMAL FREQUENCY DISTRIBUTION



© Copyright 2005, Doğan N. LEBLEBİCİ

POSITIVE SKEWNESS



NEGATIVE SKEWNESS



© Copyright 2005, Doğan N. LEBLEBİCİ

In addition to measures of central tendency, we need other indicators that summarize characteristics of scores. The most common indicators are the range, the variance, and the standart deviation.

DATA SET 1	DATA SET 2
6	4
6	6
6	8
6	3
6	6
6	9
$\overline{\mathbf{X}} = 6$	$\overline{\mathbf{X}} = 6$

THE RANGE

The simplest method for describing variability is the range. The range is simply the difference between the highest score and the lowest score. For our prior example, the range for data sets are:

Data Set 1:The range is 6-6=0Data Set 2:The range is 9-3=6

THE VARIANCE

It seems reasonable to use average deviations arround the mean as an indicator of variability. However, it is not feasible because the mean of the deviations around the mean always equals zero.

Student Name	Variable	Score	\overline{X}	$x = X - \overline{X}$
Hasan	X_{I}	60	71.25	-11.25
Deniz	X_2	70	71.25	-1.25
Mesut	X_3	60	71.25	-11.25
Emrah	X_4	80	71.25	8.75
Derya	X_5	80	71.25	8.75
Havva	X_6	60	71.25	-11.25
Cem	X_7	70	71.25	-1.25
Salih	X_8	90	71.25	18.75
				$\Sigma(X-\overline{X})=0$

THE VARIANCE

Average deviations arround the mean being always equals zero mean that sum of deviations with "+" signs equals the sum of deviations with "-" signs. If we can eliminate "-" signs, then the sum of deviations would not equal to zero.

THE VARIANCE

How can we eliminate "-" signs? We can square all of the deviations around the mean. Then we can obtain the average squared deviation around the mean. This is known as the variance. Formula is:



THE STANDARD DEVIATION

The variance gives us the result in squared units. How can we get the result back into units that will make sense to us? The answer is by taking the square root of the variance. The result gives us the standard deviation.



Calculation of Variance and Standard Deviation

Student Name	X	X	X-X	$(\overline{X-X})^2$
Hasan	60	68.57	-8.57	73.44
Deniz	70	68.57	+1.43	2.04
Mesut	60	68.57	-8.57	73.44
Emrah	80	68.57	+11.43	130.64
Derya	80	68.57	+11.43	130.64
Havva	60	68.57	-8.57	73.44
Cem	70	68.57	+1.43	2.04
	ΣX=480	X =68.57		485.68

Calculation of Variance and Standard Deviation

Variance = 485.68/7 = 69.38

Standard Deviation =
$$69.3/8 = 8.32$$