STATISTICS

Instructor:

Prof. Dr. Doğan Nadi LEBLEBİCİ

Source: Kaplan, Robert M. Basic Statistics for the Behavioral Sciences, Allyn and Bacon, Inc., Boston, 1987. SENTENCES IN THIS POWER POINT PRESENTATION ARE USUALLY BORROWED FROM KAPLAN'S BOOK.

SAMPLES AND POPULATIONS

In statistics, there is a distinction between a population and a sample. The distintion reflects itself on various statistical calculations. This means that calculation procedures are different for populations and for samples.

SAMPLES AND POPULATIONS

A population is defined as the entire collection of a set of objects, people, events,etc. Of interest in a particular context. In other words, population refers to the collection of all items that we want to make generalization about.

SAMPLES AND POPULATIONS

A sample is a subset of observations selected from a population.

SAMPLES AND POPULATIONS

We use different formula for the calculation of standard deviation for population and samples. For populations, we use "σ" to show standard deviation. For samples, we use "*S*" to show standard deviation.

SAMPLES AND POPULATIONS

Calculation of Standard Deviation

$$\sigma = \sqrt{\frac{\Sigma\left(X - \overline{X}\right)^2}{N}} \qquad S = \sqrt{\frac{\Sigma\left(X - \overline{X}\right)^2}{N-1}}$$

Finding Points Within Distribution

In distributions, we need to know what each score means for drawing results about scores. For example, what does being 65th in a race mean? Think about the population. Does it mean the same for a population of 70 and for a population of 10.000?

Finding Points Within Distribution

We need to compare a specific score with other scores in the distribution. One way to make comparisons among scores is to rank them.

Finding Points Within Distribution

Percentile Ranks replace simple ranks when we want to adjust for the number of scores in a group. A percentile rank expresses the percentage of scores that fall below a particular score (X_i).

Finding Points Within Distribution

We need to compare a specific score with other scores in the distribution. One way to make comparisons among scores is to rank them.

Finding Points Within Distribution

To calculate a percentile rank to compare a specific score you need to pursue following procedure:

1. Determine how many cases are below the score of interest.
2. Determine how many cases are in the group.
3. Divide the number of cases below by the total number of cases.
4. Multiply the result by 100.

The formula is $P_t=(B/N)x100$

Finding Points Within Distribution

Country	Male Life Expectancy
Canada	69.8
USA	69.3
Austria	68.5
Denmark	71.9
England	70.2
France	69.9
East Germany	68.9
Ireland	69.0
New Zeland	68.9

Finding Points Within Distribution

Country	Male Life Expectancy
Denmark	71.9
England	70.2
France	69.9
Canada	69.8
USA	69.3
Ireland	69.0
East Germany	68.9
New Zeland	68.9
Austria	68.5

Finding Points Within Distribution

In this population, calculation of percentile rank for USA is as follow:

1. Number of cases below USA is 4
(Ireland, East Germany, New Zeland, Austria)
2. Number of cases in group is 9
3. $P_t=(4/9)x100 = 44.4$

This means that USA is in the 44th percentile rank for life expectancy of males.

Standardized Scores and Distributions

There are less boring methods for locating a score within a distribution. One of them is Z scores.

Z Scores

One of the problem with means and standard deviation is that their meanings are not clear enough. It is still difficult to interpret the distribution.

Standardized Scores and Distributions

Z Scores

The Z score is a transformation of data into standardized units that are easier to interpret. It is the difference between a score and the mean divided by the standart deviation. The formula is

$$Z = \frac{X - \overline{X}}{S}$$

Standardized Scores and Distributions

Z Scores

Z score is the deviation of a score X from the mean in standart deviation units. If a score is equal to the mean, its Z-score is 0.

Finding Points Within Distribution

Name	Test Score	X^2
Muhittin	70	4.900
Hasan	85	7.225
Beyza	92	8.464
Sevda	65	4.225
Medet	83	6.889
Celal	98	9.604
Bican	75	5.625
Korkut	90	8.100
Cevdet	60	3.600
Nusret	78	6.084
	ORTALAMA= 79.6	$\Sigma X^2 = 64.716$

Standardized Scores and Distributions

Calculation of mean, standard deviation, and Z scores

$$\overline{X} = \frac{\sum X}{N} = \frac{796}{10} = 79.60$$

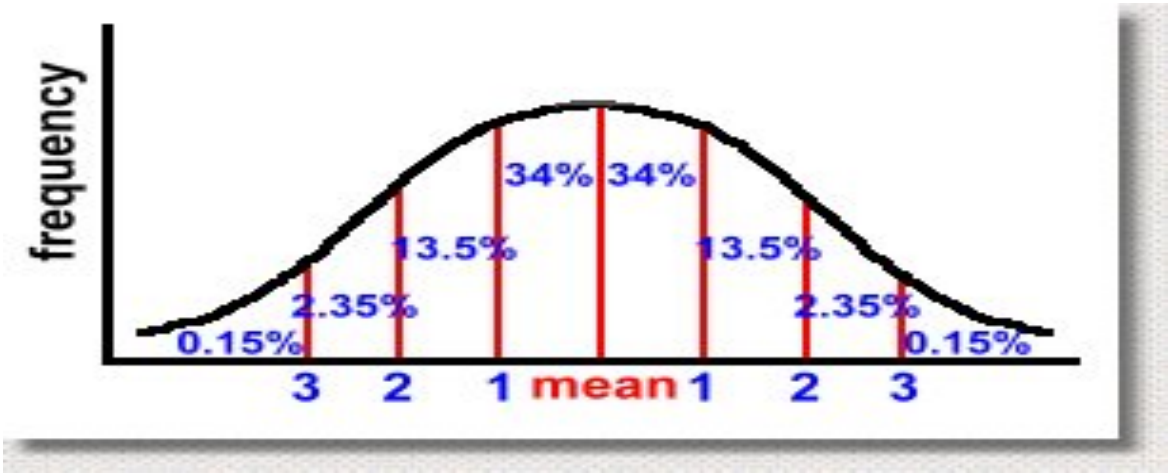
$$S = \sqrt{\frac{\sum X^2 - \frac{\left(\sum X\right)^2}{N}}{N-1}} = \sqrt{\frac{64.716 - \frac{(796)^2}{10}}{10-1}} = 12.267$$

Standardized Scores and Distributions

Calculation of mean, standard deviation, and Z scores

Beyza's Z-score

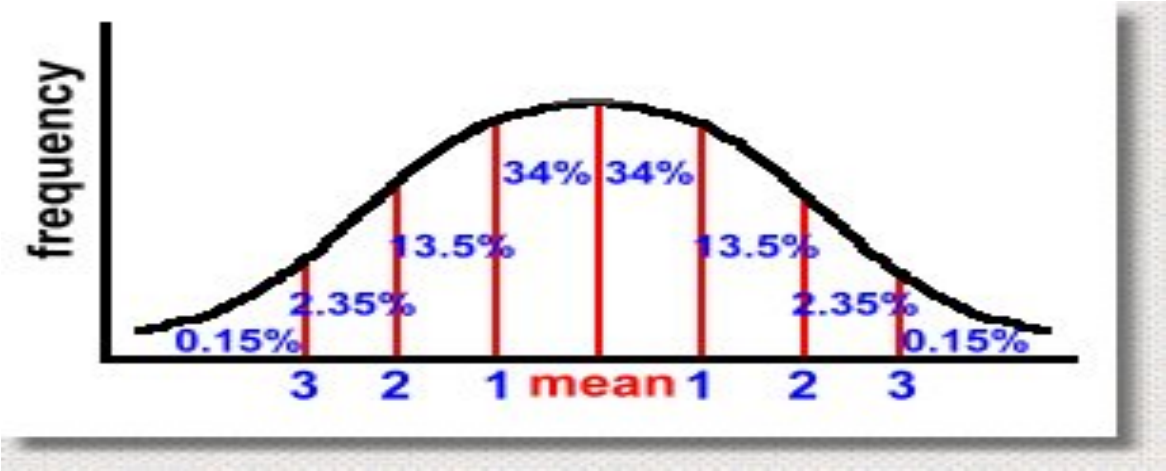
$$Z = \frac{X - \overline{X}}{S} = \frac{92 - 79.6}{12.267} = 1.01$$



Standardized Scores and Distributions

Calculation of mean, standard deviation, and Z scores

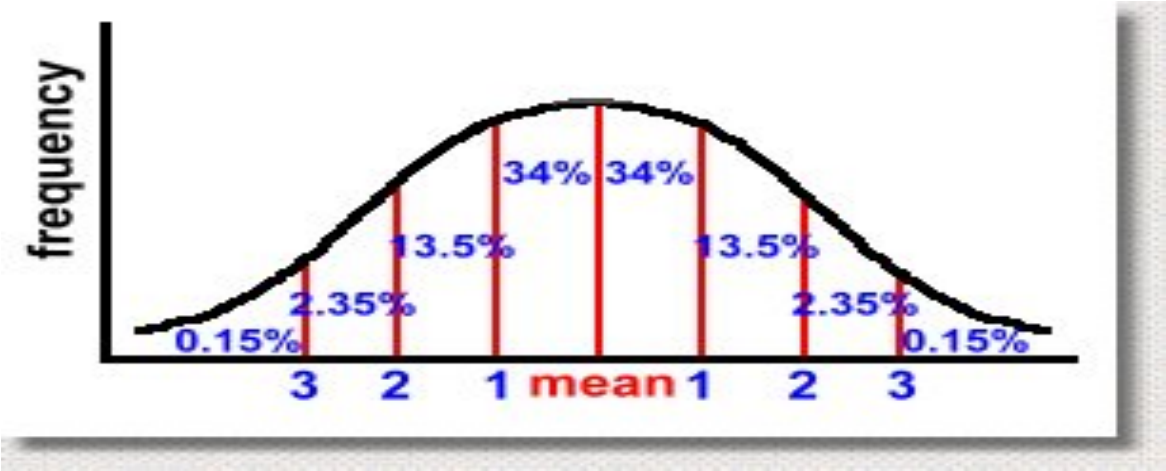
Beyza's Z-score (1.01) means that it is on the area between 1 and 2 in the normal distribution curve. It clearly means that Beyza's score is above (since Z-score is positive) average (mean) and its percentile rank is 84.38.



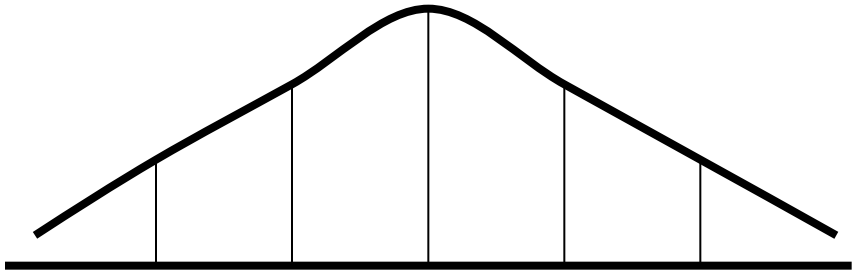
Standardized Scores and Distributions

Calculation of mean, standard deviation, and Z scores

In other words, Beyza's score is above 84.38 % of all scores.



Standardized Scores and Distributions



Raw Score	20	30	40	50	60	70	80
Z-score	-3	-2	-1	0	+1	+2	+3
Percentile rank	.13	2.28	15.87	50.00	84.13	97.72	99.87

Finding Points Within Distribution

Name	Test Score	*Z-Scores*	Z^2
Muhittin	70	-0.78	0.60
Hasan	85	0.44	0.19
Beyza	92	1.01	1.02
Sevda	65	-1.19	1.41
Medet	83	0.27	0.07
Celal	98	1.49	2.22
Bican	75	-0.37	0.13
Korkut	90	0.84	0.70
Cevdet	60	-1.59	2.52
Nusret	78	-0.13	0.01
S=12.267	μ=796/10=79.6	$\Sigma Z = 0$	$\Sigma Z^2 = 9$

Properties of Z-Scores

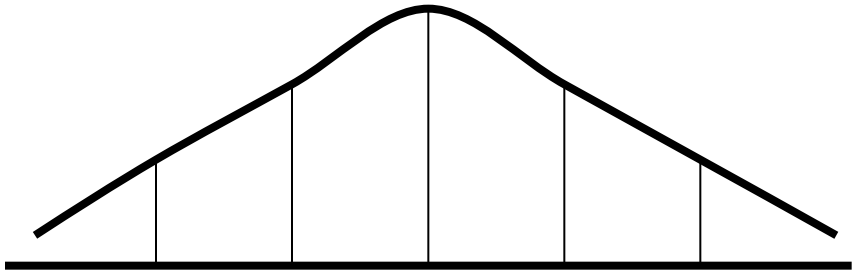
1. Sum of z-score is zero.
2. The mean of z-scores is zero.
3. The sum of squared z-scores equals N-1.
4. The standard deviation and variance of z-scores is one.

The standard normal distribution has a μ of 0, a σ of 1, and a total area equal to 1.00.

Finding Points Within Distribution

Name	Test Score	*Z-Scores*	*Percentile Rank*
Celal	98	1.49	93.19
Beyza	92	1.01	84.38
Korkut	90	0.84	79.95
Hasan	85	0.44	67.00
Medet	83	0.27	60.64
Nusret	78	-0.13	44.83
Bican	75	-0.37	35.57
Muhittin	70	-0.78	21.77
Sevda	65	-1.19	11.70
Cevdet	60	-1.59	5.59
S=12.267			

Standardized Scores and Distributions



Raw Score	60	65	70	75	80	85	90
Z-score	-3	-2	-1	0	+1	+2	+3
Percentile rank	.13	2.28	15.87	50.00	84.13	97.72	99.87
Grade	F3	D2	D1	C	B	A2	A1

McCall's T

There are variety of systems by which we can transform raw scores to give them more inuitive meaning. One system was established in 1939 by W.A. McCall. It is exactly the same as standard scores (*z*) except the mean in McCall's system is 50 rather than 0 and the standard deviation is 10 rather than 1. Thus, z-scores can be transformed into T scores by applying the lineer transformation.

$$T = 10Z + 50$$

Quartiles and Deciles

The terms quartiles and deciles refer to divisions of the percentile scale into groups. The quartile system divides percentile scale into four groups and the decile system divides the scale into ten groups.

Percentile Scale	0.00			50.00				100.00		
Quartile Scale	25			50		75			100	
Decile scale	10	20	30	40	50	60	70	80	90	100