

LING 411 Problem Session: Descriptive Statistics

Central tendency and variability

Part 1. Central tendency

Mean	Add all scores and divide by how many there are: $M = \Sigma X / n$.
Median (Mdn)	Put the scores in order and take the middle one. Two middle scores: take their average.
Mode	The score (or category) that occurs most often.
Shape	Positive skew: tail on the right, mean is the largest. Negative skew: tail on the left, mean is the smallest. The mean is always pulled toward the tail.

1. Guess the shape. Sketch each distribution.

- Word durations: mean = 410 ms, Mdn = 380 ms, mode = 360 ms.
- Accuracy on an easy test: mean = 88%, Mdn = 92%, mode = 95%.
- In a Turkish corpus most words occur only once or twice, but *ve* and *bir* occur thousands of times. Which is larger: the mean or the median frequency?

2. Followers example. Five linguistics students have 120, 150, 200, 180 and 150 Instagram followers.

- Find the mean, the median and the mode.
- A celebrity with 1,000,000 followers joins. Find the three measures again.
- Which measure still describes a “typical” student in the group?

3. Reading times (ms) of five participants: 400, 380, 1200, 420, 390.

- Find the mean and the median.
- Remove the 1200 ms participant. Find the mean and the median again. Which one changed more?

4. Headline check. “The average salary at this company is 90,000 TL, but 80% of the workers earn less than that.” Is this possible? Explain.

5. Mean, median or mode?

- The native language of each participant.
- Acceptability ratings on a 1–7 scale.
- Number of languages spoken: 1, 2, 3, or “4 or more”.
- Vowel durations with a symmetrical distribution.

6. Group A has 3 participants with a mean of 600 ms. Group B has 1 participant with 700 ms. What is the mean of all 4 participants? (Hint: it is not 650.)

7. The missing score. Five participants have a mean naming time of $M = 500$ ms. Four of the times are 480, 510, 470 and 520 ms.

- What is the fifth time?
- The participant with 470 ms is removed. Does the mean go up or down? Find the new mean.

Part 2. Variability

Range	Largest score – smallest score. It uses only two scores.
SD, step by step	1. Find the mean. 2. Subtract the mean from each score ($X - M$); these always add up to 0. 3. Square each one. 4. Add them up: this is SS . 5. Divide: $s^2 = SS/(n - 1)$ for a sample ($\sigma^2 = SS/N$ for a population). 6. Take the square root.
Constants	Add a number to every score: the mean changes, the SD does <i>not</i> . Multiply every score: both change.
z-score	$z = (X - \mu)/\sigma$: how many SDs a score is above (+) or below (–) the mean. Back to X : $X = \mu + z\sigma$.

8. Guess first, then check. Rank the three sets from smallest to largest SD without calculating. Then check B and C.

A: 5, 5, 5, 5, 5 B: 3, 4, 5, 6, 7 C: 1, 1, 5, 9, 9 (all have $M = 5$)

9. Use the reading times from Question 3: 380, 390, 400, 420, 1200.

a. Find the range with and without the 1200 ms participant.

b. The interquartile range (IQR) uses only the middle 50% of the scores. Is it affected much by the 1200 ms participant?

10. Voice onset time (VOT, ms) of /t/ for one speaker (a sample): 60, 65, 70, 75, 80. Find s .

X	$X - M$	$(X - M)^2$
60		
65		
70		
75		
80		
$M =$	sum =	$SS =$

11. Speech rate (syllables per second) of four speakers (a sample): 4, 6, 5, 9.

a. Find SS , s^2 and s .

b. A fifth speaker joins with exactly 6 syllables per second. Without a full calculation: does s get larger or smaller? Check it.

12. Reaction times have $M = 600$ ms and $s = 80$ ms. What are the new M and s if:

a. we subtract 20 ms from every score?

b. we change from ms to seconds (divide every score by 1000)?

13. Who did better? Elif scores 34 on a vocabulary test ($\mu = 30$, $\sigma = 4$) and 72 on a grammar test ($\mu = 60$, $\sigma = 16$).

a. Find her z -score on each test. On which test did she do better compared to the others?

b. Her friend Can has $z = -0.50$ on the grammar test. What was his score?

14. Myth or fact?

a. The mean is always one of the scores.

b. A standard deviation can be negative.

c. The mode can be a word, not a number.

d. The deviations from the mean always add up to 0.