Empirical Research Methods 1

Descriptive statistics and diagrams

Miguel Rejón,

Room 1.14

m.rejon@edutech.uni-saarland.de

Before we start 1/2: Info and tools

- Strawpoll.me
- MS Teams for comments, questions, etc
- MS Forms!
- Reminder: SPSS activities cool, try to give advice to students as early as possible because of the vpn thing and that SPSS takes some time to load

Agenda

- Mean and measures of dispersion
  - Symbols
  - Sensitivity
  - …with normal distribution
  - Variance and SD
  - Mode
  - Reporting Mean and SD
- Q&A
- Mid-semester feedback survey

Mean and measures of dispersion: Symbols

μ = Population mean

$\bar{x}$, x_bar, x bar = Sample mean

σ = Population standard deviation

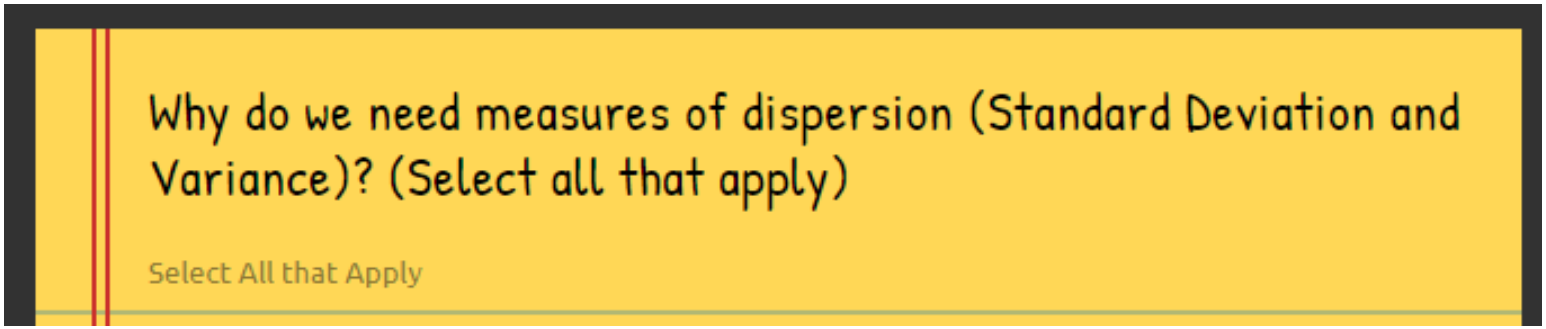
s = Sample standard deviation

N = Population size

n = Sample size

https://www.statisticssolutions.com/common-statistical-formulas/; https://www1.cgmh.org.tw/intr/intr5/c6700/OBGYN/F/Statistics/response.cfm-ID=779.htm

Mean and measures of dispersion: quick quiz



Multiple choice, alphabetically ordered answers

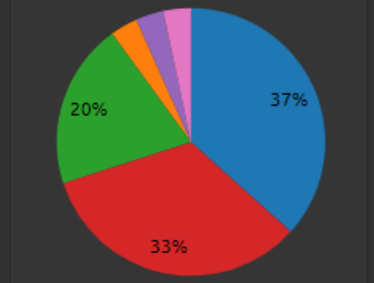
http://www.strawpoll.me/42226189

Why do we need measures of dispersion (Standard Deviation and Variance)? (Select all that apply)

Because with these measures we can better know how our data is distributed	11 Votes
	37%
Because with these measures we can better know how to convert our inferential data into descriptive	1 Votes
	3%
Measures of dispersion allow us to reject or accept our null Hypothesis	6 Votes
	20%
Measures of dispersion allow us to report the descriptive data from our study	10 Votes
	33%
Measures of dispersion are not that important, as long as we do inferential statistics	1 Votes
	3%
The measures of dispersion are only important when talking about non-normal distributions	0 Votes
	0%
The measures of dispersion are only important when talking about normal distributions	1 Votes
	3%
The measures of dispersion are only important when we are going to make a graph of our data	0 Votes
	0%

30 Votes **0** Comments

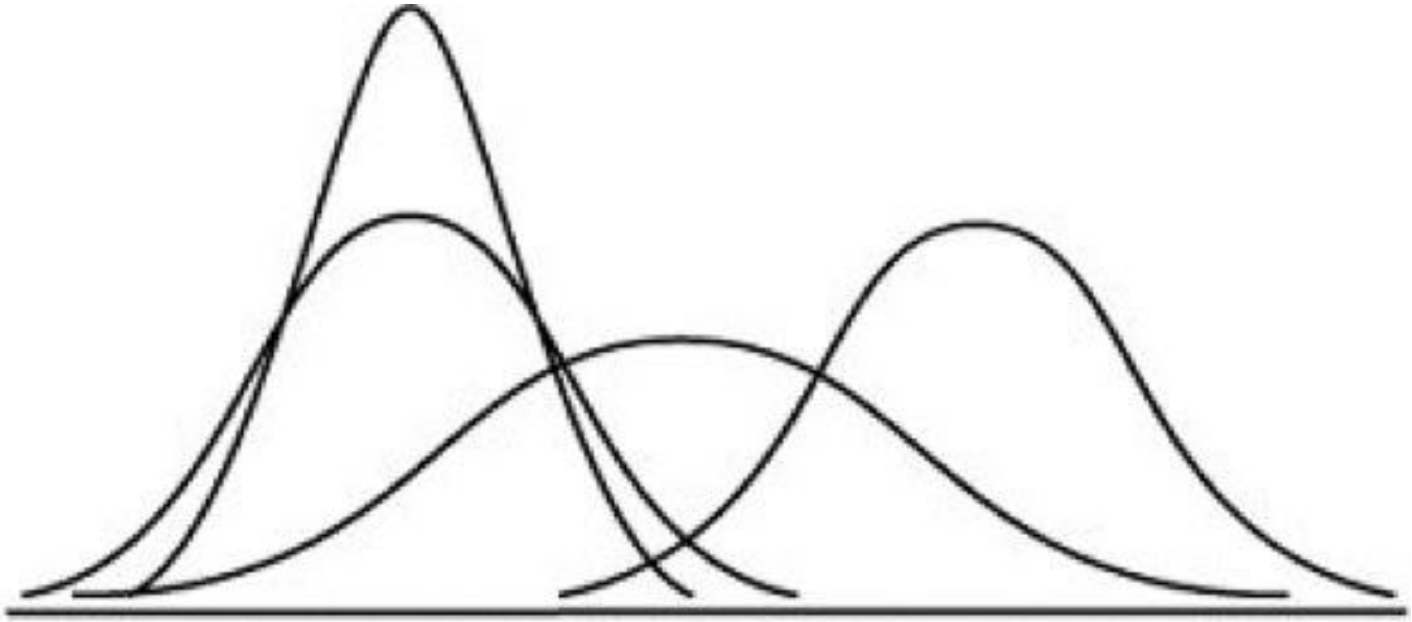
Vote Share



Mean and measures of dispersion: quick quiz answers

- Because with these measures we can better know how our data is distributed
- Because with these measures we can better know how to convert our inferential data into descriptive
- Measures of dispersion allow us to reject or accept our null Hypothesis
- Measures of dispersion allow us to report the descriptive data from our study
- Measures of dispersion are not that important, as long as we do inferential statistics
- The measures of dispersion are only important when talking about non-normal distributions
- The measures of dispersion are only important when talking about normal distributions
- The measures of dispersion are only important when we are going to make a graph of our data

Mean and measures of dispersion

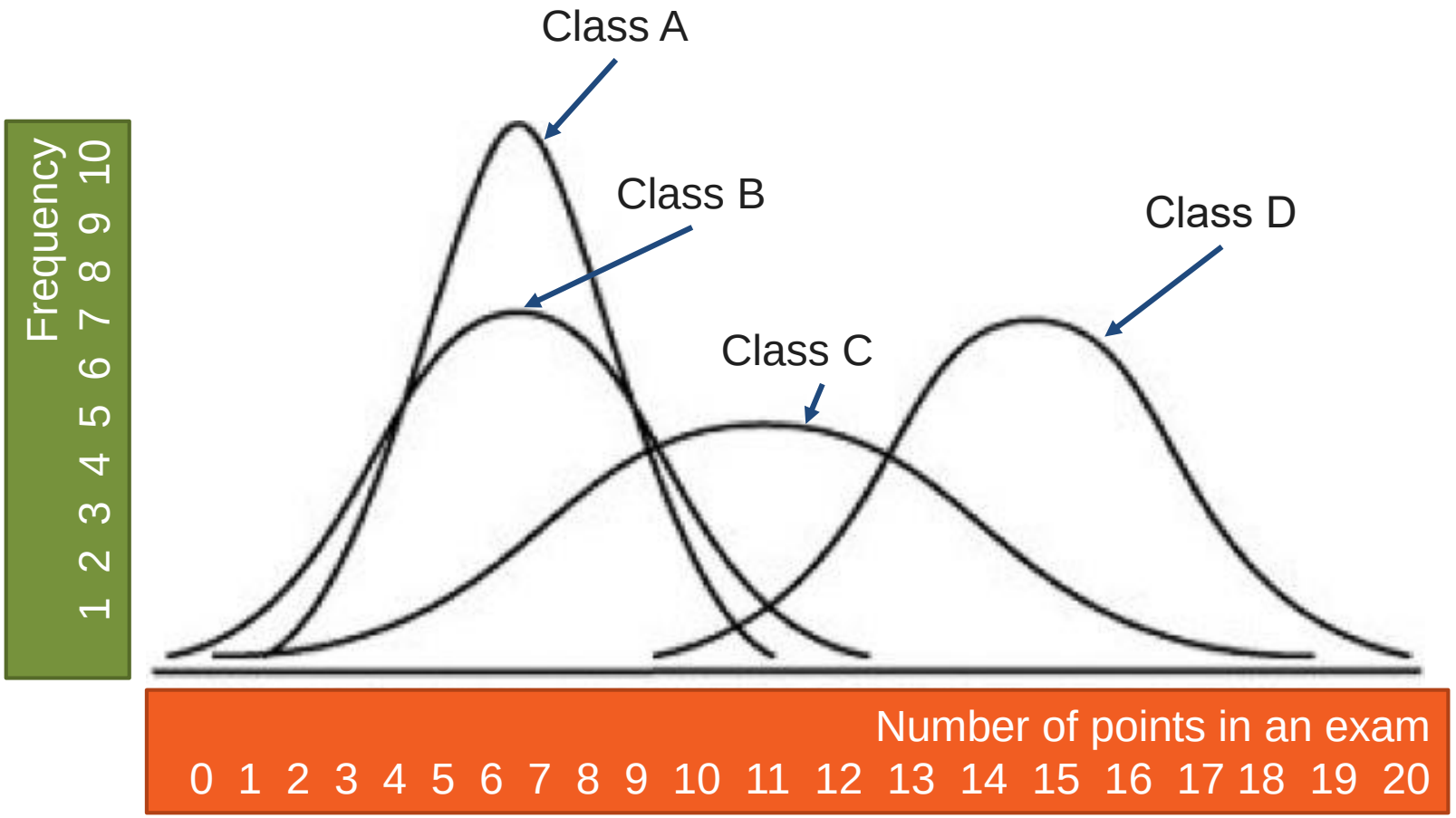


Normalverteilungen mit unterschiedlichen μ und σ

(Normal distributions with different μ und σ)

http://www.spektrum.de/lexika/images/psycho/f2f209_w.jpg

Mean and measures of dispersion



http://www.spektrum.de/lexika/images/psycho/f2f209_w.jpg

Measures of central tendency and their sensitivity to extreme observations

Data: 5, 4, 5, 2

Mean

$\frac{5+4+5+2}{4} = \frac{16}{4} = 4$

Mode

5, 4, 5, 2

→ most frequent: 5

Median

Order the values: 5, 4, 5, 2 → 2 4 5 5

→ median = 4.5

Measures of central tendency and their sensitivity to extreme observations

Data: 5, 4, 5, 2, 20

→ Mean = $\frac{5+4+5+2+20}{5}$ = 7.4

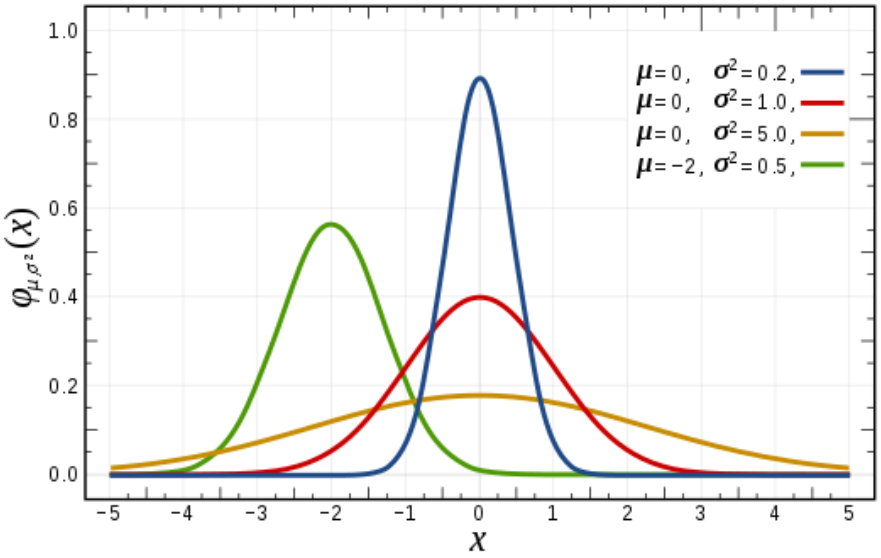
→ Mode: 5, 4, 5, 2, 20 → 5

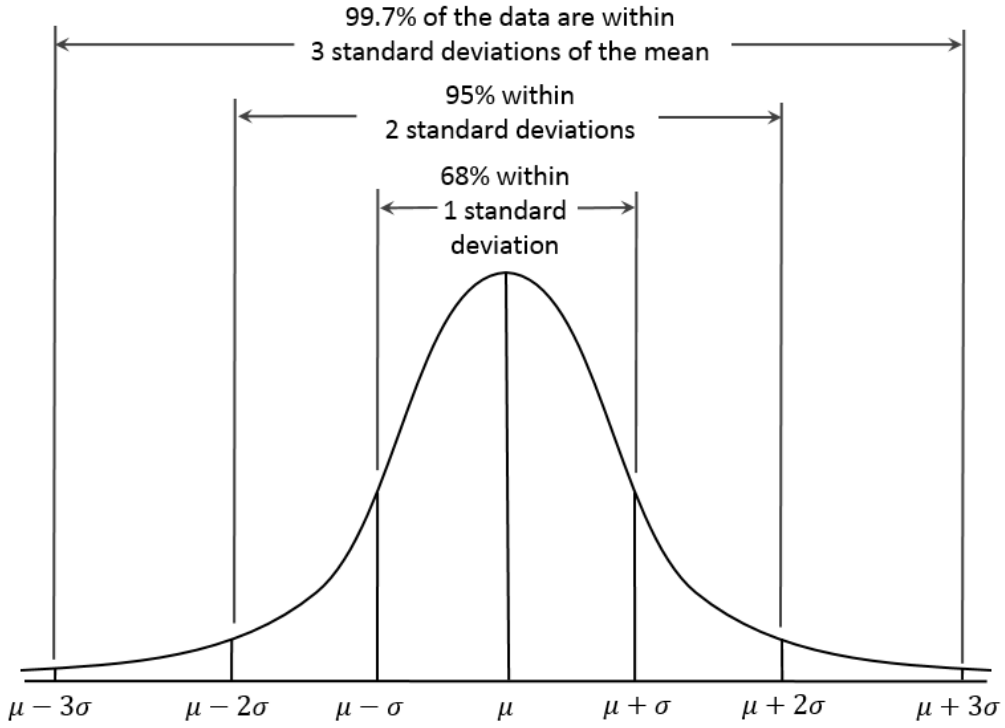
→ Median: 2 4 5 5 20 → 5

→ Median and mode are less sensitive to extreme values (compared to the mean)

Normal distribution

- A probability distribution that underlies many variables in the population (e.g. intelligence)
- Symmetric, unimodal, around 68% of the values are within the first SD, around 95% within 2 SD, and around 99% within 3 SD
- Normality is a basic assumption for many statistical tests





(**Normal distribution.** In *Wikipedia*. Retrieved December 09, 2019, from https://en.wikipedia.org/wiki/Normal_distribution)

Variance and standard deviation (SD)

- Variance (s^2) = sum of the squared deviations of the mean, divided by N

- Standard deviation (SD or s) = square root of the variance

Variance: Example

$$s^2 = \frac{\sum (X - \bar{X})^2}{N - 1}$$

Data: 5, 4, 5, 2

→ Mean = $\frac{5+4+5+2}{4} = 4$

$$s^2 = \frac{(5-4)^2+(4-4)^2+(5-4)^2+(2-4)^2}{4-1} = \frac{1+0+1+4}{3} = \frac{6}{3} = 2$$

→ Variance $s^2 = 2$

→ Standard deviation $s = \sqrt{2} = 1{,}4$

Variance and standard deviation (SD)

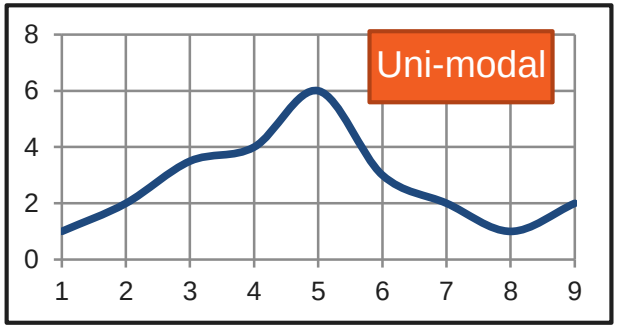
- Both variance (s^2) and Standard deviation (SD) indicate the variation (or spread) of our data in relation to the mean. So…
- Q: Why do we need variance?
  - A: so we can get a non-zero sum and positive numbers
- Q: Why do we need standard deviation?
  - A: so we can get a value of spread in the same unit as our data and mean. Examples: “age” instead of “age squared”, “euros” instead of “euros squared”

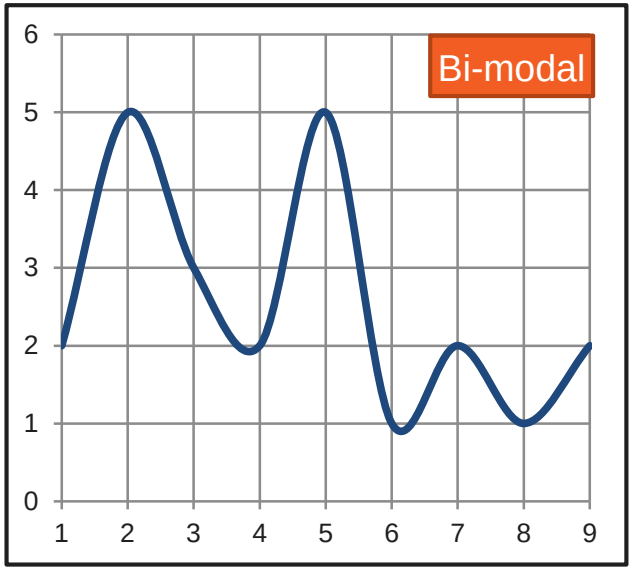
Mode

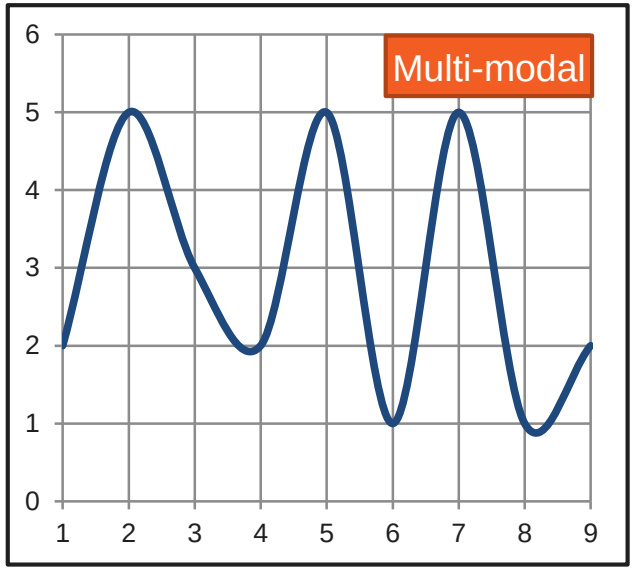
What if there are several values with the same high frequency? Which one is the mode?

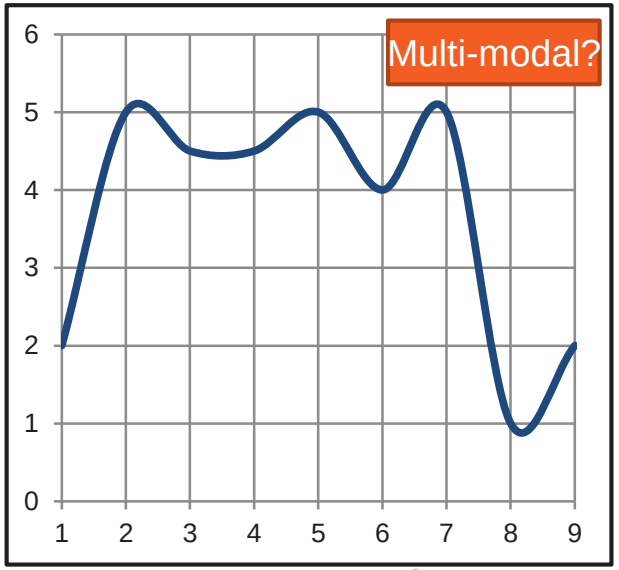
<u>Two approaches:</u>

- “bi-modal“, “multi-modal“ vs.
- “cannot be determined “









Example: Reporting Mean + SD

they posted. The control group ($M = 36.6$, $SD = 9.91$) and individual preparation-no script ($M = 30.86$, $SD = 9.26$) posted more, followed by argumentation scripts ($M = 19.19$, $SD = 6.84$) and the combination condition ($M = 15.26$, $SD = 6.11$). To take these differences

Tsovaltzi, D., Judele, R., Puhl, T., & Weinberger, A. (2017). Leveraging social networking sites for knowledge co-construction: Positive effects of argumentation structure, but premature knowledge consolidation after individual preparation. *Learning and Instruction*, *52*, 161-179.

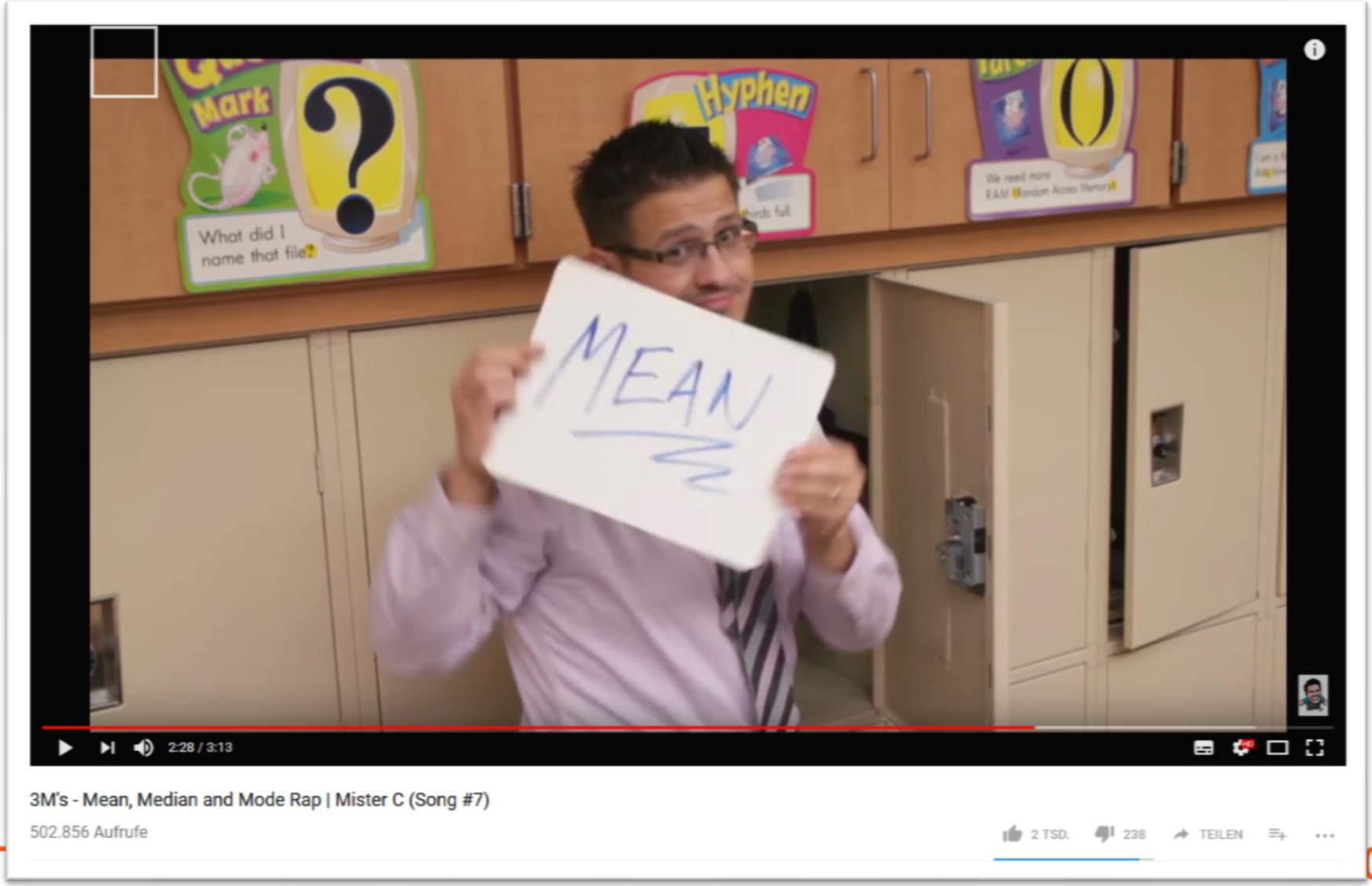
Table 1. Formal argumentative dimension by experimental group: Mean percentages and standard deviations of grounds and counterarguments.

	Grounds		*Counterarguments*	
Experimental group	*M*	*SD*	*M*	*SD*
Control group	12.08 %	11.48	2.46 %	3.67
Script for the construction of single arguments	33.80 %	11.19	5.36 %	8.07
Script for the construction of argumentation sequences	16.36 %	17.78	5.99 %	3.95
Combined condition	30.64 %	6.10	13.00 %	6.59

Weinberger, A., Stegmann, K., & Fischer, F. (2005). Computer-Supported Collaborative Learning in Higher Education : Scripts for Argumentative Knowledge Construction in Distributed Groups. In *The Next 10 Years! Proceedings of the 2005 Conference on Computer Support for Collaborative Learning, CSCL '05* (pp. 717–726).

Mean, Median & Mode

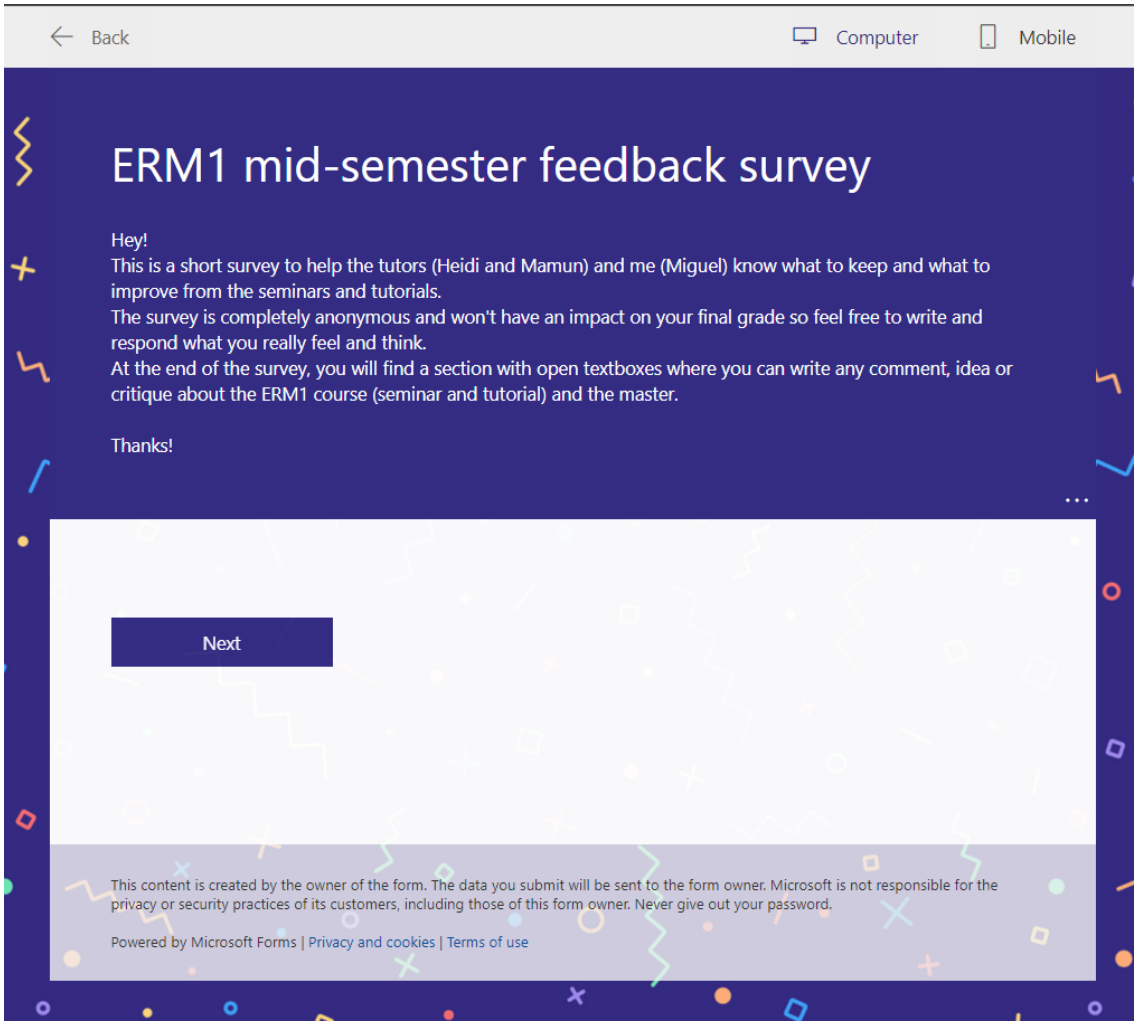
https://www.youtube.com/watch?v=A7MxGyEaN64



Mid-semester feedback survey

- Completely anonymous; your time to be heard (or read)!
- Reflection on:
  - This and the all the previous sessions
  - Workload: presentations and weekly sheets
  - Tutorials
  - How would you want ERM1 (seminar and tutorial) to improve
  - What would you want for ERM1 (seminar and tutorial) to keep

Mid-semester feedback survey



Q&A:
TODAY‘S TOPIC; ERM1 IN GENERAL

WEBCAMS ON FOR THE GOODBYE