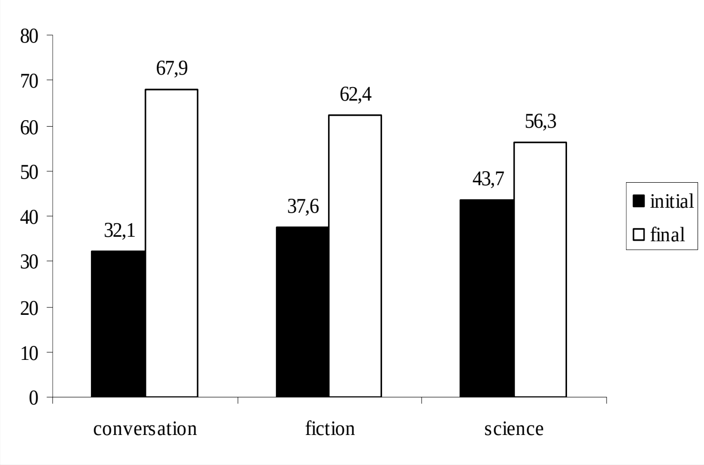
Descriptive statistics

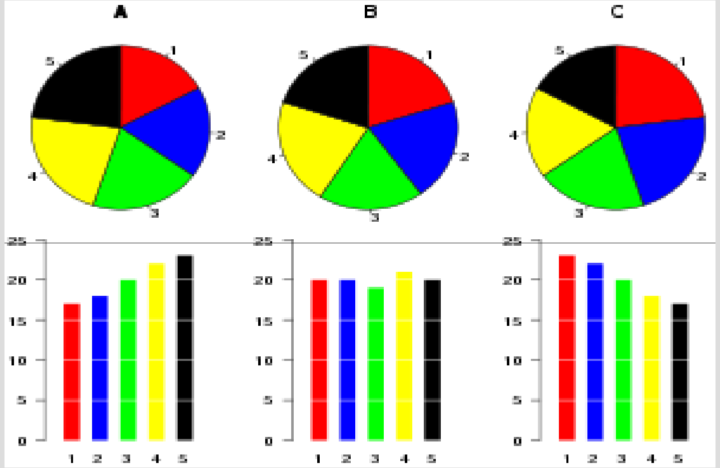
Holger Diessel

holger.diessel@uni-jena.de

Graphs



Bar chart



Pie chart

Graphs

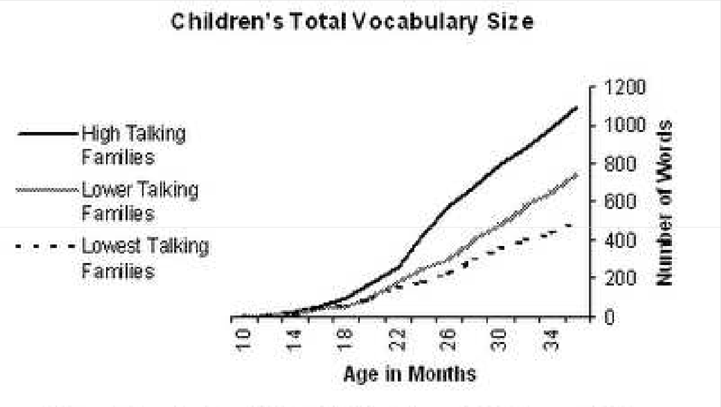
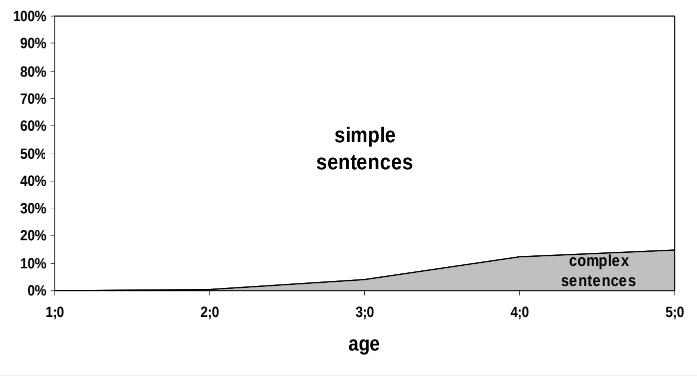

Children's Total Vocabulary Size

Figure 2. Based on data in Hart and Risley, *Meaningful Differences*, 1995

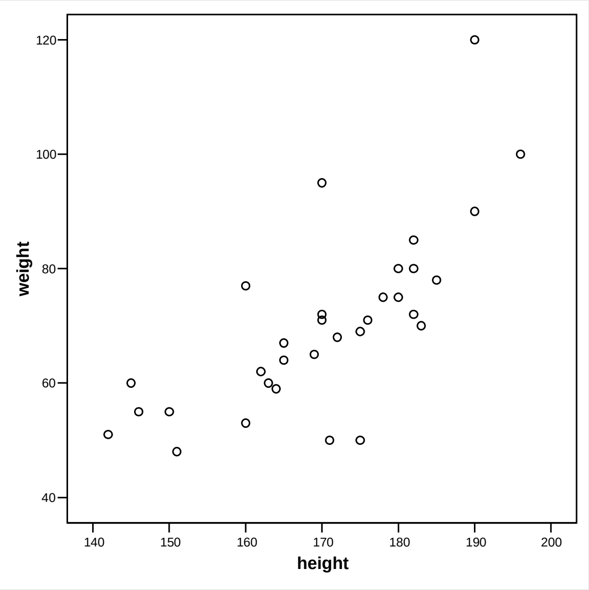
Figure 1. Proportion of complex sentences



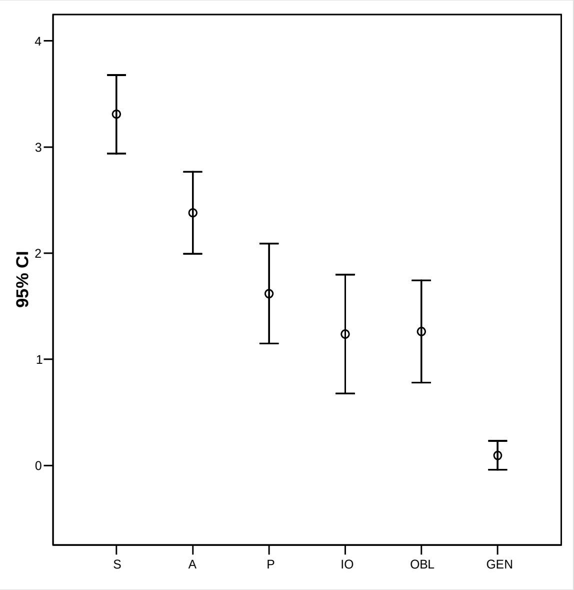
Flow chart 1

Flow chart 2

Graphs

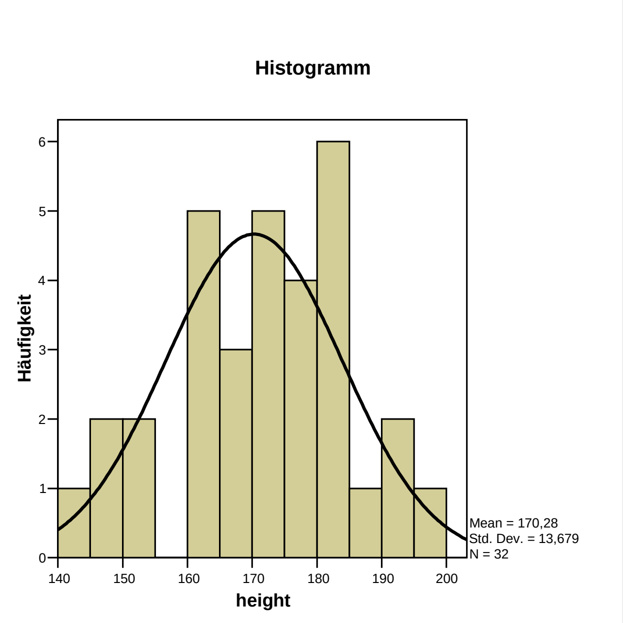


Scatter plot

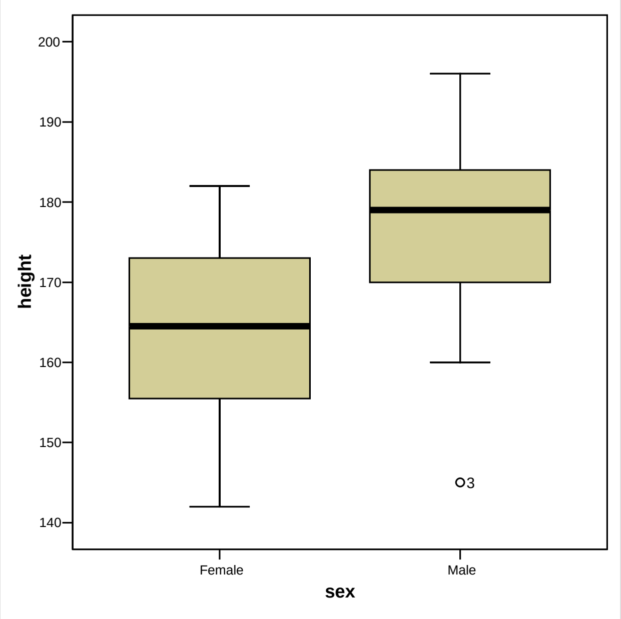


Error bars

Graphs



Histogram



Box plot

Central tendency

Data:	2,3,3,3,4,6,6,9,12,13,13
Mean	(2+3+3+3+4+6+6+9+12+13+13)/11 = 6.72
Median	Middle score: 6
Mode	Most frequent score: 3

The effect of outliers:

Mean	2+3+3+3+4+6+9+12+13+13+31/11 = 9
Median	2+3+3+3+4+6+9+12+13+13+31/11 = 6

Central tendency

A child language researcher investigates the emergence of verbal particles in the speech of two children. Verbal particles occur in five different contexts in his data.

(1) He picked me up. [Transitive verb particle construction]
(2) He walked away. [Intransitive verb particle construction]
(3) I am back. [Predicative verb particle construction]
(4) Shoes on. [Fragmented verb particle constructions]
(5) Put it on the table. [Prepositional construction]

	Peter	Eve	Total
Transitive	291	281	572
Intransitive	232	256	488
Predicative	17	25	42
Fragmented	130	70	200
Prepositional	519	754	1273
	1189	1386	2575

Central tendency

	Peter	Eve	Total	Percentage	Mean%
Transitive	291 (24.5)	281 (20.3)	572	22.2	22.4
Intransitive	232 (19.5)	256 (19.5)	488	19.0	19.0
Predicative	17 (1.4)	25 (1.4)	42	1.6	1.6
Fragmented	130 (10.9)	70 (10.9)	200	7.8	8.0
Prepositional	519 (43.7)	754 (43.7)	1273	49.4	49.1
	1189	1386	2575	100.0	100.0

	Jack	Sue	Total
Transitive	491 (44.2)	81 (12.7)	572
Intransitive	432 (38.9)	156 (24.5)	588
Predicative	37 (3.3)	29 (4.6)	66
Fragmented	30 (2.7)	50 (7.8)	80
Prepositional	121 (10.9)	321 (50.4)	442
	1111	637	1748

Central tendency

	Peter	Eve	Total	Percentage	Mean%
Transitive	291 (24.5)	281 (20.3)	572	22.2	22.4
Intransitive	232 (19.5)	256 (19.5)	488	19.0	19.0
Predicative	17 (1.4)	25 (1.4)	42	1.6	1.6
Fragmented	130 (10.9)	70 (10.9)	200	7.8	8.0
Prepositional	519 (43.7)	754 (43.7)	1273	49.4	49.1
	1189	1386	2575	100.0	100.0

	Jack	Sue	Total	Percentage	Mean
Transitive	491 (44.2)	81 (12.7)	572	32.7	28.5
Intransitive	432 (38.9)	156 (24.5)	588	33.6	31.7
Predicative	37 (3.3)	29 (4.6)	66	3.8	3.9
Fragmented	30 (2.7)	50 (7.8)	80	4.5	5.2
Prepositional	121 (10.9)	321 (50.4)	442	25.3	30.6
	1111	637	1748	100.0	

Variance

Measurements for the spread of data:

- Range
- Variance
- Standard variation

Range: 2,3,3,3,4,6,6,9,12,13,13 = 2 - 13

Standard variation

$$\sigma = \sqrt{\frac{\sum (x_i - m)^2}{n-1}}$$

Standard variation

S	words
1	3
2	7
3	4
4	9
5	12
6	9
7	11
8	4
	Σ 59 / 8 = 7.4 (mean)

Standard variation

S	words	(=$X_1 – X_{mean}$)
1	3	3 – 7.4
2	7	7 – 7.4
3	4	4 – 7.4
4	9	9 – 7.4
5	12	12 – 7.4
6	9	9 – 7.4
7	11	11 – 7.4
8	4	4 – 7.4
	Σ 59 / 8 = 7.4 (mean)	

Standard variation

S	words	$(=X_1 – X_{mean})$	d_1
1	3	3 – 7.4	–4.4
2	7	7 – 7.4	–0.4
3	4	4 – 7.4	–3.4
4	9	9 – 7.4	1.6
5	12	12 – 7.4	4.6
6	9	9 – 7.4	1.6
7	11	11 – 7.4	3.6
8	4	4 – 7.4	–3.4
	Σ 59 / 8 = 7.4 (mean)		Σ 0 / 8 = 0

Standard variation

S	words	$(=X_1 – X_{mean})$	d_1	d_1^2 (residuals)
1	3	3 – 7.4	–4.4	19.36
2	7	7 – 7.4	–0.4	0.16
3	4	4 – 7.4	–3.4	11.56
4	9	9 – 7.4	1.6	2.56
5	12	12 – 7.4	4.6	21.16
6	9	9 – 7.4	1.6	2.56
7	11	11 – 7.4	3.6	12.96
8	4	4 – 7.4	–3.4	11.56
	Σ 59 / 8 = 7.4 (mean)		Σ 0 / 8 = 0	Σ 81.87

Standard variation

S	words	(= $X_1 – X_{mean}$)	d_1	d_1^2 (residuals)
1	3	3 – 7.4	–4.4	19.36
2	7	7 – 7.4	–0.4	0.16
3	4	4 – 7.4	–3.4	11.56
4	9	9 – 7.4	1.6	2.56
5	12	12 – 7.4	4.6	21.16
6	9	9 – 7.4	1.6	2.56
7	11	11 – 7.4	3.6	12.96
8	4	4 – 7.4	–3.4	11.56
	Σ 59 / 8 = 7.4 (mean)		Σ 0 / 8 = 0	Σ 81.87

Standard variation

Variance: 81.87 / (8-1) = 11.7

The variance is a meaningless measure.

Standard deviation: $\sqrt{11.7} = 3.42$

70% of the data fall within one SD from the mean:

70% of all sentences in the sample include between 3.98 and 10.82 words.

z-scores

Scores from two different language proficiency test:

	Test 1 – candidate A			Test 2 – candidate B		
Scenario	Score	Mean	SD	Score	Mean	SD
1	41	49		53	49	

z-scores

Scores from two different language proficiency test:

	Test 1 – candidate A			Test 2 – candidate B		
Scenario	Score	Mean	SD	Score	Mean	SD
1	41	49		53	49	
2	41	49		53	58	

z-scores

Scores from two different language proficiency test:

	Test 1 – candidate A			Test 2 – candidate B		
Scenario	Score	Mean	SD	Score	Mean	SD
1	41	49		53	49	
2	41	49		53	58	
3	41	49	8	53	58	5

z-scores

$$z = \frac{x - \mu}{\sigma}$$

z-scores

S	Number of words
1	73
2	42
3	36
4	51
5	63
	Σ 265 / 5 = 53 (mean) SD = 15.12

z-scores

S	Number of words	$(=X_1 – X_{mean})$	d_1
1	73	73 – 53	20
2	42	42 – 53	–11
3	36	36 – 53	–17
4	51	51 – 53	–2
5	63	63 – 53	10
	Σ 265 / 5 = 53 (mean) SD = 15.12		

z-scores

S	Number of words	(=X_1 – X_{mean})	d_1	z = (d_1 / SD)
1	73	73 – 53	20	1.32
2	42	42 – 53	–11	–0.73
3	36	36 – 53	–17	–1.12
4	51	51 – 53	–2	–0.13
5	63	63 – 53	10	0.66
	Σ 265 / 5 = 53 (mean) SD = 15.12			

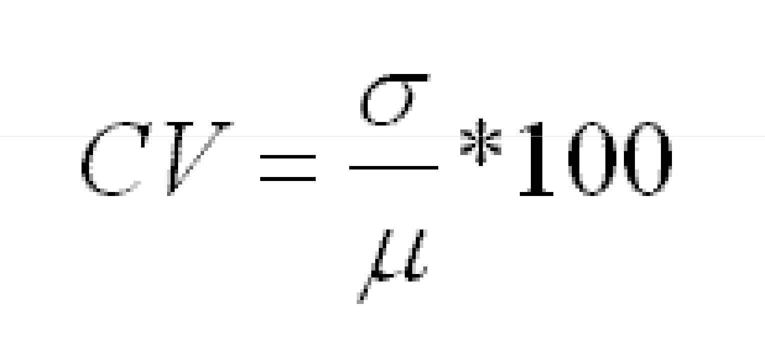
Example

Zwei Kandidaten haben an zwei unterschiedlichen Sprachtests teilgenommen. Kandidat A hat 121 Punkte erzielt, Kandidat B hat 177 Punkte erzielt. Im ersten Test (an dem Kandidat A teilgenommen hat) lag der Mittelwert bei 92 und die Standardabweichung bei 14; im zweiten Test (an dem Kandidat B teilgenommen hat) lag der Mittelwert bei 143 und die Standardabweichung bei 21. Welcher der beiden Kandidaten hat besser abschlossen (im Vergleich zu allen übrigen Kandidaten)?

$$Z_A = 121 – 92 / 14 = 2.07$$

$$Z_B = 177 – 143 / 21 = 1.62$$

Coefficient of variance



$$CV = \frac{\sigma}{\mu} * 100$$

Coefficient of variance

Over a 4 months period a mean number of 90 parking tickets was issued. The standard deviation was 5. The tickets yielded an average of $5400 per day and the SD was $775. Where do you have more variability, in the number of parking tickets that were issued each day or in the amount of money that was generate each day?

Parking tickets:	Mean = 90, SD = 5
Fines:	Mean = 5400, SD = 775

Parking tickets:	$CV1 = 5/90 \times 100 = 6\%$
Fines:	$CV2 = 775/5400 \times 100 = 14\%$