Empirical Methods of Linguistics Research

Script WS 2007/8

Holger Diessel

Criteria for empirical research

a. Objectivity
b. Reliability
c. Internally validity
d. Externally validity
e. Occam's Razor

Stages of an empirical investigation

1. Exploration phase
a. Literature research
b. Impressionistic analysis of observational data
c. Informal pretest

2. Theoretical phase
a. State your hypotheses (informally)
b. Are your hypotheses falsifiable?
c. Are your hypotheses compatible with your theoretical framework?

3. Planning phase
a. Select a method
b. Define and operationalize the variables
c. Select the appropriate statistical measure

4. Data collection
Data collection has to follow a predetermined procedure. Don't change the procedure in the middle of your study.

5. Analysis and decision
a. Describe the frequency data
b. Submit the frequency data to statistical analysis

Sample study: The acquisition of the English verb-particle construction (Diessel & Tomasello 2005)

The English verb-particle construction consists of a transitive verb and a particle that can precede or follow the direct object (1-2). Particles must be distinguished from prepositions. In contrast to particles, prepositions generally precede the noun (3-4).

(1) He looked **up** the number.
(2) He looked the number **up**.
(3) He walked **up** the hill.
(4) *He walked the hill **up**.

Previous studies have shown that in adult language the positioning of the particle varies with a variety of factors (variables): (1) the NP type of the direct object, (2) the length of the direct object, (3) the syntactic/semantic complexity of the direct object, (4) the meaning of the particle, (5) the information status of object and particle, (6) the accentuation of the direct object, (7) the occurrence of a directional prepositional phrase (cf. Bolinger 1971; Fraser 1974, 1976; Bock 1977; Dixon 1982; Chen 1986; Hawkins 1994; Peters 1999; Wasow 2002; Dehé et al. 2002; Gries 1999, 2003).

(5) He looked it **up**.	NP type
(6) *He looked **up** it.	
(7) He looked the number **up**.	Length
(8) He looked the number of his neighbour in the yellow pages **up**.	
(9) He put the ball with the blue stripes **down**.	Complexity
(10) He put the ball that Sue had given him **down**.	
(11) He pushed the chair **away**.	Meaning
(12) He ate **up** his lunch.	
(13) He turned **on** the TV.	
(14) What did she do with the ball? She picked the ball **up**.	Information
(15) What did she pick up? She picked **up** the ball.	
(16) I turn the light **on**.	
(17) I turn **on** a light.	
(18) Pick **up** HIM (not her).	Stress
(19) Peter put the cup **back**.	PP
(20) Peter put **back** the cup.	
(21) Peter put the cup **back** on the table.	
(22) Peter put **back** the cup on the table.	

Research questions

1. Does the positioning of the particle in child language vary with the same factors as in adult language?
2. Do children use the two particle positions productively?

Data collection

Table 1. Overview of the data

	Age	Files	First VPC
Peter	1;9-3;1	20	1;9
Eve	1;6-2;3	20	1;7
	>2;3	40	

Particles: on, off, back, away, in, out, down, over, around, up

Utterances including *up*:

(1) turn up hill / up hill .
(2) then wake up
(3) milk all wiped up .
(4) up .
(5) Eve stand up Mommy stool .
(6) I pick up .
(7) up wall .
(8) bobbing up an(d) down .
(9) I covered it up .
(10) well this is up in the house .

Construction types:

(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]
(4a)	Shoes on.	[Fragmented verb particle constructions]
(4b)	Down!	[Fragmented verb particle constructions]
(5)	Put it on the table.	[Prepositional construction]

Table 2. Frequency of construction types

	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

Table 3. Frequency of particle and prepositional constructions

	Peter	Eve	Total	Percentage	Mean%
Transitive	291	281	572	22.2	
Intransitive	232	256	488	19.0	
Predicative	17	25	42	1.6	
Fragmented	130	70	200	7.8	
Prepositional	519	754	1273	49.4	
	1189	1386	2575	100.0	

Table 4. Hypothetical frequencies

	Jack	Sue	Total	Percentage	Mean
Transitive	491	81	572	32.7	
Intransitive	432	156	588	33.6	
Predicative	37	29	66	3.8	
Fragmented	30	50	80	4.5	
Prepositional	121	32	442	25.3	
	1111	637	1748	100.0	

Table 5. Particles in the VPC

	Peter Frequency	First	Eve Frequency	First	Total Frequency	Mean%
on	59	2;0	49	1;7	108	18.9
off	73	1;11	33	1;9	106	18.4
back	61	1;11	39	1;9	100	17.5
up	21	1;11	44	1;7	65	11.5
in	9	1;11	46	1;9	55	9.8
away	19	1;11	35	1;9	54	9.5
out	24	1;11	19	1;8	43	7.6
down	20	1;10	13	1;10	33	5.8
over	5	1;9	2	2;3	7	1.2
around	0	–	1	2;1	1	0.2
	291		281		572	100.0

Table 6. The ten most frequent verbs in the VPC

	Peter	Eve	Total in VPCs	Total in the entire corpus
put	120	144	264	597
take	72	30	102	178
turn	48	1	49	111
blow	0	20	20	33
get	4	14	18	447
have	1	14	15	429
push	6	8	14	41
pick	6	6	12	26
move	8	0	8	42
pull	1	6	7	25
	291	281	572	1929

Omitted direct object:

	VPC	**Age**	**CHILD**
1	turn down	1;10	Peter
2	pick up … thank you .	1;11	Peter
3	mm … put down …	1;11	Peter
4	take off this .	1;11	Peter
5	one ... rinse off ...	1;11	Peter
6	rinse off ...	1;11	Peter
7	put the screw in .	1;11	Peter
8	Put milk in .	1;11	Peter
9	Plug in .	1;11	Peter
10	Plug in.	1;11	Peter
11	take out ...	1;11	Peter
12	I put them back .	1;11	Peter
13	yyy screwdriver ... put back .	1;11	Peter
14	xxx put yyy back .	1;11	Peter
15	put back	1;11	Peter
16	I put back .	1;11	Peter
17	put back .	1;11	Peter
18	part ... put back .	1;11	Peter
19	put back .	1;11	Peter

Table 7. The occurrence of an overt direct object in the VPC

	Peter	Eve	Total	Mean%
Overt object	210	240	450	78.8
No overt object	81	41	122	21.1
Total	291	281	572	100.0

Coding

	Spalte1	Construction	Lenght	Complexity	Definiteness	Meaning
1	take off this .					
2	put the screw in .					
3	Put milk in .					
4	I put them back .					
5	put em back ... ok .					
6	turn it over .					
7	turn it over .					
8	turn on a light off .					
9	... turn on a light off .					
10	pick up my cup .					
11	take off wheels .					
12	telephone ... put back ... telephone .					
13	pick am up .					
14	close it up .					
15	close it … up .					
16	I turn the light on .					
17	turn it on …					
18	turn it on … xxx .					
19	uhhuh … we have to put our coats on.					
20	take this … put it on .					
21	my put barrette on .					
22	a turn on a light !					
23	turn it on …					
24	there put it on .					
25	right there … turn the light on .					
26	put it on … .					
27	a put it on …					
28	my put it on .					
29	..... got engine off .					
30	don't take a wheels off ...					
31	turn it off ... there .					
32	I turn that off .					
33	push it off ...					
34	get your hand off .					
35	take it out .					
36	take a out .					
37	turn a light out ... xxx .					
38	let’s take a piece out .					
39	put it back .					
40	put a back .					
41	I put the pen back in my pocketbook .					
42	put more back .					
43	my out put ... xxx ... put more back					
44	in a box ... gonna put a back too .					
45	put it back .					
46	let's put it back right there .					
47	... put xxx away					
48	put these away .					
49	put toys away ... go home ...					
50	turn it over ?					
51	a turn it over ... uhoh .					
52	all finished put it on a tape recorder on.					
53	I do it … turn on the light on .					

Coding scheme

Column	Category	Values	Example
A	CHILD	1 = Peter 2 = Eve	
B	CONSTRUCTIO	1 = V NP P 2 = V P NP	Look the number up Look up the number
C	LENGTH	1 = 1 word 2 = 2 words 3 = 3+ words	Pick **it** up Pick **the ball** up Pick **the blue ball** up
D	COMPLEXITY	0 = simple 1 = intermediate 2 = complex	Pick it up [PRO, (ART)-N] Pick my ball up [GEN-N, N-PP, N and N Pick up the one I want [N-CLAUSE]
E	DEFINITENESS	0 = no determiner 1 = def. determiner 2 = indef. determin	Pick book up [PRO, bare N] Pick the ball up [the, my, this, etc.] Pick a ball up [a]
F	NP-Type	1 = unstressed PRO 2 = stressed PRO 3 = lexical	Pick it up Pick this up Pick the ball up
G	MEANING	0 = spatial 1 = non-spatial	Pick me up Eat it up
H	PP	0 = no PP 1 = PP	Put it back **on the floor** Put it back
I	PARTICLE	1 = up 2 = down 3 = on 4 = off 5 = in 6 = out 7 = back 8 = away 9 = over 0 = around	Pick it up Put it down Put it on Put it off Put in it Take it out Put it back Put it away Turn it over Turn it around

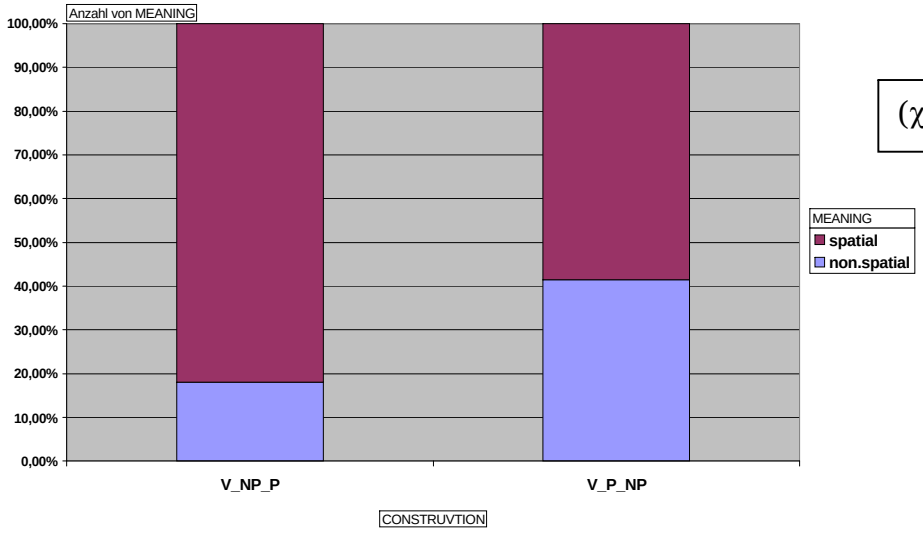
Results

Table 1. Frequency of the different particle positions in the VPC

	Peter	Eve	Total	Mean%
V NP P	195	226	29	93.5
V P NP	15	14	421	6.5
	210	240	572	100.0

Table 2. Distribution of particle position relative to *meaning*

	VP NP P Frequency	VP P NP Frequency	Total
Spatial	345	17	362
Non-spatial	76	12	88
	421	29	450



$(\chi^2(1) = 9.385; p < .004)$

Table 3. Distribution of particle position relative to *complexity*

	VP NP P Frequency	VP P NP Frequency	Total
simple	406	26	432 (96.0%)
intermediate	15	1	16 (3.6%)
complex	0	2	2 (0.4%)
	421	29	450

Simple vs. intermediate vs. complex: $(\chi^2(2) = 29.16; p < .004)$
Simple vs (intermediate + complex): $(\chi^2(1) = 3.25; p > .102)$

Four of the six factors vary with the position of the particle:

1. the length of the direct object
2. the complexity of the direct object
3. the NP type of the direct object, and
4. the meaning of the particle

The two other variables, i.e. the definiteness of the direct object and the occurrence of a directional PP, are not significant.

Multifactorial analysis: Logistic regression

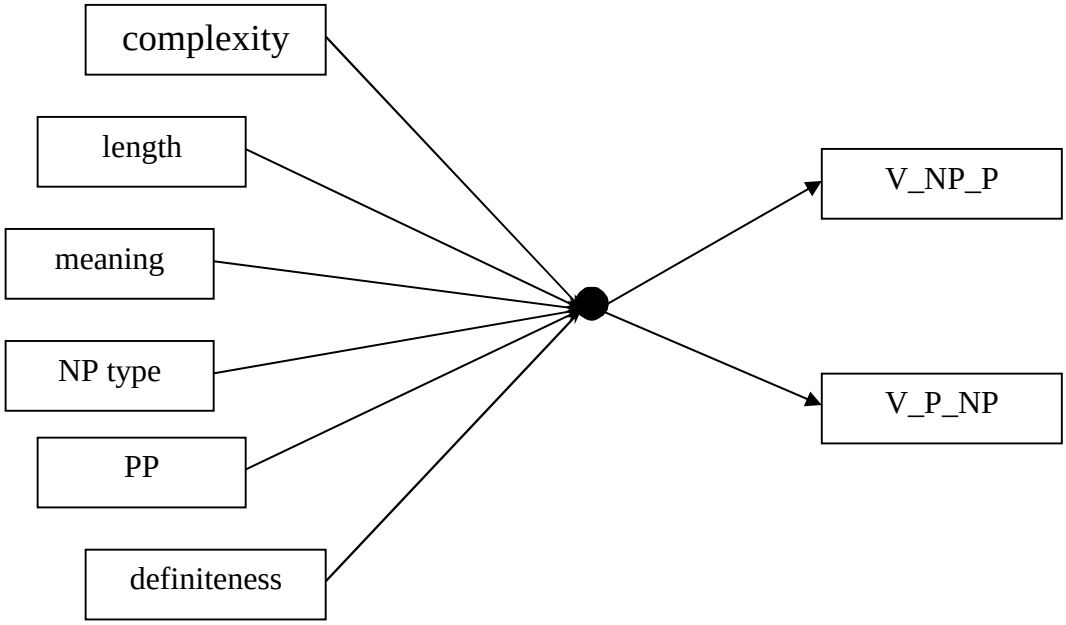


Table 1. Result of the logistic regression

Factor	Odds ratio		p value
NP type	lexical Ns vs. personal PROs	= 72.46	.001
	other PROs vs. personal PROs	= 22.04	.029
	lexical Ns vs. other PROS	= 3.29	.156
Meaning	spatial vs. non-spatial	= 7.1	.001

Discussion

Hypothesis 1:

Children as young as 2,0 years of age process the verb-particle construction in the same way as adult speakers (except maybe that they do not take all factors into account).

Hypothesis 2:

Children 're-produce' the verb-particle constructions they encounter in the ambient language without processing the factors that influence particle placement in adult language.

Structure of an empirical research paper

A. Introduction

1. Background and preview
   Why do you conduct this study? What makes it interesting? Short statement of purpose/hypothesis.
2. Linguistic phenomenon
   Describe the linguistic phenomenon you investigate (e.g. relative clauses, resultative construction): define the category, describe its features (e.g. inflectional variation, word order variation), give examples.
3. Literature review
   If there are previous studies, summarize the main findings and say what you intend to do in your study based on previous work (e.g. look at a particular phenomenon that has not been investigated thus far; challenge a previous hypothesis; replicate a previous study to see if the results of that study carry over to other data; etc.). If there are no previous studies, make that clear: "This is the first study to investigate …."
4. Preview / explicit hypothesis
   State your hypothesis (or hypotheses) and if necessary explain it/them in more detail. This may include a preview of your most important results.

B. Method

1. Subjects, corpus, and materials
   Describe the data you investigate. If you conduct an experimental or questionnaire study, characterize your subjects and describe the materials you have used. If you conduct a corpus study, characterize the corpus (e.g. size, kind). You may want to include summary tables of your data, but don't present the results of your analysis at this stage.
2. Procedure
   Describe the way you have collected the data. How did you conduct the experiment? How did your search for particular constructions in the corpus?
3. Coding
   Describe how you have categorized the data. Give an overview of all categories and state how you assigned a particular response (in an experimental study) or a particular instance (in a corpus study) to a particular category.

C. Results

1. Descriptive summary of results
   (i) In the social sciences (e.g. psychology, sociology) you first present your results and then discuss them. In linguistics, the results and the discussion are often combined in one section, but if you find it appropriate you can separate them.
   (ii) Present tables and figures to summarize your findings, but don't present your findings twice, i.e. first in a table and then in a figure.
   (iii) Preparing tables and figures can be difficult. Don't give long tables including hundreds of numbers; nobody has the time to look at them. The tables and figures in the text serve to provide easy access to your most important findings. The appendix may include a more detailed summary of your results presented in more comprehensive tables.
   (iv) You need to discuss the results presented in tables and figures! The figures/tables alone are not sufficient. Say what the descriptive statistics suggest.

2. Inferential statistics

   Once you have described your data, submit them to statistical analysis. Say what type of test you have used and present the relevant measures (e.g. p-value, F-value, degrees of freedom, effect size, confidence intervals). If it is not obvious, why you used a particular test, explain your decision, but don't describe obvious choices (e.g. I have used a chi-square test because the data is frequency data). Say also what the statistical analysis suggests, i.e. how the result should be interpreted.

D. Discussion

1. Provide a short summary of your results
2. Theoretical implications: If possible consider your paper from a broader theoretical perspective and mention implications of your study for related questions.
4. Future direction of research: Mention open questions: What should be done in the next step? Ideas for an experiment. Etc.

Appendix

If the data are too comprehensive to be included in the text, include them in the appendix. If the data are very comprehensive, you might only present parts of your data in the appendix.

References

List all articles and books you have cited.

Introduction to Excel

Raw data

Table 1. Brandt, Diessel, and Tomasello (2007)

RC's Leo	case	file	wordorder	head	focus	age
CHI: [D] Glocken # die laeuten gerade nicht .	4	le020223	vs	NP	subj	2;5
CHI: [D] ein Schiff # das nehmen wir !	5	le020224	vs	NP	obj	2;5
CHI: mit [MA] (G)locke(n), die laeuten .	6	le020224	am	OBL	subj	2;5
CHI: [D] Sterne # die lachen auch .	7	le020227	vs	NP	subj	2;5
CHI: eine Sonne, die lacht Sonnen(sch)ein [/2].	8	le020228	vs	NP	subj	2;5
CHI: Loewe der bruellt .	9	le020228	am	NP	subj	2;5
CHI: [D] Flamingos # die machen laut !	10	le020229	vs	NP	subj	2;5
CHI: Muscheln, die kann man essen .	11	le020301	vs	NP	obj	2;5
CHI: Forellen [/2], die [/3] haben xx .	12	le020301	vs	NP	am	2;5

Table 2. Diessel and Tomasello (2005)

	S	A	P	IO	OBL	GEN
English						
Katie	2,5	2,0	1,5	1,0	0,0	0,0
Stepanie	3,0	1,5	1,5	0,5	1,5	0,0
Olivia	4,0	3,5	2,0	2,0	2,5	0,0
Elle Mae	3,0	1,5	1,0	0,0	1,0	0,0
Cara	4,0	3,5	2,0	1,0	0,0	0,0
Antonia	4,0	2,5	3,5	0,5	1,0	0,0
Francesca	2,5	1,0	1,0	0,0	0,0	0,0
Peter	4,0	4,0	1,5	2,5	1,0	0,0
Rebecca	4,0	3,0	2,0	3,0	1,5	1,0
Jessica	3,5	2,0	2,0	2,5	2,5	1,0
Isabella	3,0	3,5	0,0	0,0	1,0	0,0
Luke	3,0	2,0	1,0	2,0	1,0	0,0
Elliot	4,0	2,5	4,0	3,5	4,0	0,0
Joseph	2,0	1,0	0,5	0,0	0,5	0,0
Megan	3,5	3,0	2,0	1,0	2,0	0,0
Nicholas	3,0	2,5	2,0	0,0	0,0	0,0
Sean	4,0	2,5	3,0	2,0	2,0	0,0
Kate	3,5	2,0	2,0	0,0	1,5	0,0
Maeve	4,0	3,0	1,0	3,5	2,5	0,0
Elizabeth	4,0	2,0	0,5	1,0	1,0	0,0
Charlotte	1,0	1,5	0,0	0,0	0,0	0,0

Zahlen und Texteingabe in Excel

1. Pluszahlen ohne Vorzeichen; Minuszahlen mit einem Minuszeichen.
2. Text steht linkbündig in der Zelle, Zahlen stehen rechtsbündig in der Zelle.
3. Wenn eine Zahl als Text erkannt werden soll: Apostroph vor Zahl (’200).
4. Bruch: 0 ¾
5. Kommentare.

Funktionen

- Daten sortieren
- Taschenrechnerfunktionen
- Funktionsassistent

Pivot-Bericht

Aufgabe 1 (Daten: Gries 2003)

1. Erstellen Sie eine Graphik, die die Anzahl von belebten und unbelebten Objekten in den beiden Konstruktionen zeigt.
2. Erstellen Sie eine Graphik, die den prozentualen Anteil von lexikalischen, pronominalen, semi-pronominalen, und Eigennamen in den beiden Konstruktionen zeigt.
3. Erstellen Sie eine Graphik, die die mittlere Länge der Objekte in den beiden Konstruktionen zeigt.

Aufgabe 2 (Daten: Brandt, Diessel, and Tomasello in press)

Head

(1) Der Mann, den ich gesehen habe, war blond.
(2) Peter sieht den Jungen, mit dem ich gesprochen habe.
(3) Peter sitzt auf dem Stuhl, auf dem ich gestern gesessen habe.
(4) Das ist der Junge, der mir die Tür aufgemacht hat.
(5) Das Bild, das ich gemalt habe.

Focus

(1) Der Mann, der mich gesehen hat.
(2) Der Mann, den ich gesehen habe.
(3) Der Mann, mit dem ich gesprochen habe.

Word order	Head	focus	Age
vf (verb final)	SUBJ	subj	2;5
vs (verb second)	OBJ	obj	3;0
am (ambiguous)	OBL	obl	3;5
	PN	am	4;0
	NP		4;5

1. Erstellen Sie ein Liniendiagramm, das die Entwicklung der beiden Wortstellungskonstruktionen zeigt.
2. Erstellen Sie ein Säulendiagramm, das die Anzahl der verschiedenen Relativsatzarten (Focus) im Gesamtkorpus zeigt.
3. Erstellen Sie ein Säulendiagramm, das den prozentualen Anteil der Foci in den beiden Wortstellungskonstruktionen zeigt.

Aufgaben 3 (Daten: Brandt, Diessel, and Tomasello in press)

1. Erstellen Sie ein Säulendiagramm, das den Anteil der verschiedenen Köpfe im Gesamtkorpus zeigt.
2. Erstellen Sie ein Liniendiagramm, das die Entwicklung der verschiedenen Köpfe in den verschiedenen Alterstufen zeigt.
3. Erstellen Sie ein Säulendiagramm, das den prozentualen Anteil der Köpfe in den beiden Wortstellungskonstruktionen zeigt.

Lassen Sie in allen Graphiken die Variablenausprägung ‘am’ unberücksichtigt.

Research design

Independent and dependent variables

1. Grammaticality judgment task
   Subjects are given two types of constructions and are asked to decide whether the given sentence is grammatical:

(1)	a.	I gave him it.	Construction 1
	b.	I gave her the book.	
	c.	…	
(2)	a.	I gave to him it.	Construction 2
	b.	I gave to her the note you sent me.	
	c.	…	

2. Sentence completion task
   Subjects are asked to complete copular sentences with a relative clause. The predicate nominals of the copular clauses belong to three different semantic types: (1) animate/human (2) inanimate/object (3) place

(1)	a.	This is the man __
	b.	This is the ball __
	c.	This is the place __

Types of data

1. Nominal/categorical data
2. Ordinal data
3. Interval data

Difference test vs. correlational analysis

1. Correlational analysis
2. Difference tests

Sampling

1. Simple random sampling
2. Stratified random sampling
3. Cluster sampling (Stichklumpenprobe)
4. Systematic sampling

Related vs. independent research designs

within subjects – related design –repeated measures design
between subjects – unrelated design – independent design

Experimental design

A child language researcher wants to find out if the acquisition of relative clauses is affected by its syntactic structure. The structure of a relative clause is defined by two features: The syntactic function of the head, i.e. the main clause element that is modified by a relative clause, and the syntactic function of the gap, i.e. the element in the relative clause that is omitted. As can be seen in examples (1) to (4), both head and gap can function as subject or object.

(1)	Peter saw the man who talked to Sally last night.	OS
(2)	Jack noticed the man who Sally met yesterday.	OO
(3)	The man who talked to Sally last night saw Peter.	SS
(4)	The man who Sally met yesterday noticed Jack.	SO

A very simple, but frequently used experimental method to test children's comprehension of linguistic structures is the sentence repetition task, in which children have to repeat a linguistic structure they may find difficult to understand. If a child is not able to understand the structure, s/he is not able to repeat it. Nobody, neither children nor adults, are able to memorize more than 7 randomly combined words. In order to memorize and repeat longer word sequences, the sequence must make sense, i.e. you must be able to understand it. If you are not able to understand it, you either don't respond or you change the given structure to a structure you understand.

Table 1. Schematic token set

schematic token set		Head	
		SUBJ	OBJ
Gap	SUBJ	a (s+s)	b (s+b)
	OBJ	c (b+s)	d (b+b)

Table 2. Concrete token sets

token set 1		HEAD	
		subject	object
REL PRO	SUBJ	The man who met Sally talked to me.	I talked to the man who met Sally.
	OBJ	The man who Sally met talked to me.	I talked to the man who Sally met.

token set 2		HEAD	
		subject	object
REL PRO	SUBJ	The girl who saw Jack kissed Bill.	Bill saw the girl who kissed Jack.
	OBJ	The girl who Jack saw kissed Bill.	Bill saw the girl who Jack kissed.

token set 3		HEAD	
		subject	object
REL PRO	SUBJ	The cat that chased the dog scared the horse.	The cat chased the dog that scared the horse.
	OBJ	The cat that the dog chased scared the horse.	The cat chased the dog that the horse scared.

token set 4		HEAD	
		subject	object
REL PRO	SUBJ	The car that hit the bus pumped into the van.	The car hit the bus that pumped into the van.
	OBJ	The car that the bus hit pumped into the van.	The car hit the bus that pumped into the van.

Table 3. Test items divided into blocks

# (=number)	blocks	REL PRO	HEAD	experimental condition	stimulus
1	1	SUBJ	subj	a	The man who met Sally talked to me.
2	1	SUBJ	obj	b	I talked to the man who met Sally.
3	1	OBJ	subj	c	The man who Sally met talked to me.
4	1	OBJ	obj	d	I talked to the man who Sally met.
5	2	SUBJ	subj	a	The girl who saw Jack kissed Bill.
6	2	SUBJ	obj	b	Bill saw the girl who kissed Jack.
7	2	OBJ	subj	c	The girl who Jack saw kissed Bill.
8	2	OBJ	obj	d	Bill saw the girl who Jack kissed.
9	3	SUBJ	subj	a	The cat that chased the dog scared the horse.
10	3	SUBJ	obj	b	The cat chased the dog that scared the horse.
11	3	OBJ	subj	c	The cat that the dog chased scared the horse.
12	3	OBJ	obj	d	The cat chased the dog that the horse scared.
13	4	SUBJ	subj	a	The car that hit the bus pumped into the van.
14	4	SUBJ	obj	b	The car hit the bus that pumped into the van.
15	4	OBJ	subj	c	The car that the bus hit pumped into the van.
16	4	OBJ	obj	d	The car hit the bus that pumped into the van.

Table 4. Section of the final experimental design

# (=number)	blocks	REL PRO	HEAD	experimental condition	stimulus	randomizer
1	1	SUBJ	subj	a	*The man who …*	0,017871058
13	4	SUBJ	obj	a	***The man who …***	0,12021337
41	filler 25	-	-	a	[...]	0,429620655
5	2	SUBJ	subj	a	*The cat that …*	0,462384624
45	filler 29	-	-	a	[...]	0,536353922
33	filler 17	-	-	a	[...]	0,563961285
9	3	SUBJ	obj	a	*The girl who …*	0,639759497
37	filler 21	-	-	a	[...]	0,648439236
21	filler 5	-	-	a	[...]	0,803529118
17	filler 1	-	-	a	[...]	0,804351424
25	filler 9	-	-	a	[...]	0,844548334
29	filler 13	-	-	a	[...]	0,899483715

Aufgabe

Erstellen sie das Experimentaldesign für eine Studie mit einem 2×3 Token Set (jeweils vier konkrete Token Sets für jede Variablenausprägung und 1,5 Mal so viele Distraktoren wie Testsätze).

Mean and variance

Central tendency

Data: 2,3,3,3,4,6,6,9,12,13,13

1. Mean
   (2+3+3+3+4+6+6+9+12+13+13)/11 = 6.72
2. Median
   The middle score (i.e. 6). If there is no data value that represents the middle score, you add the two closest data points and divide them by 2.
3. Mode
   The most frequent score (i.e. 3).

Variance and standard deviation

Exercise: Determine the mean number of words and their 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

Variance: 81.87 / (8-1) = 11.7
Standard deviation: √11.7 = 3.42

z-scores

Suppose you want to compare the relative success of two individuals, who have been tested in two different language proficiency tests. A has a score of 41 in test 1, and B has a score of 53 in test B. Which candidate performed better?

Table 1. Data

	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

Table 2. Calculating 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			

$$Z = X_1 - X_{mean} / SD$$

Aufgabe

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 Punkten und die Standard-abweichung bei 14 Punkten; im zweiten Test (an dem Kandidat B teilgenommen hat) lag der Mittelwert bei 143 Punkten und die Standardabweichung bei 21 Punkten. Welcher der beiden Kandidaten hat besser abschlossen (im Vergleich zu allen anderen Kandidaten)?

Coefficient of variance

$$CV = \frac{SD}{Mean} \times 100$$

Je größer der CV Wert, desto größer die relative Varianz.

Aufgabe

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?

Introduction to SPSS

Dateneingabe

Die Variablen werden auf dem Blatt ‘Variablenansicht’ eingegeben. Dabei sind bestimmte Konventionen zu beachten. Die Eingaben dürfen

1. nicht länger als 8 Zeichen sein
2. nicht mit einer Zahl beginnen (z.B. 1group)
3. nicht mit einem Punkt enden
4. nicht ein Leerzeichen beinhalten
5. nicht die folgenden Zeichen beinhalten: !, ?, *

Correlational analysis

Case	Head size	Intelligence
1	55	125
2	59	132
3	48	94
4	60	110
5	62	140
6	50	96

Experiment (within-subjects)

CASE	CONDITION 1	CONDITION 2	CONDITION 3
1	220	300	260
2	250	290	300
3	260	280	290
4	230	340	190
5	190	300	250
6	220	270	240
7	250	320	270
8	280	290	260
9	270	340	250
10	240	300	350

Experiment (between-subjects)

Case	Group	Score
1	1	3
2	1	5
3	1	3
4	1	2
5	1	4
6	1	6
7	1	9
8	1	3
9	1	8
…	…	…

Frequency data

	Disapprove	
Sex	Yes	No
Female	30	20
Male	10	40

SEX	DISAPPROVE	FREQUENCY
1	1	30,00
1	2	20,00
2	1	10,00
2	2	40,00

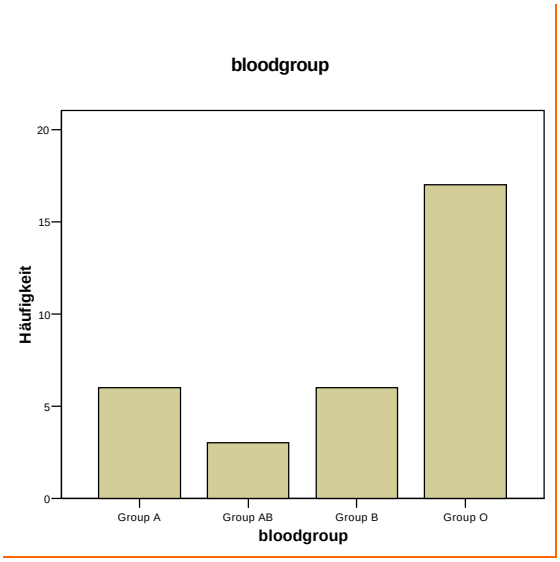
Tables and Figures

bloodgroup

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Group A	6	18,8	18,8	18,8
	Group AB	3	9,4	9,4	28,1
	Group B	6	18,8	18,8	46,9
	Group O	17	53,1	53,1	100,0
	Gesamt	32	100,0	100,0	

sex

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	Female	16	50,0	50,0	50,0
	Male	16	50,0	50,0	100,0
	Gesamt	32	100,0	100,0	

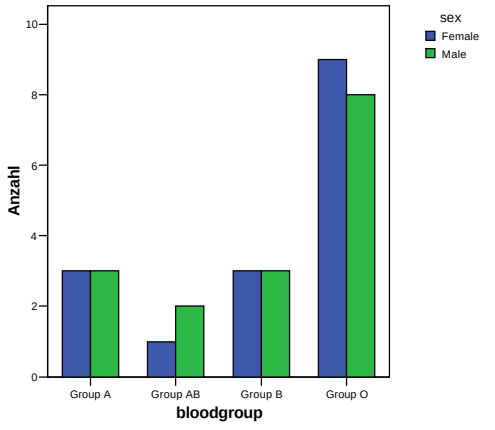


bloodtype * sex Kreuztabelle

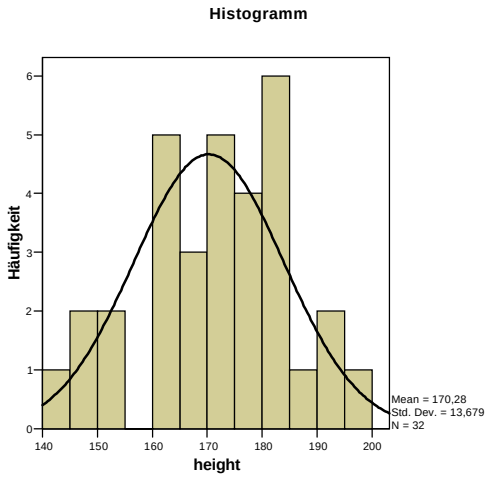
Anzahl

		sex		Gesamt
		Female	Male	
bloodtype	Group A	3	3	6
	Group AB	1	2	3
	Group B	3	3	6
	Group O	9	8	17
Gesamt		16	16	32

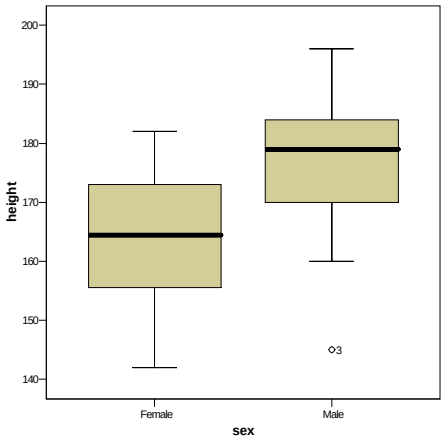
Säulendiagramm



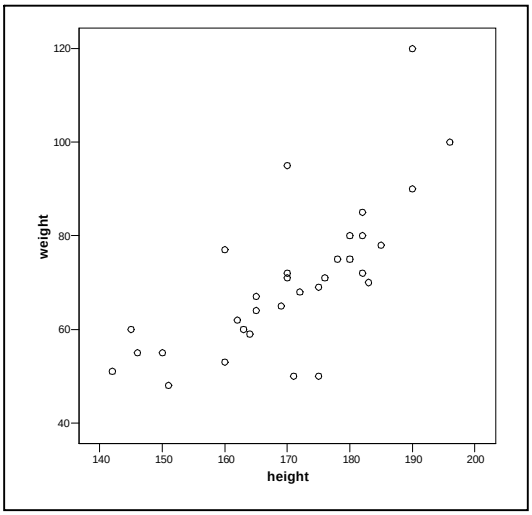
Histogramm



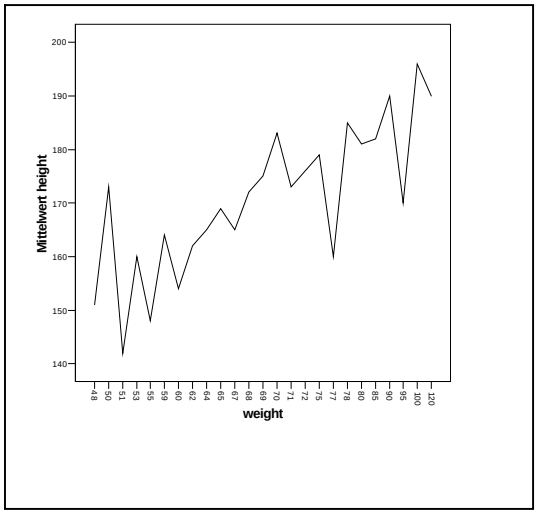
Box-and-leaf Plots



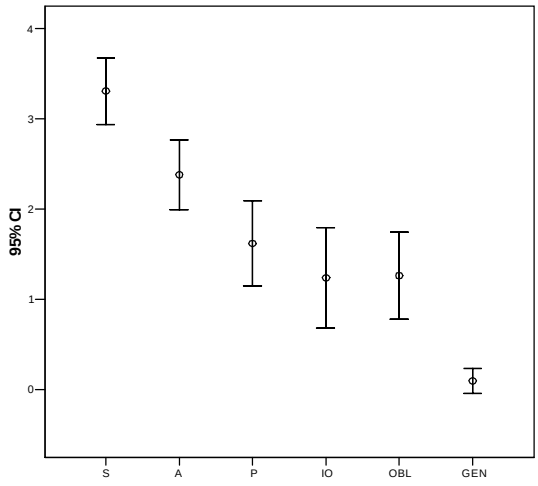
Streudiagramm



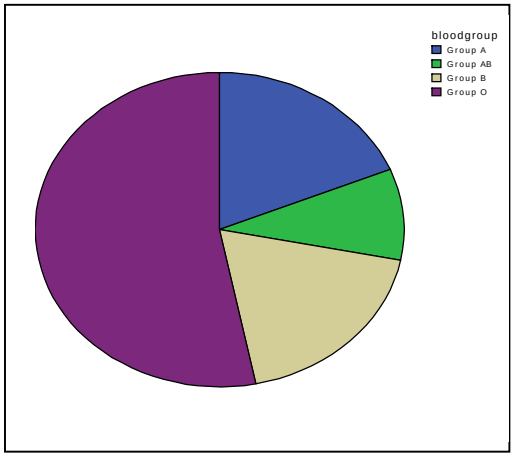
Liniendiagramm



Error Bars



Kreisdiagramm



Probability and statistical hypothesis testing

Probability

Prior probability

(1) $P(6) = 1/6 = 0.166$

- Probability values range from 0 to 1.
- Adding all probabilities of the sample yields 1.
- The probability that an event A will not occur is 1 minus the probability of A.
- If two events are independent, the probability that one or the other event occurs is the sum of their individual probabilities.
- Two events A and B are independent if knowing that the occurrence of A does not change the probability of the occurrence of B.

Joint probability

(2) a. $P(A,B) = P(A) \times P(B)$
b. $P(5,6) = (0.166) \times P(0.166) = 0.0277$

Conditional probability

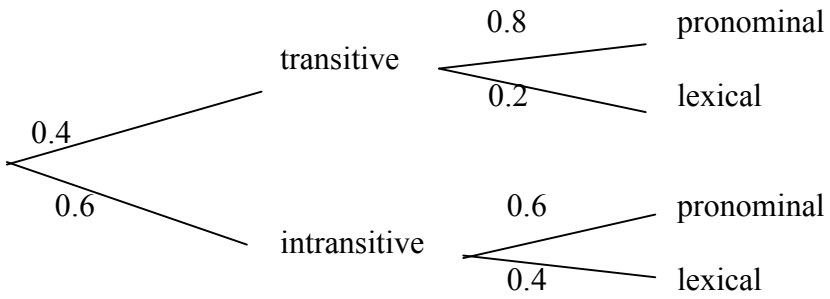
(3) $P(A \mid B) = P(A \cap B) / P(A)$

(4) a. $P(N \mid A) = P(A \cap N) / P(N)$
b. $P(N \mid A) = 2.000 / 12.000 = 0.1666$

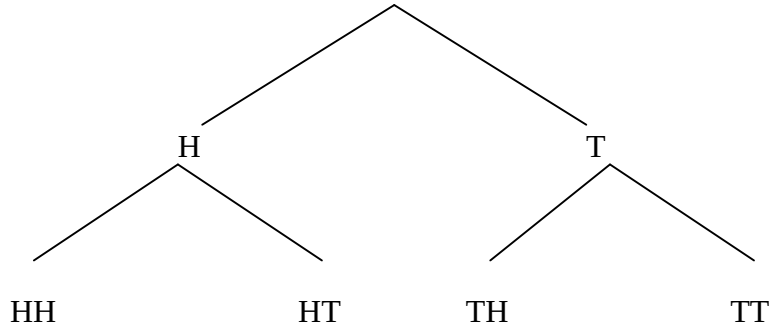
Exercises

(1) In a corpus including 12.000 nouns and 3.500 adjectives, 2.000 adjectives precede a noun. (1) What is the likelihood that a noun occurs after an adjective? (2) What is the likelihood that an adjective precedes a noun?

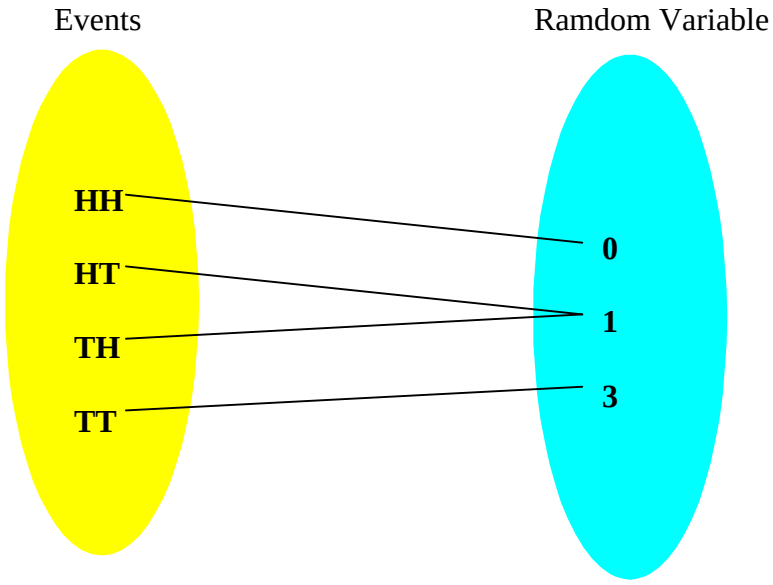
(2) Determine the probability that a transitive sentence includes a pronominal subject.



Probability distributions



1. 0 girl = HH
2. 1 girl = HT + TH
3. 2 girls = TT



Cumulative outcome	Probability
0 1 2	0.25 0.50 0.25
	Σ P(x) = 1

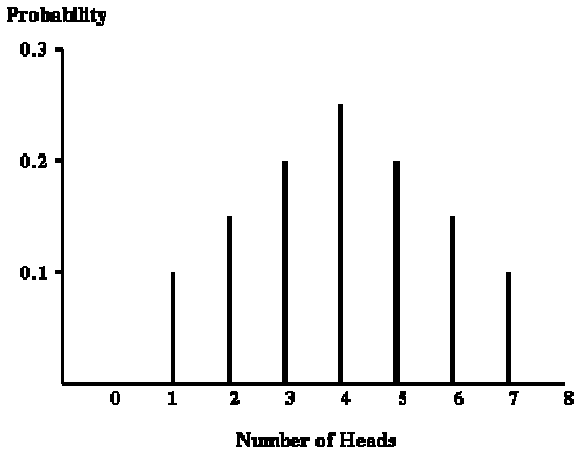
Exercise

A coin is tossed three times. Determine the probability of obtaining exactly 2 tails.

Binomial distribution

A Bernouli trail has the following properties:

1. two possible outcomes on each trail
2. the outcomes are independent of each other
3. the probability ratio is constant across trails

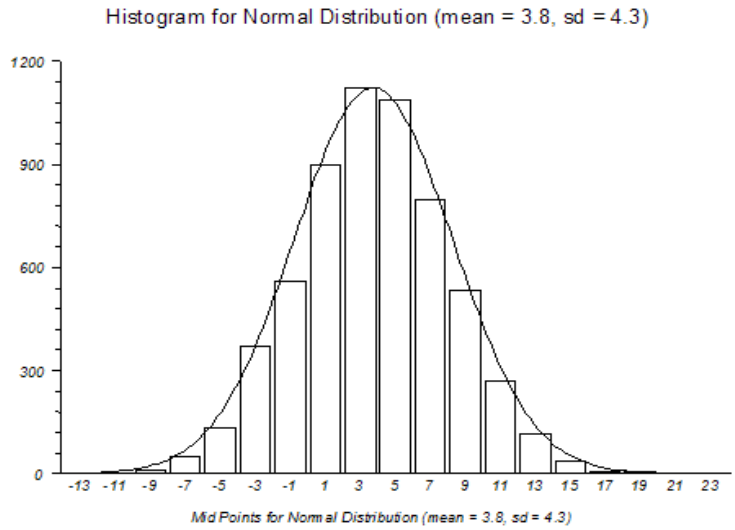


Wenn sie eine Münze 8 Mal werfen, wie hoch ist die Wahrscheinlichkeit, dass sie 8 Mal Kopf bekommen, 7 Mal Kopf, 6 Mal Kopf … ? Die Graphik zeigt, dass die Wahrscheinlichkeit, dass sie 4 Mal Kopf und 4 Mal Zahl bekommen, am größten ist.

The binomial distribution has the following properties:

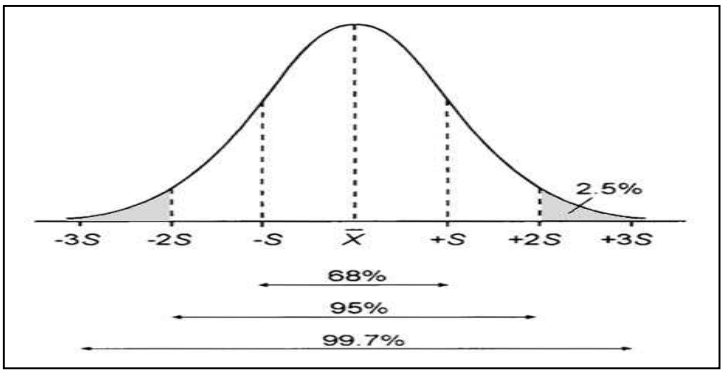
- It is based on categorical / nominal data.
- There are exactly two outcomes for each trail.
- All trials are independent.
- The probability of the outcomes is the same for each trail.
- A sequence of Bernoulli trails gives us the binominal distribution.

Normal distribution



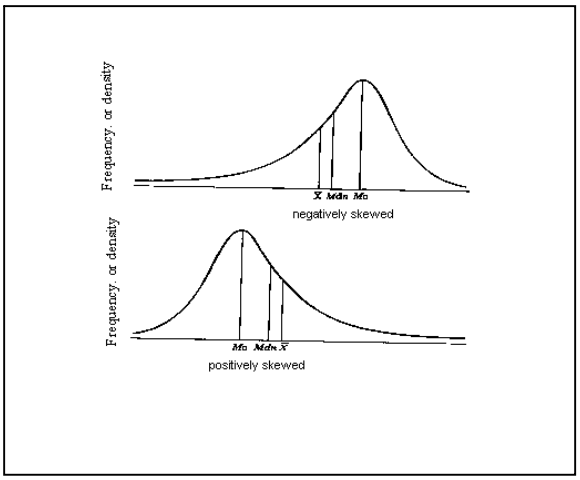
The normal distribution has the following properties:

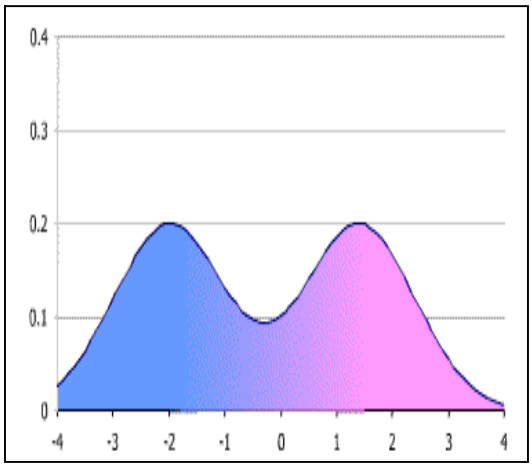
- The center of the curve represents the mean, median, and mode.
- The curve is symmetrical around the mean.
- The tails meet the x-axis in infinity.
- The curve is bell-shaped.
- The total area under the curve is equal to 1 (by definition).

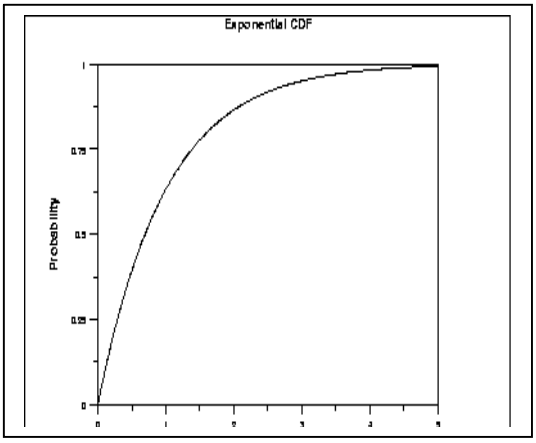


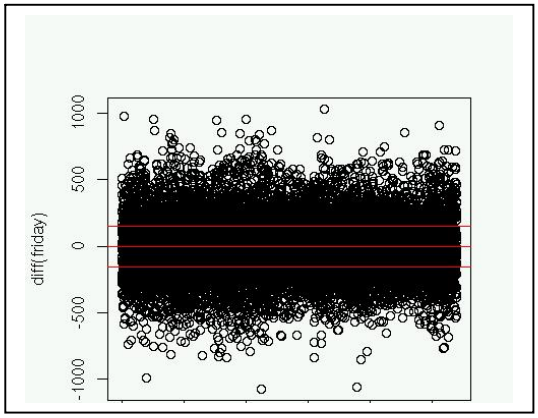
Empirical rule

Other distributions

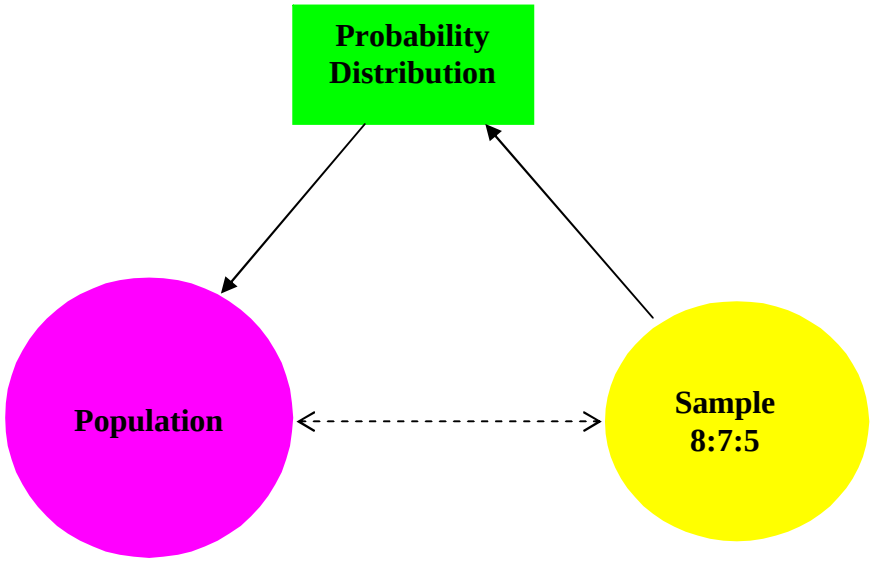






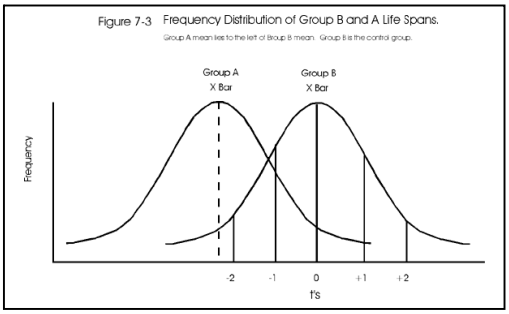


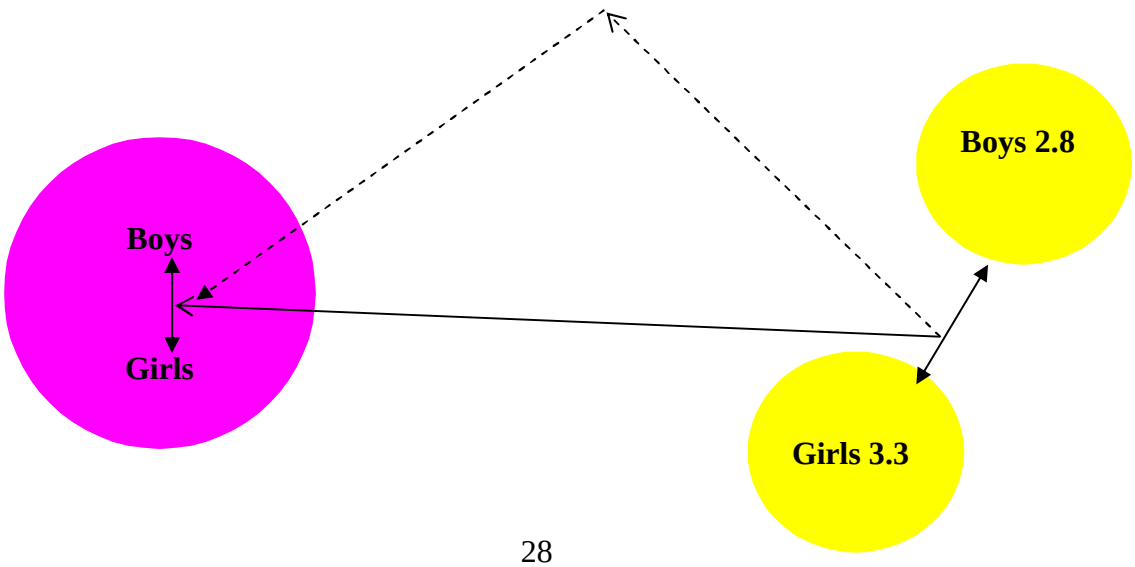
The role of probability models in statistical analysis



Example:

We want to know if the mean length of utterances of English-speaking girls and English-speaking boys is the same or if it is different. To answer this question, we collect a sample of utterances from several three year old boys and several three year old girls. The boys' sample has an MLU of 2.8 words, while the girls' sample has an MLU of 3.3 words. The sample means are different, but are they different enough to assume that boys and girls have different MLUs in the true population? To answer this question, we need a probability model.





The logic of statistical hypothesis testing

Statistical hypothesis testing involves two hypotheses: (1) the null hypothesis, and (2) the experimental (or alternative) hypothesis.

- The null hypothesis says that there is no relationship between the variables in the population (or that there is no difference between two or more groups in the population).
- The alternative hypothesis says that there is a relationship between the variables in the population (or that there is a difference between two or more groups in the population).

Type 1 and Type 2 errors

Type 1 error:
You reject the null hypothesis although there is no correlation between X and Y.
Type 2 error:
You accept the null hypothesis although there is a difference between X and Y.

	p < 0.05	**p > 0.05**
There is a correlation between X and Y in the Population and you **reject** the Null hypothesis	**Correct**	**Type 1 error**
There **is no** a correlation between X and Y in the Population and you **accept** the Null hypothesis	**Type 2 error**	**Correct**

Two-tailed and one-tailed hypotheses

Figure 8-3 One-tailed t-test.

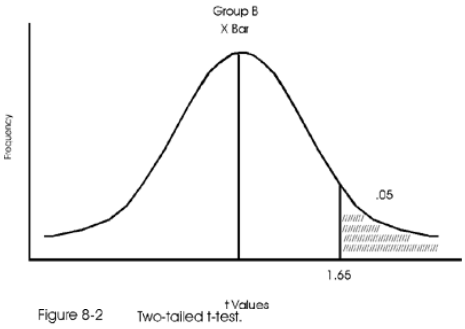
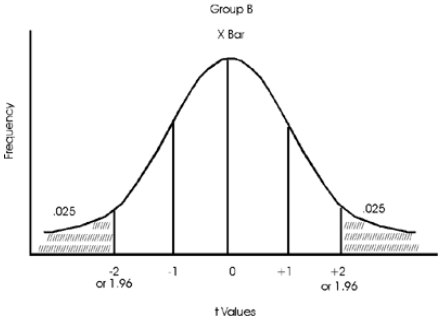


Figure 8-2 Two-tailed t-test.



Classification of statistical tests

- Correlational tests vs. difference test
- Parametric tests vs. non-parametric tests
- Between subjects tests vs. within subjects tests
- Number of conditions
- Number of independent variables

Power

Factors determining the power level of a test:

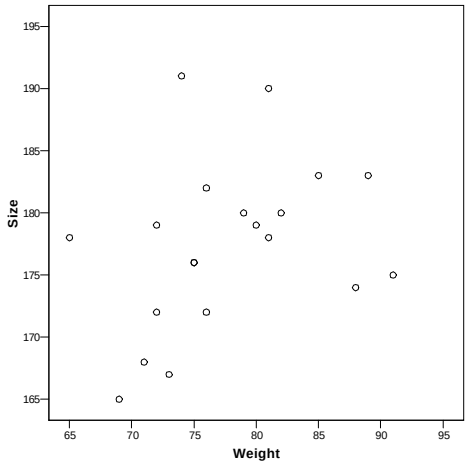
- [the effect size (which is what we are interested in)]
- [the level of significance that we assume (e.g. $p < 0.05$)]
- the sample size: the more subjects the more power
- the type of test: parametric tests are more powerful than non-parametric test
- the experimental design: a within-subject design is more powerful than a between-subject design
- the direction of the experimental hypothesis: a test involving a one-tailed hypothesis is more powerful than a test involving a two-tailed hypothesis

Aufgabe

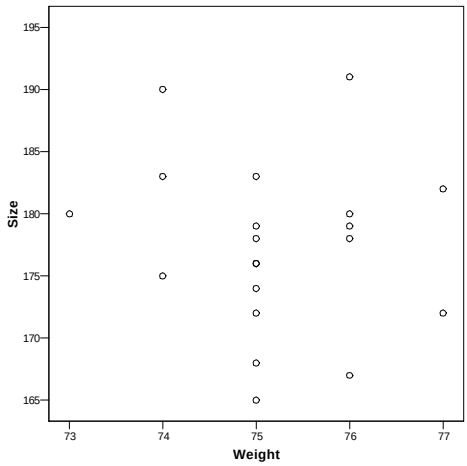
1 (p 152)
2 (p 153)

Correlational analysis

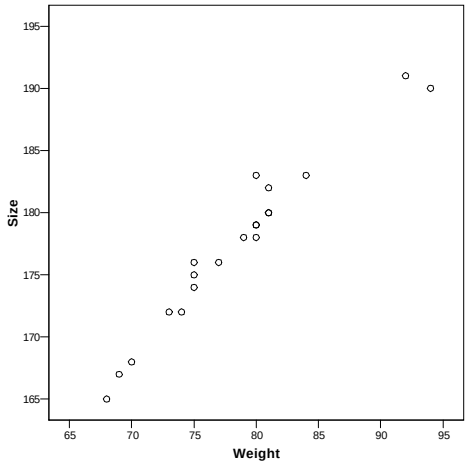
Inspection of the data



Correlation possible



Correlation very unlikely



Correlation very likely

Correlational measures

Nominal	1. Phi coefficient 2. Cramer's V 3. Pearson $\chi 2$ test 4. Loglinear analysis
Ordinal	1. Spearman's Rho 2. Kendall's Tau
Interval	1. Pearson's r

Pearson's r

Correlation coefficient	Shared variance (effect size)	
r = 0.0	0.00	Kein Zusammenhang
r = 0.1	0.01 (1%)	Geringe Korrelation
r = 0.2	0.04 (4%)	
r = 0.3	0.09 (9%)	Mittlere Korrelation
r = 0.4	0.16 (16%)	
r = 0.5	0.25 (25%)	
r = 0.6	0.36 (36%)	Hohe Korrelation
r = 0.7	0.49 (49%)	
r = 0.8	0.64 (64%)	Sehr hohe Korrelation
r = 0.9	0.81 (81%)	
r = 1.0	1.00 (100%)	

Aufgabe

Ein wichtiges Maß für den Entwicklungsstand eines Kindes ist die mittlere Länge von Äußerungen (MLU = Mean length of Utterances). Man geht davon aus, dass die Anzahl der Wörter und/oder Silben, die in einer Äußerung auftauchen, mit dem Alter zunehmen. Überprüfen Sie diese Hypothese an den folgenden Daten.

Table 1. Age and MLU

Child	Age in months	MLU
1	24	2.10
2	23	2.16
3	32	2.25
4	20	1.93
5	43	2.64
6	58	5.63
7	28	1.96
8	34	2.23
9	53	5.19
10	46	3.45
11	49	3.21
12	36	2.84

Kendell's Tau and Spearman's Rho

Beispiel:

In einer Studie zur lexikalischen Semantik mussten 15 Probanden eine List von 10 Wörtern danach beurteilen, wie typisch das jeweilige Wort für die Kategorie 'Fahrzeug' ist. Dazu mussten sie die Wörter sortieren: 1=sehr typisch, 10=sehr untypisch. Jedes Fahrzeug durfte nur einer Zahl zugewiesen werden. Parallel dazu wurde eine zweite Liste erstellt, in der die Wörter aufgrund ihrer Häufigkeit in einem Korpus geordnet wurden (von 1 bis 10). Frage: Gibt es einen Zusammenhang zwischen der Häufigkeit der Wörter und dem Urteil der Probanden?

Table 1. Data

Words	Typicality rank	Frequency rank
car	1	1
truck	2	2
sports car	3	6
motor bike	4	5
train	5	3
bicycle	6	4
ship	7	8
boat	8	7
scat board	9	9
space shuttle	10	10

Korrelationen

			Typicality	Frequency
Kendall-Tau-b	Typicality	Korrelationskoeffizient	1,000	**,733(**)**
		Sig. (2-seitig)	.	**,003**
		N	10	10
	Frequency	Korrelationskoeffizient	,733(**)	1,000
		Sig. (2-seitig)	,003	.
		N	10	10
Spearman-Rho	Typicality	Korrelationskoeffizient	1,000	**,879(**)**
		Sig. (2-seitig)	.	**,001**
		N	10	10
	Frequency	Korrelationskoeffizient	,879(**)	1,000
		Sig. (2-seitig)	,001	.
		N	10	10

** Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Partial correlation

Beispiel:

Zipfs Gesetz besagt, dass die Länge eines Wortes mit seiner Häufigkeit korreliert: Häufig gebrauchte Wörter sind in der Regel kürzer als selten gebrauchte Wörter. Ein Forscher möchte die Gültigkeit des Zipfschen Gesetzes empirisch überprüfen. Zu diesem Zweck untersucht er die Länge und die Verwendungshäufigkeit von 10 willkürlich ausgewählten Wörtern. Als Indikator für die Verwendungshäufigkeit der Wörter ermittelt der Forscher deren Anzahl in einem großen Korpus. Für die Länge der Wörter verwendet er zwei Maße: (i) Anzahl der Phoneme, (ii) Anzahl der Silben.

Wort	Anzahl der Phoneme	Anzahl der Silben	Häufigkeit des Wortes
1	8	3	10,989
2	5	1	8,549
3	4	1	4,209
4	9	3	9,897
5	10	3	8,897
6	4	1	5,345
7	5	2	5,289
8	6	2	6,001
9	8	3	7,987
10	8	2	8,988

Korrelationen

		Phoneme	Silben	Häufigkeit
Phoneme	Korrelation nach Pearson	1	**,898(**)**	**,795(**)**
	Signifikanz (2-seitig)		**,000**	**,006**
	N	10	**10**	**10**
Silben	Korrelation nach Pearson	,898(**)	1	**,677(*)**
	Signifikanz (2-seitig)	,000		**,031**
	N	10	10	10
Häufigkeit	Korrelation nach Pearson	,795(**)	,677(*)	1
	Signifikanz (2-seitig)	,006	,031	
	N	10	10	10

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

* Die Korrelation ist auf dem Niveau von 0,05 (2-seitig) signifikant.

Korrelationen (Silben konstant)

Kontrollvariablen			Phoneme	Häufigkeit
Silben	Phoneme	Korrelation	1,000	**,578**
		Signifikanz (zweiseitig)	.	**,103**
		Freiheitsgrade	0	7
	Häufigkeit	Korrelation	,578	1,000
		Signifikanz (zweiseitig)	,103	.
		Freiheitsgrade	7	0

Assoziationsmaße für nominale Daten

Example:

50 boys and 50 girls are asked to select toys from a cupboard. The available toys were previously categorized as mechanical and non-mechanical. The hypothesis is that boys prefer mechanical toys and girls prefer non-mechanical toys. True?

Table 1. Data

	Mechanical	Non-mechanical	Total
Boys	30	20	50
Girls	15	35	50
Total	45	55	100

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	9,091[b]	1	,003		
Kontinuitätskorrektur[a]	7,919	1	,005		
Likelihood-Quotient	9,240	1	,002		
Exakter Test nach Fisher				,005	,002
Anzahl der gültigen Fälle	100				

a. Wird nur für eine 2x2-Tabelle berechnet

b. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 22,50.

Symmetrische Maße

		Wert	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,302	,003
	Cramer-V	,302	,003
Anzahl der gültigen Fälle		100	

a. Die Null-Hyphothese wird nicht angenommen.

b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.

Regression

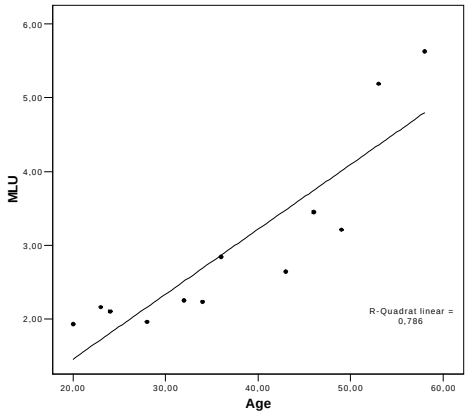
Correlational analysis gives us a measure that represents how closely the data points are associated.

Regression analysis measures the effect of the predictor variable x on the criterion y. – How much does y change if you change x.

A correlational analysis is purely descriptive, whereas a regression analysis allows us to make predictions.

	Predictor variable	Criterion (target) variable
Linear regression	1 interval	1 interval
Multiple regression	2+ (some of the variables can be categorical)	1 interval
Logistic regression Discriminant analysis	1+ (some of the variables can be categorical)	1 categorical 1 categorical

Linear regression



Regression equation: $y = bx + a$

y	=	variable to be predicted
x	=	given value on the variable x
b	=	value of the slope of the line (note that b is the same score as r)
a	=	the intercept (or constant), which is the place where the line-of best-fit intercepts the y-axis.

Example

The table below shows the age of 10 children and their MLU scoure at this age. Use these data to predict how much the MLU score increases relative to the children's age.

Table 1. Children's age and MLU

Child	Age in months	MLU
1	24	2.10
2	23	2.16
3	32	2.25
4	20	1.93
5	43	2.64
6	58	5.63
7	28	1.96
8	34	2.23
9	53	5.19
10	46	3.45
11	49	3.21
12	36	2.84

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,887[a]	,786	,765	,60242

a. Einflußvariablen : (Konstante), Age

ANOVA[b]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	13,369	1	13,369	36,838	,000[a]
	Residuen	3,629	10	,363		
	Gesamt	16,998	11			

a. Einflußvariablen : (Konstante), Age

b. Abhängige Variable: MLU

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	-,304	,566		-,536	,603
	Age	,088	,014	,887	6,069	,000

a. Abhängige Variable: MLU

Multiple regression

$$y = b1x1 + b2x2 + b3x3 + \ldots a$$

1. Simultaneous multiple regression

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,875[a]	,765	,731	16,913

a. Einflußvariablen : (Konstante), Project, Age, IQ, Entrance

ANOVA[b]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	26067,689	4	6516,922	22,783	,000[a]
	Residuen	8009,220	28	286,044		
	Gesamt	34076,909	32			

a. Einflußvariablen : (Konstante), Project, Age, IQ, Entrance

b. Abhängige Variable: Final

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	-267,843	48,263		-5,550	,000
	Entrance	2,492	,459	,576	5,431	,000
	Age	1,057	1,059	,099	,998	,327
	IQ	1,511	,328	,447	4,606	,000
	Project	,503	,355	,141	1,417	,168

a. Abhängige Variable: Final

2. Stepwise partial regression

Aufgenommene/Entfernte Variablen(a)

Modell	Aufgenommene Variablen	Entfernte Variablen	Methode
1	Entrance	.	Schrittweise Auswahl (Kriterien: Wahrscheinlichkeit von F-Wert für Aufnahme <= ,050, Wahrscheinlichkeit von F-Wert für Ausschluß >= ,100).
2	IQ	.	Schrittweise Auswahl (Kriterien: Wahrscheinlichkeit von F-Wert für Aufnahme <= ,050, Wahrscheinlichkeit von F-Wert für Ausschluß >= ,100).

a Abhängige Variable: Final

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,729[a]	,531	,516	22,705
2	,855[b]	,732	,714	17,458

a. Einflußvariablen : (Konstante), Entrance

b. Einflußvariablen : (Konstante), Entrance, IQ

ANOVA[c]

Modell		Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
1	Regression	18096,329	1	18096,329	35,104	,000[a]
	Residuen	15980,580	31	515,503		
	Gesamt	34076,909	32			
2	Regression	24933,067	2	12466,534	40,901	,000[b]
	Residuen	9143,842	30	304,795		
	Gesamt	34076,909	32			

a. Einflußvariablen : (Konstante), Entrance

b. Einflußvariablen : (Konstante), Entrance, IQ

c. Abhängige Variable: Final

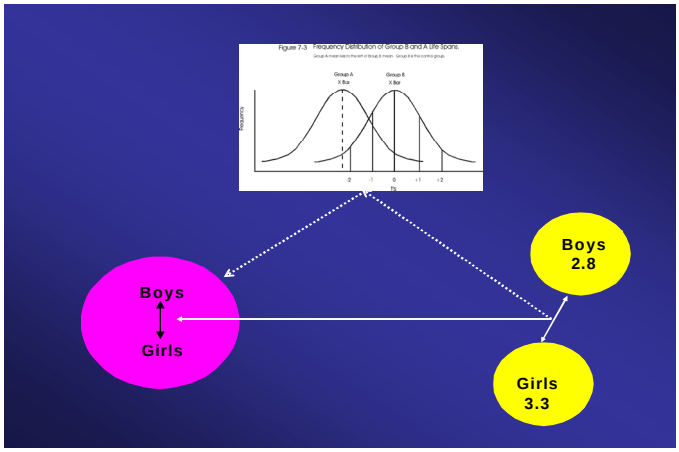
Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	-46,305	25,477		-1,817	,079
	Entrance	3,155	,532	,729	5,925	,000
2	(Konstante)	-221,659	41,888		-5,292	,000
	Entrance	2,521	,431	,582	5,854	,000
	IQ	1,591	,336	,471	4,736	,000

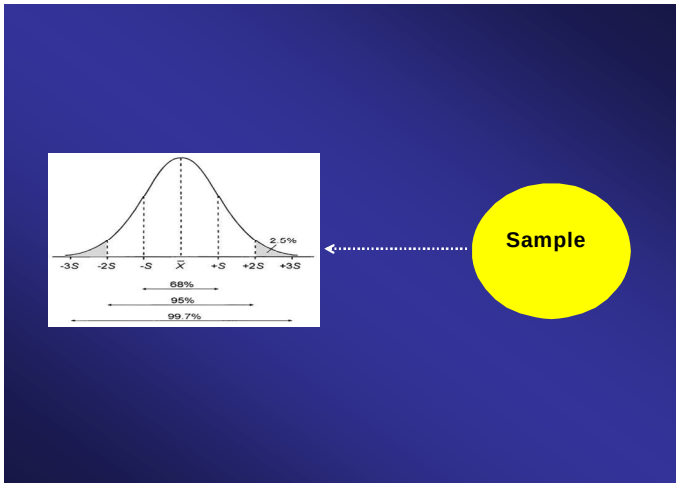
a. Abhängige Variable: Final

One-sample tests

A difference test is often used to find out if there is a difference between two or more groups.



However, sometimes you do not want to know if there is a difference between two samples (or groups), but rather if a sample matches a known distribution.



The Binomial test

A linguist has collected a sample of sentences including ditransitive verbs from a corpus. Overall, there are 46 sentences in his sample; in 27 sentences the verb occurs with two NP objects, in 19 sentences the verb occurs with an NP and a PP.

(1) a. He gives Peter the ball. V NP NP
b. He gives the ball to Peter. V NP PP

Test auf Binomialverteilung

		Kategorie	N	Beobachteter Anteil	Testanteil	Asymptotische Signifikanz (2-seitig)
Frequency	Gruppe 1	27,00	27	,59	,50	,302(a)
	Gruppe 2	19,00	19	,41		
	Gesamt		46	1,00		

a Basiert auf der Z-Approximation.

Aufgabe

A researcher examined the effectiveness of two new drugs on chronic pain. The first drug was given and pain assessed (Pain1); then a month later the second drug was given and again pain assessed (Pain2). Following the study the researcher wanted to know if the proportions of men and women in the sample used were what would be expected by chance, i.e. 50:50. 15 subjects participated in the study, 10 male and 5 female.

Case	Group	Pain 1	Pain 2
1	1	2	2
2	1	5	3
3	1	2	2
4	1	3	2
5	2	4	2
6	1	5	4
7	2	2	3
8	1	3	3
9	1	4	4
10	2	1	2
11	2	2	4
12	1	3	3
13	1	4	2
14	2	2	1
15	1	3	1

Runs test

a. ABABABAABABBABABABABAAABABA
b. AAAAAAABAAAAAAABBBBBBBBABBB

Kolmogorov-Smirnov

Used to test if the distribution in your sample is still in the range of what you would expect if the sample is drawn from a normal distribution.

$\chi 2$ test for goodness-of-fit

Statistik für Test

	Frequency
Chi-Quadrat[a]	1,391
df	1
Asymptotische Signifikanz	,238

a. Bei 0 Zellen (,0%) werden weniger als 5 Häufigkeiten erwartet. Die kleinste erwartete Zellenhäufigkeit ist 23,0.

Exercise

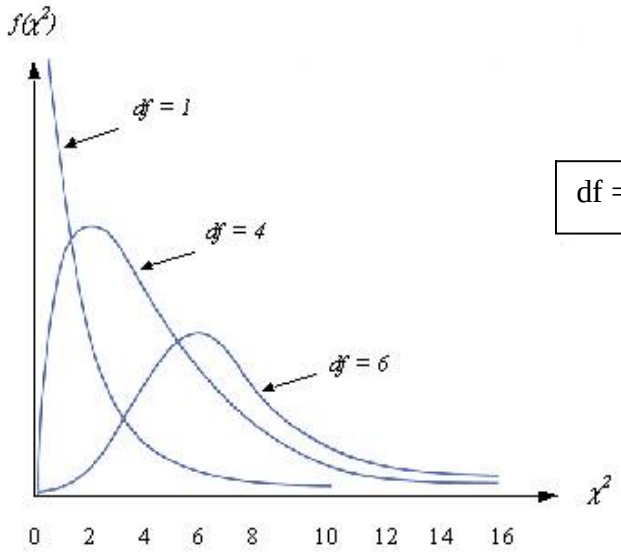
A linguist has collected relative clauses from a corpus, which he divided into four types: (1) subjects relatives, (2) object relatives, and (3) oblique relatives, (4) genitive relatives. Are the differences between the four types of relative clauses in the sample sufficient to posit a frequency difference between the four types in the population?

Table 1. Data

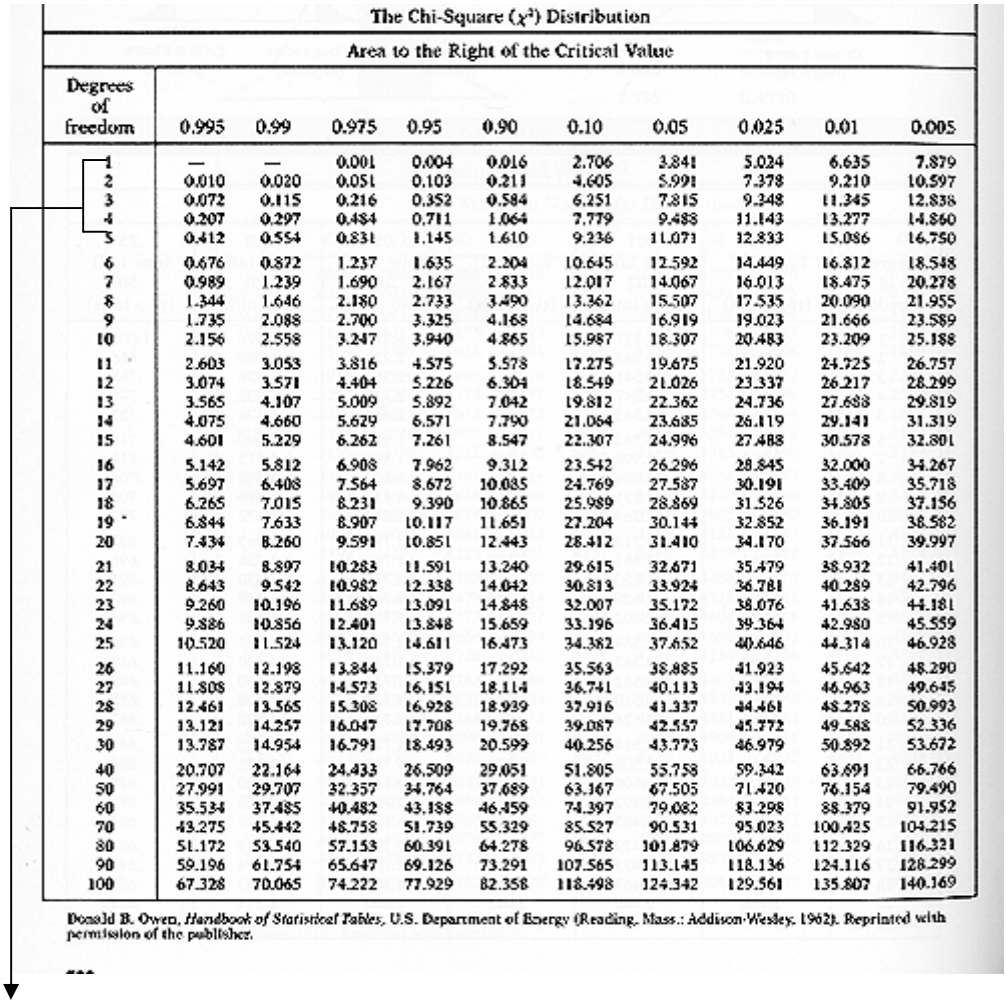
	Subject	Object	Oblique	Genitive	Total
Frequency	55	53	39	4	151

Table 2. Chi-square value

observed	Expected	Difference (Residuals)	Square	Sum	Divided by expected frequency
55	37.75	17.25	297.56	1670	$\chi 2 = 44.25$
53	37.75	15.25	232.56		
39	37.75	1.25	1.56		
4	37.75	-33.75	1139.06		



df = [number of levels] – [1] (in a one-sample test)



The Chi-Square (χ^2) Distribution

Area to the Right of the Critical Value

Degrees of freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.299
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.772	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

Donald B. Owen, *Handbook of Statistical Tables*, U.S. Department of Energy (Reading, Mass.: Addison-Wesley, 1962). Reprinted with permission of the publisher.

	.995	.99	.975	.95	.90	.10	**.05**	.025	.01	.005
1 df										
2 df										
3 df	0.072	0.115	0.216	0.352	0.584	6.25	**7.81**	9.35	11.34	12.84
4 df										

frequency

	Beobachtetes N	Erwartete Anzahl	Residuum
4,00	4	37,8	-33,8
39,00	39	37,8	1,3
53,00	53	37,8	15,3
55,00	55	37,8	17,3
Gesamt	151		

Statistik für Test

	frequency
Chi-Quadrat [a]	44,258
df	3
Asymptotische Signifikanz	,000

a. Bei 0 Zellen (,0%) werden weniger als 5 Häufigkeiten erwartet. Die kleinste erwartete Zellenhäufigkeit ist 37,8.

$\chi 2$ test for independence: 2×2

Example

A linguist wants to find out if subject and object are expressed by the same type of nouns. Specifically, he wants to know if lexical and pronominal NPs are equally distributed. In order to test this hypothesis, he has collected the following frequency data from a small corpus.

	Subject	Object	Total
Pronominal	47	17	64
Lexical	41	52	93
Total	88	69	157

$$\text{Expected frequency} = \frac{X \times Y}{\text{total}}$$

	Subject	Object	Total
Pronominal	**47** 64×88/157 = **35.9**	**17** 64×69/157 = **28.1**	64
Lexical	**41** 93×88/157 = **52.1**	**52** 93×69/157 = **40.9**	93
Total	88	69	157

$$\chi 2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$$\chi 2 = \frac{(47-35.9)^2}{35.9} + \frac{(17-28.1)^2}{28.1} + \frac{(41-52.1)^2}{52.1} + \frac{(52-40.9)^2}{40.9} = 13.18$$

$$df = (\text{rows} - 1) \times (\text{columns} - 1)$$

NPtype * Role Kreuztabelle

			Role		
			OBJ	SUBJ	Gesamt
NPtype	N	Anzahl	52	41	93
		Erwartete Anzahl	40,9	52,1	93,0
	PRO	Anzahl	17	47	64
		Erwartete Anzahl	28,1	35,9	64,0
Gesamt		Anzahl	69	88	157
		Erwartete Anzahl	69,0	88,0	157,0

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)	Exakte Signifikanz (2-seitig)	Exakte Signifikanz (1-seitig)
Chi-Quadrat nach Pearson	13,258[b]	1	,000		
Kontinuitätskorrektur[a]	12,094	1	,001		
Likelihood-Quotient	13,628	1	,000		
Exakter Test nach Fisher				,000	,000
Anzahl der gültigen Fälle	157				

a. Wird nur für eine 2x2-Tabelle berechnet

b. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 28,13.

Symmetrische Maße

		Wert	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,291	,000
	Cramer-V	,291	,000
Anzahl der gültigen Fälle		157	

a. Die Null-Hyphothese wird nicht angenommen.

b. Unter Annahme der Null-Hyphothese wird der asymptotische Standardfehler verwendet.

Prerequisites of the $\chi 2$ test of independence:

1. each subject provides a score for only one cell
2. none of the cells is empty
3. not more than 25% of the cells has an expected frequency of less than 5 (which is one cell in a 2 by 2 table)

Fisher exact

http://www.matforsk.no/ola/fisher.htm

McNemar

Beispiel:

Wir wollen herausfinden, ob der Sprachgebrauch (d.h. die Erfahrung mit sprachlichen Einheiten) einen Einfluss auf den Erwerb einer grammatischen Konstruktion hat. Dazu bitten wir 100 Kinder, einen ditransitiven Satz mit 10 Wörtern nachzusprechen (*Der Mann gibt dem kleinen Jungen einen sehr großen Ballon*). Alle Kinder müssen den Satz zwei Mal nachsprechen: (1) zu Beginn der Studie, und (2) nach einer Trainingsphase, in der die Kinder ähnliche ditransitive Sätze 5 Mal scheinbar beiläufig in einer einstündigen Konversation gehören. Nach der Datenerhebung werden die Kinder in vier Gruppen eingeteilt:

(i) Kinder, die den Satz vor und nach der Trainingsphase richtig wiederholen (Zelle a)
(ii) Kinder, die den Satz vorher falsch und nachher richtig wiederholen (Zelle b)
(iii) Kinder, die den Satz vorher richtig und nachher falsch wiederholen (Zelle c)
(iv) Kinder, die den Satz vorher falsch und nachher falsch wiederholen (Zelle d)

		vorher		
		richtig	falsch	Total
nachher	richtig	31 (a)	39 (b)	70
	falsch	13 (c)	17 (d)	30
	Total	44	56	100

Statistik für Test(b)

	Vor & Nach
N	100
Chi-Quadrat(a)	12,019
Asymptotische Signifikanz	,001

a Kontinuität korrigiert
b McNemar-Test

Der McNemar Test ist vielseitig einsetzbar (sie Bortz, Lienert, Boehnke 161-5). Er ist aber auf 2×2 Tabelle beschränkt und hat die gleichen Voraussetzungen wie der $\chi 2$-Test (keine Zelle mit einer erwarteten Häufigkeit von weniger als 5).

Es gibt zwei Erweiterungen des McNemar:

1. Wenn die Antworten der Probanden nicht zwei sondern drei+ mögliche Antworten zulassen (z.B. 1. richtig und verständlich, 2. falsch aber verständlich, 3. falsch und unverständlich), benutzt man den **Bowker Test** (siehe Bortz, Lienert, Boehnke 165-8).
2. Wenn die Probanden nicht nur zwei Mal sondern mehrmals zu verschiedenen Zeiten getestet werden, benutzt man den **Q-test von Cochran** (siehe Bortz, Lienert, Boehnke 169-71).

$\chi 2$ test for independence: r×c

Beispiel

Wir wollen prüfen, ob es einen Zusammenhang zwischen Rauchen und Trinken gibt. Dazu befragen wir 337 Probanden, die wir jeweils in drei Gruppen einteilen:

1. Heavy drinker
2. Light drinker
3. Non-drinker

1. Heavy smoker
2. Light smoker
3. Non-smoker

Drinker * Smoker Kreuztabelle

Anzahl

		Smoker			
		heavy	leight	non	Gesamt
Drinker	heavy	33	32	35	100
	leight	56	23	34	113
	non	42	28	54	124
Gesamt		131	83	123	337

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	11,282[a]	4	,024
Likelihood-Quotient	10,949	4	,027
Zusammenhang linear-mit-linear	,665	1	,415
Anzahl der gültigen Fälle	337		

a. 0 Zellen (,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 24,63.

Nominaldaten mit zwei oder mehr unabhängigen Variablen

	Declarative clauses		**Questions**		
	Transitive	**Intransitive**	**Transitive**	**Transitive**	**Total**
Written	23	45	56	12	136
Spoken	34	56	32	22	144
Total	57	101	88	34	280

Drei Variablen: 1. spoken – written
2. declarative – interrogative
3. transitive – intansitive

Um Nominaldaten mit zwei oder mehr unabhängigen Variablen bearbeiten zu können werden besondere statistische Verfahren verwendet:

1. Konfigurationsfrequenzanalyse (KFA)
2. Log-lineare Analyse

Konfigurationsfrequenzanalyse (KFC)

Beispiel

Im Englischen gibt es die Möglichkeit, Relationen zwischen Possessor und Possessed mit zwei Konstruktionen auszudrücken:

1. of-Genitive The cover of the book
2. s-Genitive The book's cover

Die Wahl der Konstruktionen wird durch verschiedene Faktoren bestimmt. Ganz wichtig ist die Semantik der beiden beteiligten NPs. Gries (2002) unterscheidet drei Typen:

1. Abstrakte NPs
2. Konkrete NPs
3. Menschliche NPs

Daraus ergibt sich folgende Variablenausprägung:

Possessed	**Abstract**		**Concrete**		**Human**		**Total**		
Possessor	*of*	*s*	*of*	*s*	*of*	*s*	*of*	*s*	Total
Abstract	80	37	9	8	3	2	92	47	139
Concrete	22	0	20	1	0	0	42	1	43
Human	9	58	1	35	6	9	16	102	118
Total	111	95	30	44	9	11	150	150	300
	206		74		20				

Eine KFC prüft, ob die Häufigkeiten bestimmter Konfigurationen höher oder niedriger sind als man vom Zufall her erwarten würde. Ist eine Konfiguration signifikant häufiger, wird sie als **Typ** bezeichnet; ist eine Konfiguration signifikant seltener, wird sie als **Antityp** bezeichnet. Die folgende Tabelle zeigt, dass die KFA auf einem Chi-Quadrat Test basiert.

Possessor	Possessed	Type	Observed	Expected	Residuals
Abstract	**Abstract**	**of**	80	$\frac{130\times206\times150}{300^2}$ = 47.72	$\frac{(80-47.72)^2}{47.72}$ = **21.83**
Abstract	**Abstract**	**s**	37	47.72	2,41
Abstract	**Concrete**	**of**	9	17.14	3.87
Abstract	**Concrete**	**s**	8	17.14	4.88
Abstract	**Human**	**of**	3	4.63	0.58
Abstract	**Human**	**s**	2	4.63	1.5
Concrete	**Concrete**	**s**	22	14.76	3.55
Concrete	**Concrete**	**of**	0	14.76	**14.76**
Concrete	**Abstract**	**s**	20	5.3	**40.73**
Concrete	**Abstract**	**of**	1	5.3	3.49
Concrete	**Human**	**s**	0	1.43	1.43
Concrete	**Human**	**s**	0	1.43	1.43
Human	**Abstract**	**s**	9	40.51	**24.51**
Human	**Abstract**	**of**	58	40.51	7.55
Human	**Concrete**	**s**	1	14.55	**12.62**
Human	**Concrete**	**of**	35	14.55	**28.73**
Human	**Human**	**s**	5	3.93	1.09
Human	**Human**	**s**	9	3.93	6.53
Summen			300	300	181,54 (χ2)

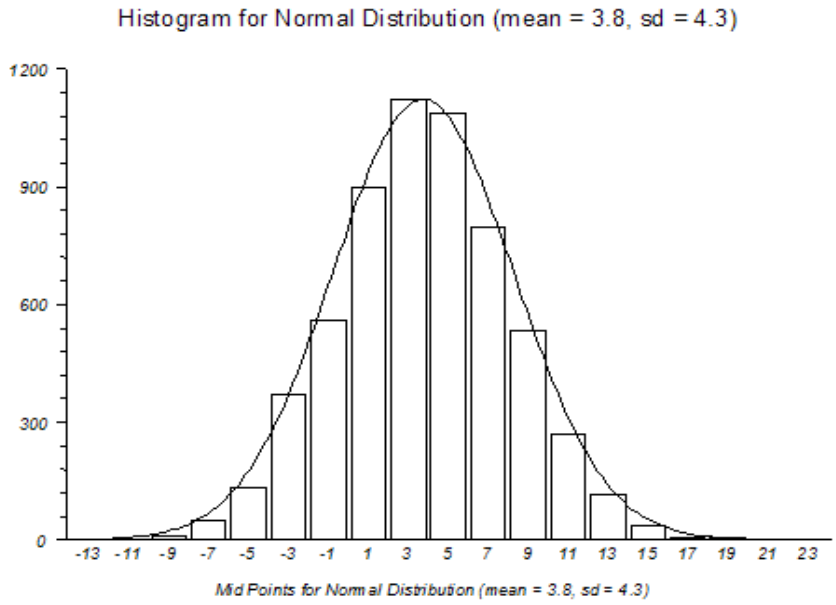
In der Spalte *expected Frequency* sind die erwarteten Häufigkeiten ausgerechnet worden, indem man wie beim χ2-test alle Randhäufigkeiten einer Konfiguration multipliziert und durch N dividiert. Die χ2-Werte, die über einem bestimmten Signifikanzniveaus liegen, sind signifikante Typen bzw. Antitypen:

Typen
1. [abstract possessor] [abstract possessed] of
2. [concrete possessor] [concrete possessed] of
3. [human possessor] [concrete possessed] s

Antitypen
1. [concrete possessor] [abstrcat possessed] s
2. [human possessor] [concrete possessed] of
3. [humn possessor] [abstract possessed] of

Anders als beim einfachen χ2-Test sind die χ2-Komponenten der einzelnen Felder bei der KFA nicht voneinander unabhängig. Die KFA ist deshalb in erster Linie ein heuristisches Verfahren, das dazu dient besonders auffällige Typen und Antitypen aufzuspüren, die dann anhand von neuen Daten nochmals überprüft werden müssen (siehe Bortz, Lienert, Boehnke 155-157).

Normal distribution and confidence intervals

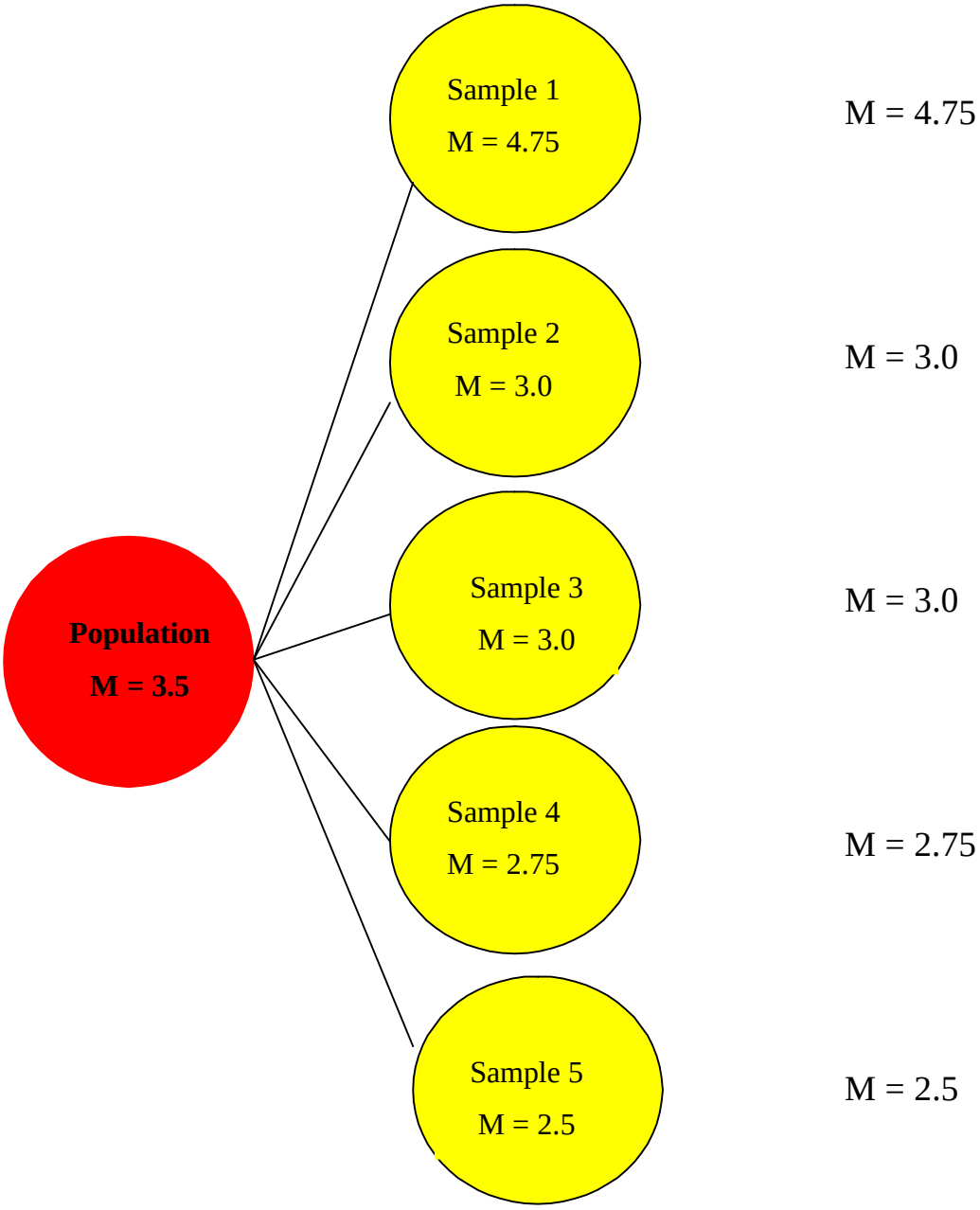


The normal distribution has the following characteristics:

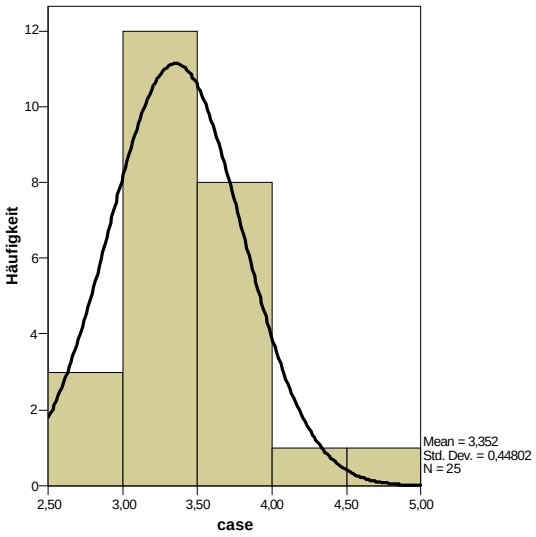
- The center of the curve represents the mean (and median and mode).
- The curve is symmetrical around the mean.
- The tails meet the x-axis in infinity.
- The curve is bell-shaped.
- The are under the curve is 1 (by definition).

Central Limit Theoren

	X_1	X_2	X_3	X_4	$\overline{M}$
Sample 1	6	2	5	6	4.75
Sample 2	2	3	1	6	3
Sample 3	1	1	4	6	3
Sample 4	6	2	2	1	2.75
Sample 5	1	5	1	3	2.5

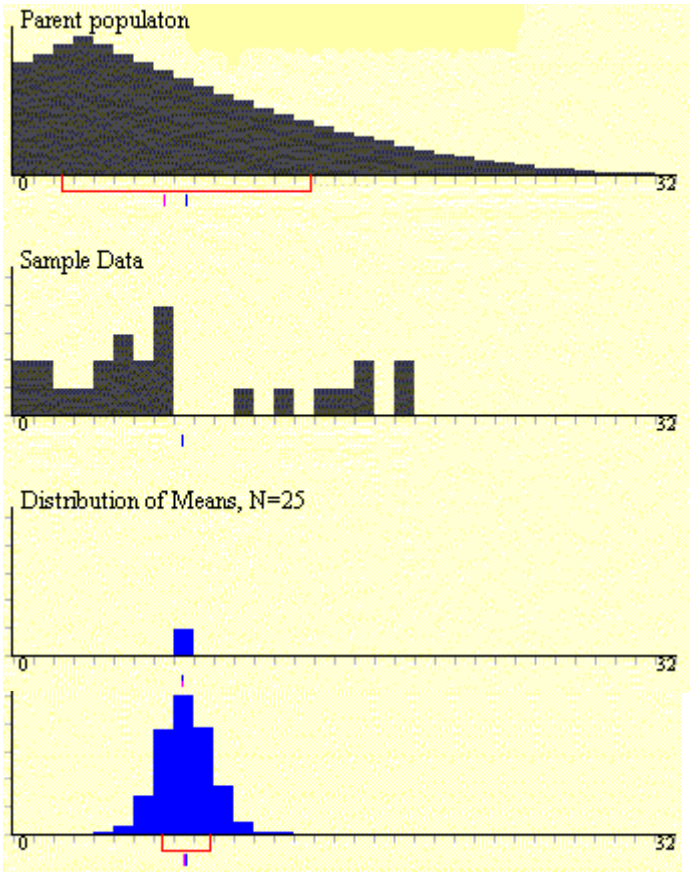


Mean of sample means: 4.75 + 3.0 + 3.0 + 2.75 + 2.5/ 5 = **3.2**

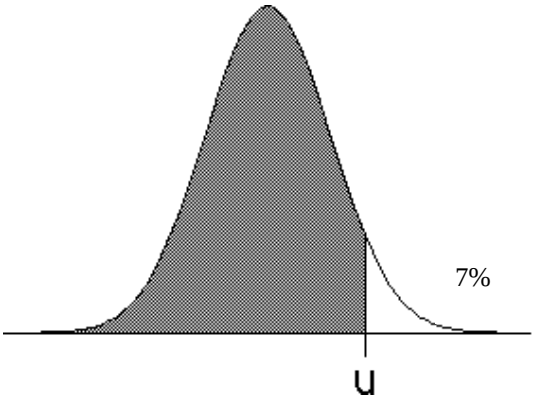


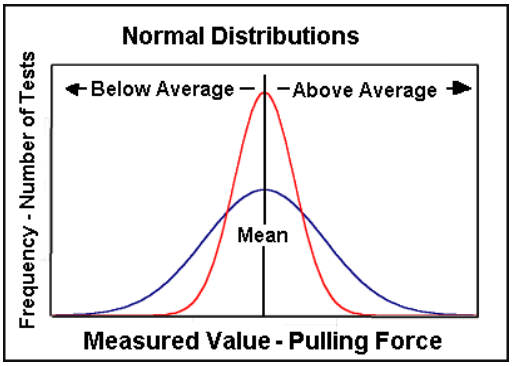
Die Mittelwerte der Stichproben ist normalverteilt, selbst wenn das Phänomen, das wir untersuchen, in der wahren Population nicht normalverteilt ist.

- Graph 1 shows the distribution in the true population.
- Graph 2 shows the distribution in an individual sample.
- Graph 3 shows the mean of the distribution in this sample.
- Graph 4 shows the distribution of the means in the sampling distribution.



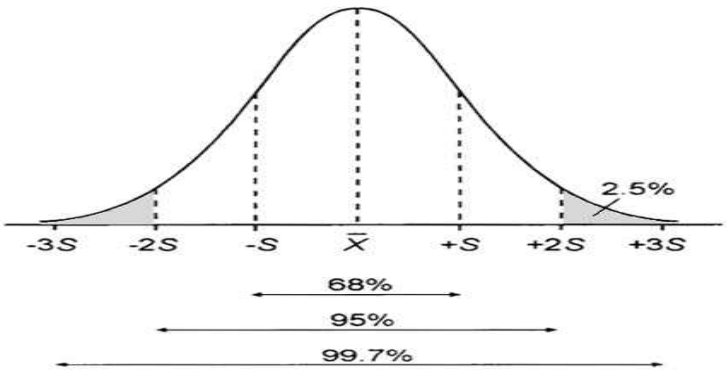
Graphical representation of probability





Every normal distribution can be transformed into the standard normalized distribution. In fact, we have already seen how this works when we talked about z-scores. Z-scores refer to a particular place on the standard normal distribution. Z-scores are calculated by subtracting the mean of a particular sample from an individual data point A and dividing the resultant score by the SD:

$$\text{z-score} = \frac{x_1 - M}{SD}$$



Standard error and confidence intervals

Samples	Mean	Mean – Mean of means		Zum Quadrat
1	1.5	1.5 / 1.66	0.16	0.0256
2	1.8	1.8 / 1.66	0.14	0.0196
3	1.3	4 / 1.66	– 0.36	0.1296
4	2.0	9 / 1.66	– 0.36	0.1156
5	1.7	12 / 1.66	0.04	0.0016
	Σ 8.3 / 5 = 1.66 (mean)			Σ 0.292

Confidence intervals: [degree of certainty] × [standard error] = **x**
[sample mean] +/–**x** = confidence interval

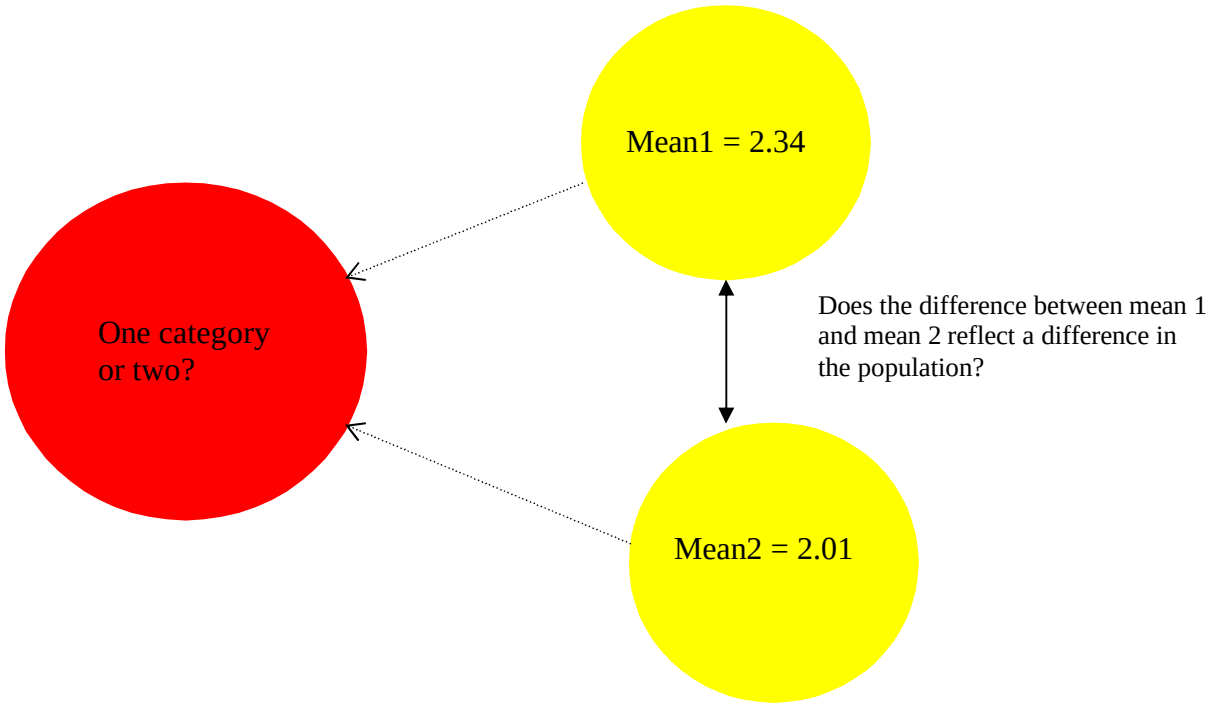
Approximation of standard error:

$$\frac{SD}{\sqrt{N}}$$

Aufgabe

Given the sample (2,5,6,7,10,12) calculate a 95% confidence interval for the population mean.

T-test, U-test, Wilcoxon



Overview of tests

	Parametric	Non-parametric
between / independent / unrelated	Independent t-test	Mann-Whitney U
within / dependent / related / repeated measures	Paired t-test	Wilcoxon

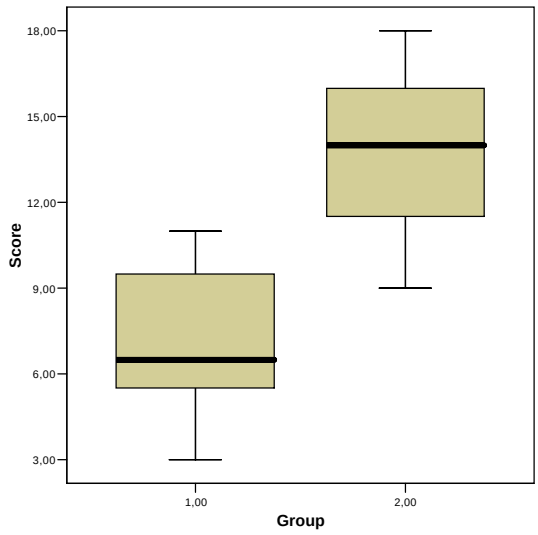
Independent t-test

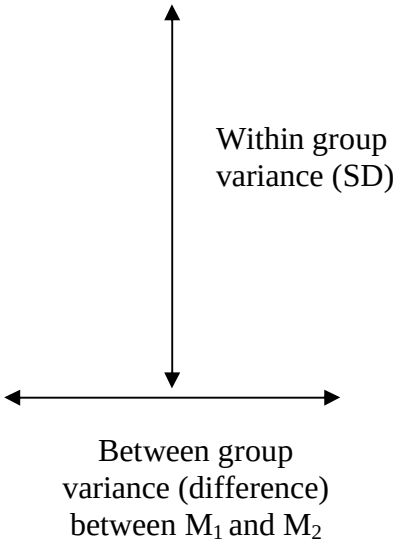
Example (from the textbook):

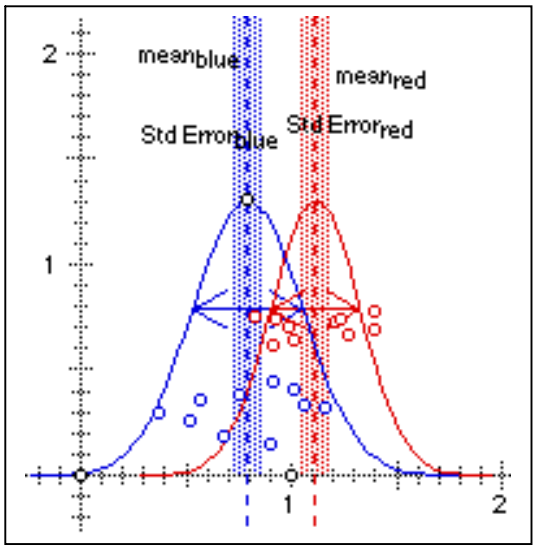
24 people were involved in an experiment to determine whether background noise (e.g. music) affects short-term memory (recall of words). Half of the sample was randomly allocated to the NOISE condition, and half to the NO NOISE condition. The participants in the NOISE condition tried to memorize a list of 20 words in two minutes, while listening to pre-recorded noise through earphones. The other participants wore earphones but heard no noise as they attempted to memorize the words. Immediately after this, they were tested to see how many words they recalled. Table 1 shows the raw results.

Table 1. Data for a t-test

NOISE (group 1)	NO NOISE (group 2)
5.00	15.00
10.00	9.00
6.00	16.00
6.00	15.00
7.00	16.00
3.00	18.00
6.00	17.00
9.00	13.00
5.00	11.00
10.00	12.00
11.00	13.00
9.00	11.00
Σ=87 X=7.3 SD=2.5	Σ=166 X=13.8 SD=2.8







Effect size of t-test

M1 – M2: 7.3 – 13.8 = –6.5

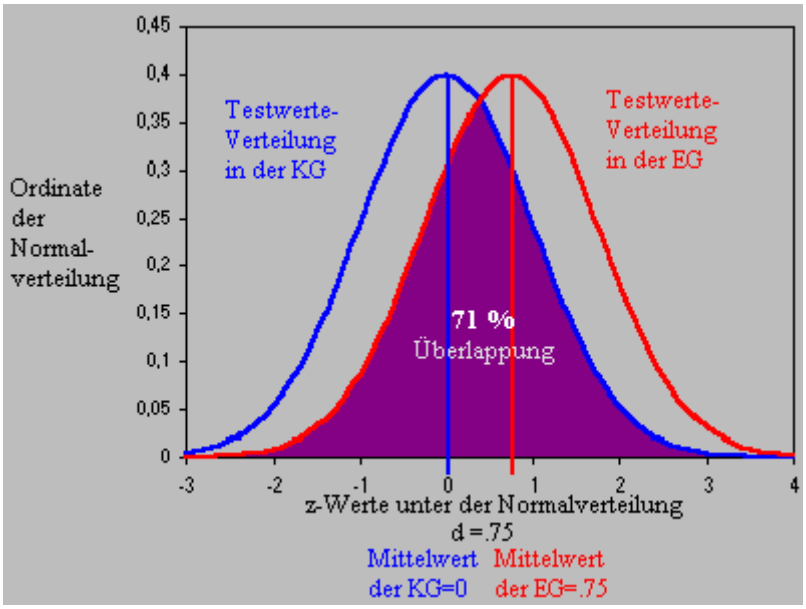
Mean SD: SD 1 – SD 2 / 2 = 2.5 × 2.8 / 2 = 2.65

$$d = \frac{M1 - M2}{\text{Mean SD}}$$

$$d = \frac{6.5}{2.65} = = 2.45$$

Table 1. Effect size of t-test

Effect size	d	Percentage of overlap
Small	0.2	85
Medium	0.5	67
Large	0.8	53



Prerequisites of the t-test

1. For small samples (i.e. N < 15), the data must be normally distributed.
2. The data must be continuous (though the t-test is also often used for ordinal data.
3. Homogeneity-of-variance (Levene's test)

Example

The word *that* is ambiguous. Among other things, it can be a demonstrative (e.g. *That's my car*) and a complementizer (e.g. *I regret that I didn't go*). The two categories tend to occur in different contexts. At the beginning of a sentence, *that* tends to be a demonstrative and is only rarely a complementizer, but after verbs *that* is usually a complementizer and only rarely a demonstrative. A psycholinguist wants to know if the different frequencies of the demonstrative and complementizer affect the interpretation of *that* in different contexts. In order to test this hypothesis, he measures the reading times (i.e. the time it takes to move from one word to another while reading a sentence) of the complementizer and the demonstrative after verbs that frequently occur with sentential complements but may also occur with an NP including a demonstrative (e.g. *find, know, regret*). Since the complementizer is more frequent in this context than the demonstrative, it is reasonable to assume that the complementizer has shorter reading times than the demonstrative. Twenty subjects were tested: 10 subjects listened to sentences in which the verbs were followed by a *that*-clause, and 10 subjects listened to sentences that were followed by an NP including a *that*-determiner. Table 1 shows the reading times.

(1) Peter know that she was coming.
(2) Peter know that guy.

Group 1	**That-clauses**	**Group 2**	**That-NPs**
1	500	11	392
2	513	12	445
3	300	13	271
4	561	14	523
5	483	15	421
6	502	16	489
7	539	17	501
8	467	18	388
9	420	19	411
10	480	20	467

Test bei unabhängigen Stichproben

		Levene-Test		T-Test für die Mittelwertgleichheit						
		F	Sig	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
Score	Varianzen sind gleich	,068	,797	1,401	18	,178	45,70000	32,61903	-22,83004	114,23004
	Varianzen sind nicht gleich			1,401	18,00	,178	45,70000	32,61903	-22,83012	114,23012

Test bei unabhängigen Stichproben

		Levene-Test		T-Test für die Mittelwertgleichheit						
		F	Sig	T	df	Sig. (2-seitig)	Mittlere Differenz	Standard-fehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
Score	Varianzen sind gleich	,997	,333	2,229	16	,040	47,55556	21,33500	2,32738	92,78373
	Varianzen sind nicht gleich			2,229	15,504	,041	47,55556	21,33500	2,20960	92,90151

Paired t-test

Exercise

Analyze the same data using a paired t-Test.

Statistik bei gepaarten Stichproben

		Mittelwert	N	Standardabweichung	Standardfehler des Mittelwertes
Paaren 1	Condition1	476,5000	10	73,08329	23,11096
	Condition2	430,8000	10	72,79316	23,01922

Korrelationen bei gepaarten Stichproben

		N	Korrelation	Signifikanz
Paaren 1	Condition1 & Condition2	10	,899	,000

Test bei gepaarten Stichproben

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
		Mittelwert	Standard-abweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere	Obere			
Paaren 1	Condition1 Condition2	45,70000	32,72121	10,34736	22,29265	69,10735	4,417	9	,002

Non-parametric equivalents of the t-test

Non-parametric equivalents of the t-tests must be used:

1. when we have ordinal data
2. when we have interval data, but the samples are very small
3. when we have interval data, but the distribution is skewed (i.e. not normal)
4. when we have interval data, but unequal sample sizes.

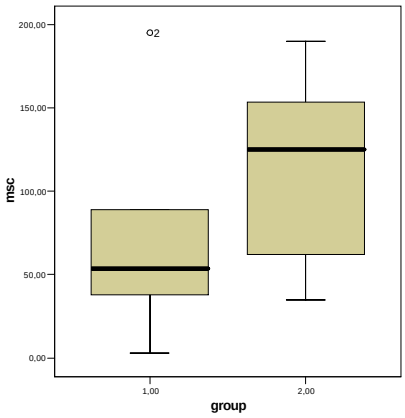
Mann-Whitney

Example:

When children begin to speak, there is often great variation in the pronunciation of particular speech sounds. A researcher wants to find out if a two-year old child pronounces /g/ and /k/ differently or if the two speech sounds are basically pronounced in the same way at this age. In adult language, /g/ and /k/ are primarily distinguished by voice onset time, i.e. VOT. In order to decide if two-year old children pronounce /g/ and /k/ differently, the researcher collects a (very small) corpus of 13 words, six words including /g/ in adult language and seven words including /k/ in adult language. The words were selected such that /g/ and /k/ are surrounded by the same speech sounds (i.e. they occur in the same phonetic environment). For each word, the researcher measured the voice onset time in milliseconds. Are the two speech sounds significantly different in the child's speech?

Table 1. Data for Mann-Whitney

Speech sound	VOT in msc	Ranks
/g/	38	3
/g/	195	13
/g/	56	6
/g/	3	1
/g/	51	4.5
/g/	89	8
/k/	125	9
/k/	73	7
/k/	138	10
/k/	35	2
/k/	51	4.5
/k/	190	12
/k/	169	11



Ränge

	group	N	Mittlerer Rang	Rangsumme
msc	1,00	6	5,92	35,50
	2,00	7	7,93	55,50
	Gesamt	13		

Statistik für Test(b)

	msc
Mann-Whitney-U	14,500
Wilcoxon-W	35,500
Z	**-,930**
Asymptotische Signifikanz (2-seitig)	**,352**
Exakte Signifikanz [2*(1-seitig Sig.)]	,366(a)

a Nicht für Bindungen korrigiert.
b Gruppenvariable: group

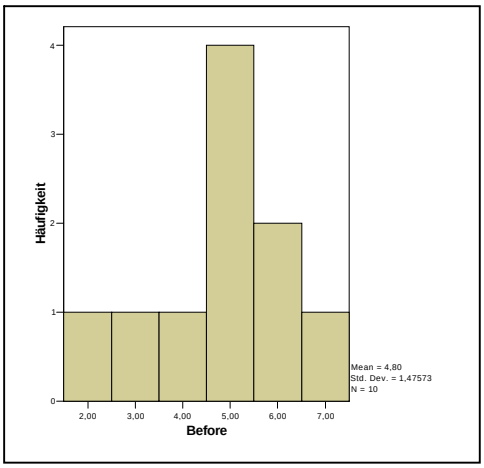
Wilcoxon

Example:

Nurses were asked to rate their sympathy on a scale between 1 and 10 for MS patients before and after talking to them. Table 1 shows the nurses' sympathy scores before and after they have talked to them. Have the scores changed?

Table 1. Data for the Wilcoxon

Before	After
5.00 6.00 2.00 4.00 6.00 7.00 3.00 5.00 5.00 5.00	7.00 6.00 3.00 8.00 7.00 6.00 7.00 8.00 5.00 8.00
Mean: 4.8 SD: 1.48 Median: 5	Mean: 6.5 SD: 1.58 Median: 7



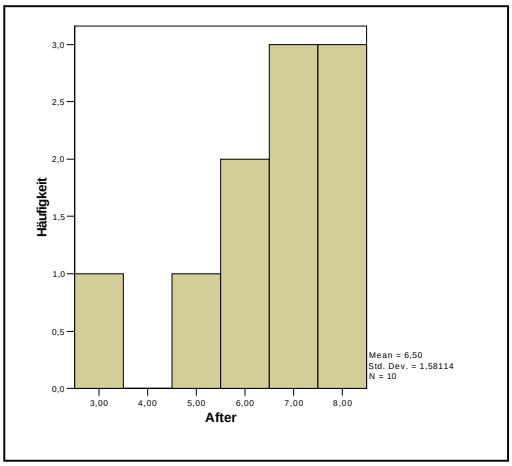


Table 2. Intermediate results of the Wilcoxon

Difference between scores from individual subjects	Ranking of the scores in the left-hand column
–2 0 –1 –4 –1 **+1** –4 –3 0 –3	4 – 2 7.5 2 **2** 7.5 5.5 – 5.5
7 minuses, 1 plus	

Ränge

		N	Mittlerer Rang	Rangsumme
After - Before	Negative Ränge	1[a]	2,00	2,00
	Positive Ränge	7[b]	4,86	34,00
	Bindungen	2[c]		
	Gesamt	10		

a. After < Before

b. After > Before

c. After = Before

Statistik für Test [b]

	After - Before
Z	-2,257[a]
Asymptotische Signifikanz (2-seitig)	,024

a. Basiert auf negativen Rängen.

b. Wilcoxon-Test

Experimental design

The t-test is a parametric test that should only be used for interval data. Many psycholinguistic experiments involve interval data: reaction time studies, eye-tracking studies, etc. Interval data provides more information than ordinal data or simple frequency counts and it can be analyzed by means of the most powerful tests. Therefore psychologists always try to work with interval data. But of course, not all experimental tasks generate interval data. There are tasks in which subjects are asked to make a choice between alternatives (forced choice tasks) and there are tasks in which the experimenter divides the subjects' responses into discrete categories (e.g. correct vs. incorrect).

In such a case, we have to use statistical tests for the analysis of frequency data or ordinal data; but that should always be the last resort. If possible we would like to base our analysis on normally distributed interval data. So what can we do?

We can design the experiment in such a way that frequency or ordinal data approximates interval data. There are several things we can do:

1. We can try to find a reasonable coding scheme that distinguishes more than two categories.
2. We can test subjects multiple times on the same condition.

For instance, in a study on the acquisition of relative clauses, we wanted to know if subject and object relative clause cause different problems for pre-school children. So we designed an experiment in which each child had to repeat subject and object relative clauses. We then assigned one of three possible scores to each response:

1.0 score for correct repetition
0.5 for partially correct repetitions
0.0 for incorrect repetitions

Each child was given 4 subject relative clauses and 4 object relative clauses (presented in random order and separated by filler sentences). Thus, four each of the two categories (SUBJ and OBJ) a child could obtain a score between zero and four on a scale with 9 possible outcomes:

0.0 – 0.5 – 1.0 – 1.5 – 2.0 – 2.5 – 3.0 – 3.5 – 4.0

This procedure yields a scale that approximates interval data and allows you to calculate a mean score for each child. For instance, assume a child obtained the following 4 scores:

1 – 0 – 0.5 – 1 = 2.5/4 = 0.625 (mean score on each trail)

In this way, you get a mean score for each category for each child. Moreover, if you look at the data, you find that they are approximately normally distributed.

Exercise:

First, determine if the A and P relative clauses are normally distributed in the Diessel&Tomasello (2005) study. Second, determine if the children of this study performed differently on A and P relative clauses.

One-sample t-test

Example:

Previous research has shown that English-speaking children have an MLU of 3.0 at age 3;2. A researcher wants to know if SLI children (i.e. children with a specific language impairment) have a lower (or higher MLU) at this age. We know that SLI children have difficulties in processing morphological units, but it is unclear if their MLUs are lower than in normally developing children. In order to test this hypothesis, the researcher collected data from 24 SLI children aged 3;1 to 3;3 and determined the MLU for each child.

Table 1. MLU of 24 SLI children aged 3;1-3;3

Child	MLU
1	2,7
2	3,0
3	2,8
4	2,9
5	3,1
6	3,0
7	3,1
8	2,5
9	3,2
10	3,1
11	2,9
12	2,9
13	2,8
14	3,1
15	3,2
16	2,4
17	2,3
18	2,8
19	3,1
20	2,5
21	2,7
22	2,9
23	2,9
24	3,0

Univariate Statistiken (confidence intervals)

			Statistik	Standardfehler
Score	Mittelwert		2,8708	,05089
	95% Konfidenzintervall des Mittelwerts	Untergrenze	**2,7656**	
		Obergrenze	**2,9761**	
	5% getrimmtes Mittel		2,8833	
	Median		2,9000	
	Varianz		,062	
	Standardabweichung		,24931	
	Minimum		2,30	
	Maximum		3,20	
	Spannweite		,90	
	Interquartilbereich		,38	
	Schiefe		-,812	,472
	Kurtosis		-,029	,918

Statistik bei einer Stichprobe

	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Score	24	2,8708	,24931	,05089

Test bei einer Sichprobe

	Testwert = 3.1					
					95% Konfidenzintervall der Differenz	
	T	df	Sig. (2-seitig)	Mittlere Differenz	Untere	Obere
Score	-4,503	23	,000	-,22917	-,3344	-,1239

ANOVA

Table 1. Overview of one-way tests

	Parametric	Non-parametric
Between subjects	Independent ANOVA	Kruskal Wallis
Within subjects	Repeated measures ANOVA	Friedman's ANOVA

Independent one-way ANOVA

Example

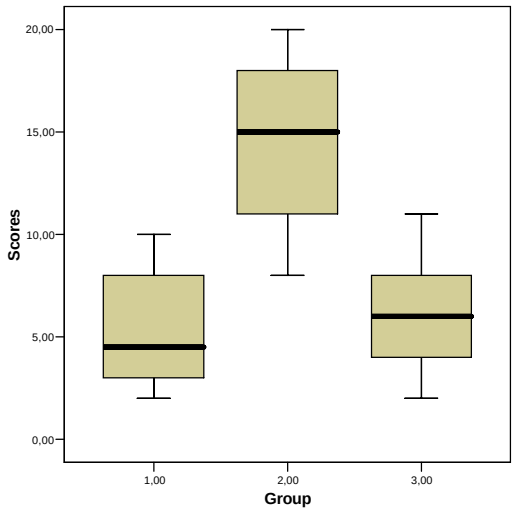
A child language researcher wants to know if semantic factors can influence children's understanding of passive sentences or if their understanding is solely based on structural features. Specifically, he wants to find out if reversible passive sentences are more difficult to understand than irreversible passive sentences. Reversible passive sentences are sentences in which the two NPs of a transitive sentence are equally likely to function as agent, whereas irreversible passive sentences are sentences in which one of the two NPs is more likely to serve as the agent:

(1) Peter was seen by Mary.
(2) The car was seen by Mary.

Example (1) is reversible because *Peter* could also be the agent (i.e. the 'seer'); example (2) is irreversible because *the car* could not really serve as the agent of *see*. In order to test whether semantic reversibility influences children's understanding of passive sentences, the researcher asked children to act-out reversible and irreversible passive sentences. As a control, he also collected data on (transitive) active sentences. Each child was exposed to a list of 20 test items from one condition plus filler sentences (i.e. each child was only tested under one condition). Table 1 shows the number of errors.

Table 1. Number of errors

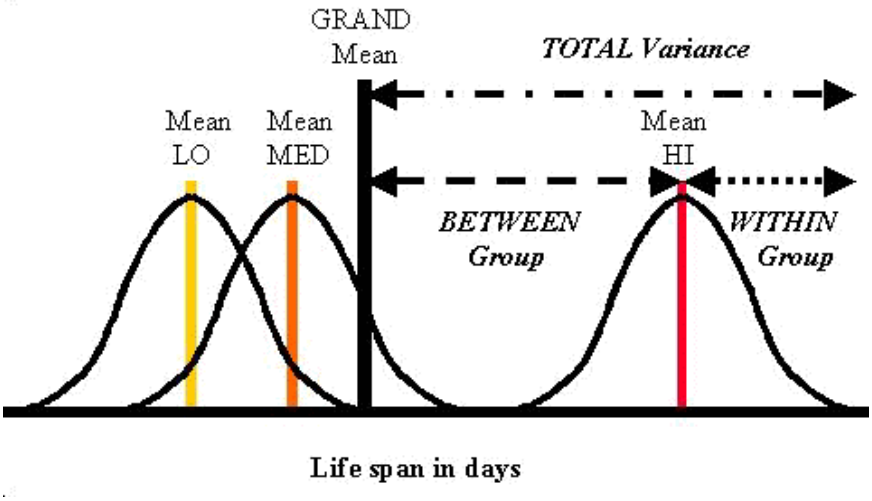
Active	3	5	3	2	4	6	9	3	8	10
Reversible Passives	20	15	14	15	17	10	8	11	18	19
Irreversible Passives	2	8	5	4	4	7	9	4	7	11



The ANOVA is based on the following measures:

1. mean of each group
2. the grand mean, i.e. the mean of the three means
3. the variation within each group
4. the deviation of each group mean from the grand mean
5. the total variance, which is the combined variability of within-group variation and between-group variation.

The ANOVA calculates a test statistics, i.e. the F-value, which represents something like the average of (1) the within groups variance, (2) the between groups variance, and (3) the total variance of the three group.



You can also think of the F-value as the ratio of the between-groups variance and the within-groups variance:

$$\text{F ratio} = \frac{\text{Between-groups variance}}{\text{Within groups variance}}$$

Post hoc tests

(i) Planned comparisons:

1. active – reversible passive
2. active – irreversible passive
3. reversible passive – irreversible passive

1 test: .05 / 1 = .050
2 tests .05 / 2 = .025
3 tests .05 / 3 = .016

(ii) Post hoc tests

(1) Tukey Honstly Significance Difference (HSD)
(2) Least Significant Difference (LSD)

ONEWAY deskriptive Statistiken

	N	Mittelwert	Standardabweichung	Standardfehler	95%-Konfidenzintervall für den Mittelwert		Minimum	Maximum
					Untergrenze	Obergrenze		
1,00	10	5,3000	2,83039	,89505	3,2753	7,3247	2,00	10,00
2,00	10	14,7000	4,00139	1,26535	11,8376	17,5624	8,00	20,00
3,00	10	6,1000	2,76687	,87496	4,1207	8,0793	2,00	11,00
Total	30	8,7000	5,34435	,97574	6,7044	10,6956	2,00	20,00

ONEWAY ANOVA

Scores

	Quadratsumme	df	Mittel der Quadrate	F	Signifikanz
Zwischen den Gruppen	543,200	2	271,600	25,722	,000
Innerhalb der Gruppen	285,100	27	10,559		
Gesamt	828,300	29			

Mehrfachvergleiche

Abhängige Variable: Scores

	(I) Group	(J) Group	Mittlere Differenz (I-J)	Standardfehler	Signifikanz	95%-Konfidenzintervall	
						Untergrenze	Obergrenze
Tukey-HSD	1,00	2,00	-9,40000*	1,45322	,000	-13,0031	-5,7969
		3,00	-,80000	1,45322	,847	-4,4031	2,8031
	2,00	1,00	9,40000*	1,45322	,000	5,7969	13,0031
		3,00	8,60000*	1,45322	,000	4,9969	12,2031
	3,00	1,00	,80000	1,45322	,847	-2,8031	4,4031
		2,00	-8,60000*	1,45322	,000	-12,2031	-4,9969
LSD	1,00	2,00	-9,40000*	1,45322	,000	-12,3818	-6,4182
		3,00	-,80000	1,45322	,587	-3,7818	2,1818
	2,00	1,00	9,40000*	1,45322	,000	6,4182	12,3818
		3,00	8,60000*	1,45322	,000	5,6182	11,5818
	3,00	1,00	,80000	1,45322	,587	-2,1818	3,7818
		2,00	-8,60000*	1,45322	,000	-11,5818	-5,6182

*. Die mittlere Differenz ist auf der Stufe .05 signifikant.

Zusammenhangsmaße

	Eta	Eta-Quadrat
Scores * Group	,810	,656

Kruskal-Wallis

Beispiel

Ein Linguist möchte wissen, ob sich die Länge von vorangestellten Adverbialsätzen mit dem Diskurstyp verändert. Drei Diskurstypen werden untersucht: (1) informeller mündlicher Diskurs, (2) akademischer mündlicher Diskurs, (3) Verkaufsgespräch. Um diese Frage zu beantworten, wertet der Linguist Transkriptionen von 15 Personen aus: 4 akademische Diskurse, 7 informelle Diskurse, und 4 Verkaufsgespräche.

Gesprächstype	Durchschnittliche Länge des ADV-Satzes
Akademisch	13
Akademisch	15
Akademisch	11
Akademisch	12
Informell	4
Informell	4
Informell	6
Informell	2
Informell	5
Informell	3
Informell	4
Verkaufsgespräch	9
Verkaufsgespräch	10
Verkaufsgespräch	10
Verkaufsgespräch	5

Ränge

	Group	N	Mittlerer Rang
Length	1,00	4	13,50
	2,00	7	4,21
	3,00	4	9,13
	Gesamt	15	

Statistik für Test [a,b]

	Length
Chi-Quadrat	11,442
df	2
Asymptotische Signifikanz	,003

a. Kruskal-Wallis-Test

b. Gruppenvariable: Group

Statistik für Test[b]

	Length
Mann-Whitney-U	,000
Wilcoxon-W	28,000
Z	-2,670
Asymptotische Signifikanz (2-seitig)	,008
Exakte Signifikanz [2*(1-seitig Sig.)]	,006[a]

a. Nicht für Bindungen korrigiert.

b. Gruppenvariable: Group

Akademisch vs. Informell

Statistik für Test(b)

	Length
Mann-Whitney-U	,000
Wilcoxon-W	10,000
Z	-2,323
Asymptotische Signifikanz (2-seitig)	,020
Exakte Signifikanz [2*(1-seitig Sig.)]	,029(a)

a Nicht für Bindungen korrigiert.

b Gruppenvariable: Group

Akademisch vs. Verkaufsgespräch

Statistik für Test[b]

	Length
Mann-Whitney-U	1,500
Wilcoxon-W	29,500
Z	-2,395
Asymptotische Signifikanz (2-seitig)	,017
Exakte Signifikanz [2*(1-seitig Sig.)]	,012[a]

a. Nicht für Bindungen korrigiert.

b. Gruppenvariable: Group

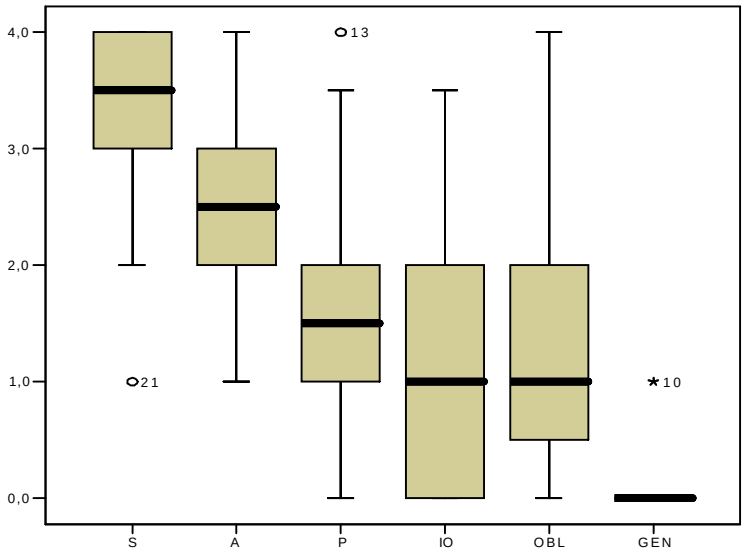
Informell vs. Verkaufsgespräch

1. Akademisch – Informell $z=2.67$; $p<0.006$
2. Akademisch – Verkaufsgespräch $z=2.32$; $p<0.020$
3. Informell – Verkaufsgespräch $z=2.40$; $p<0.012$

Repeated measures ANOVA

Beispiel

In einer Studie zum Erwerb von englischen und deutschen Relativsätzen wurden sechs verschiedenen Relativsatztypen untersucht: S, A, P, IO, OBL, GEN (Diessel and Tomasello 2005). Insgesamt wurden in der englischen Studie 21 Kinder untersucht (in der deutschen Studie 24). Um zu testen, wie die Kinder mit den verschiedenen Relativsatztypen klar kommen, mussten sie die Sätze nachsprechen. Insgesamt, mussten alle Kinder alle 6 Relativsatztypen jeweils 4 Mal nachsprechen. Die Fehler wurden nach einem vorher spezifizierten System kodiert und anschließend (statistisch) analysiert. Ermitteln sie, ob es sich die Fehlerzahl für die verschiedenen Relativsatztypen unterscheidet.



Deskriptive Statistiken

	Mittelwert	Standardabweichung	N
S	3,310	,8136	21
A	2,381	,8501	21
P	1,619	1,0357	21
IO	1,238	1,2310	21
OBL	1,262	1,0562	21

Mauchly-Test auf Sphärizität(b)

Innersubjekteffekt	Mauchly-W	Approximiertes Chi-Quadrat	df	Signifikanz	Epsilon(a)		
					Greenhouse-Geisser	Huynh-Feldt	Untergrenze
relativ	,447	14,840	9	**,097**	,768	,923	,250

Prüft die Nullhypothese, daß sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a Kann zum Korrigieren der Freiheitsgrade für die gemittelten Signifikanztests verwendet werden. In der Tabelle mit den Tests der Effekte innerhalb der Subjekte werden korrigierte Tests angezeigt.

b Design: Intercept Innersubjekt-Design: relativ

Tests der Innersubjekteffekte

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz	Partielles Eta-Quadrat
relativ	**Sphärizität angenommen**	**65,586**	**4**	**16,396**	**32,060**	**,000**	**,616**
	Greenhouse-Geisser	65,586	3,071	21,359	32,060	,000	,616
	Huynh-Feldt	65,586	3,691	17,770	32,060	,000	,616
	Untergrenze	65,586	1,000	65,586	32,060	,000	,616
Fehler(relativ)	Sphärizität angenommen	40,914	**80**	,511			
	Greenhouse-Geisser	40,914	61,413	,666			
	Huynh-Feldt	40,914	73,817	,554			
	Untergrenze	40,914	20,000	2,046			

Schätzungen

relativ	Mittelwert	Standardfehler	95% Konfidenzintervall	
			Untergrenze	Obergrenze
1	3,310	,178	2,939	3,680
2	2,381	,186	1,994	2,768
3	1,619	,226	1,148	2,090
4	1,238	,269	,678	1,798
5	1,262	,230	,781	1,743

Paarweise Vergleiche

(I) relativ	(J) relativ	Mittlere Differenz (I-J)	Standardfehler	Signifikanz(a)	95% Konfidenzintervall für die Differenz(a)	
					Untergrenze	Obergrenze
1	2	,929(*)	,152	**,000**	,451	1,406
	3	1,690(*)	,184	,000	1,110	2,271
	4	2,071(*)	,213	,000	1,398	2,745
	5	2,048(*)	,195	,000	1,433	2,662
2	1	-,929(*)	,152	,000	-1,406	-,451
	3	,762	,253	**,068**	-,035	1,559
	4	1,143(*)	,244	,001	,372	1,913
	5	1,119(*)	,253	,003	,320	1,918
3	1	-1,690(*)	,184	,000	-2,271	-1,110
	2	-,762	,253	,068	-1,559	,035
	4	,381	,269	**1,000**	-,468	1,230
	5	,357	,221	1,000	-,341	1,055
4	1	-2,071(*)	,213	,000	-2,745	-1,398
	2	-1,143(*)	,244	,001	-1,913	-,372
	3	-,381	,269	1,000	-1,230	,468
	5	-,024	,194	**1,000**	-,634	,587
5	1	-2,048(*)	,195	,000	-2,662	-1,433
	2	-1,119(*)	,253	,003	-1,918	-,320
	3	-,357	,221	1,000	-1,055	,341
	4	,024	,194	1,000	-,587	,634

Basiert auf den geschätzten Randmitteln

* Die mittlere Differenz ist auf dem Niveau ,05 signifikant

a Anpassung für Mehrfachvergleiche: Bonferroni.

Test bei gepaarten Stichproben

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz Untere	Obere			
Paaren 1	S - A	,9286	,6944	,1515	,6125	1,2446	6,128	20	**,000**
Paaren 2	A - P	,7619	1,1578	,2527	,2349	1,2889	3,016	20	**,007**
Paaren 3	P - IO	,3810	1,2339	,2693	-,1807	,9426	1,415	20	**,173**
Paaren 4	IO - OBL	-,0238	,8871	,1936	-,4276	,3800	-,123	20	**,903**

Friedman's ANOVA

Beispiel

Ein Linguist möchte wissen, ob sich die Stellung von Adverbialsätzen mit ihrer Bedeutung verändert. Dafür untersucht er die Adverbialsätze von 15 Personen in Transkriptionen gesprochener Sprache. Die Adverbialsätze werden in drei große semantische Klassen eingeteilt: (1) Konditionalsätze, (2) Temporalsätze, (3) Kausalsätze. Für jede Klasse wurde ermittelt, wie häufig ein Sprecher den Satz jeweils vor und nach dem Hauptsatz verwendet hat. Eingeschobene Adverbialsätze, die ohnehin nur sehr selten vorkommen, wurden ignoriert.

	Konditional		Kausal		Temporal	
	Vor	Nach	Vor	Nach	Vor	Nach
1	42,70	57,30	12,50	87,50	28,10	71,90
2	72,70	27,30	9,10	90,90	36,60	63,40
3	75,00	25,00	,00	100,00	26,70	73,30
4	63,60	36,60	20,00	80,00	10,30	89,70
5	88,90	11,10	,00	100,00	41,70	58,30
6	70,00	30,00	,00	100,00	36,00	64,00
7	72,70	27,30	,00	100,00	56,00	44,00
8	60,00	40,00	37,50	62,50	28,60	71,40
9	66,70	33,30	,00	100,00	52,60	47,40
10	64,00	36,00	,00	100,00	15,40	84,60
11	57,10	42,90	10,00	90,00	10,90	89,10
12	45,50	54,50	23,10	76,90	26,20	73,80
13	70,60	29,40	33,30	66,70	66,60	33,40
14	100,00	,00	,00	100,00	41,70	58,30
15	55,50	45,50	7,70	92,30	34,30	65,70

Ränge

	Mittlerer Rang
Conditional	3,00
Causal	1,13
Temporal	1,87

Statistik für Test[a]

N	15
Chi-Quadrat	26,533
df	2
Asymptotische Signifikanz	,000

a. Friedman-Test

Statistik für Test(b)

	Causal - Conditional
Z	-3,408(a)
Asymptotische Signifikanz (2-seitig)	,001

a Basiert auf positiven Rängen.
b Wilcoxon-Test

Conditional-Causal

Statistik für Test(b)

	Temporal - Conditional
Z	-3,408(a)
Asymptotische Signifikanz (2-seitig)	,001

a Basiert auf positiven Rängen.
b Wilcoxon-Test

Conditional-Temporal

Statistik für Test[b]

	Temporal - Causal
Z	-3,011[a]
Asymptotische Signifikanz (2-seitig)	,003

a. Basiert auf negativen Rängen.

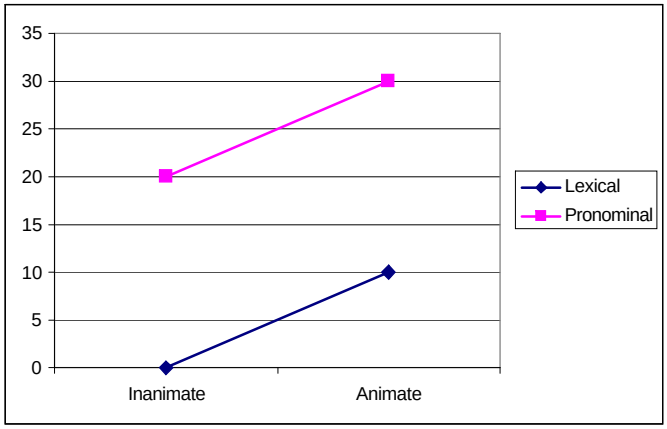
b. Wilcoxon-Test

Causal-Temporal

Multifactorial ANOVA

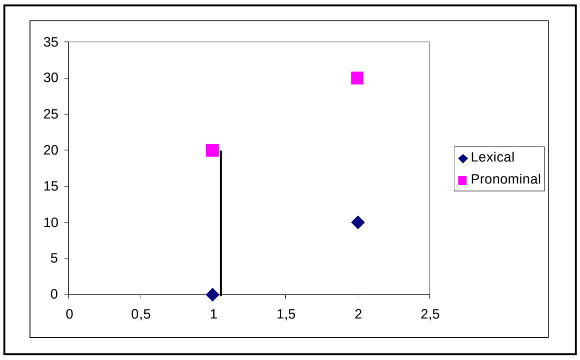
In einem faktorial Design untersucht man einmal den Effekt jeder einzelnen IV (main effects) und zum anderen die Interaktionen zwischen den beiden IVs (interaction). Außerdem werden für jede IV die Effekte berechnet, die erzielt werden, wenn man die anderen IV auf einem bestimmten Level konstant hält (simple main effects).

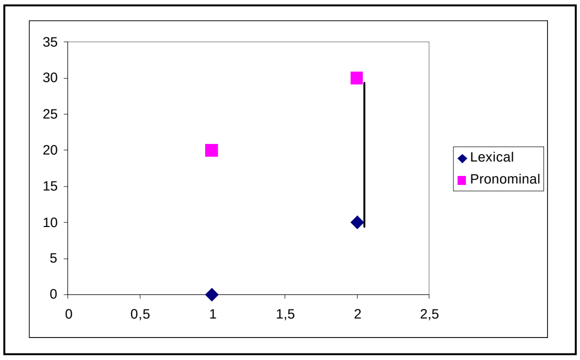
	Animacy		
Type of NP	Inanimate	Animate	average
Lexical	00	10	5
Pronominal	20	30	25
average	10	20	15

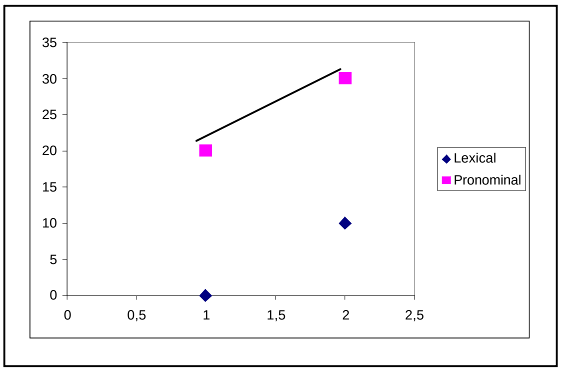


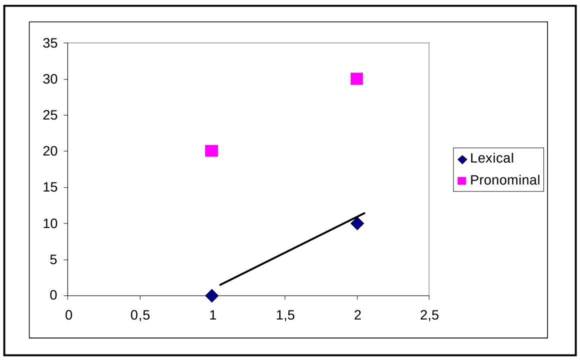
- Main Effect beschreibt den Unterschied zwischen den beiden Levels der beiden IVs:
    1. Main Effect (NP type): 25 – 5 = 20
    2. Main Effect (Syntactic Function): 20 – 10 = 10
- Simple Main Effect beschreibt den Effekt eines bestimmten Faktors, auf einem bestimmten Level des anderen Faktors. Bei einem 2×2 Design gibt es 4 Simple Main Effects:
    1. Wenn wir uns nur die Subjekte anschauen, dann sind pronominale NPs um 20 Punkte häufiger als lexikalische NPs: 20 – 0 = 20.
    2. Wenn wir uns nur die Objekte anschauen, dann sind pronominale NPs auch um 20 Punkte häufiger als lexikalische NPs: 30 – 10 = 20.
    3. Wenn wir uns nur Lexikalische NPs anschauen, dann sind Objekte um 10 Punkte häufiger als Subjekte: 30 – 20 = 10.
    4. Wenn wir uns nur pronominale NPs anschauen, dann sind Objekte auch um 10 Punkte häufiger als Subjekte: 10 – 0 = 10.
- Interaction: Wenn die Simple Main Effects auf den beiden Levels der anderen IV konstant bleiben, dann gibt es zwischen den beiden IVs keine Interaktion.

Graphische Darstellung der Simple Main Effects

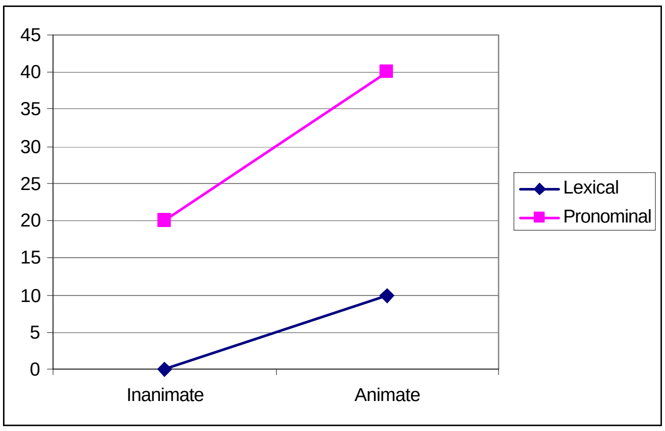




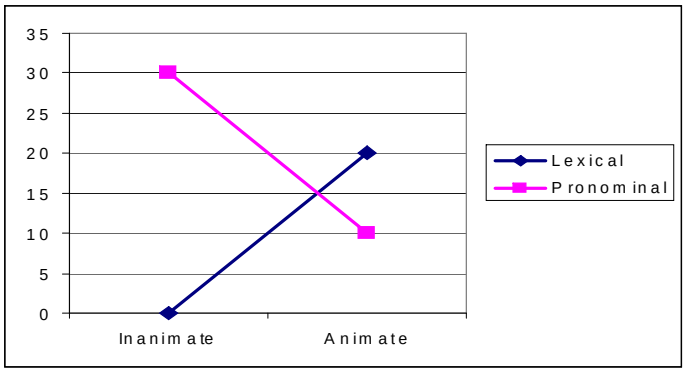




	Animacy		
Type of NP	Inanimate	Animate	average
Lexical	00	10	5
Pronominal	20	40	25
average	10	25	17.5



	Animacy		
Type of NP	Inanimate	Animate	average
Lexical	00	20	10
Pronominal	30	10	20
average	15	15	15



Analyse der Varianz

Das Zusammenwirken von mehr als 3 IVs lässt sich nur schwer analysieren:

1 Varianz, die von Faktor 1 herrührt
2 Varianz, die von Faktor 2 herrührt
3 Varianz, die von der Interaktion zwischen 1 und 2 herrührt
4 Varianz, die von der Interaktion zwischen 1 und 3 herrührt
5 Varianz, die von der Interaktion zwischen 2 und 3 herrührt
6 Varianz, die von der Interaktion zwischen 1, 2 und 3 herrührt
7 Varianz die vom Sampling herrührt

Independent factorial ANOVA

Beispiel

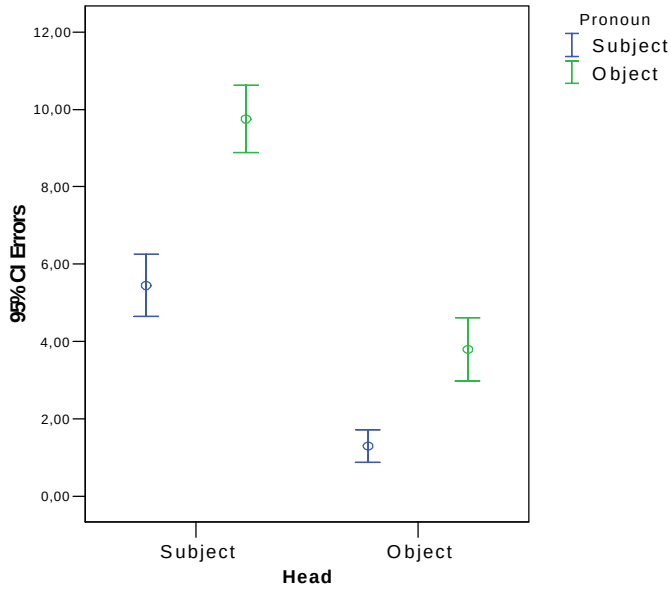
Kinder haben große Schwierigkeiten mit Relativsätzen. Dabei bereiten unterschiedliche Relativsatztypen unterschiedlich große Schwierigkeiten. Ein Spracherwerbsforscher möchte herausfinden, welche Faktoren, die größten Schwierigkeiten bereiten. Zu diesem Zweck werden 40 Kinder untersucht, die die Bedeutung von 4 verschiedene Typen von Relativsätzen mit Spielfiguren darstellen sollen: (1) Subjekt-Subjekt-Relativsätze, (2) Subjekt-Objekt-Relativsätze, (3) Objekt-Subjekt-Relativsätze, (4) Objekt-Objekt-Relativsätze.

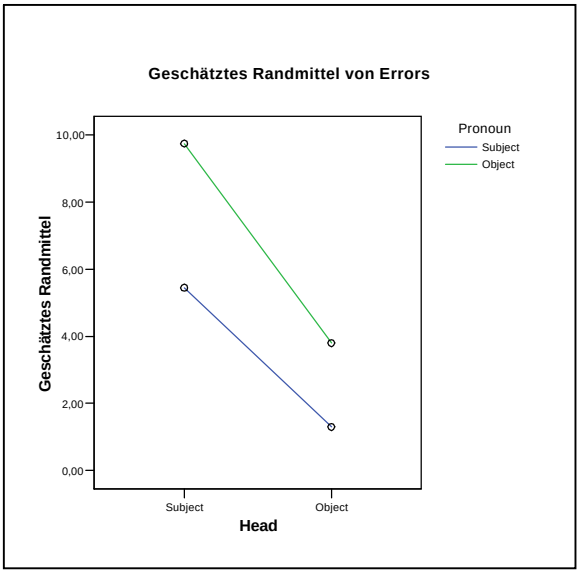
(1) Der Mann, der den Jungen gesehen hat, umarmt die Frau. SS
(2) Der Mann, den der Junge gesehen hat, umarmt die Frau. SO
(3) Der Mann umarmt die Frau, die der Junge gesehen hat. OS
(4) Der Mann umarmt die Frau, die den Jungen gesehen hat. OO

In der Literatur gibt es zwei Hypothesen, die in diesem Zusammenhang von Interesse sind: (a) die Non-Interruption-Hypothesis, die besagt, dass Relativsätze, die dem Hauptsatz folgen, weniger Schwierigkeiten bereiten, als Relativsätze, die in den Hauptsatz eingebetet sind (OS+OO einfacher als SS+SO); (b) die Parallel-Function-Hypothesis, die besagt, dass Relativsätze, in denen der Kopf des Relativsatzes und das Relativpronomen die gleiche syntaktische Funktion haben, leichter zu verstehen sind als Relativsätze, in denen der Kopf des Relativsatzes und das Relativpronomen unterschiedliche syntaktische Funktionen haben (SS+OO einfacher als SO+OS). Die folgende Tabelle zeigt die Anzahl der Fehler, die die Kinder bei der Darstellung der vier Relativsatztypen gemacht haben. Die 40 Kinder wurden in vier Gruppen aufgeteilt und jedes Kind hat jeweils 15 Relativsätze eines bestimmten Typs nachgespielt. Zusammen mit den Relativsätzen mussten die Kinder 30 Distraktoren nachspielen. Die folgende Tabelle zeigt die Anzahl der Fehler.

	Subjekt-Kopf	Objekt-Kopf
Subjekt-Pronomen	4,5	0
	4,5	1
	6	1
	5	1
	5,5	1,5
	4,5	1,5
	4	1,5
	6,5	2
	7	2
	7	1,5
Objekt-Pronomen	11	1
	10	3,5
	11	5,5
	9	3,5
	10,5	4
	10,5	4
	9,5	4
	7	4
	9	4
	10	4,5

SS 1:1	OS 2:1
SO 1:2	OO 2:2





Deskriptive Statistiken

Abhängige Variable: Errors

Head	Pronoun	Mittelwert	Standardabweichung	N
Subject	Subject	5,4500	1,11679	10
	Object	9,7500	1,20761	10
	Gesamt	7,6000	2,47939	20
Object	Subject	1,3000	,58689	10
	Object	3,8000	1,13529	10
	Gesamt	2,5500	1,55513	20
Gesamt	Subject	3,3750	2,29917	20
	Object	6,7750	3,25849	20
	Gesamt	5,0750	3,27295	40

Tests der Zwischensubjekteffekte

Abhängige Variable: Errors

Quelle	Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz	Partielles Eta-Quadrat
Korrigiertes Modell	378,725(a)	3	126,242	116,382	,000	,907
Konstanter Term	1030,225	1	1030,225	949,759	,000	,963
Head	**255,025**	**1**	**255,025**	**235,106**	**,000**	**,867**
Pronoun	**115,600**	**1**	**115,600**	**106,571**	**,000**	**,747**
Head * Pronoun	**8,100**	**1**	**8,100**	**7,467**	**,010**	**,172**
Fehler	39,050	36	1,085			
Gesamt	1448,000	40				
Korrigierte Gesamtvariation	417,775	39				

a R-Quadrat = ,907 (korrigiertes R-Quadrat = ,899)

Planned comparisons:

Test bei unabhängigen Stichproben[a]

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
									95% Konfidenzintervall der Differenz	
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	Untere	Obere
Errors	Varianzen sind gleich	,031	,862	-8,267	18	,000	-4,30000	,52015	-5,39279	-3,20721
	Varianzen sind nicht gleich			-8,267	17,891	,000	-4,30000	,52015	-5,39327	-3,20673

a. Head = Subject

Test bei unabhängigen Stichproben[a]

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
									95% Konfidenzintervall der Differenz	
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	Untere	Obere
Errors	Varianzen sind gleich	,637	,435	-6,186	18	,000	-2,50000	,40415	-3,34908	-1,65092
	Varianzen sind nicht gleich			-6,186	13,490	,000	-2,50000	,40415	-3,36989	-1,63011

a. Head = Object

Test bei unabhängigen Stichproben[a]

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
									95% Konfidenzintervall der Differenz	
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	Untere	Obere
Errors	Varianzen sind gleich	6,959	,017	10,402	18	,000	4,15000	,39896	3,31182	4,98818
	Varianzen sind nicht gleich			10,402	13,619	,000	4,15000	,39896	3,29207	5,00793

a. Pronoun = Subject

Test bei unabhängigen Stichproben[a]

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
									95% Konfidenzintervall der Differenz	
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	Untere	Obere
Errors	Varianzen sind gleich	,363	,554	11,352	18	,000	5,95000	,52414	4,84882	7,05118
	Varianzen sind nicht gleich			11,352	17,932	,000	5,95000	,52414	4,84852	7,05148

a. Pronoun = Object

Paired factorial ANOVA

Beispiel

Das gleiche Experiment kann natürlich auch mit einem within-subjects Design durchgeführt werden. Dazu müssen die Daten jedoch anders eingegeben werden.

	SS	SO	OS	OO
1	4,5	11	0	1
2	4,5	10	1	3,5
3	6	11	1	5,5
4	5	9	1	3,5
5	5,5	10,5	1,5	4
6	4,5	10,5	1,5	4
7	4	9,5	1,5	4
8	6,5	7	2	4
9	7	9	2	4
10	7	10	1,5	4,5

Mauchly-Test auf Sphärizität[b]

Maß: MASS_1

Innersubjekteffekt	Mauchly-W	Approximiertes Chi-Quadrat	df	Signifikanz	Epsilon[a] Greenhouse-Geisser	Huynh-Feldt	Untergrenze
head	1,000	,000	0	.	1,000	1,000	1,000
pro	1,000	,000	0	.	1,000	1,000	1,000
head * pro	1,000	,000	0	.	1,000	1,000	1,000

Prüft die Nullhypothese, daß sich die Fehlerkovarianz-Matrix der orthonormalisierten transformierten abhängigen Variablen proportional zur Einheitsmatrix verhält.

a. Kann zum Korrigieren der Freiheitsgrade für die gemittelten Signifikanztests verwendet werden. In der Tabelle mit den Tests der Effekte innerhalb der Subjekte werden korrigierte Tests angezeigt.

b.
Design: Intercept
Innersubjekt-Design: head+pro+head*pro

Tests der Innersubjekteffekte

Maß: MASS_1

Quelle		Quadratsumme vom Typ III	df	Mittel der Quadrate	F	Signifikanz	Partielles Eta-Quadrat
head	Sphärizität angenommen	255,025	1	255,025	283,361	,000	,969
	Greenhouse-Geisser	255,025	1,000	255,025	283,361	,000	,969
	Huynh-Feldt	255,025	1,000	255,025	283,361	,000	,969
	Untergrenze	255,025	1,000	255,025	283,361	,000	,969
Fehler(head)	Sphärizität angenommen	8,100	9	,900			
	Greenhouse-Geisser	8,100	9,000	,900			
	Huynh-Feldt	8,100	9,000	,900			
	Untergrenze	8,100	9,000	,900			
pro	Sphärizität angenommen	115,600	1	115,600	101,255	,000	,918
	Greenhouse-Geisser	115,600	1,000	115,600	101,255	,000	,918
	Huynh-Feldt	115,600	1,000	115,600	101,255	,000	,918
	Untergrenze	115,600	1,000	115,600	101,255	,000	,918
Fehler(pro)	Sphärizität angenommen	10,275	9	1,142			
	Greenhouse-Geisser	10,275	9,000	1,142			
	Huynh-Feldt	10,275	9,000	1,142			
	Untergrenze	10,275	9,000	1,142			
head * pro	Sphärizität angenommen	8,100	1	8,100	7,458	,023	,453
	Greenhouse-Geisser	8,100	1,000	8,100	7,458	,023	,453
	Huynh-Feldt	8,100	1,000	8,100	7,458	,023	,453
	Untergrenze	8,100	1,000	8,100	7,458	,023	,453
Fehler(head*pro)	Sphärizität angenommen	9,775	9	1,086			
	Greenhouse-Geisser	9,775	9,000	1,086			
	Huynh-Feldt	9,775	9,000	1,086			
	Untergrenze	9,775	9,000	1,086			

Main effect head: $F(1,9) = 283, 36; p < 0.001$
Main effect pro: $F(1,9) = 101,26; p < 0.001$
Interaction: $F(1,9) = 7,46; p < 0.023$

Test bei gepaarten Stichproben

		Gepaarte Differenzen					T	df	Sig. (2-seitig)
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz Untere	Obere			
Paaren 1	SS - SO	-4,30000	1,91775	,60645	-5,67188	-2,92812	-7,090	9	,000
Paaren 2	OS - OO	-2,50000	,88192	,27889	-3,13089	-1,86911	-8,964	9	,000
Paaren 3	SS - OS	4,15000	,94428	,29861	3,47450	4,82550	13,898	9	,000
Paaren 4	SO - OO	5,95000	1,75515	,55503	4,69444	7,20556	10,720	9	,000

Aufgabe

Ermitteln sie, ob es in der experimentellen Studie von Diessel&Tomasello (2005) einen signifikanten Unterschied zwischen den englischsprachigen und den deutschsprachigen Kindern gegeben hat (lassen sie die GEN-Relativsätze außer Acht). Erstellen sie außerdem ein Boxplot-Diagramm und prüfen sie, ob sich einzelne Relativsatztypen voneinander unterscheiden.