

***t*-test**

## *t*-test

---

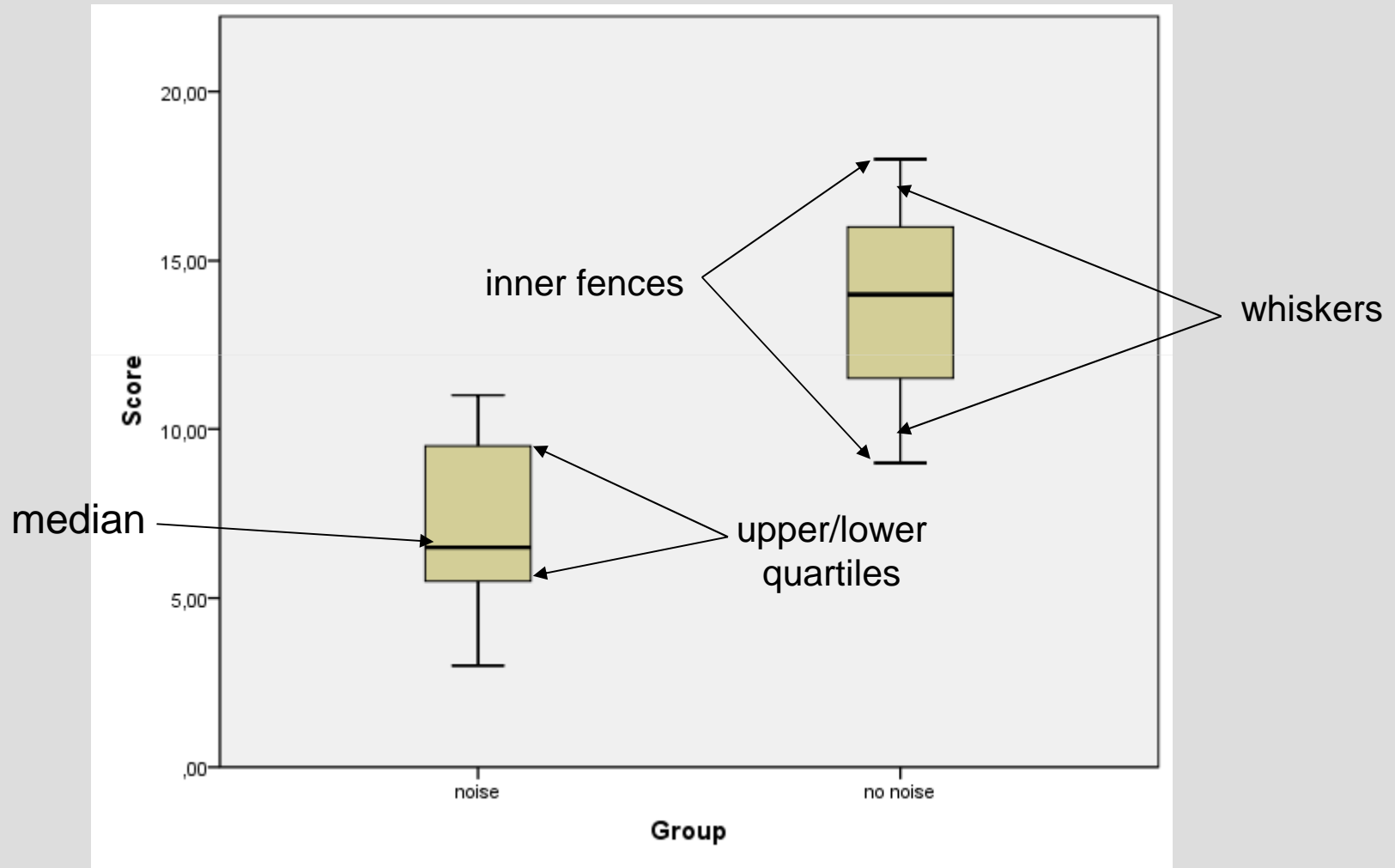
24 people were involved in an experiment to determine whether background noise (e.g. music) affects short-term memory (recall of words). Half of the participants were randomly allocated to the NOISE condition, and half to the NO NOISE condition. The participants in the NOISE condition tried to memorize 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.

<b>NOISE (group 1)</b>	<b>NO NOISE (group 2)</b>
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

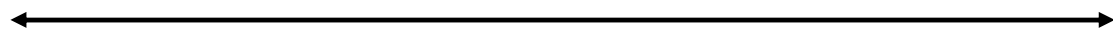
<b>NOISE (group 1)</b>	<b>NO NOISE (group 2)</b>
5.00 10.00 6.00 6.00 7.00 3.00 6.00 9.00 5.00 10.00 11.00 9.00	15.00 9.00 16.00 15.00 16.00 18.00 17.00 13.00 11.00 12.00 13.00 11.00
$\Sigma=87$ $X=7.3$ $SD=2.5$	$\Sigma=166$ $X=13.8$ $SD=2.8$

# Box-and-leaf plot

---



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



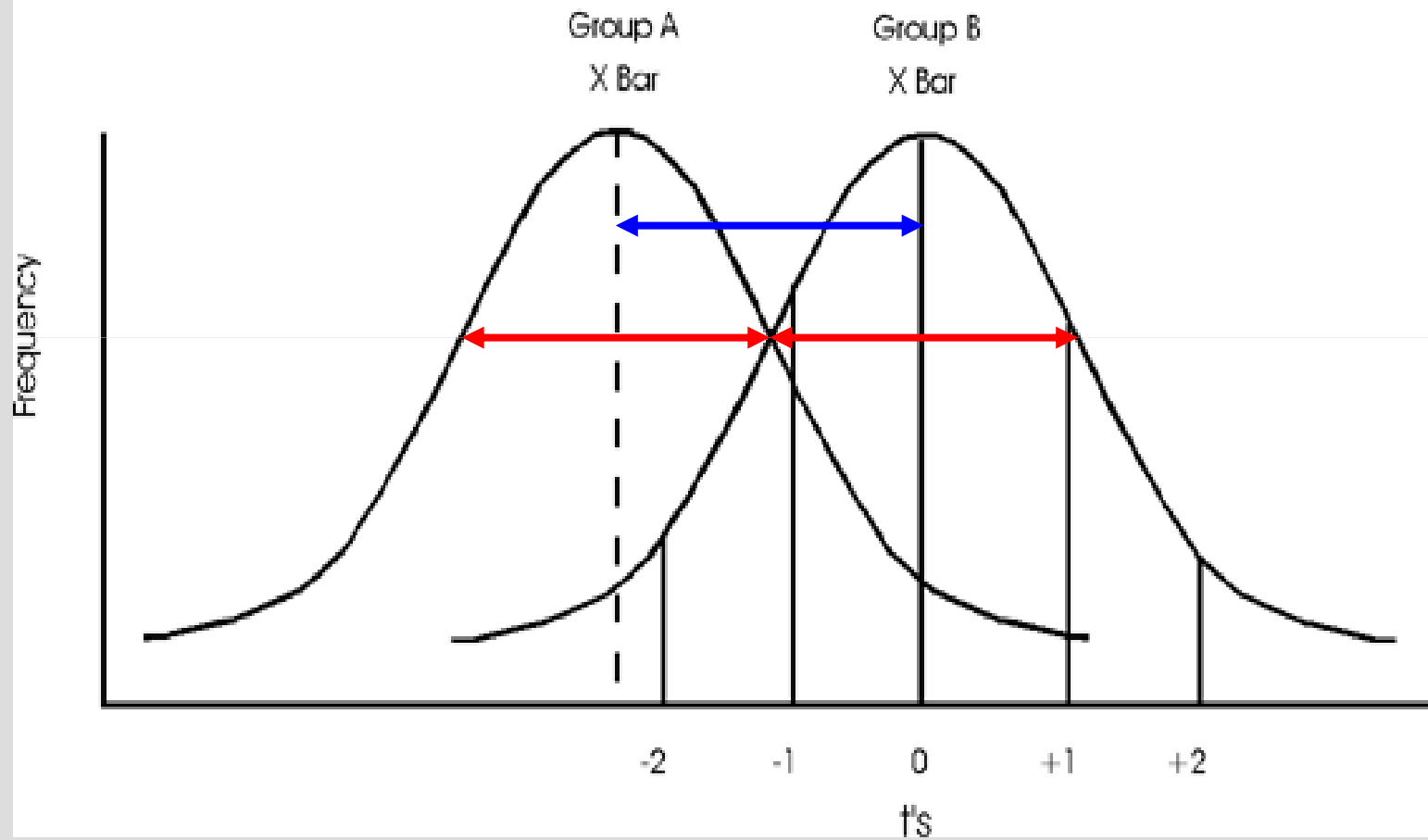
Between group variance (difference between M1 and M2)

Within group variance

Within group variance

Figure 7-3 Frequency Distribution of Group B and A Life Spans.

Group A mean lies to the left of Group B mean. Group B is the control group.



# Preconditions for $t$ -test

---

- Interval data
- For small data sets ( $N < 15$ ) the distribution of the means must be normally distributed (check histograms)
- Homogeneity-of-variance (**Levene's test** )



### Gruppenstatistiken

	Group	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Score	noise	12	7,2500	2,49089	,71906
	no noise	12	13,8333	2,75791	,79614

### Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
									Untere	Obere
Score	Varianzen sind gleich	,177	,678	-6,137	22	,000	-6,58333	1,07279	-8,80817	-4,35850
	Varianzen sind nicht gleich			-6,137	21,776	,000	-6,58333	1,07279	-8,80950	-4,35717

# Tests

---

	Parametric	Non-parametric
between / independent / unrelated		
within / dependent / related / repeated measures		

# Tests

---

	Parametric	Non-parametric
between / independent / unrelated	Independent t-test	
within / dependent / related / repeated measures		

# Tests

---

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

# Tests

---

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

# Tests

---

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

# Effect size

---

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

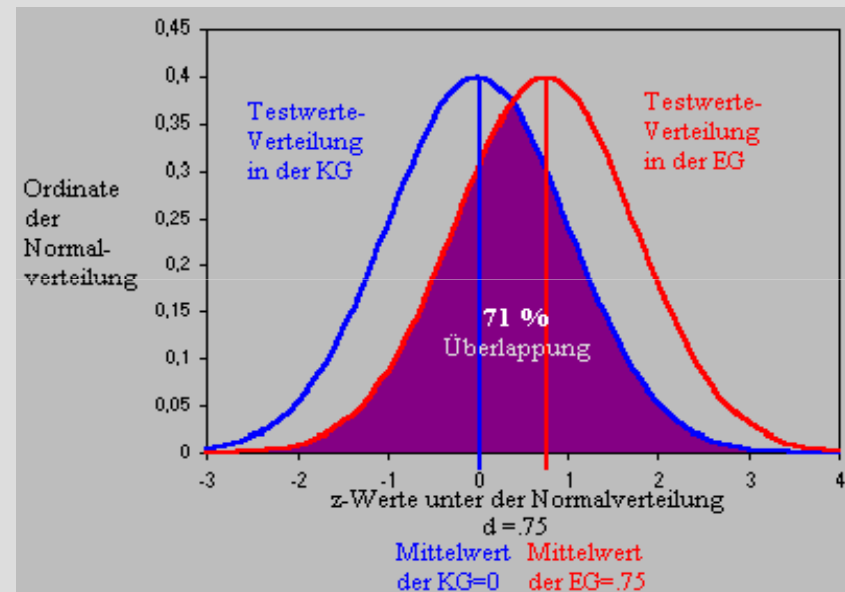
Difference between means:  $7.3 - 13.8 = -6.5$

$$\text{Mean SD} = \frac{\text{SD 1} + \text{SD 2}}{2} = \frac{2.5 + 2.8}{2} = 2.65$$

$$d = \frac{M1 - M2}{\text{Mean SD}} = \frac{6.5}{2.65} = 2.45$$

# Effect size

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





Exercise: The word *that* is ambiguous. Among other things, it can be a demonstrative (e.g. *That's my car*) or 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*).

- (1) Peter knows that she was coming.
- (2) Peter knows that guy.

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.

# Independent *t*-test

---

<b>Group 1</b>	<b>That-NPs</b>	<b>Group 2</b>	<b>That-clause</b>
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

# Independent *t*-test

**Gruppenstatistiken**

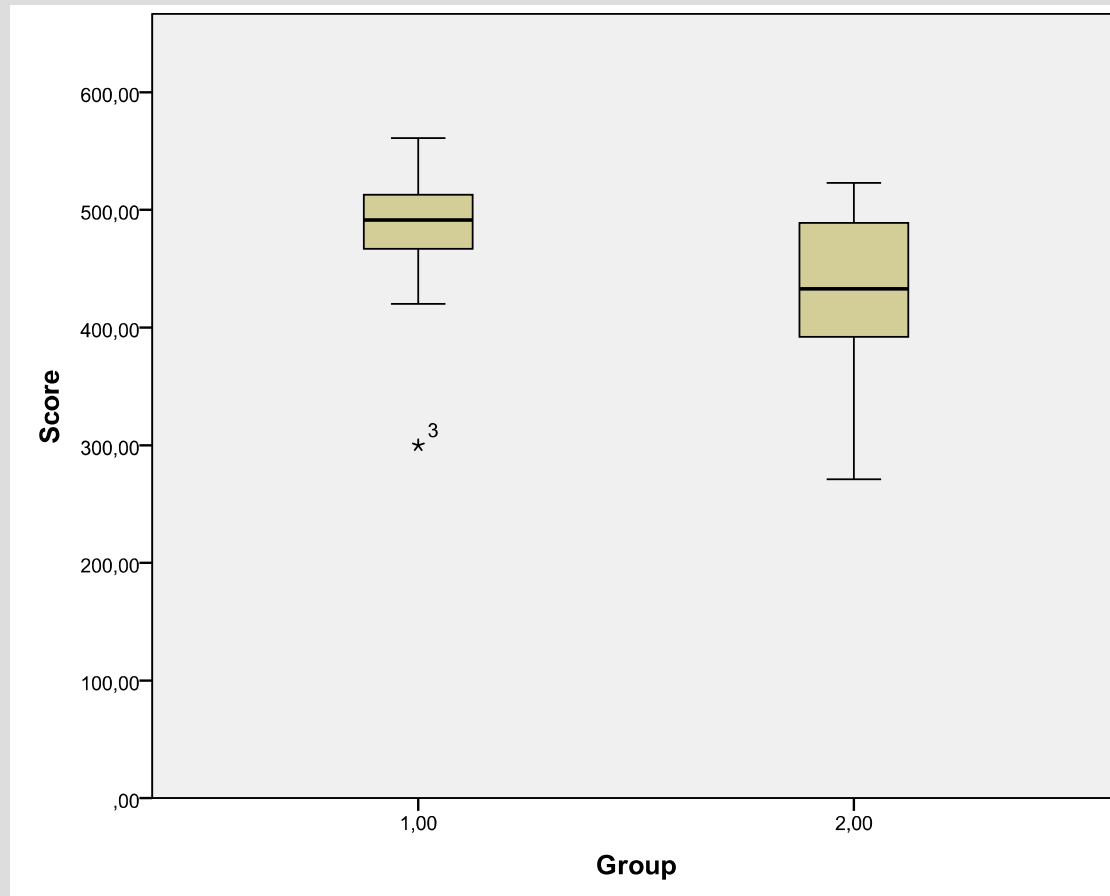
	Group	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Score	1,00	10	476,5000	73,08329	23,11096
	2,00	10	430,8000	72,79316	23,01922

**Test bei unabhängigen Stichproben**

	Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit						
	F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler der Differenz	95% Konfidenzintervall der Differenz	
								Untere	Obere
Score	,068	,797	1,401	18	,178	45,70000	32,61903	-22,83004	114,23004
Varianzen sind nicht gleich			1,401	18,000	,178	45,70000	32,61903	-22,83012	114,23012

# Independent $t$ -test

---



# Independent *t*-tests

---

Group 1	That-NPs	Group 2	That-clause
1	500	11	392
2	513	12	445
3	<b>300</b>	13	<b>271</b>
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

# Independent *t*-test

Results when outliers are excluded

Group	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
Score 1,00	9	496,1111	41,01355	13,67118
2,00	9	448,5556	49,13784	16,37928

## Test bei unabhängigen Stichproben

	Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit							
	F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz	Standardfehler 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

# Paired *t*-test

---

<b>Condition 1</b>	<b>That-NPs</b>	<b>Condition 2</b>	<b>That-clause</b>
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



# 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

**Test bei gepaarten Stichproben**

		Gepaarte Differenzen				T	df	Sig. (2-seitig)	
		Mittelwert	Standardabweichung	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

# Mann-Whitney U-test

# U-test

---

Exercise: 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 VOT. In order to decide if two-year old children pronounce /g/ and /k/ differently, the researcher collects a corpus of 13 words, six words including /g/ and seven words including /k/. 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.

# U-test

---

Speech sound	VOT in msc
/g/	38
/g/	195
/g/	56
/g/	3
/g/	51
/g/	89
/k/	125
/k/	73
/k/	138
/k/	35
/k/	51
/k/	190
/k/	169

# U-test

---

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

# U-test

---

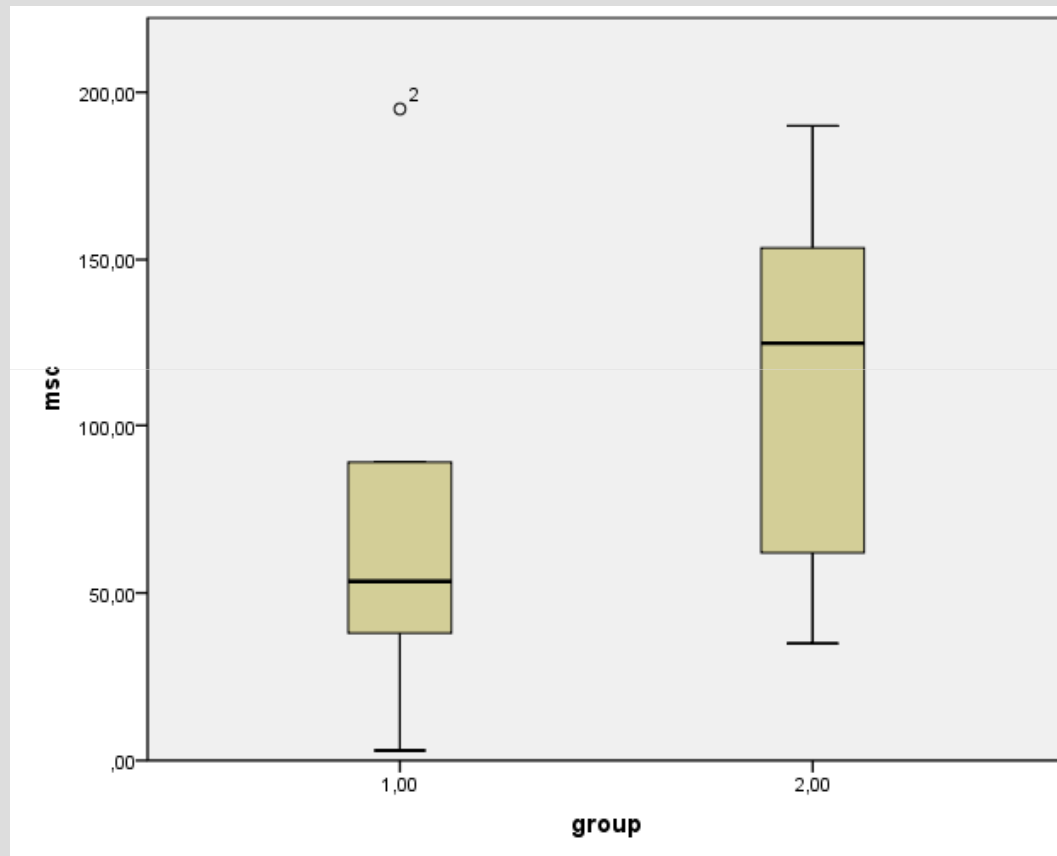
## Übersicht über Hypothesentest

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von msc ist über Kategorien von group gleich.	Mann-Whitney-U-Test unabhängiger Stichproben	,352	Nullhypothese behalten.

Asymptotische Signifikanz werden angezeigt. Das Signifikanzniveau ist .05.

# U-test

---



# U-test

---

Results when outlier excluded

## Übersicht über Hypothesentest

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von msc ist über Kategorien von group gleich.	Mann-Whitney-U-Test unabhängiger Stichproben	,104	Nullhypothese behalten.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist .05.



Wilcoxon

# Wilcoxon

---

Exercise: Nurses were asked to rate their sympathy on a scale between 1 and 10 for MS patients before and after talking to these patients. Table 1 shows the nurses' sympathy scores before and after they had talked to them.

# Wilcoxon

---

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

# Wilcoxon

---

## Übersicht über Hypothesentest

	Nullhypothese	Test	Sig.	Entscheidung
<b>1</b>	Der Medianwert der Unterschiede zwischen Before und After ist gleich 0.	Wilcoxon-Vorzeichen-Rang-Test verbundener Stichproben	,024	Nullhypothese ablehnen.

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist .05.

# One sample $t$ -test

# One sample *t*-test

---

Exercise. Previous research has shown that English-speaking children have an MLU of 301 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.

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

# One sample *t*-test

---

		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	



# One sample *t*-test

---

**Statistik bei einer Stichprobe**

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

**Test bei einer Stichprobe**

	Testwert = 0					
	T	df	Sig. (2-seitig)	Mittlere Differenz	95% Konfidenzintervall der Differenz	
					Untere	Obere
Score	56,412	23	,000	2,87083	2,7656	2,9761