

04

Statistische Tests

Dominic Schmitz & Janina Esser

Statistische Tests

- Einfachster Teil der **inferenziellen Statistik**:
wir nehmen unsere Daten und leiten etwa aus ihnen ab
- Geschieht meist anhand des „**Null-Hypothesis Significance Testing**“
- Resultat ist oftmals der berühmte **p-Wert** (*probability value*)

Tests für verschiedene Datentypen

Ziel	normalverteilte Daten	nicht normalverteilte Daten	kategoriale Daten
Beschreibung einer Gruppe	Durchschnitt, Standardabweichung	Median, Interquartial-Spannweite	Proportion
Testen auf Normalverteilung	Shapiro-Wilk Test		-
Vergleich zweier unabhängiger Gruppen	t-Test independent samples	Wilcoxon-Mann-Whitney Test	Fisher's Test
Vergleich zweier abhängiger Gruppen	t-Test dependent samples	Wilcoxon Signed-Rank Test	McNemar's Test
Vergleich dreier oder mehr unabhängiger Gruppen	ANOVA	Kruskal-Wallis Test	Chi-Quadrat Test
Vergleich dreier oder mehr abhängiger Gruppen	ANOVA	Friedman Test	Cochrane Q
Korrelation	Pearson	Spearman	

Tests für verschiedene Datentypen

Ziel	normalverteilte Daten	nicht normalverteilte Daten	kategoriale Daten
Beschreibung einer Gruppe	Durchschnitt, Standardabweichung	Median, Interquartial-Spannweite	Proportion
Testen auf Normalverteilung	Shapiro-Wilk Test		-
Vergleich zweier unabhängiger Gruppen	t-Test independent samples	Wilcoxon-Mann-Whitney Test	Fisher's Test
Vergleich zweier abhängiger Gruppen	t-Test dependent samples	Wilcoxon Signed-Rank Test	McNemar's Test
Vergleich dreier oder mehr unabhängiger Gruppen	ANOVA	Kruskal-Wallis Test	Chi-Quadrat Test
Vergleich dreier oder mehr abhängiger Gruppen	ANOVA	Friedman Test	Cochrane Q
Korrelation	Pearson	Spearman	

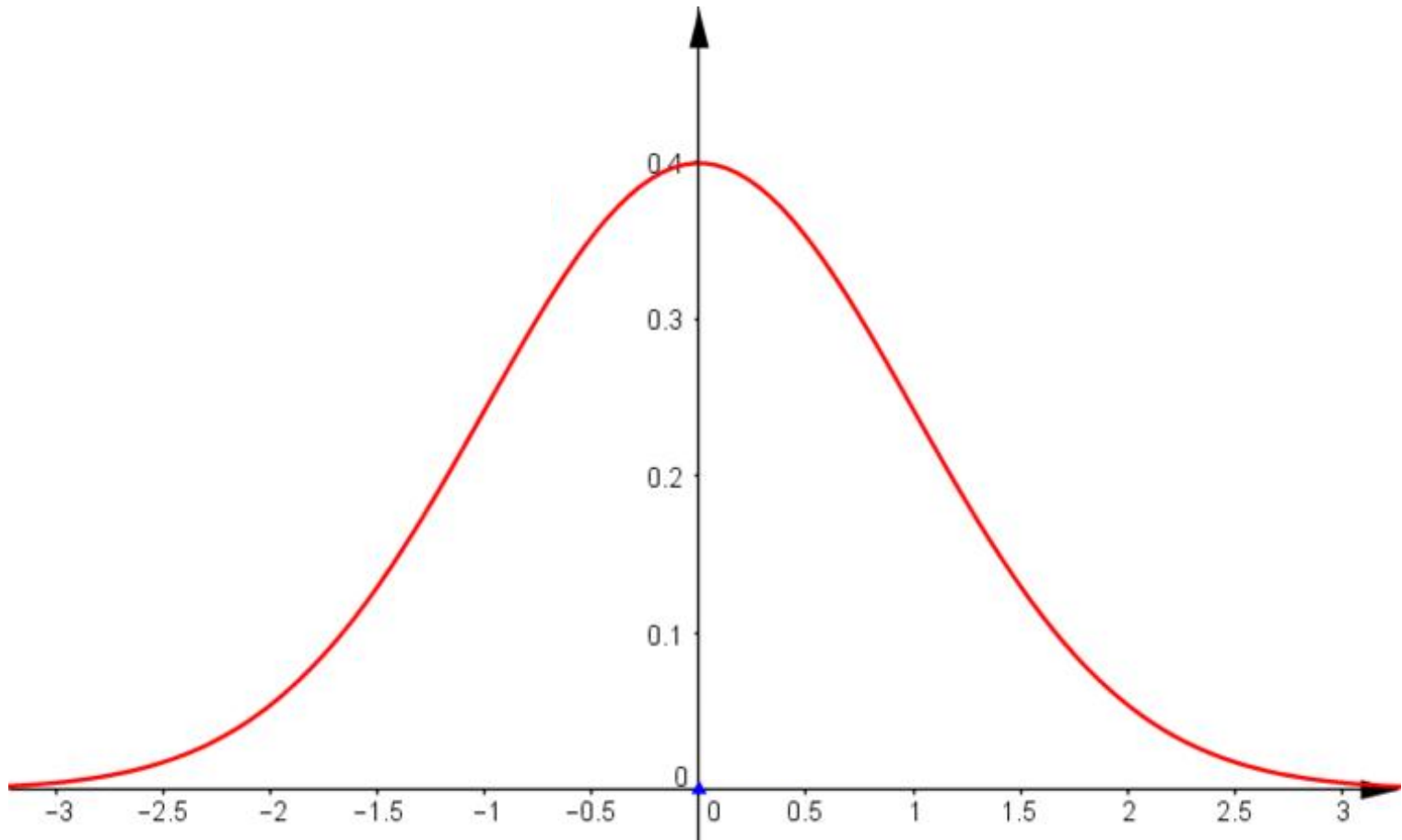
Tests für verschiedene Datentypen

Ziel	normalverteilte Daten	nicht normalverteilte Daten	kategoriale Daten
Beschreibung einer Gruppe	Durchschnitt, Standardabweichung	Median, Interquartial-Spannweite	Proportion
Testen auf Normalverteilung	Shapiro-Wilk Test		-
Vergleich zweier unabhängiger Gruppen	t-Test independent samples	Wilcoxon-Mann-Whitney Test	Fisher's Test
Vergleich zweier abhängiger Gruppen	t-Test dependent samples	Wilcoxon Signed-Rank Test	McNemar's Test
Vergleich dreier oder mehr unabhängiger Gruppen	ANOVA	Kruskal-Wallis Test	Chi-Quadrat Test
Vergleich dreier oder mehr abhängiger Gruppen	ANOVA	Friedman Test	Cochrane Q
Korrelation	Pearson	Spearman	

Shapiro-Wilk Test

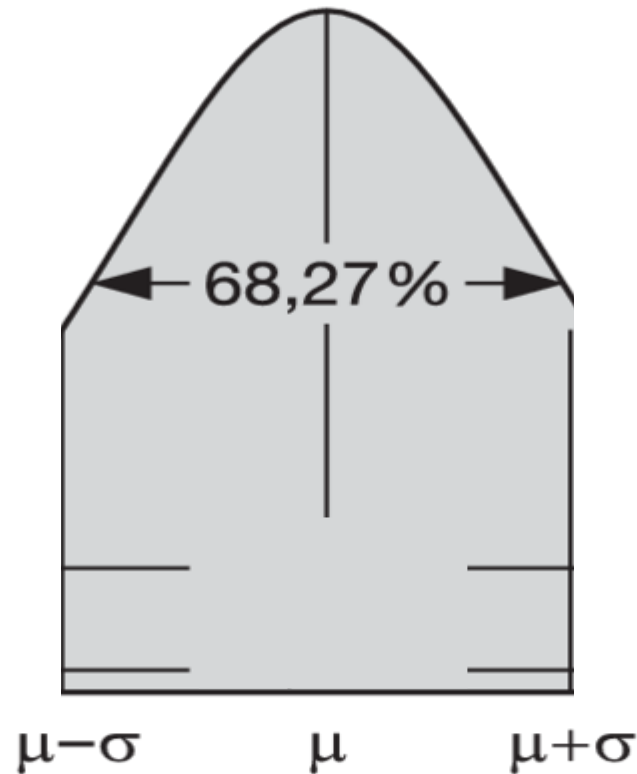
- mit einem **Shapiro-Wilk Test** kann man feststellen, ob eine Stichprobe **normalverteilt** ist
- die Normalverteilung ist eine um den Durchschnitt symmetrische Verteilung, d.h. Werte in der Nähe des Durchschnitts treten häufiger auf als solche, welche weiter vom Durchschnitt entfernt liegen

Exkurs: Normalverteilung



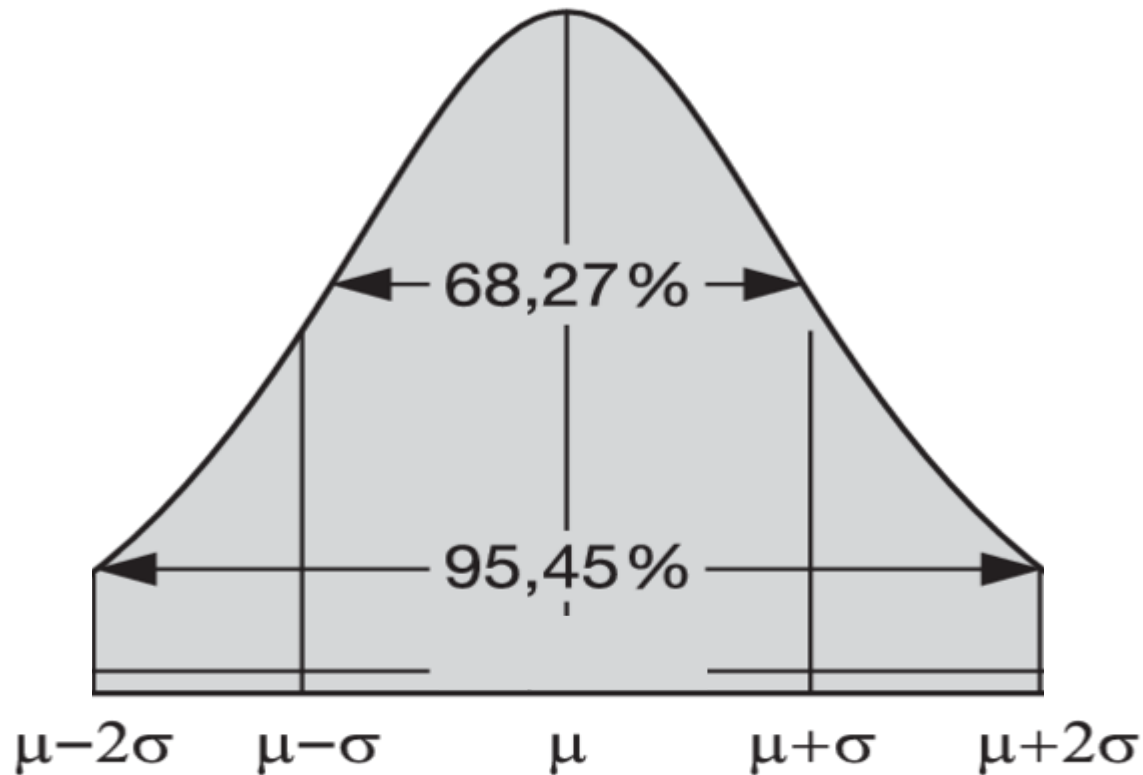
Exkurs: Normalverteilung

- 68.27 % der Datenpunkte befinden sich innerhalb 1 Standardabweichung vom Durchschnitt



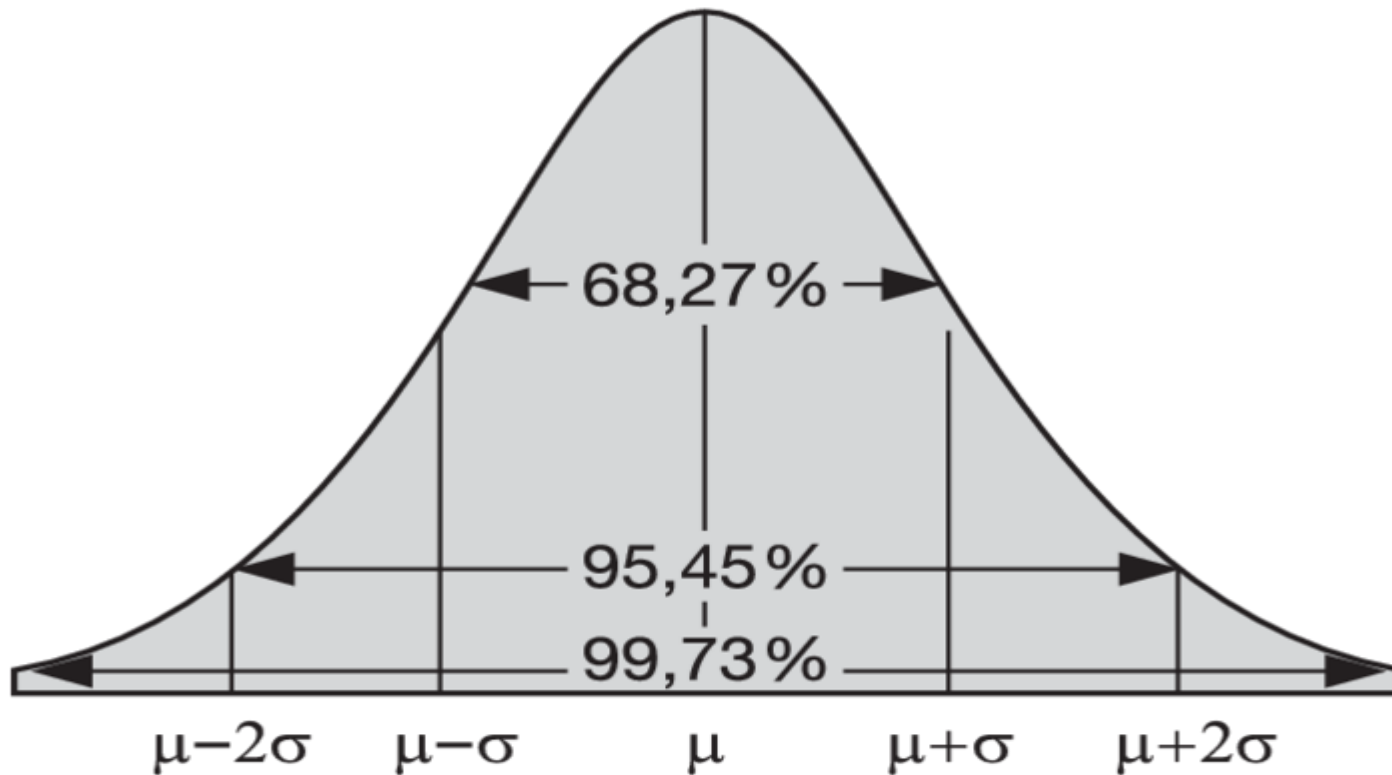
Exkurs: Normalverteilung

- 94.45 % der Datenpunkte befinden sich innerhalb 2 Standardabweichungen vom Durchschnitt



Exkurs: Normalverteilung

- 99.73 % der Datenpunkte befinden sich innerhalb 3 Standardabweichungen vom Durchschnitt

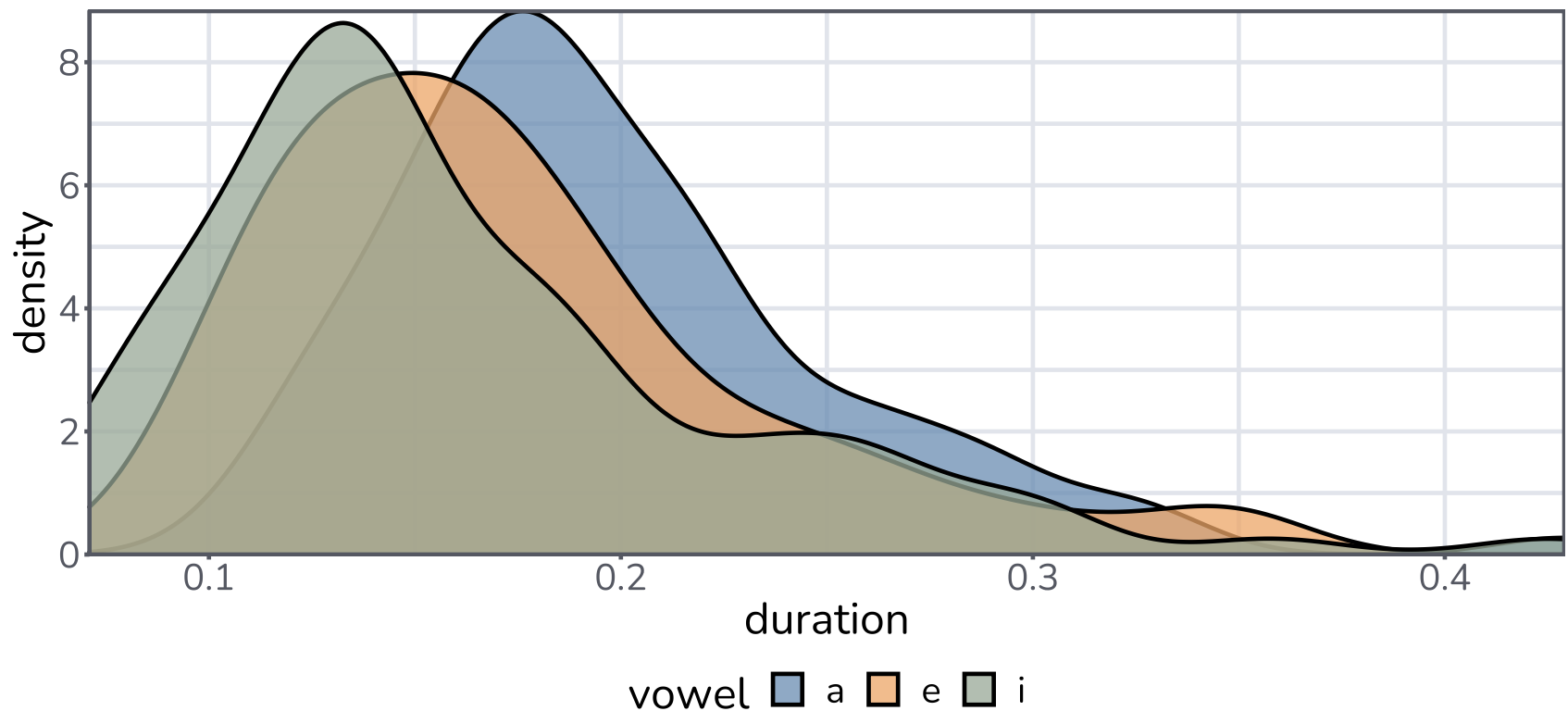


Shapiro-Wilk Test

- mit einem **Shapiro-Wilk Test** kann man feststellen, ob eine Stichprobe **normalverteilt** ist
- die Normalverteilung ist eine um den Durchschnitt symmetrische Verteilung, d.h. Werte in der Nähe des Durchschnitts treten häufiger auf als solche, welche weiter vom Durchschnitt entfernt liegen
- diese Info ist wichtig, da verschiedene andere Tests nur dann funktionieren, wenn Daten (annähernd) normalverteilt sind
- Daten müssen voneinander unabhängig sein; die Datenmenge sollte zwischen 3 und 5000 liegen
- als Beispiel nutzen wir das „Vowel Shortening in German“ Datenset aus dem SfL Package

Shapiro-Wilk Test

- Sind die Vokaldauern von /a/, /e/ und /i/ normalverteilt?



Shapiro-Wilk Test

- Sind die Vokaldauern von /a/, /e/ und /i/ normalverteilt?
- Der Shapiro-Wilk Test kommt zu folgenden Ergebnissen:

	p-Wert
/a/	$p < 0.001$
/e/	$p < 0.001$
/i/	$p < 0.001$

- Da die p -Werte kleiner 0.05 sind, sind die Daten **nicht normalverteilt**

Tests für verschiedene Datentypen


Ziel	normalverteilte Daten	nicht normalverteilte Daten	kategoriale Daten
Beschreibung einer Gruppe	Durchschnitt, Standardabweichung	Median, Interquartial-Spannweite	Proportion
Testen auf Normalverteilung	Shapiro-Wilk Test		-
Vergleich zweier unabhängiger Gruppen	t-Test independent samples	Wilcoxon-Mann-Whitney Test	Fisher's Test
Vergleich zweier abhängiger Gruppen	t-Test dependent samples	Wilcoxon Signed-Rank Test	McNemar's Test
Vergleich dreier oder mehr unabhängiger Gruppen	ANOVA	Kruskal-Wallis Test	Chi-Quadrat Test
Vergleich dreier oder mehr abhängiger Gruppen	ANOVA	Friedman Test	Cochrane Q
Korrelation	Pearson	Spearman	

t-Test

- Es gibt **verschiedene Arten** des t-Tests
- Wichtig dabei:
Stammen meine Daten aus dem gleichen Sample?
- Ja – z.B. falls zwei Experimente mit gleichen TN durchgeführt werden
→ **dependent samples t-test**
- Nein – z.B. falls zwei Experimente mit verschiedenen TN durchgeführt werden
→ **independent samples t-test**

t-Test – dependent samples

Beispiel: Blutdruck

$$df = 10 - 1$$


Blutdruck	1	2	3	4	5	6	7	8	9	10
Placebo x	168	184	172	173	150	155	163	164	151	146
Medikament y	176	145	150	163	136	168	164	139	145	112
Differenz z	8	-39	-22	-10	-14	13	1	-25	-6	-34

- $\bar{z} = -12.8$
- $s = 17.36$
- $t = -2.332$

t-Test – independent samples

Beispiel: f0 bei Männern

f0	1	2	3	4	5	6	7	8	9	10
Gruppe 1 x	55	69	64	70	75	70	83	69	75	69
Gruppe 2 y	61	60	62	58	75	63	52	66	59	

- $n_y = 10, n_x = 9$
- $\bar{x} = 69.00, \bar{y} = 61.78$
- $s_x = 7.972, s_y = 6.280$

$$s_p = 7.226$$

$$t = -2.175$$

Tests für verschiedene Datentypen

Ziel	normalverteilte Daten	nicht normalverteilte Daten	kategoriale Daten
Beschreibung einer Gruppe	Durchschnitt, Standardabweichung	Median, Interquartial-Spannweite	Proportion
Testen auf Normalverteilung	Shapiro-Wilk Test		-
Vergleich zweier unabhängiger Gruppen	t-Test independent samples	Wilcoxon-Mann-Whitney Test	Fisher's Test
Vergleich zweier abhängiger Gruppen	t-Test dependent samples	Wilcoxon Signed-Rank Test	McNemar's Test
Vergleich dreier oder mehr unabhängiger Gruppen	ANOVA	Kruskal-Wallis Test	Chi-Quadrat Test
Vergleich dreier oder mehr abhängiger Gruppen	ANOVA	Friedman Test	Cochrane Q
Korrelation	Pearson	Spearman	

Chi-Quadrat-Test

- mit **Chi-Quadrat-Tests** können wir bestimmen, ob zwei kategoriale Variablen zusammenhängen
- als Beispiel nutzen wir das „Age and Looks“ Datenset aus dem SfL Package

	blue	brown	green
blonde	3	7	3
brunette	5	15	2
red	1	3	1

Chi-Quadrat-Test

- nun können wir mit einem **Chi-Quadrat-Test** testen, ob Haar- und Augenfarbe in unserem Sample zusammenhängen
- Ergebnis: $p = 0.84 > 0.05$, d.h. nein, kein Zusammenhang

	blue	brown	green
blonde	3	7	3
brunette	5	15	2
red	1	3	1

Tests für verschiedene Datentypen

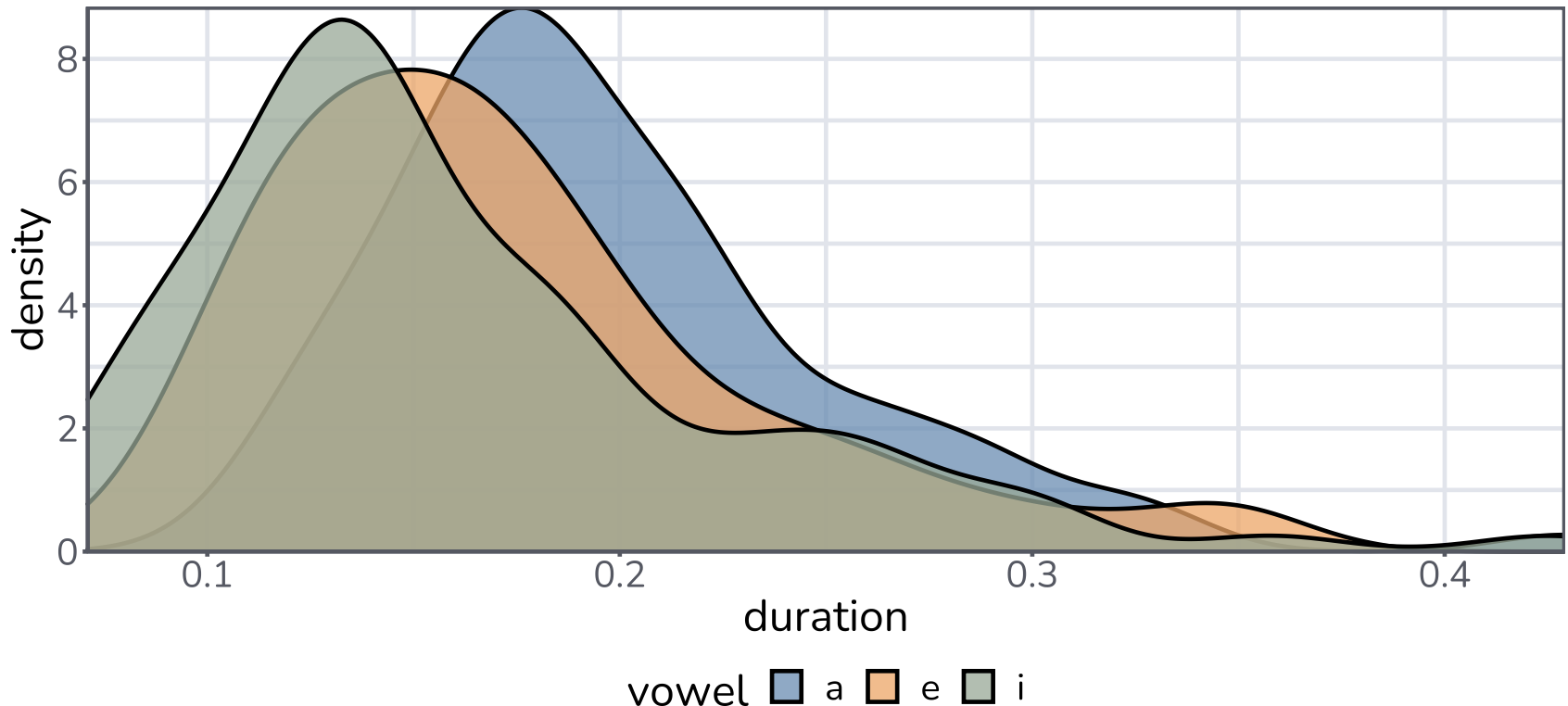
Ziel	normalverteilte Daten	nicht normalverteilte Daten	kategoriale Daten
Beschreibung einer Gruppe	Durchschnitt, Standardabweichung	Median, Interquartial-Spannweite	Proportion
Testen auf Normalverteilung	Shapiro-Wilk Test		-
Vergleich zweier unabhängiger Gruppen	t-Test independent samples	Wilcoxon-Mann-Whitney Test	Fisher's Test
Vergleich zweier abhängiger Gruppen	t-Test dependent samples	Wilcoxon Signed-Rank Test	McNemar's Test
Vergleich dreier oder mehr unabhängiger Gruppen	ANOVA	Kruskal-Wallis Test	Chi-Quadrat Test
Vergleich dreier oder mehr abhängiger Gruppen	ANOVA	Friedman Test	Cochrane Q
Korrelation	Pearson	Spearman	

Wilcoxon-Mann-Whitney Test

- Reminder: t-Tests setzen eine (annähernde) Normalverteilung der Daten voraus
- der **Wilcoxon-Mann-Whitney Test** kann auch mit **nicht-normalverteilten Daten** umgehen
- als Beispiel nutzen wir das „Vowel Shortening in German“ Datenset aus dem SfL Package

Wilcoxon-Mann-Whitney Test

- die Vokaldauern von /a/, /e/ und /i/ sind nicht normalverteilt (siehe Shapiro-Wilk Test)



Wilcoxon-Mann-Whitney Test

- Ergebnis:
ja, die Vokale haben unterschiedliche Dauern

	/a/ vs. /e/	/a/ vs. /i/	/e/ vs. /i/
t-Test	<0.001	<0.001	0.00568
WMW-Test	<0.001	<0.001	0.00241

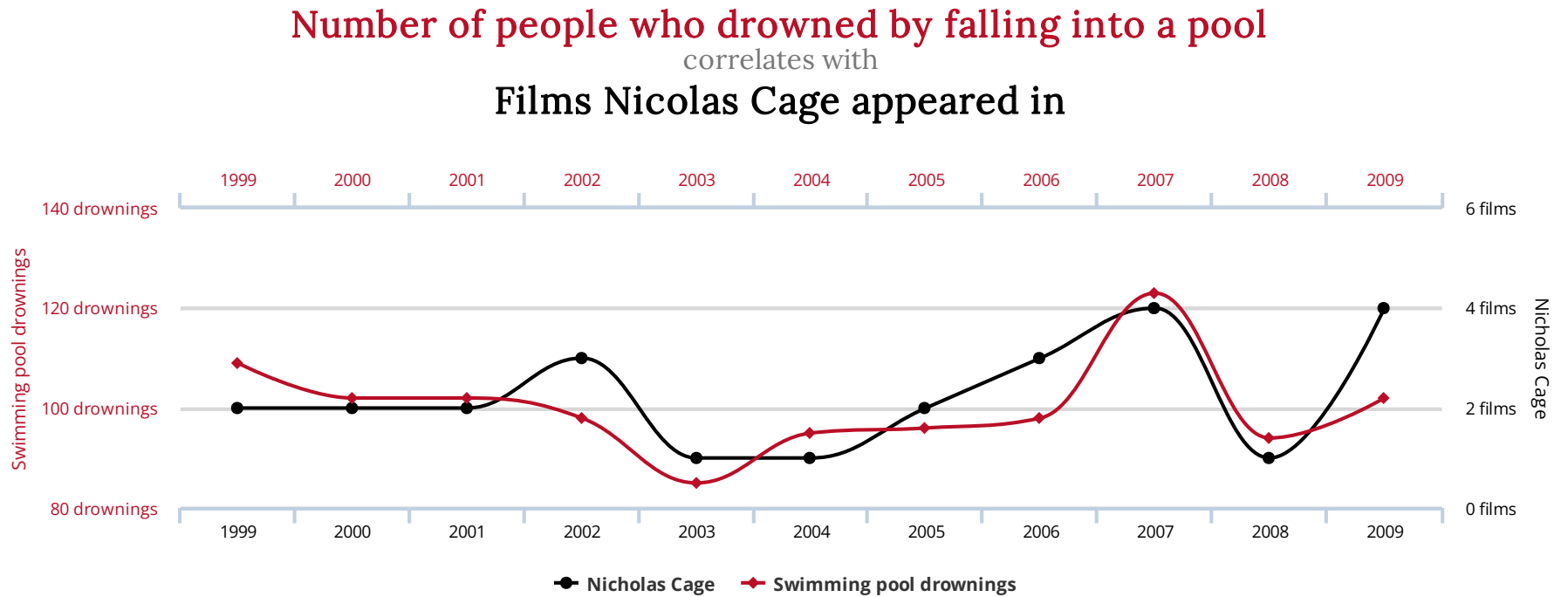
Tests für verschiedene Datentypen

Ziel	normalverteilte Daten	nicht normalverteilte Daten	kategoriale Daten
Beschreibung einer Gruppe	Durchschnitt, Standardabweichung	Median, Interquartial-Spannweite	Proportion
Testen auf Normalverteilung	Shapiro-Wilk Test		-
Vergleich zweier unabhängiger Gruppen	t-Test independent samples	Wilcoxon-Mann-Whitney Test	Fisher's Test
Vergleich zweier abhängiger Gruppen	t-Test dependent samples	Wilcoxon Signed-Rank Test	McNemar's Test
Vergleich dreier oder mehr unabhängiger Gruppen	ANOVA	Kruskal-Wallis Test	Chi-Quadrat Test
Vergleich dreier oder mehr abhängiger Gruppen	ANOVA	Friedman Test	Cochrane Q
Korrelation	Pearson	Spearman	

Korrelation

- die **Korrelation** beschreibt eine Beziehung zwischen zwei oder mehr Variablen
- Korrelation **bedeutet nicht** Kausalität!
 - zwei Variablen können korreliert sein
 - ohne dabei in kausaler Verbindung zu stehen

Korrelation

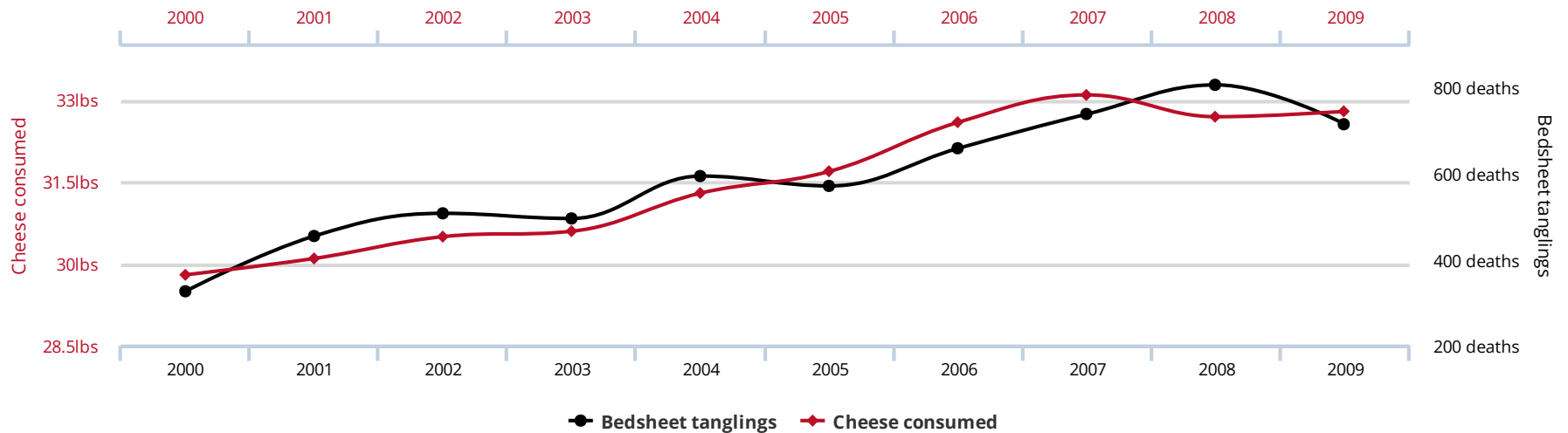


Korrelation

Per capita cheese consumption

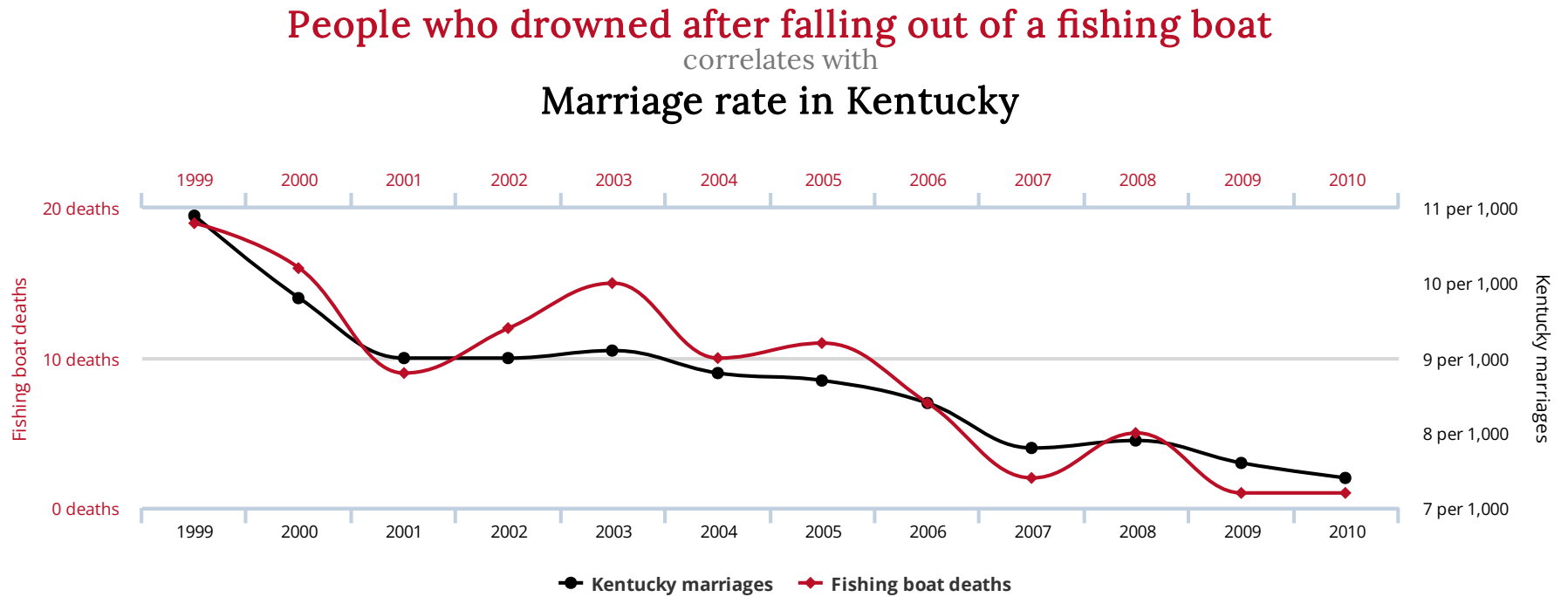
correlates with

Number of people who died by becoming tangled in their bedsheets



tylervigen.com

Korrelation



tylervigen.com

Korrelation

- sind die zu vergleichenden Daten normalverteilt und numerisch, nutzen wir **Pearson's r**
- sind die zu vergleichenden Daten nicht normalverteilt und/oder nicht numerisch, nutzen wir **Spearman's rho**
- als Beispiel nutzen wir das „Duration of word-final /s/ in English“ Datenset aus dem SfL Package

Korrelation

- Wann sprechen wir von Korrelation?

→ 4-stufige Version

Korrelationskoeffizient			Label	Richtung
0.7	< r ≤	1.0	sehr hoch	positive Korrelation
0.5	< r ≤	0.7	hoch	
0.2	< r ≤	0.5	mittel	
0.0	< r ≤	0.2	niedrig	
r ≈ 0			keine Korrelation	
0.0	> r ≥	-0.2	niedrig	negative Korrelation
-0.2	> r ≥	-0.5	mittel	
-0.5	> r ≥	-0.7	hoch	
-0.7	> r ≥	-1.0	sehr hoch	

Korrelation

- Wann sprechen wir von Korrelation?

→ 3-stufige Version

Korrelationskoeffizient			Label	Richtung
0.6	< r ≤	1.0	hoch	positive Korrelation
0.3	< r ≤	0.6	mittel	
0.0	< r ≤	0.3	niedrig	
r ≈ 0			keine Korrelation	
0.0	> r ≥	-0.3	niedrig	negative Korrelation
-0.3	> r ≥	-0.6	mittel	
-0.6	> r ≥	-1.0	hoch	

Korrelation

- Wann sprechen wir von Korrelation?

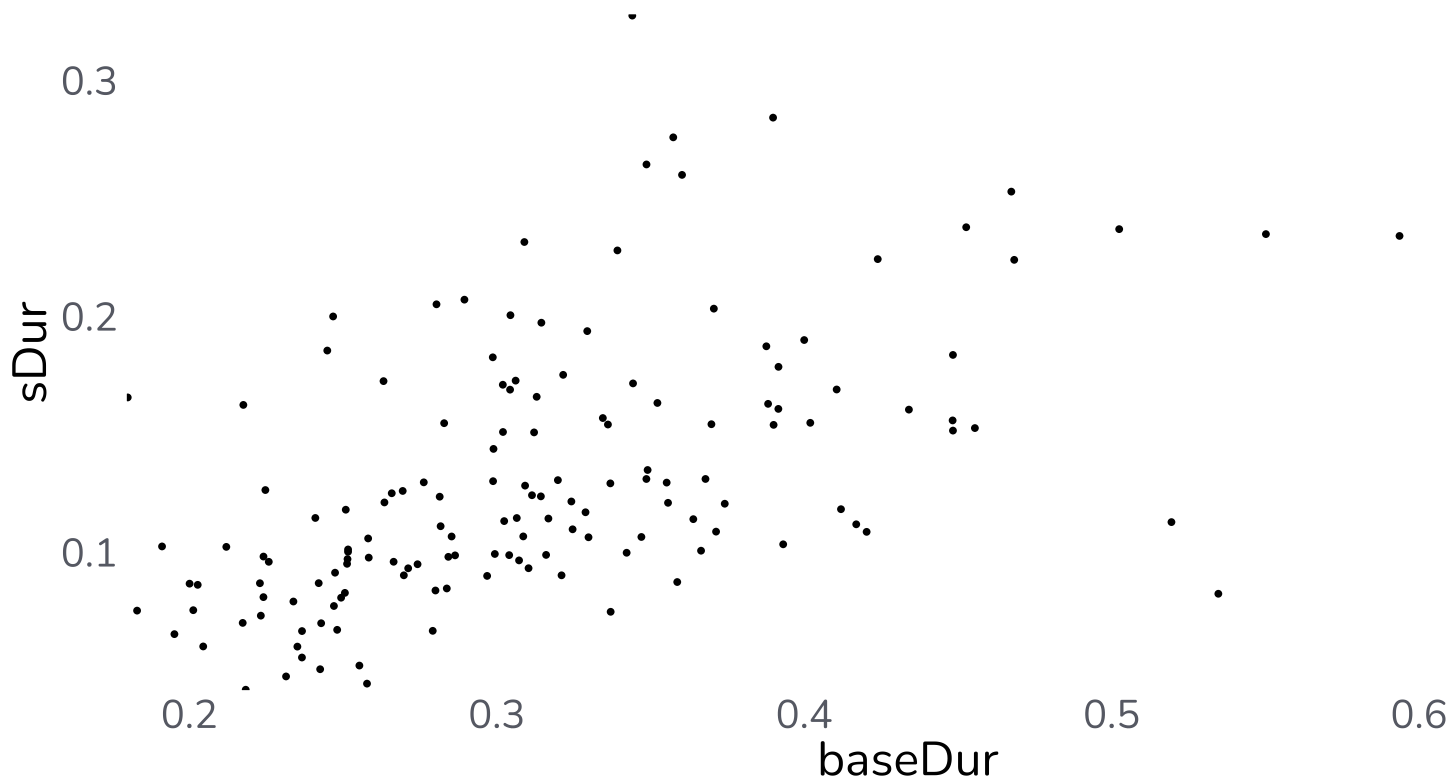
→ 2-stufige Version

Korrelationskoeffizient			Label	Richtung
0.6	< r ≤	1.0	hoch	positive Korrelation
0.2	< r ≤	0.5	mittel	
-0.2 ≤ r ≤ 0.2			niedrig bis keine Korrelation	
-0.2	> r ≥	-0.5	mittel	negative Korrelation
-0.5	> r ≥	-1.0	hoch	

- generell gilt: es gibt so viele Versionen wie wissenschaftliche Aufsätze

Korrelation

Frage: sind /s/-Dauer und base-Dauer korreliert?



Antwort: ja, da $r = 0.47$

Tests für verschiedene Datentypen

Ziel	normalverteilte Daten	nicht normalverteilte Daten	kategoriale Daten
Beschreibung einer Gruppe	Durchschnitt, Standardabweichung	Median, Interquartial-Spannweite	Proportion
Testen auf Normalverteilung	Shapiro-Wilk Test		-
Vergleich zweier unabhängiger Gruppen	t-Test independent samples	Wilcoxon-Mann-Whitney Test	Fisher's Test
Vergleich zweier abhängiger Gruppen	t-Test dependent samples	Wilcoxon Signed-Rank Test	McNemar's Test
Vergleich dreier oder mehr unabhängiger Gruppen	ANOVA	Kruskal-Wallis Test	Chi-Quadrat Test
Vergleich dreier oder mehr abhängiger Gruppen	ANOVA	Friedman Test	Cochrane Q
Korrelation	Pearson	Spearman	