

Session 04: Statistische Tests

Dominic Schmitz & Janina Esser

Verein für Diversität in der Linguistik

Statistische Tests



- Einfachster Teil der **inferentiellen 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*)

Statistische Tests



1. Shapiro-Wilk Test
 2. t-Test
 3. Chi-Quadrat-Test
 4. Wilcoxon-Mann-Whitney Test
 5. ANOVA
 6. ANCOVA
 7. Korrelation
- etc.

1. Shapiro-Wilk Test



- mit einem **Shapiro-Wilk Test** kann man feststellen, ob eine Stichprobe **normalverteilt** ist
- diese Info ist wichtig, da verschiedene andere Tests nur dann funktionieren, wenn Daten (annähernd) normalverteilt sind
- aber was ist eine Normalverteilung?

1. Shapiro-Wilk Test

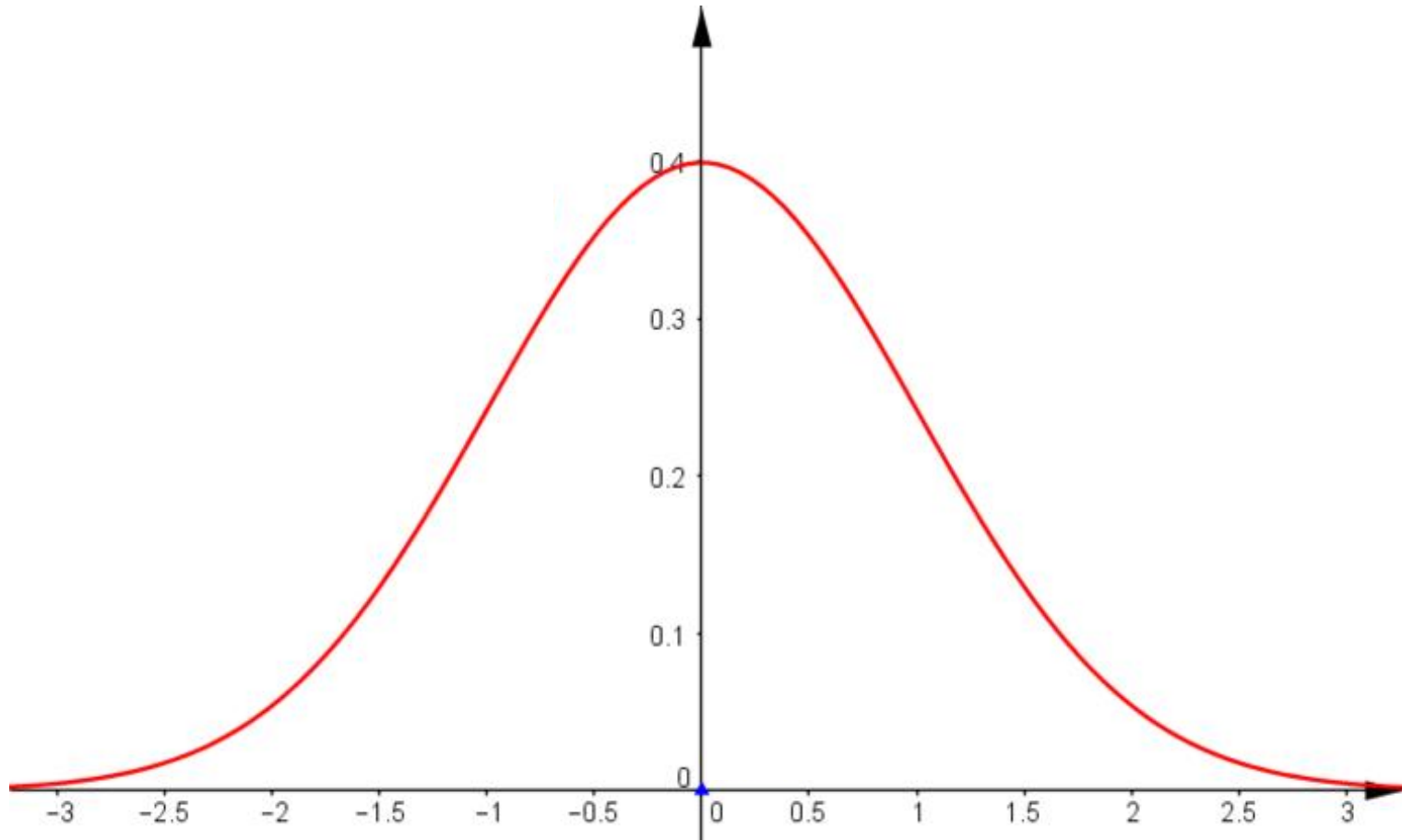


- Die Normalverteilung ist eine um den Durchschnitt symmetrische Verteilung, d.h. Werte in der Nähe des Durchschnitts treten häufiger auf als solche, welche weiter vom Durchschnitt entfernt liegen.

1. Shapiro-Wilk Test



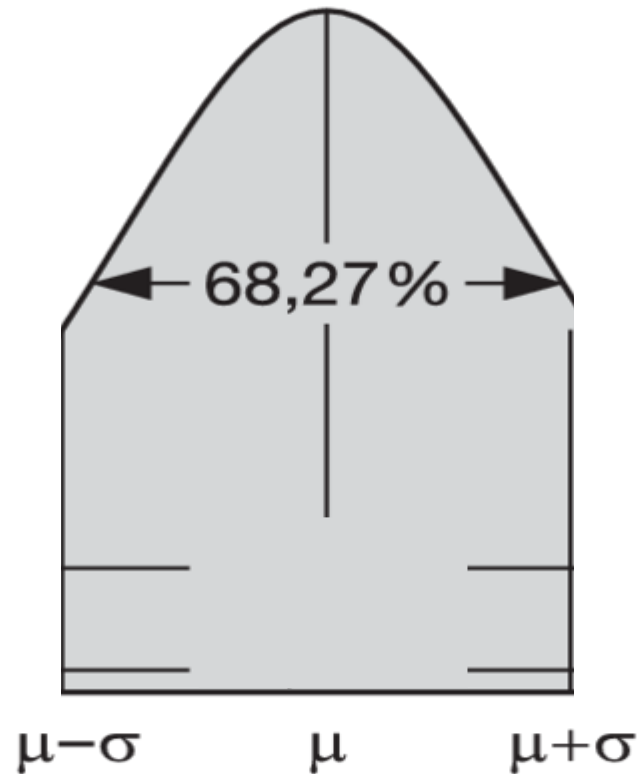
- Normalverteilung



1. Shapiro-Wilk Test



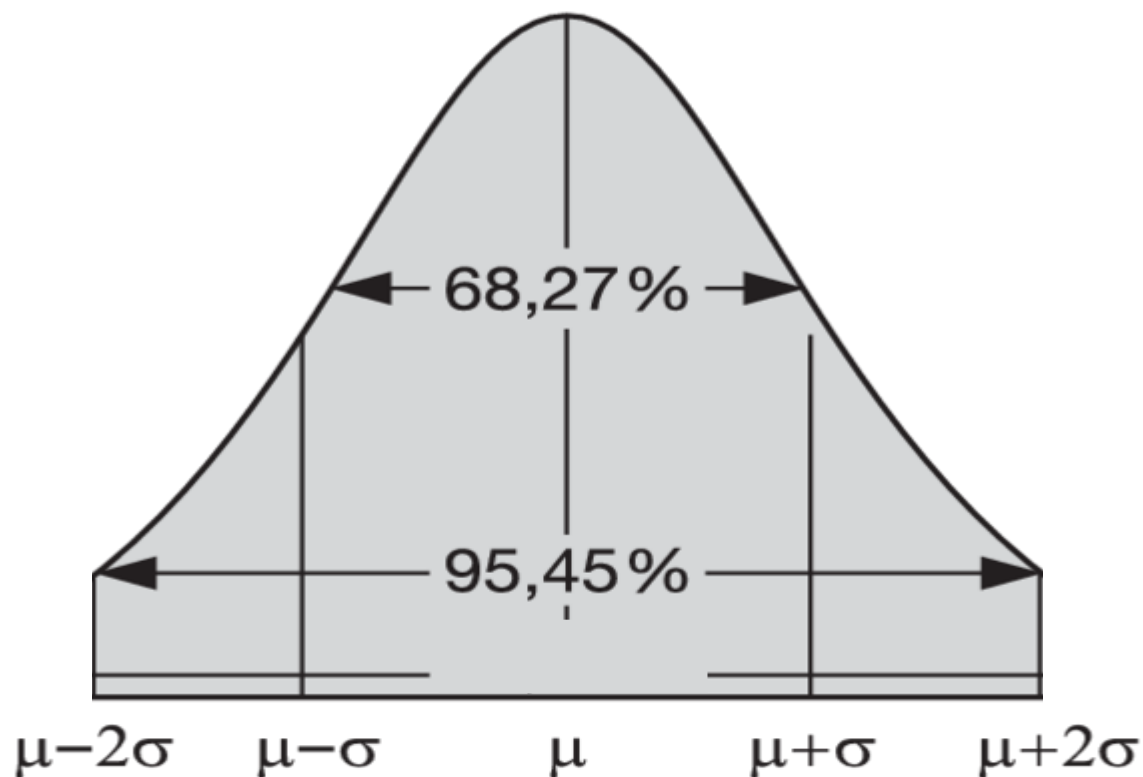
- 68.27 % der Datenpunkte befinden sich innerhalb 1 Standardabweichung vom Durchschnitt



1. Shapiro-Wilk Test



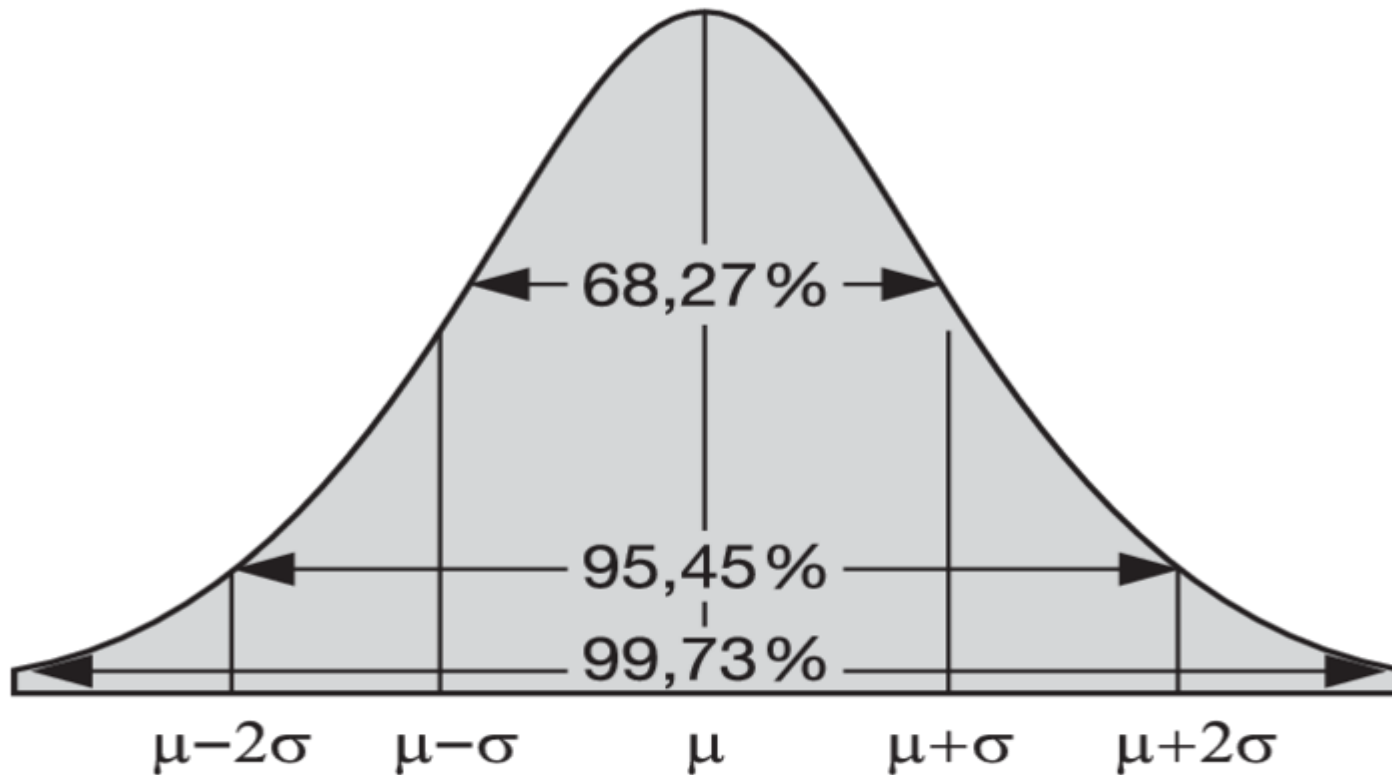
- 94.45 % der Datenpunkte befinden sich innerhalb 2 Standardabweichungen vom Durchschnitt



1. Shapiro-Wilk Test



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



1. Shapiro-Wilk Test

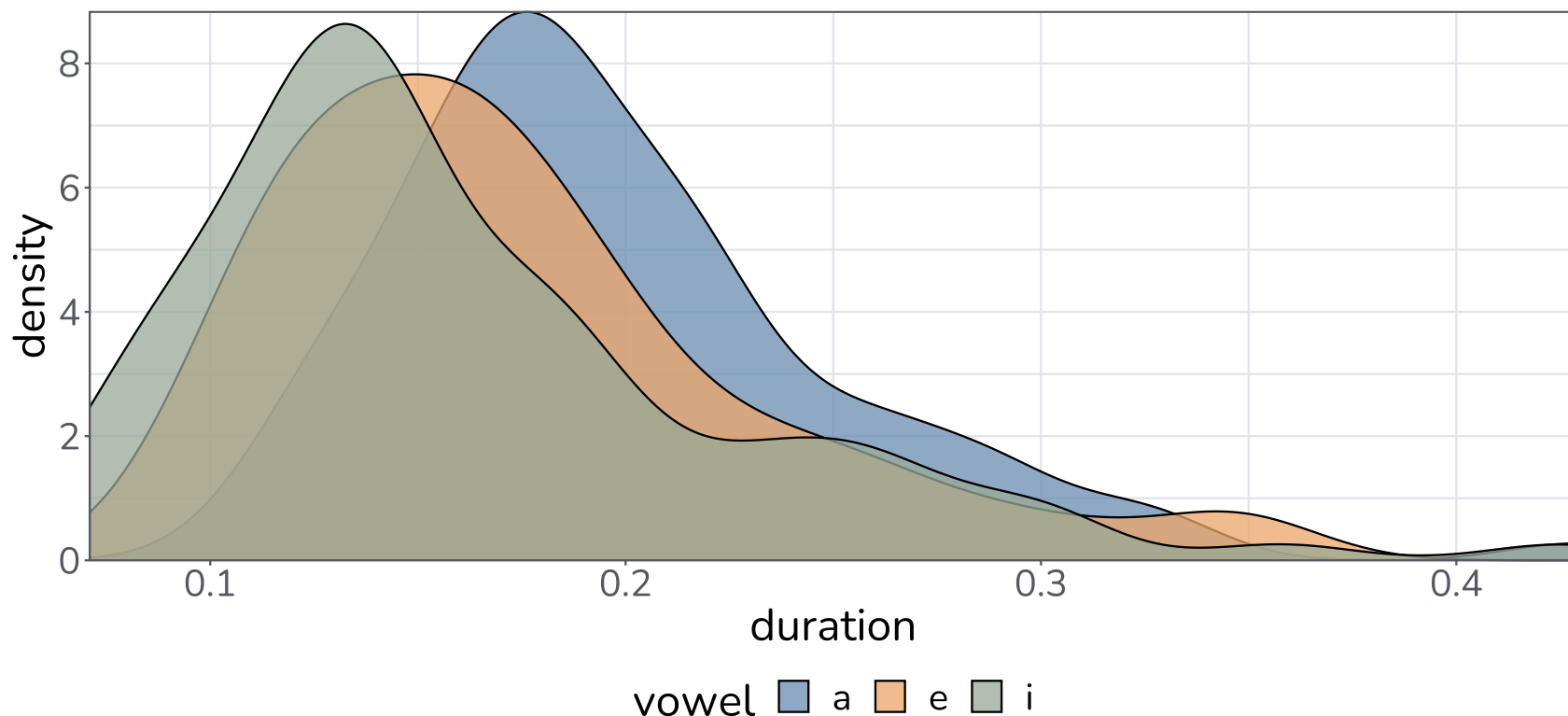


- mit einem **Shapiro-Wilk Test** kann man feststellen, ob eine Stichprobe **normalverteilt** ist
- diese Info ist wichtig, da verschiedene andere Tests nur dann funktionieren, wenn Daten (annähernd) normalverteilt sind
- als Beispiel nutzen wir das „Vowel Shortening in German“ Datenset aus dem SfL Package

1. Shapiro-Wilk Test



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





1. 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**

Statistische Tests



1. Shapiro-Wilk Test ✓
 2. t-Test
 3. Chi-Quadrat-Test
 4. Wilcoxon-Mann-Whitney Test
 5. ANOVA
 6. ANCOVA
 7. Korrelation
- etc.

2. 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**
- Das Ergebnis des t-Tests gibt an, ob zwei Messreihen signifikant verschieden sind

Statistische Tests



1. Shapiro-Wilk Test ✓
 2. t-Test ✓
 3. Chi-Quadrat-Test
 4. Wilcoxon-Mann-Whitney Test
 5. ANOVA
 6. ANCOVA
 7. Korrelation
- etc.

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 zusammengehören
- 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

Statistische Tests



1. Shapiro-Wilk Test ✓
 2. t-Test ✓
 3. Chi-Quadrat-Test ✓
 4. Wilcoxon-Mann-Whitney Test
 5. ANOVA
 6. ANCOVA
 7. Korrelation
- etc.

Wilcoxon-Mann-Whitney Test

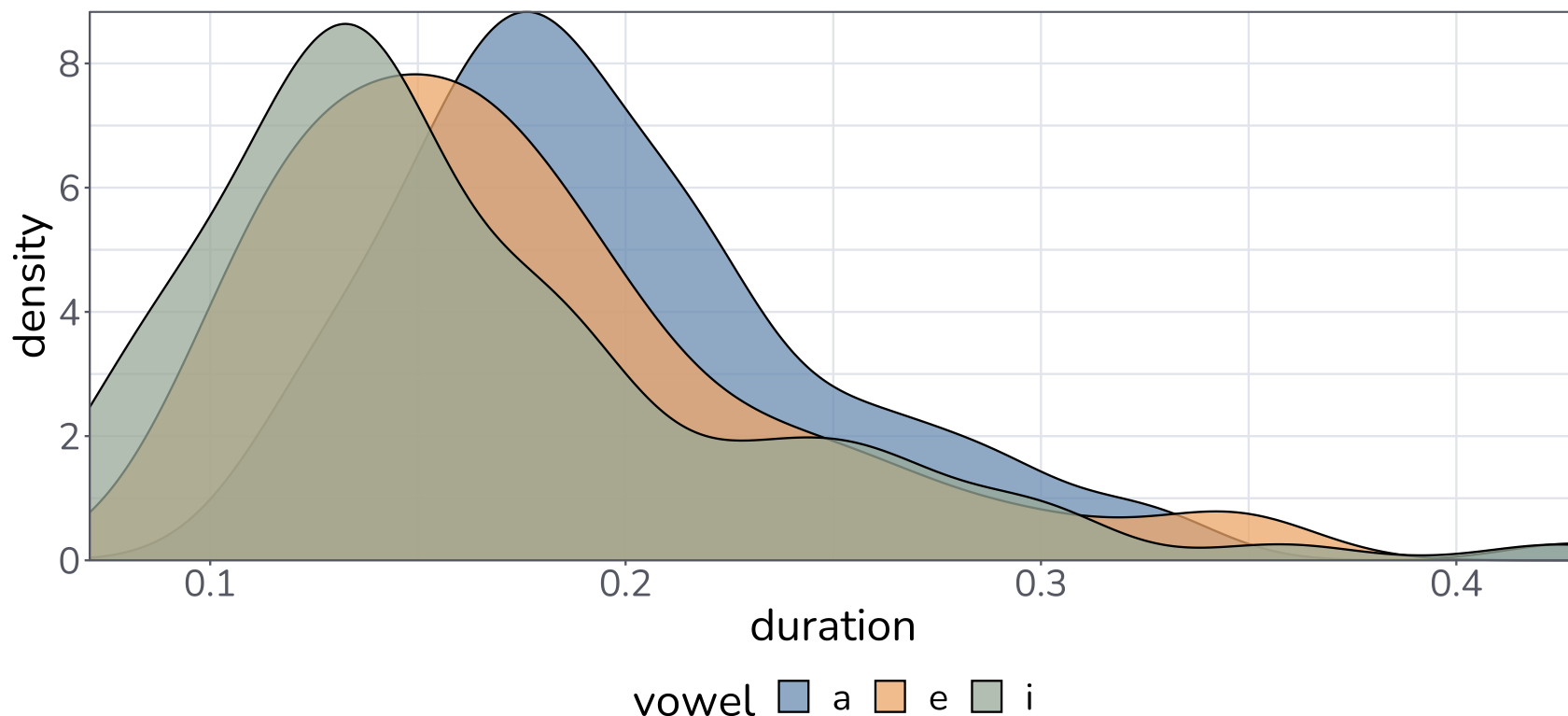


- t-Tests setzen eine (annähernde) Normalverteilung der Daten voraus
- der Wilcoxon-Mann-Whitney Test kann auch mit nicht-normalverteilten Daten umgehen
- als Beispiel nutzen wir das das „Vowel Shortening in German“ Datenset aus dem SfL Package

Wilcoxon-Mann-Whitney Test



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



Wilcoxon-Mann-Whitney Test



- Ergebnis:
die Vokale haben (teils) 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

Statistische Tests



1. Shapiro-Wilk Test ✓
 2. t-Test ✓
 3. Chi-Quadrat-Test ✓
 4. Wilcoxon-Mann-Whitney Test ✓
 5. ANOVA
 6. ANCOVA
 7. Korrelation
- etc.

ANOVA

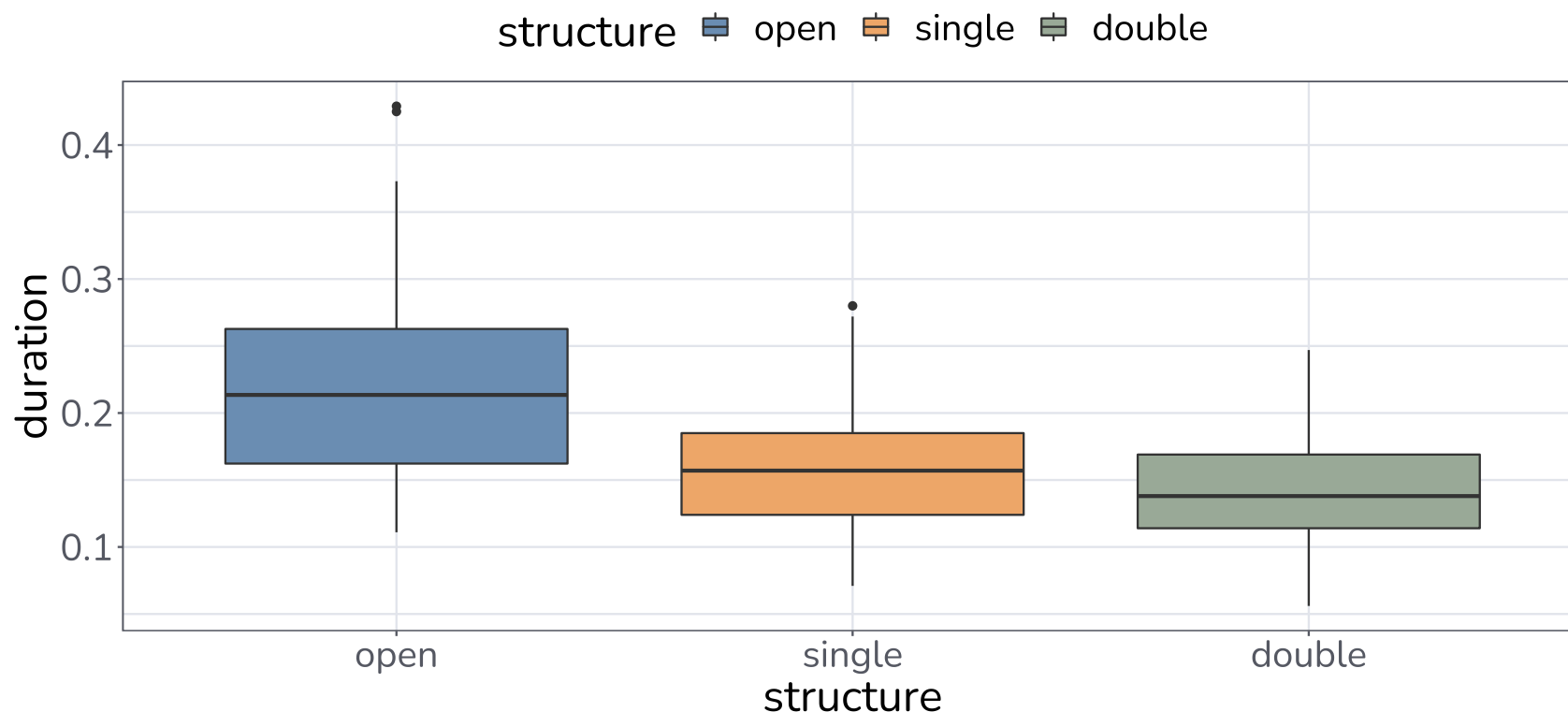


- die **analysis of variance**, d.h. die Varianzanalyse, kann dann genutzt werden, wenn man die Durchschnitte mehrerer Gruppen in einem Rutsch miteinander vergleichen möchte
- als Beispiel nutzen wir das „Vowel Shortening in German“ Datenset aus dem SfL Package
- unsere Gruppen: Silbenstruktur (offen, single, double)

ANOVA



- Hypothese: Die Vokaldauern unterscheiden sich je nach Silbenstruktur.



ANOVA



- führen wir nun eine ANOVA durch, müssen wir den Inhalt unserer Hypothese spezifizieren:

Dauer ~ Silbenstruktur

- die ANOVA gibt dann mit einem p -Wert an, ob eine signifikante Abhängigkeit besteht

$$p < 0.001$$

- wir wissen allerdings noch nicht, ob die Unterschiede zwischen allen Silbenstrukturen signifikant verschieden sind

ANOVA



- hierzu müssen wir einen **Post-Hoc-Test** nutzen
- wir nutzen den **Tukey-Test**, einen der meist genutzten Post-Hoc-Test
- dieser liefert uns folgende Ergebnisse:

	p-Wert
single – open	<0.001
double – open	<0.001
double – single	0.005

- **Ergebnis:** alle Unterschiede sind signifikant

Statistische Tests



1. Shapiro-Wilk Test ✓
 2. t-Test ✓
 3. Chi-Quadrat-Test ✓
 4. Wilcoxon-Mann-Whitney Test ✓
 5. ANOVA ✓
 6. ANCOVA
 7. Korrelation
- etc.

ANCOVA

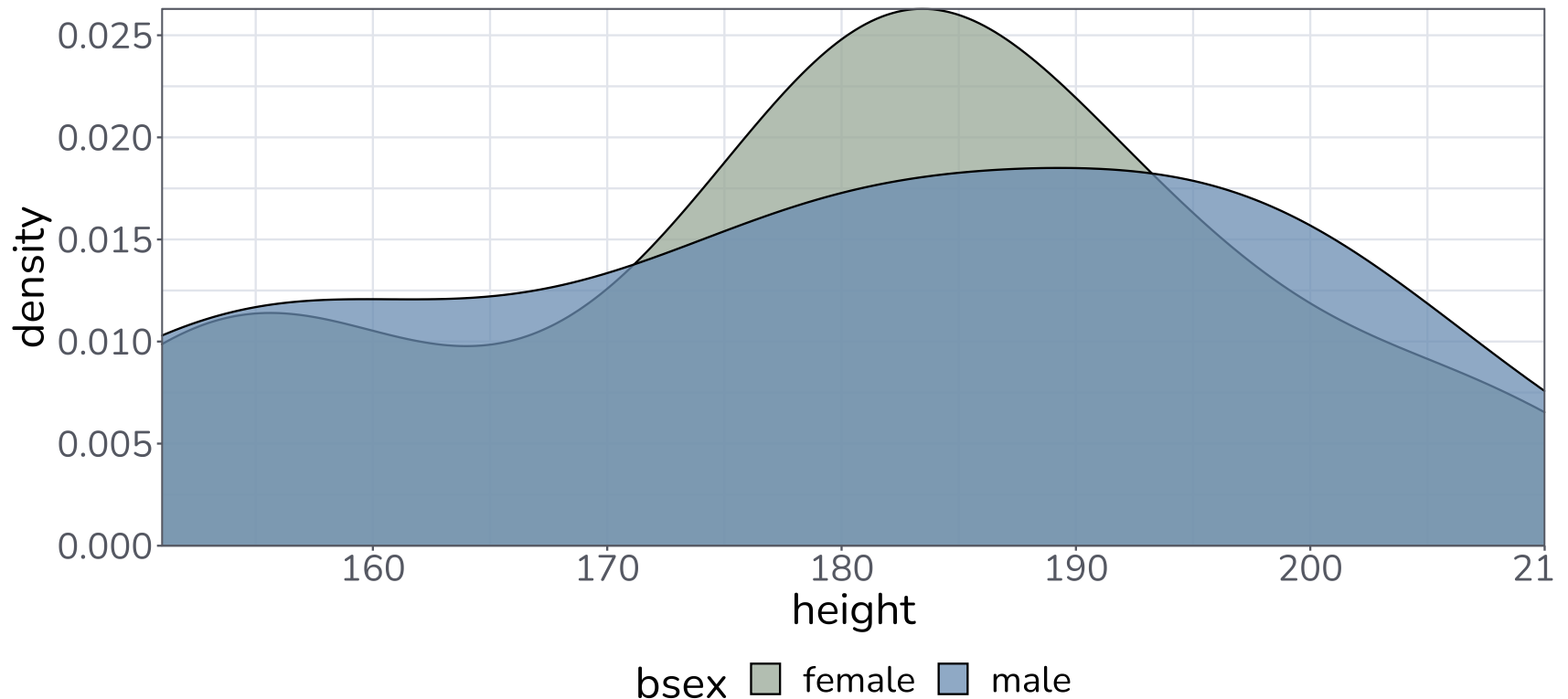


- die **analysis of covariance**, d.h. die Kovarianzanalyse, kann dann genutzt werden, wenn man den Durchschnittswert einer Variable in potentieller Abhängigkeit von einer kategorischen Variable und ihrer Levels herausfinden möchte
- als Beispiel nutzen wir das „Age and Looks“ Datenset aus dem SfL Package

ANCOVA



- Hypothese: Männer sind größer als Frauen.



ANCOVA



- führen wir nun eine ANCOVA durch, müssen wir den Inhalt unserer Hypothese spezifizieren:

$$\textit{Größe} \sim \textit{Geschlecht}$$

- die ANCOVA gibt dann mit einem p -Wert an, ob eine signifikante Abhängigkeit besteht

$$p = 0.957$$

- in diesem Fall: die Hypothese wird nicht bestätigt
- bei $p < 0.05$ muss ein Post-Hoc-Test durchgeführt werden

Statistische Tests



1. Shapiro-Wilk Test ✓
 2. t-Test ✓
 3. Chi-Quadrat-Test ✓
 4. Wilcoxon-Mann-Whitney Test ✓
 5. ANOVA ✓
 6. ANCOVA ✓
 7. Korrelation
- etc.

Korrelation

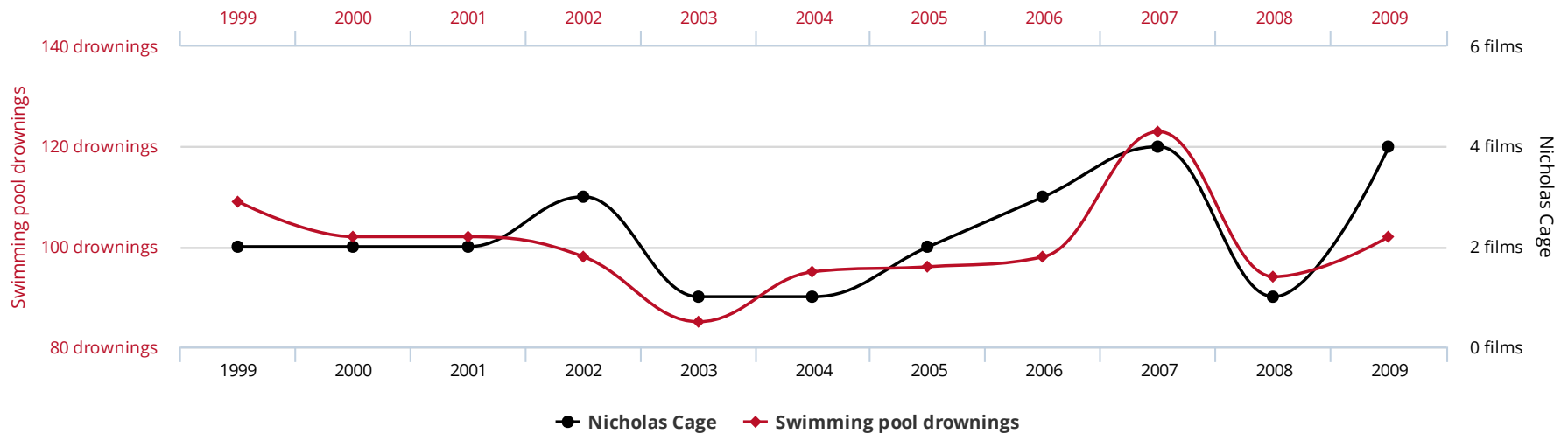


- die Korrelation beschreibt eine Beziehung zwischen zwei oder mehr Variablen
- Korrelation **bedeutet nicht** Kausalität:
 - zwei Variablen können korreliert sein
 - ohne dabei in kausaler Verbindung zu stehen

Korrelation



Number of people who drowned by falling into a pool
correlates with
Films Nicolas Cage appeared in



tylervigen.com

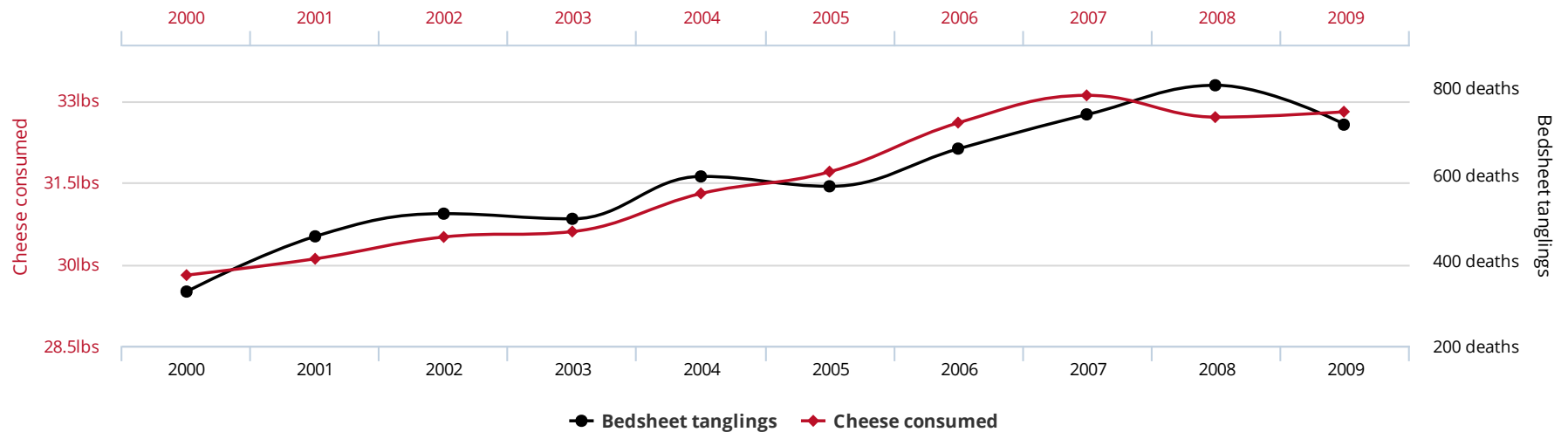
Korrelation



Per capita cheese consumption

correlates with

Number of people who died by becoming tangled in their bedsheets

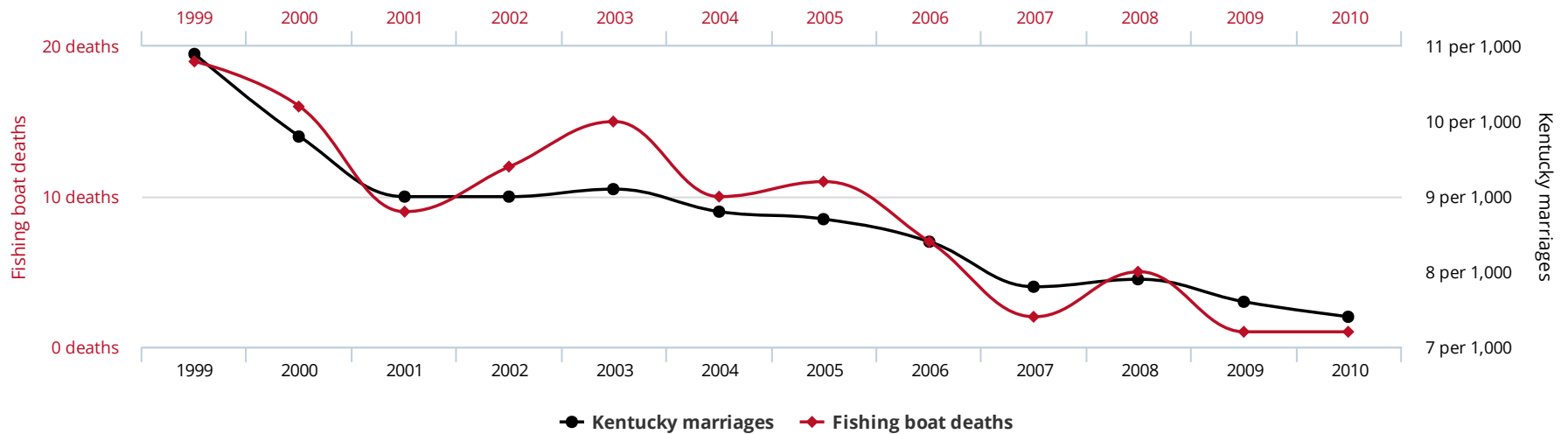


tylervigen.com

Korrelation



People who drowned after falling out of a fishing boat
correlates with
Marriage rate in Kentucky



tylervigen.com

Korrelation



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

Korrelation



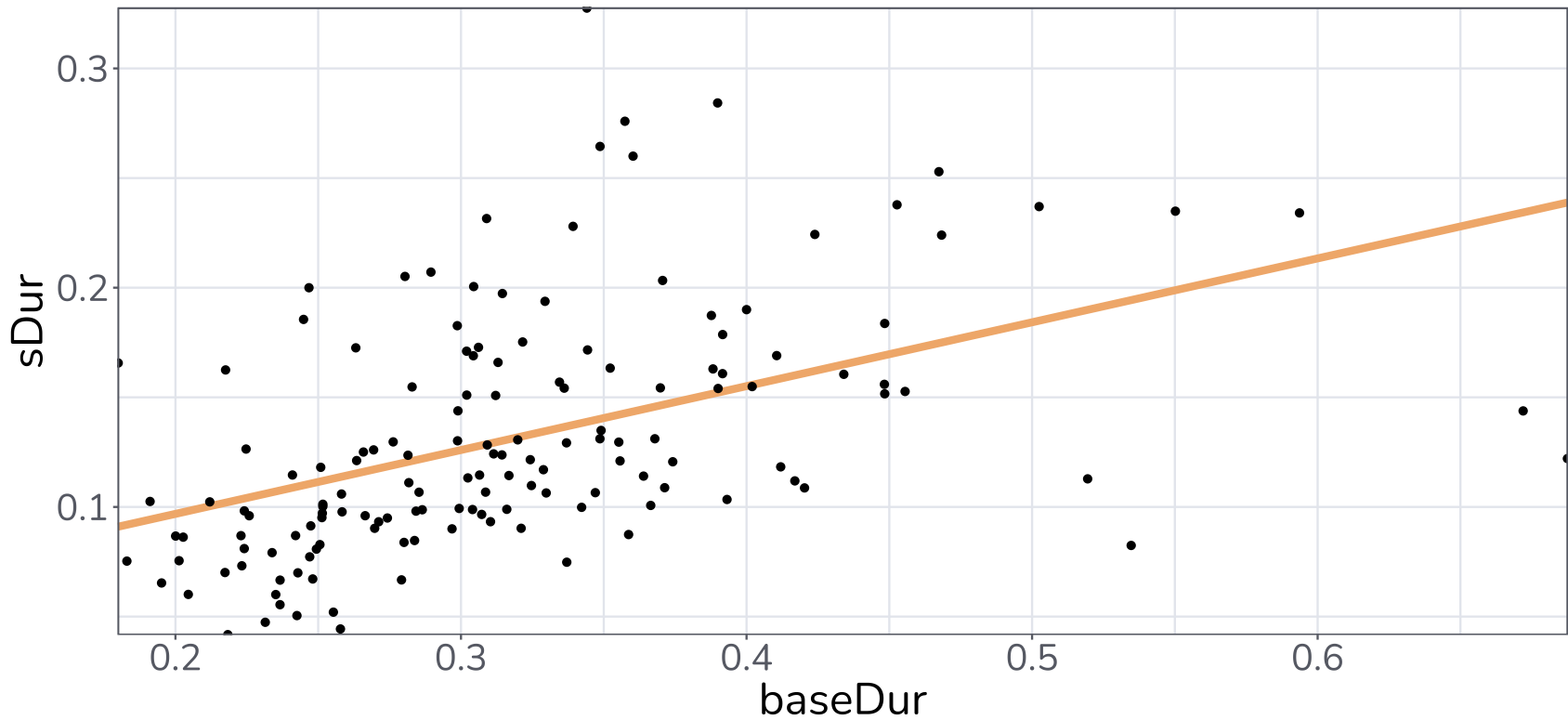
- Wann sprechen wir von vorhandener Korrelation?

correlation coefficient			labeling	kind of correlation
0.7	< r ≤	1	very high	positive correlation
0.5	< r ≤	0.7	high	
0.2	< r ≤	0.5	intermediate	
0	< r ≤	0.2	low	
r ≈ 0			no statistical correlation	
0	> r ≥	-0.2	low	negative correlation
-0.2	> r ≥	-0.5	intermediate	
-0.5	> r ≥	-0.7	high	
-0.7	> r ≥	-1	very high	

Korrelation



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



- **Antwort:** ja, da $r = 0.47$

Statistische Tests



1. Shapiro-Wilk Test ✓
 2. t-Test ✓
 3. Chi-Quadrat-Test ✓
 4. Wilcoxon-Mann-Whitney Test ✓
 5. ANOVA ✓
 6. ANCOVA ✓
 7. Korrelation ✓
- etc.