



GRaP Independence Test

Marcus Vollmer

Weihnachtskolloquium

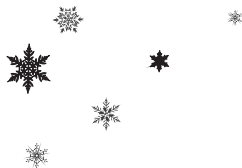
15. Dezember 2010





- 1 Problemstellung
- 2 Beispiel: Datensatz Oregon
- 3 Geometrie Randomisierter Permutationen (GRaP)
- 4 Power Analyse
- 5 Beispiele
- 6 Hauptresultate
- 7 Referenzen
- 8 Anhang





Problemstellung





Problemstellung

- stetige Merkmale X , Y

z.B. Tanne:



X - Höhe des Baumes

Y - Länge der Baumkrone





Problemstellung

- stetige Merkmale X , Y

z.B. Tanne:



X - Höhe des Baumes

Y - Länge der Baumkrone

- Testen die Hypothesen:

H_0 : X und Y sind unabhängig

H_1 : X und Y sind nicht unabhängig





Problemstellung

- stetige Merkmale X , Y

z.B. Tanne:



X - Höhe des Baumes

Y - Länge der Baumkrone

- Testen die Hypothesen:

H_0 : X und Y sind unabhängig

H_1 : X und Y sind nicht unabhängig

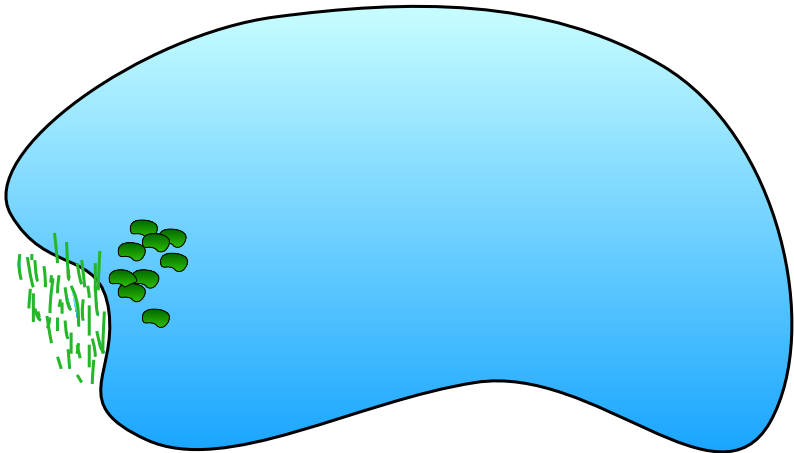
- empirische Ausprägungen x_1, \dots, x_n und y_1, \dots, y_n

z.B.: Southwest Oregon Data



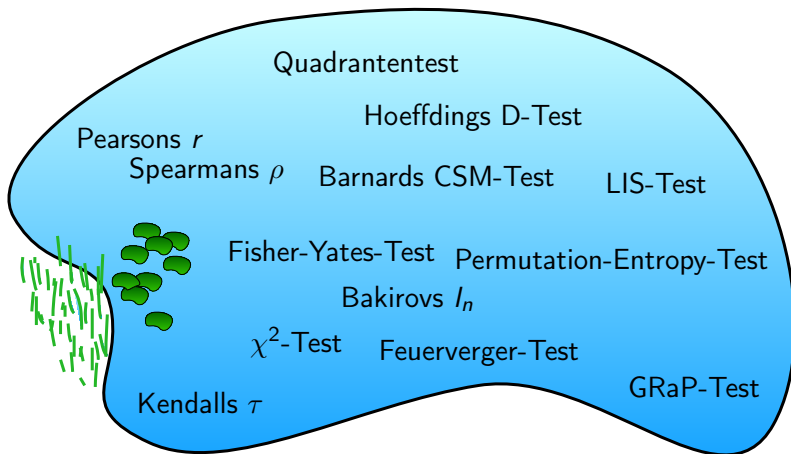


Pool mathematischer Unabhängigkeitstests



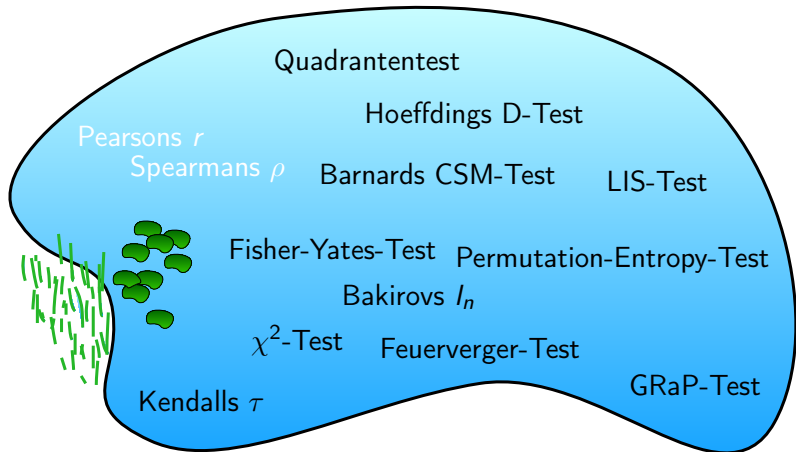


Pool mathematischer Unabhängigkeitstests



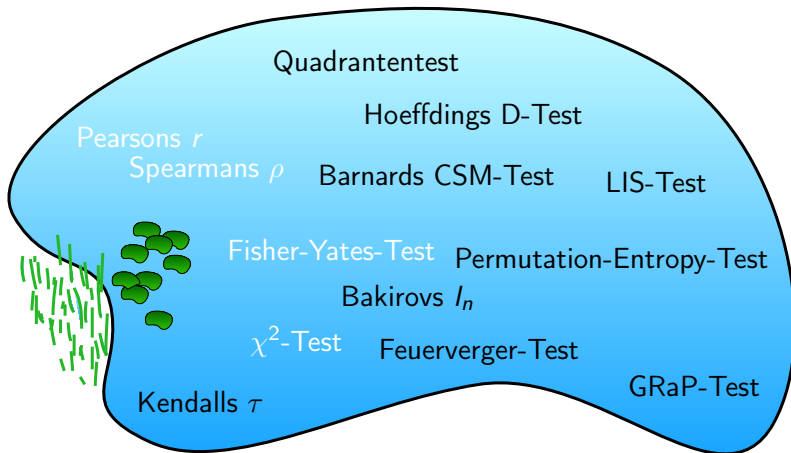


Pool mathematischer Unabhängigkeitstests



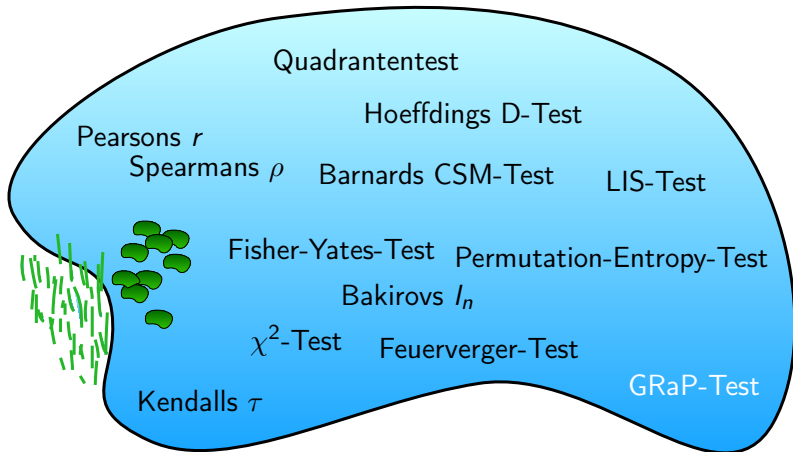


Pool mathematischer Unabhängigkeitstests



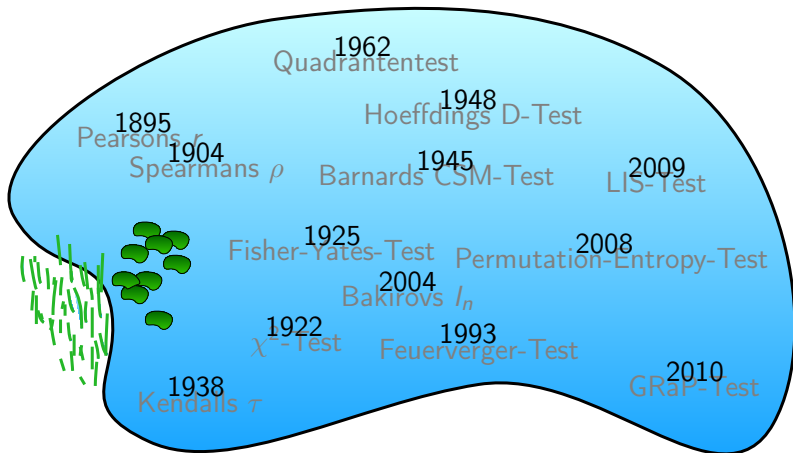


Pool mathematischer Unabhängigkeitstests





Pool mathematischer Unabhängigkeitstests





Was ist ein guter Test?

- Test akzeptiert H_0 bei unabhängigen Merkmalen
- Test lehnt H_0 ab bei abhängigen Merkmalen





Was ist ein guter Test?

- Test akzeptiert H_0 bei unabhängigen Merkmalen
- Test lehnt H_0 ab bei abhängigen Merkmalen

		Wirklichkeit	
		H_0	H_1
Test	H_0	☺	☹
	H_1	☹	☺





Was ist ein guter Test?

- Test akzeptiert H_0 bei unabhängigen Merkmalen
- Test lehnt H_0 ab bei abhängigen Merkmalen

		Wirklichkeit	
		H_0	H_1
Test	H_0	☺	☹
	H_1	☹	☺

- $P(H_1|H_0)$ Fehler 1.Art (α -Fehler, p -Wert)





Was ist ein guter Test?

- Test akzeptiert H_0 bei unabhängigen Merkmalen
- Test lehnt H_0 ab bei abhängigen Merkmalen

		Wirklichkeit	
		H_0	H_1
Test	H_0	😊	😞
	H_1	😞	😊





- $P(H_1|H_0)$ Fehler 1.Art (α -Fehler, p -Wert)
- $P(H_0|H_1)$ Fehler 2.Art (β -Fehler)





Was ist ein guter Test?

- Test akzeptiert H_0 bei unabhängigen Merkmalen
- Test lehnt H_0 ab bei abhängigen Merkmalen

		Wirklichkeit	
		H_0	H_1
Test	H_0		
	H_1		





- $P(H_1|H_0)$ Fehler 1.Art (α -Fehler, p -Wert)
- $P(H_0|H_1)$ Fehler 2.Art (β -Fehler)
- $P(H_0|H_0)$ Spezifität





Was ist ein guter Test?

- Test akzeptiert H_0 bei unabhängigen Merkmalen
- Test lehnt H_0 ab bei abhängigen Merkmalen

		Wirklichkeit	
		H_0	H_1
Test	H_0		
	H_1		

- $P(H_1|H_0)$ Fehler 1.Art (α -Fehler, p -Wert)
- $P(H_0|H_1)$ Fehler 2.Art (β -Fehler)
- $P(H_0|H_0)$ Spezifität
- $P(H_1|H_1)$ Sensitivität





Beispiel: Datensatz Oregon





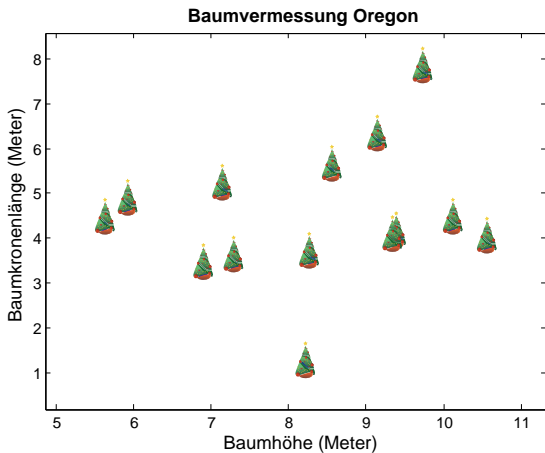
Datensatz

- Daten aus dem Organon Growth and Yield Project
- Baumbestand im Südwesten Oregons
- Variablen:
 - Standort
 - Baumgattung
 - Durchmesser auf Brusthöhe
 - Baumhöhe
 - Verhältnis Baumkrone-Baumhöhe
 - Anzahl d. Bäume pro Acre ($4046m^2$)
 - Radiales Wachstum





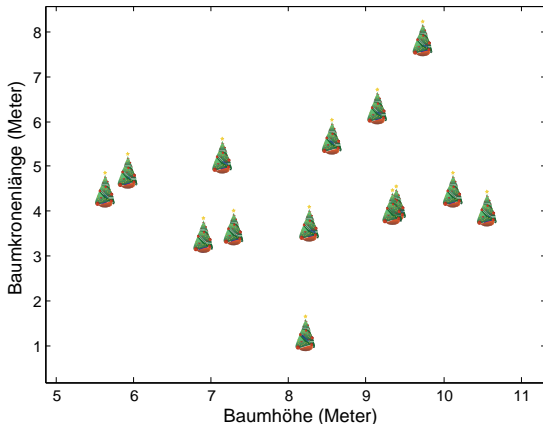
Nadelgehölze im Südwesten Oregons





Nadelgehölze im Südwesten Oregons

Baumvermessung Oregon



Pearsonkorrelation: $r = 0.19$
Spearmankorrelation: $r_s = 0.14$
Kendalls Tau: $\tau = 0.16$

p-Werte:

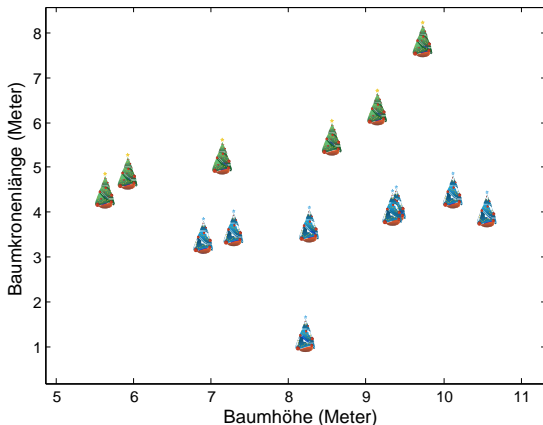
Pearson = 0.5157
Spearman = 0.6266
Chi2-Ind. = 0.5930
Fisher = 0.5000
Kendall = 0.4506
Quadranten = 0.5957
Barnard = 0.3953
Hoeffding = 0.2085
Feuerverger = 0.4604
Bakirov = 0.5550
Entropie-Test = 0.9680
LIS = 0.8073
GRaP = 0.0280





Nadelgehölze im Südwesten Oregons

Baumvermessung Oregon



Pearsonkorrelation: $r = 0.19$
Spearmankorrelation: $r_s = 0.14$
Kendalls Tau: $\tau = 0.16$

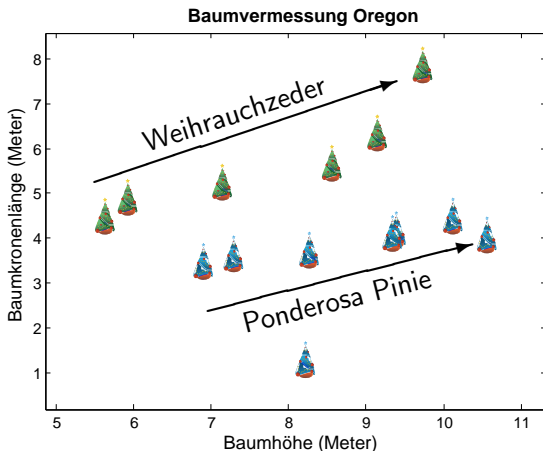
p-Werte:

Pearson = 0.5157
Spearman = 0.6266
Chi2-Ind. = 0.5930
Fisher = 0.5000
Kendall = 0.4506
Quadranten = 0.5957
Barnard = 0.3953
Hoeffding = 0.2085
Feuerverger = 0.4604
Bakirov = 0.5550
Entropie-Test = 0.9680
LIS = 0.8073
GRaP = 0.0280





Nadelgehölze im Südwesten Oregons

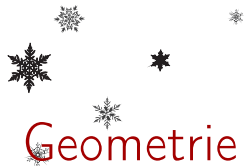


Pearsonkorrelation: $r = 0.19$
 Spearmankorrelation: $r_s = 0.14$
 Kendalls Tau: $\tau = 0.16$

p-Werte:

Pearson = 0.5157
 Spearman = 0.6266
 Chi2-Ind. = 0.5930
 Fisher = 0.5000
 Kendall = 0.4506
 Quadranten = 0.5957
 Barnard = 0.3953
 Hoeffding = 0.2085
 Feuerverger = 0.4604
 Bakirov = 0.5550
 Entropie-Test = 0.9680
 LIS = 0.8073
 GRaP = 0.0280





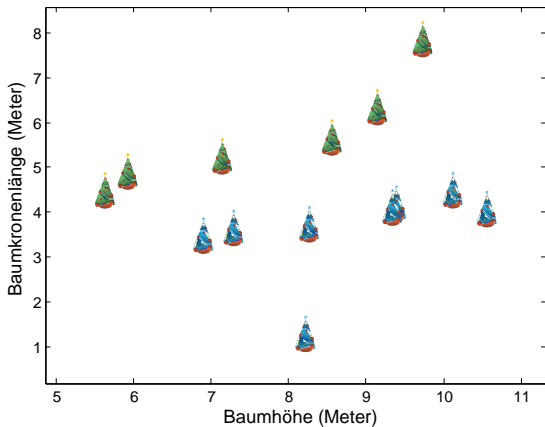
Geometrie Randomisierter Permutationen (GRaP)





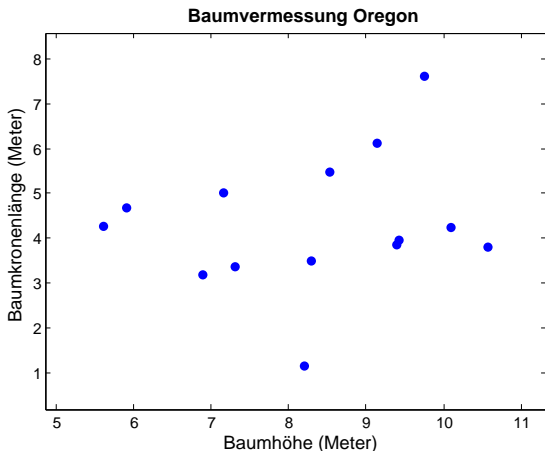
GRaP Independence Test

Baumvermessung Oregon



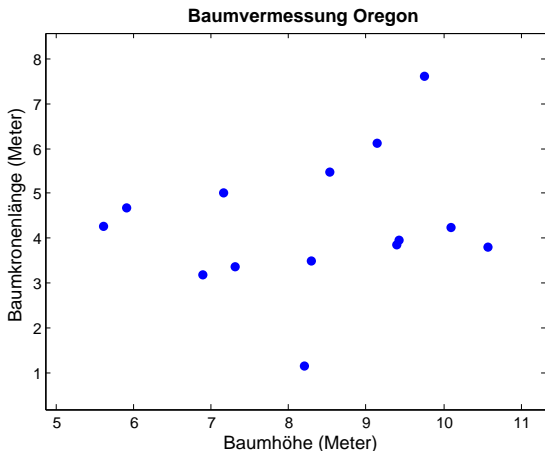


GRaP Independence Test



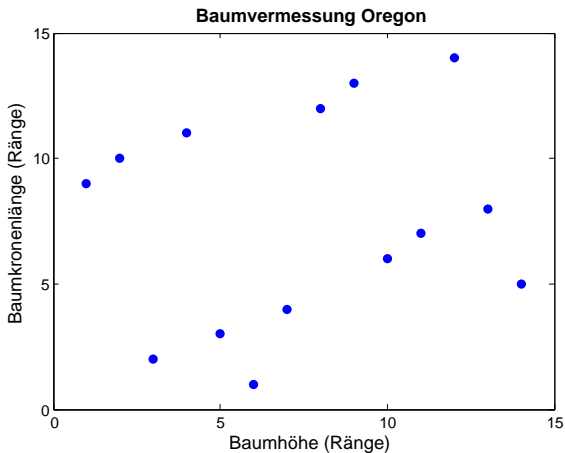


GRaP Independence Test



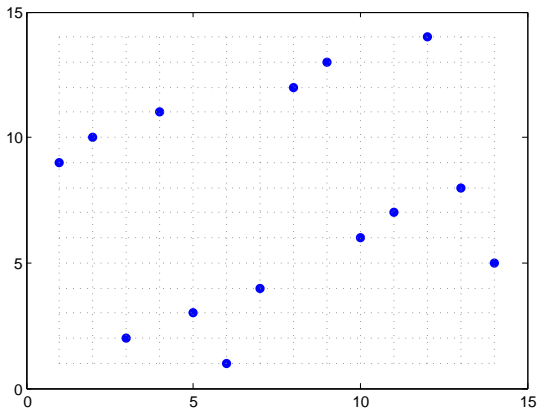


GRaP Independence Test





GRaP Independence Test

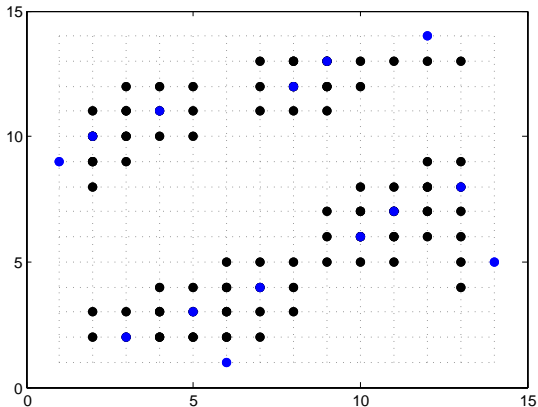


→ Ränge





GRaP Independence Test

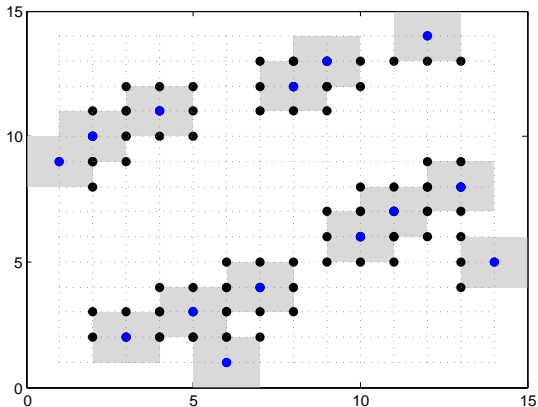


→ Ränge
 → Punktumgebungen





GRaP Independence Test

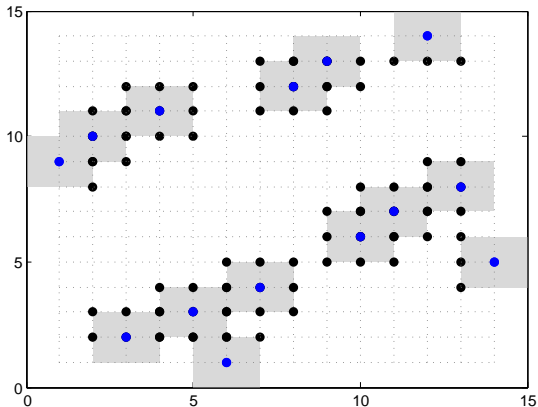


→ Ränge
→ Punktumgebungen





GRaP Independence Test

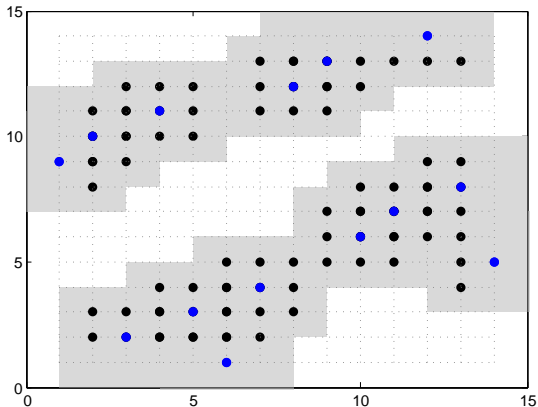


- Ränge
 - Punktumgebungen
- Anzahl: $S_1=70$





GRaP Independence Test

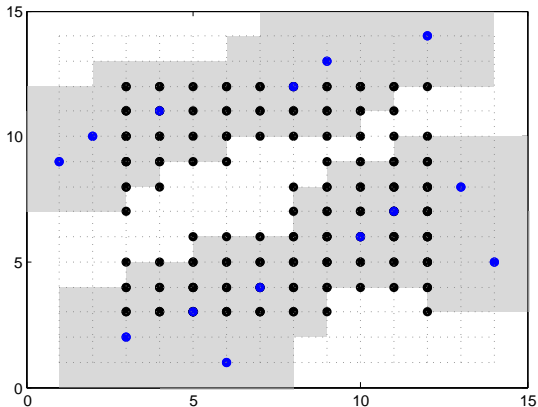


- Ränge
 - Punktumgebungen
- Anzahl: $S_1=70$





GRaP Independence Test

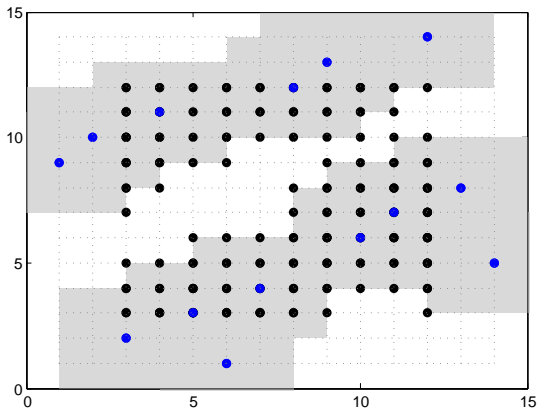


- Ränge
 - Punktumgebungen
- Anzahl: $S_1=70$





GRaP Independence Test

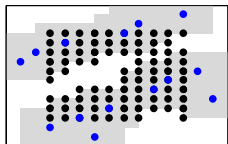


- Ränge
 - Punktumgebungen
- Anzahl: $S_1=70$
 $S_2=86$

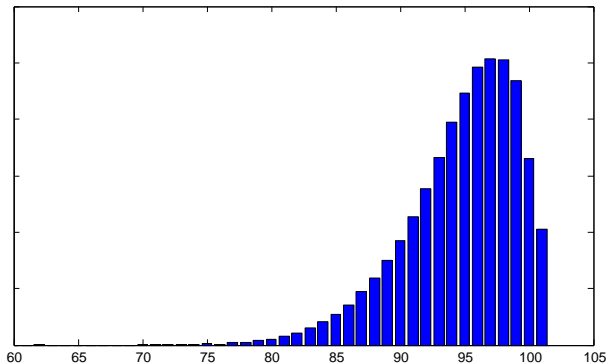




Verteilung zufälliger Graphen

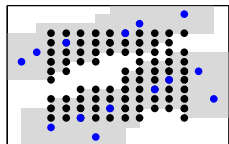


$$\rightarrow S_2=86$$

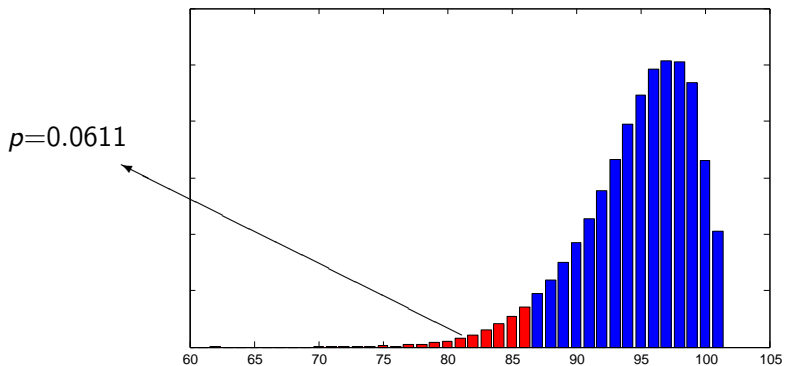




Verteilung zufälliger Graphen

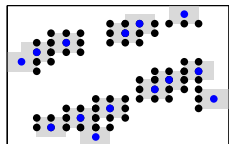


→ $S_2=86$

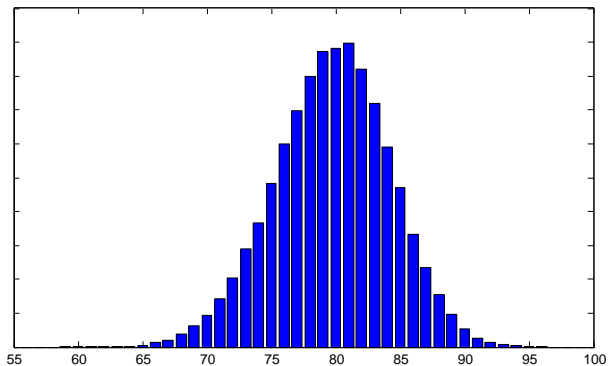




Verteilung zufälliger Graphen

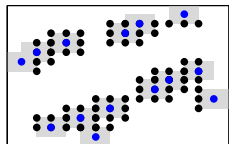


$$\rightarrow S_1=70$$



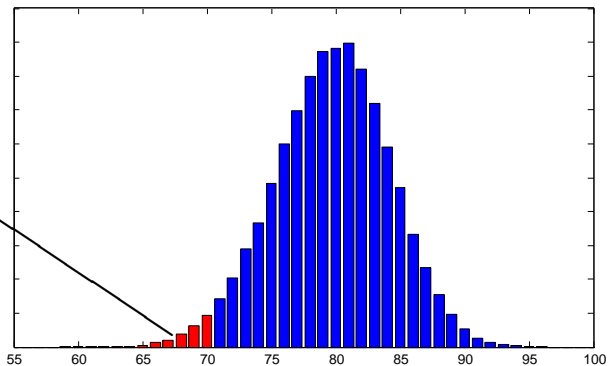


Verteilung zufälliger Graphen



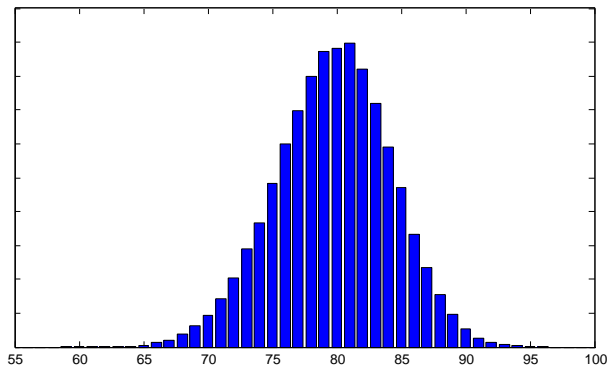
→ $S_1=70$

$p=0.0280$



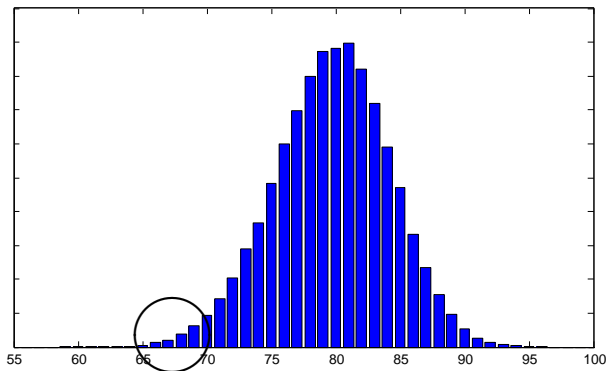


Verteilung zufälliger Graphen



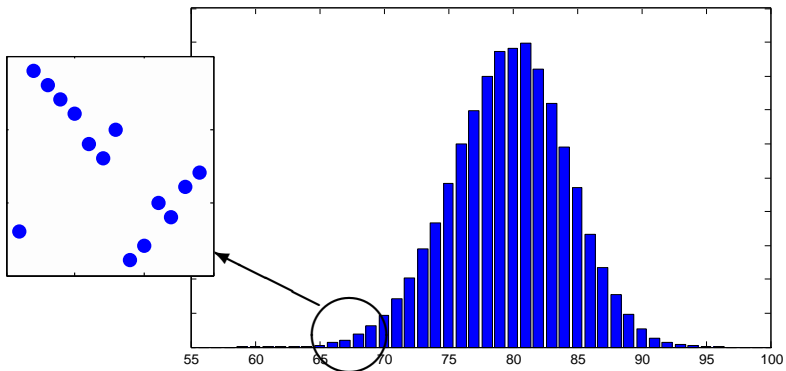


Verteilung zufälliger Graphen



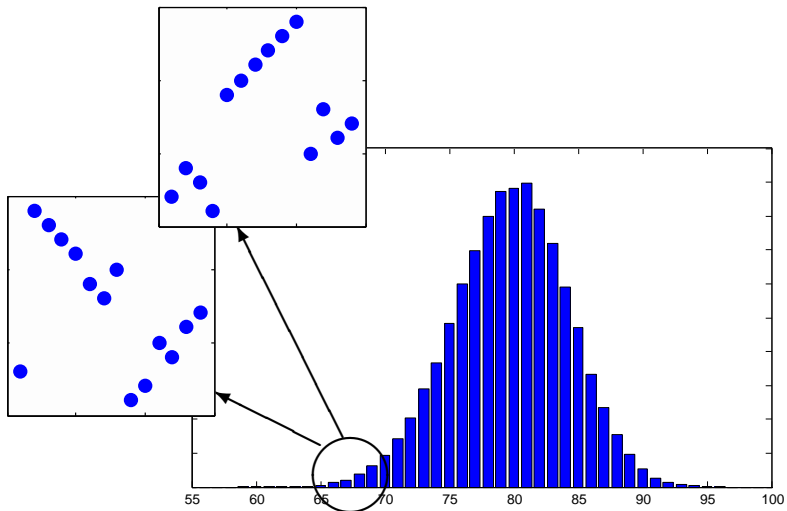


Verteilung zufälliger Graphen



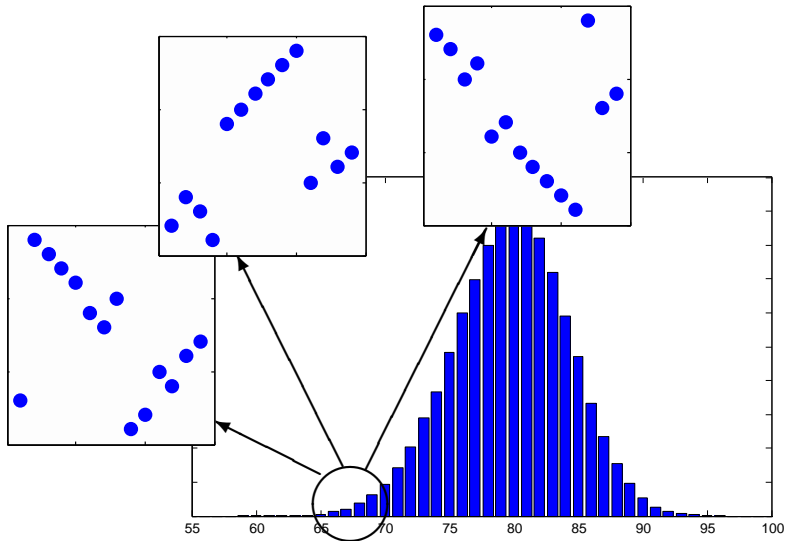


Verteilung zufälliger Graphen



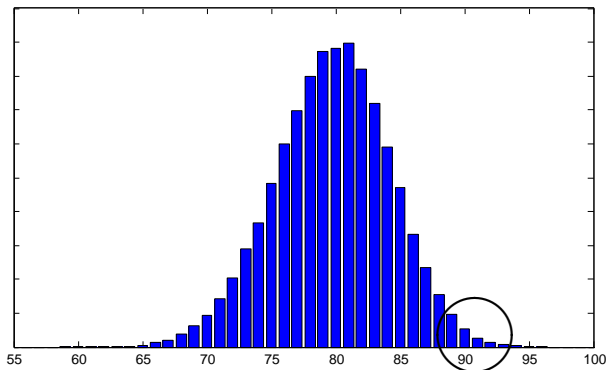


Verteilung zufälliger Graphen



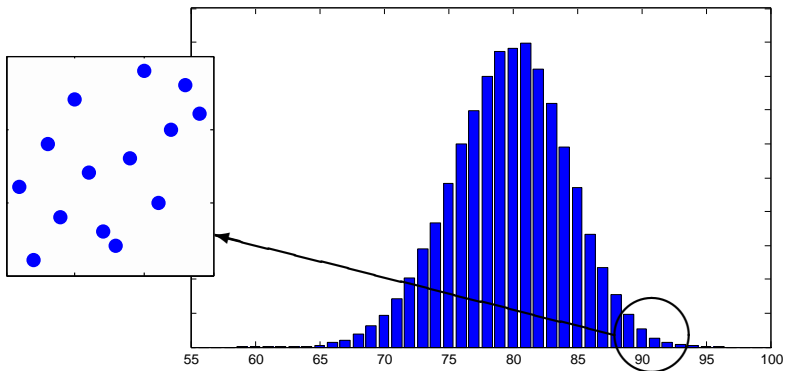


Verteilung zufälliger Graphen



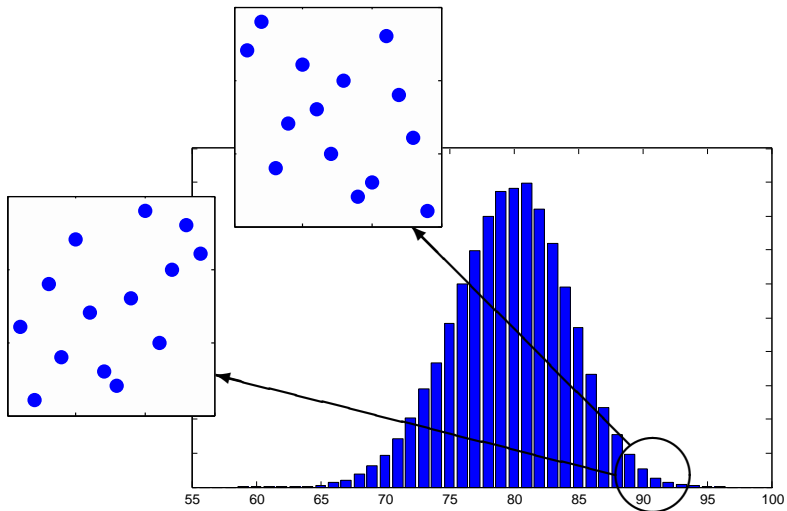


Verteilung zufälliger Graphen



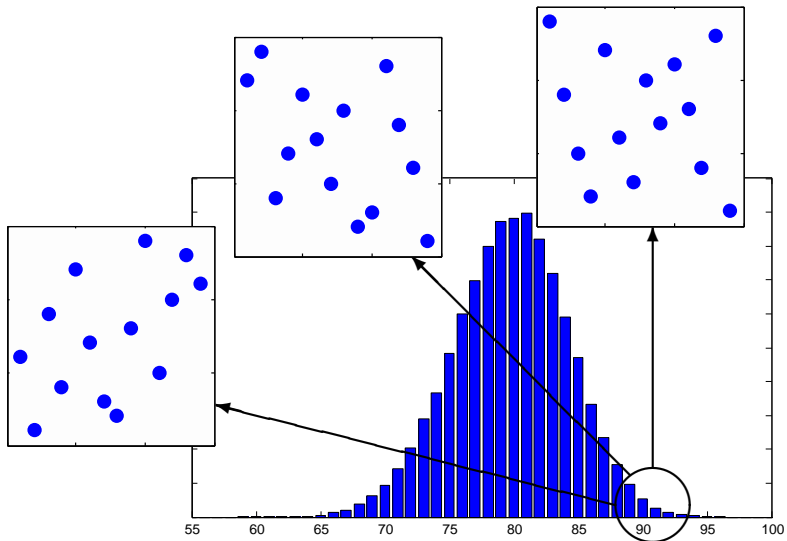


Verteilung zufälliger Graphen





Verteilung zufälliger Graphen

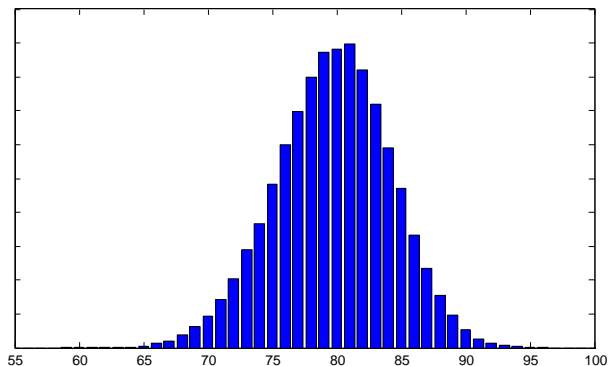




Verteilung zufälliger Graphen

H_0 : X und Y sind unabhängig

H_1 : X und Y sind nicht unabhängig

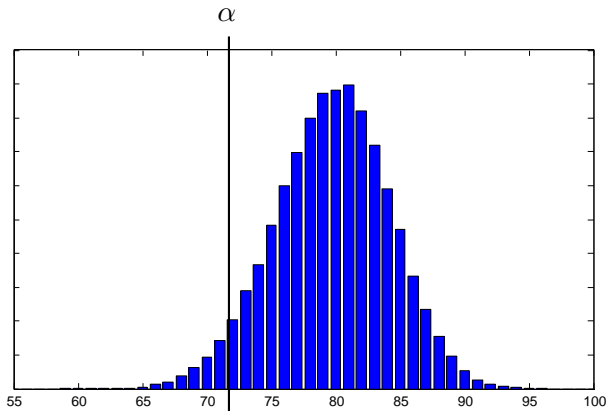




Verteilung zufälliger Graphen

H_0 : X und Y sind unabhängig

H_1 : X und Y sind nicht unabhängig

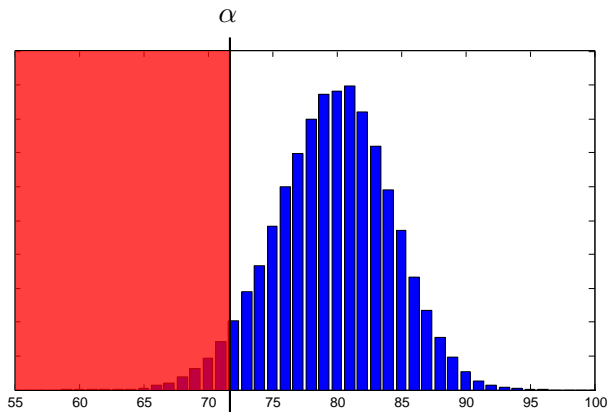




Verteilung zufälliger Graphen

H_0 : X und Y sind unabhängig

H_1 : X und Y sind nicht unabhängig

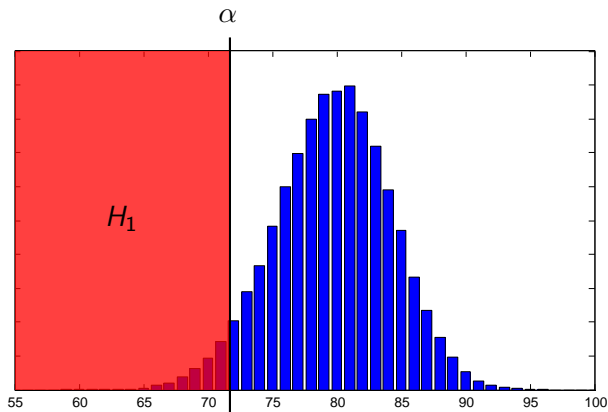




Verteilung zufälliger Graphen

H_0 : X und Y sind unabhängig

H_1 : X und Y sind nicht unabhängig

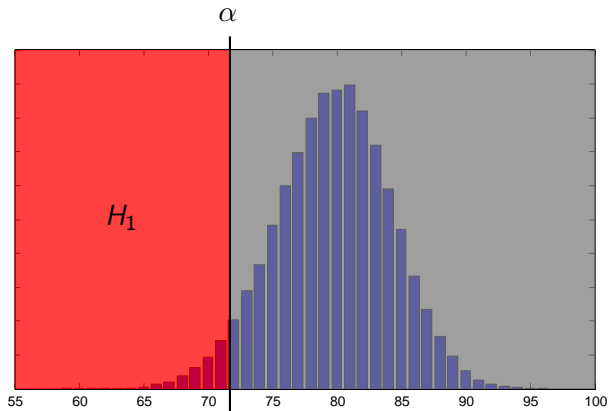




Verteilung zufälliger Graphen

H_0 : X und Y sind unabhängig

H_1 : X und Y sind nicht unabhängig

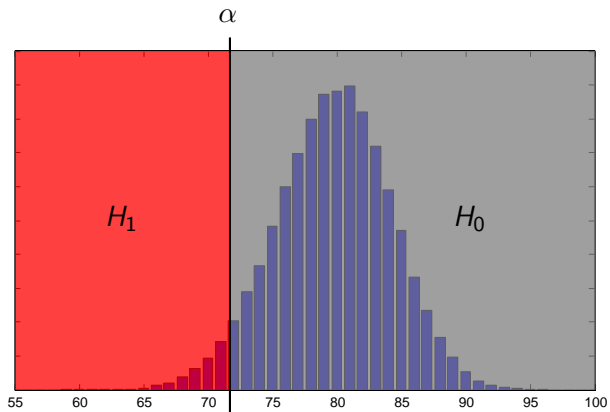




Verteilung zufälliger Graphen

H_0 : X und Y sind unabhängig

H_1 : X und Y sind nicht unabhängig





Kenngößen der Verteilung

Mittelwert

$$M(S_k) = (n-2k)^2 - \frac{(n-2k)^{2k+2}}{n^{2k}}$$

Minimum

$$\min(S_k) = (n-4k)(4k+1)$$

n - Stichprobenumfang

k - Radius der Umgebung





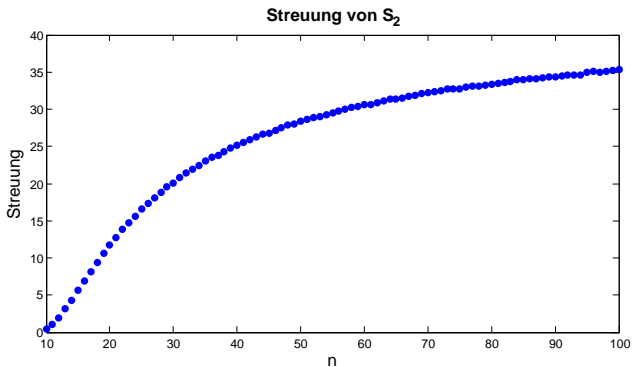
Kenngößen der Verteilung

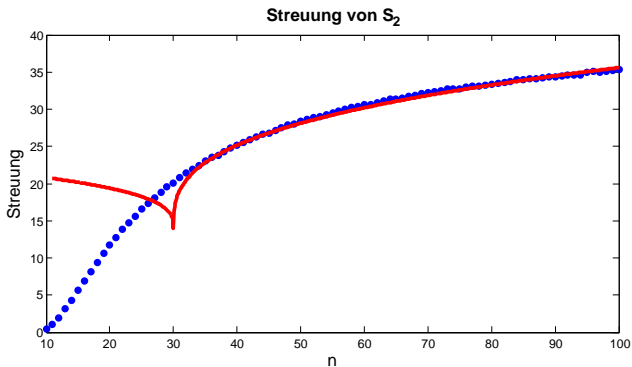
Maximum

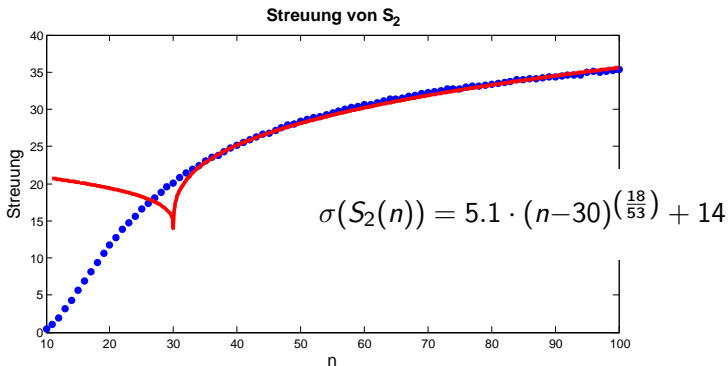
$$\max(S_k) = \begin{cases} (n-2)(2k+1)^2 + 1 - 4k(4k^2-1) & , \text{für } \frac{n+j}{2k+1} > 2k+2 \text{ und } j=2k \\ (n-2k)^2 & , \text{für } \frac{n+j}{2k+1} < 2k+2 \\ n(2k+1)^2 - 4k(1+2k+4k^2) & , \text{sonst} \end{cases}$$

mit $\frac{n+j}{2k+1} \in \mathbb{N}$ und $j=0, \dots, 2k$



Streuung von S_2 

Streuung von S_2 

Streuung von S_2 



Der p-Wert

Berechnung des p-Wertes

- 1 Quantiltabelle





Der p-Wert

Berechnung des p-Wertes

1 Quantiltabelle

- Ablesen des Signifikanzniveaus aus Tabellen
- Grundlage dieser Tabellen ist eine Simulation der Verteilungen von S_k





Der p-Wert

Berechnung des p-Wertes

- 1 Quantiltabelle
 - Ablesen des Signifikanzniveaus aus Tabellen
 - Grundlage dieser Tabellen ist eine Simulation der Verteilungen von S_k
- 2 Normalverteilung





Der p-Wert

Berechnung des p-Wertes

1 Quantiltabelle

- Ablesen des Signifikanzniveaus aus Tabellen
- Grundlage dieser Tabellen ist eine Simulation der Verteilungen von S_k

2 Normalverteilung

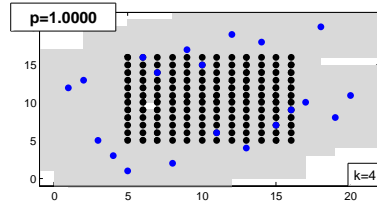
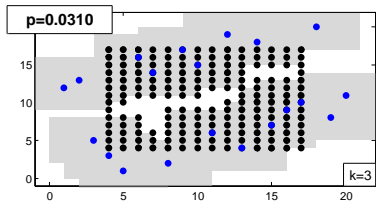
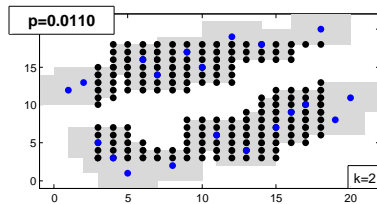
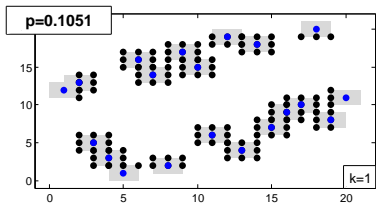
- Nutzung des exakten Mittelwertes
- Nutzung der geschätzten Streuung

$$\hat{p}_k = \Phi \left(\frac{s_k - \mu_k}{\hat{\sigma}(S_k)} \right)$$



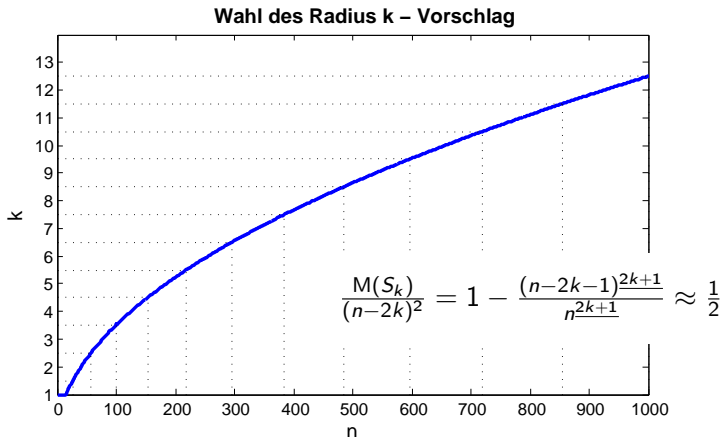


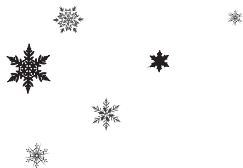
Die richtige Wahl des Radius k





Die richtige Wahl des Radius k





Power Analyse



Betrachtete Verteilungen

- 1 Unabhängigkeit
- 2 Abhängigkeit

Unabhängigkeit

Abhängigkeit



Betrachtete Verteilungen

- 1 Unabhängigkeit
 - Bivariate Gleichverteilung
 - Bivariate Normalverteilung
- 2 Abhängigkeit

Unabhängigkeit

Abhängigkeit



Betrachtete Verteilungen

1 Unabhängigkeit

- Bivariate Gleichverteilung
- Bivariate Normalverteilung

2 Abhängigkeit

- Lineare Verteilung
- Exponentielle Verteilung
- "Doppelt lineare" Verteilung
- Bivariater Normalverteilungsmix
- Sinusoide Verteilung
- Kreisförmige Verteilung

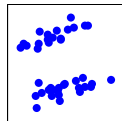
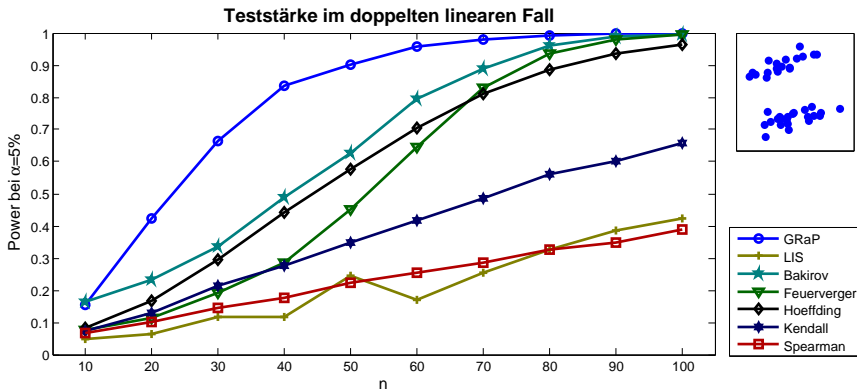
Unabhängigkeit

Abhängigkeit





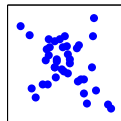
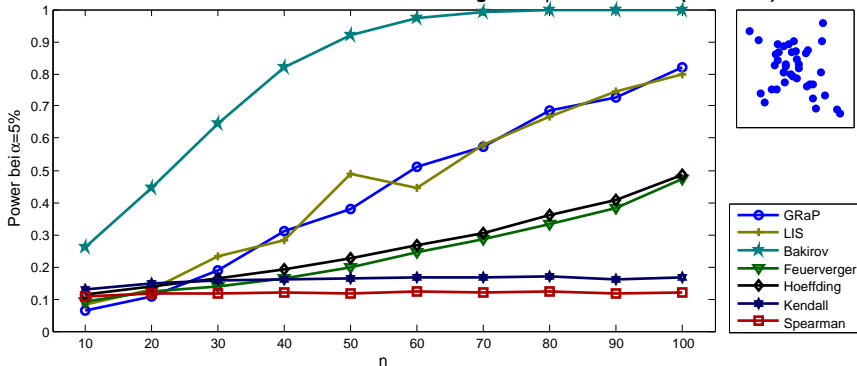
Teststärke bei Abhängigkeit





Teststärke bei Abhängigkeit

Teststärke im Fall zweier bivariater Normalverteilungen mit $\rho=0.9$ und -0.9 (50:50 Mix)



$\rho = 0.9$

$\rho = 0.8$

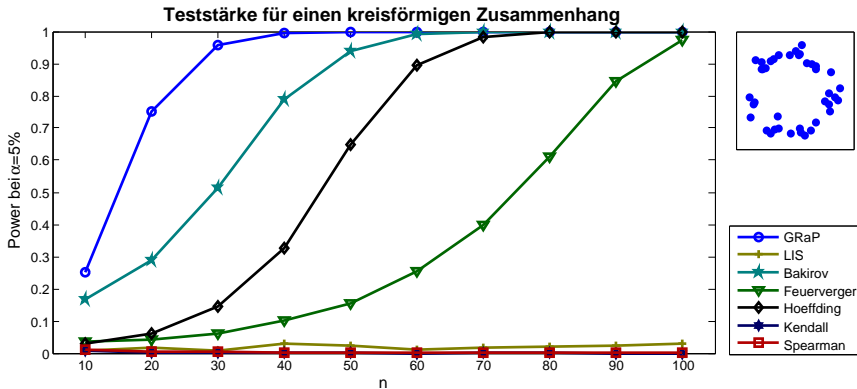
$\rho = 0.7$

$\rho = 0.6$





Teststärke bei Abhängigkeit

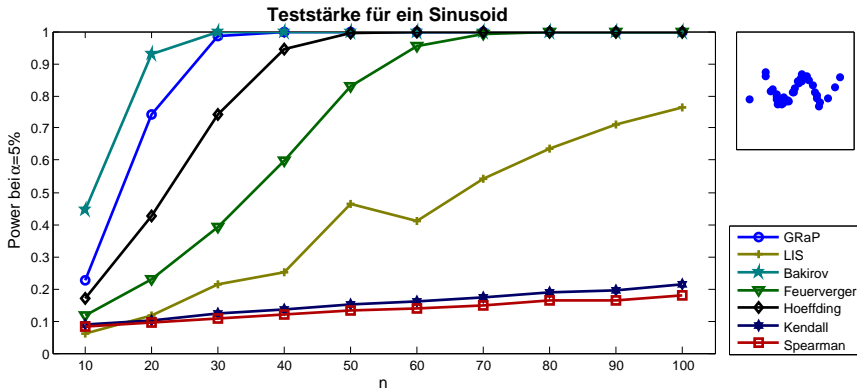


mehr Streuung



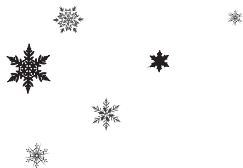


Teststärke bei Abhängigkeit



mehr Streuung





Beispiele

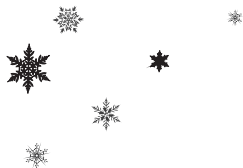




IndepTestTool

Weine





Hauptresultate





Resultate

GRaP Independence Test

- Hohe Teststärke





Resultate

GRaP Independence Test

- Hohe Teststärke
- Verständliche und leicht zu berechnende Teststatistik





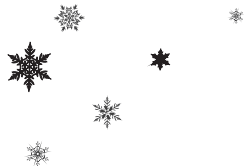
Resultate

GRaP Independence Test

- Hohe Teststärke
- Verständliche und leicht zu berechnende Teststatistik
- Gute asymptotische Eigenschaften











Referenzen









ausgewählte Literatur

- K. Pearson (1895), *Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material*, Philosophical Transactions of the Royal Society of London. A, Vol. 186, pp. 343-414. 
- C. Spearman (1904), *The Proof and Measurement of Association between Two Things*, The American Journal of Psychology, Vol. 15, pp. 72-101. 
- R. A. Fisher (1922), *On the Interpretation of χ^2 from Contingency Tables, and the Calculation of P*, Journal of the Royal Statistical Society, Vol. 85, pp. 87-94. 
- M. G. Kendall (1938), *A New Measure Of Rank Correlation*, Biometrika, Vol. 30, pp. 81-93. 








ausgewählte Literatur

- G. A. Barnard (1947), *Significance Tests for 2×2 Tables*, Biometrika, Vol. 34, pp. 123-138. 
- W. Hoeffding (1948), *A Non-Parametric Test of Independence*, The Annals of Mathematical Statistics, Vol. 19, pp. 546-557. 
- RCE1962R. C. Elandt (1962), *Exact and Approximate Power Function of the Non-Parametric Test of Tendency*, The Annals of Mathematical Statistics, Vol. 33, pp. 471-481. 
- F1993A. Feuerverger (1993), *A Consistent Test for Bivariate Dependence*, International Statistical Review / Revue Internationale de Statistique, Vol. 61, pp. 419-433. 

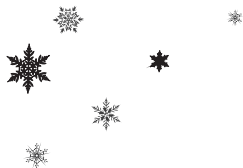




ausgewählte Literatur

- N. K. Bakirov, M. L. Rizzo, G. J. Székely (2006), *A multivariate nonparametric test of independence*, Journal of Multivariate Analysis, Vol. 97, pp. 1742-1756. 
- M. Matilla-García, M. R. Marín (2008), *A non-parametric independence test using permutation entropy*, Journal of Econometrics, Vol. 144, pp. 139-155. 
- J. E. García, V. A. González-López (2009), *A Nonparametric Independence Test using Random Permutations*, Preprint, arXiv:0908.2794v2. 



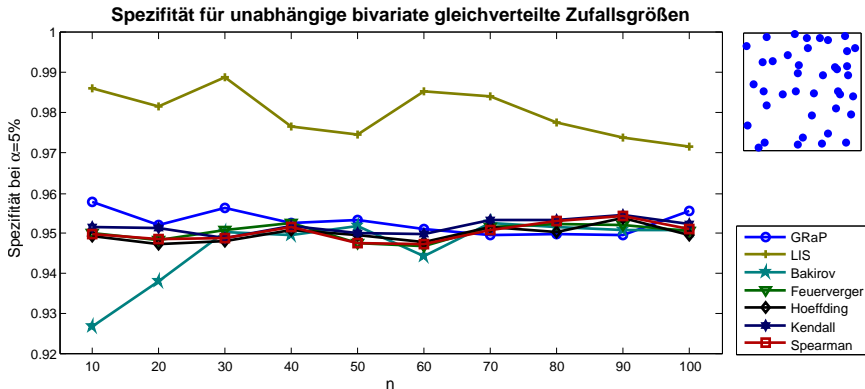


Anhang





Spezifität bei Unabhängigkeit

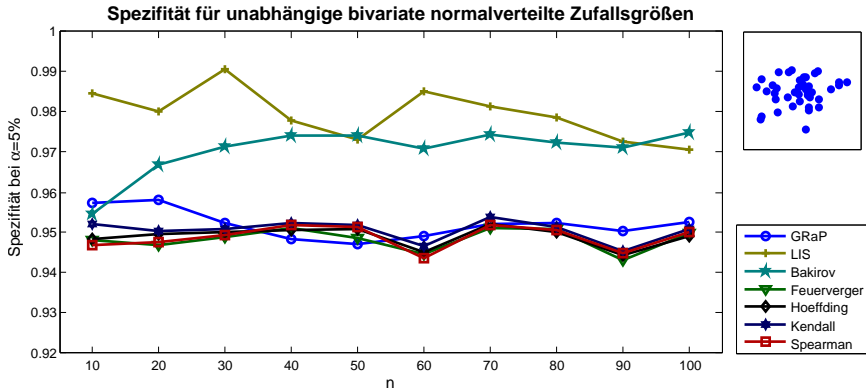


zurück





Spezifität bei Unabhängigkeit

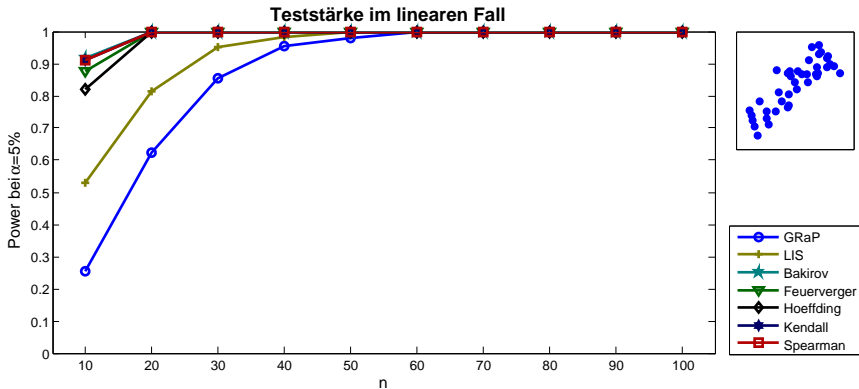


zurück





Teststärke bei Abhängigkeit

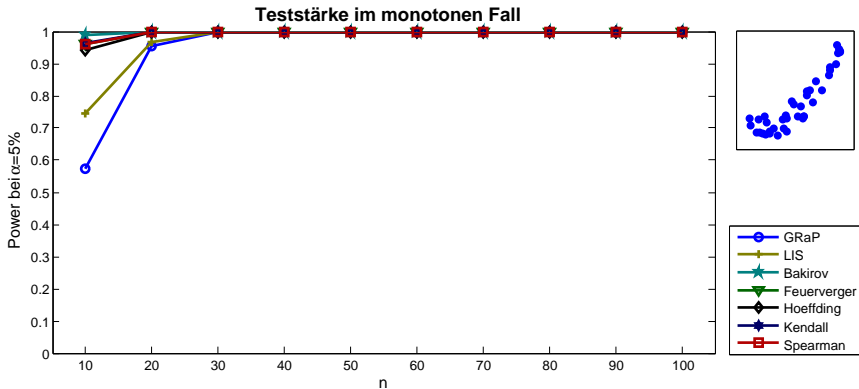


zurück





Teststärke bei Abhängigkeit

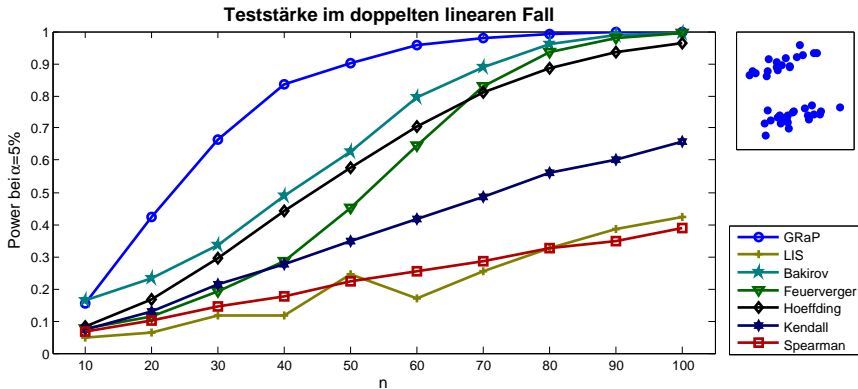


zurück





Teststärke bei Abhängigkeit



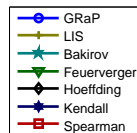
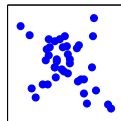
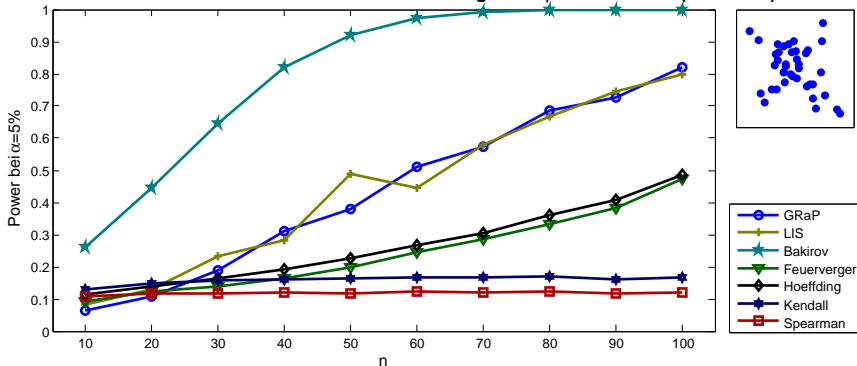
zurück





Teststärke bei Abhängigkeit

Teststärke im Fall zweier bivariater Normalverteilungen mit $\rho=0.9$ und -0.9 (50:50 Mix)



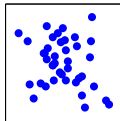
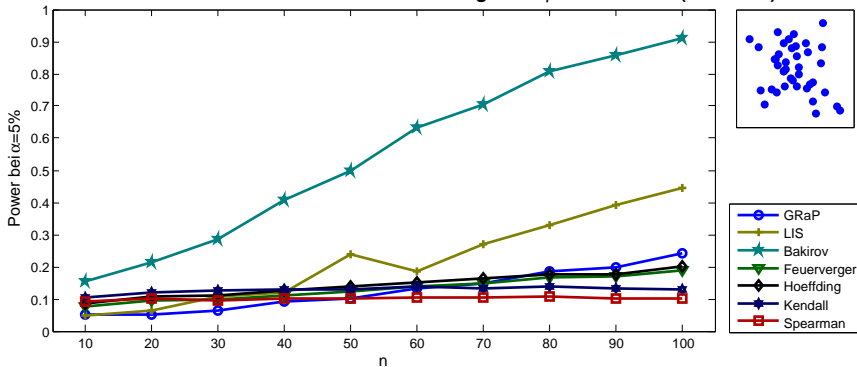
zurück





Teststärke bei Abhängigkeit

Teststärke im Fall zweier bivariater Normalverteilungen mit $\rho=0.8$ und -0.8 (50:50 Mix)



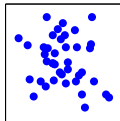
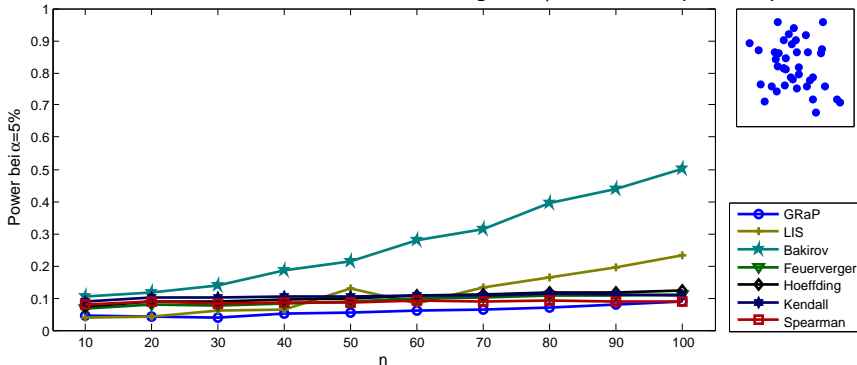
zurück





Teststärke bei Abhängigkeit

Teststärke im Fall zweier bivariater Normalverteilungen mit $\rho=0.7$ und -0.7 (50:50 Mix)



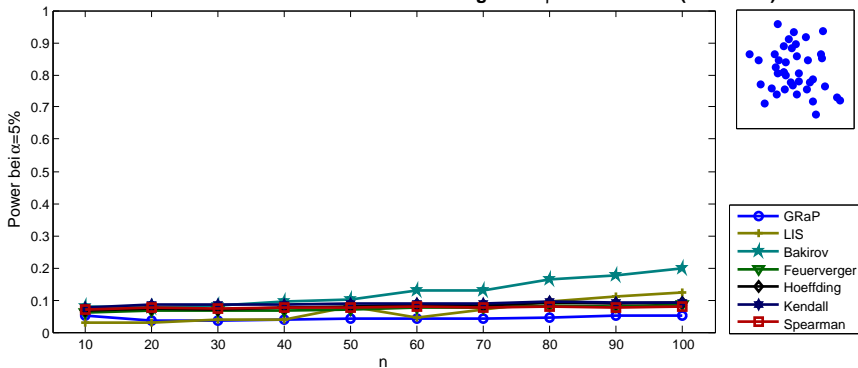
zurück





Teststärke bei Abhängigkeit

Teststärke im Fall zweier bivariater Normalverteilungen mit $\rho=0.6$ und -0.6 (50:50 Mix)

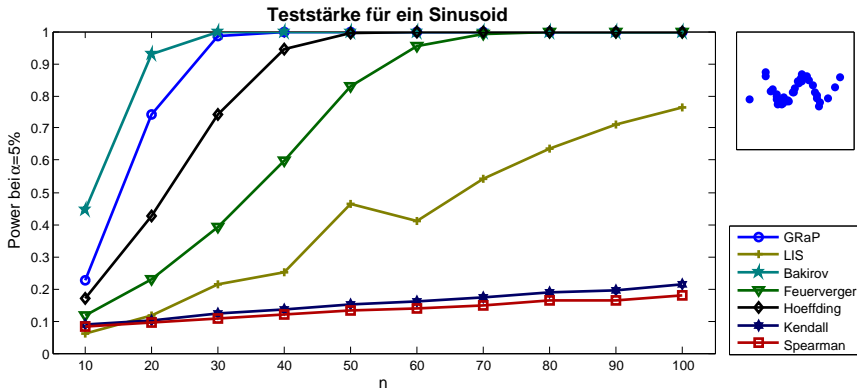


zurück





Teststärke bei Abhängigkeit

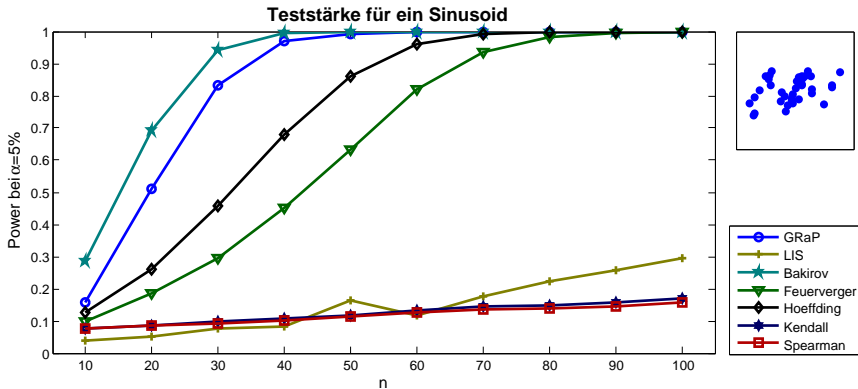


zurück





Teststärke bei Abhängigkeit

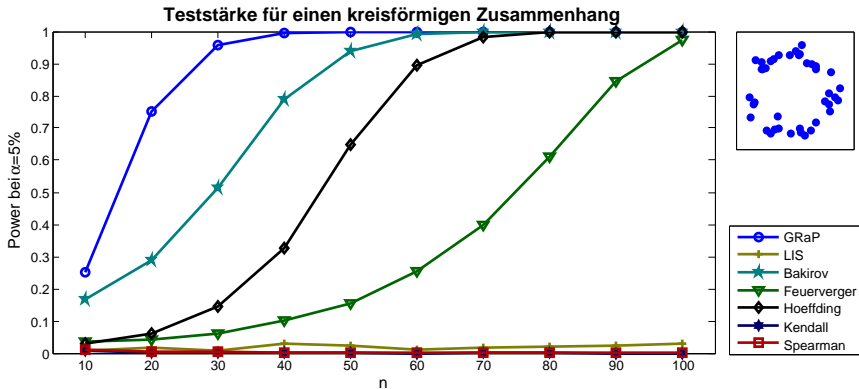


zurück





Teststärke bei Abhängigkeit

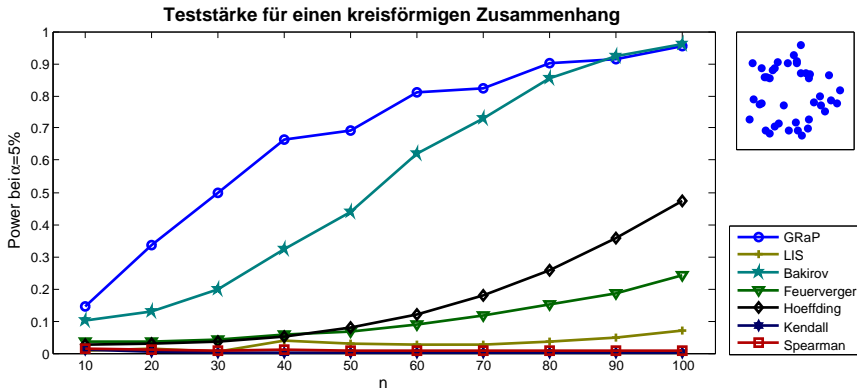


zurück





Teststärke bei Abhängigkeit



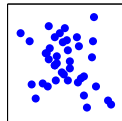
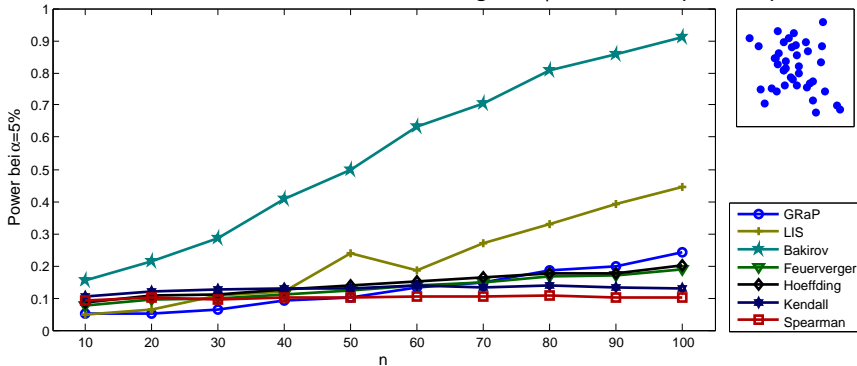
zurück





Teststärke bei Abhängigkeit

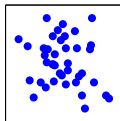
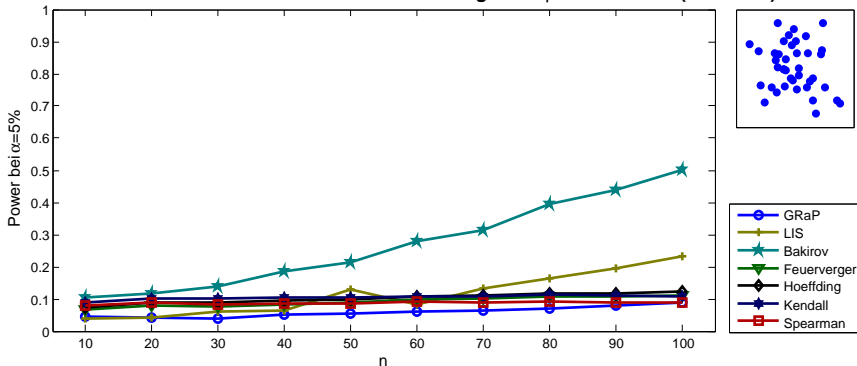
Teststärke im Fall zweier bivariater Normalverteilungen mit $\rho=0.8$ und -0.8 (50:50 Mix)

 $\rho = 0.9$ $\rho = 0.8$ $\rho = 0.7$ $\rho = 0.6$ 



Teststärke bei Abhängigkeit

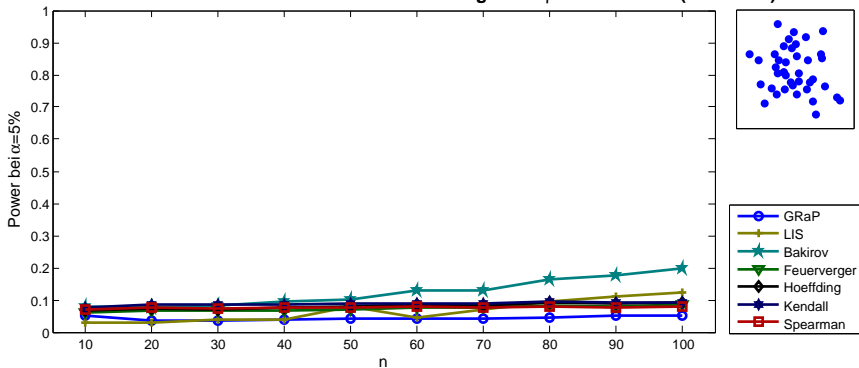
Teststärke im Fall zweier bivariater Normalverteilungen mit $\rho=0.7$ und -0.7 (50:50 Mix)

 $\rho = 0.9$ $\rho = 0.8$ $\rho = 0.7$ $\rho = 0.6$ 



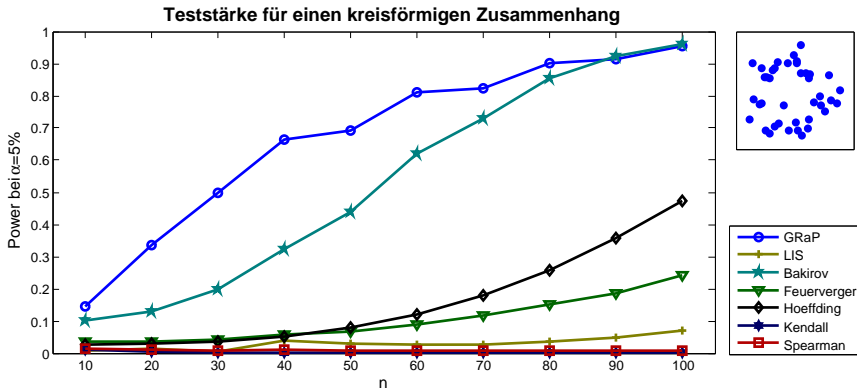
Teststärke bei Abhängigkeit

Teststärke im Fall zweier bivariater Normalverteilungen mit $\rho=0.6$ und -0.6 (50:50 Mix)

 $\rho = 0.9$ $\rho = 0.8$ $\rho = 0.7$ $\rho = 0.6$ 



Teststärke bei Abhängigkeit

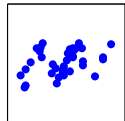
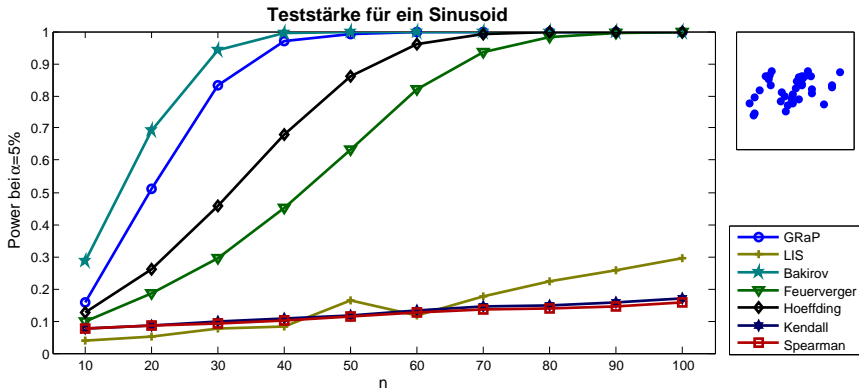


zurück





Teststärke bei Abhängigkeit

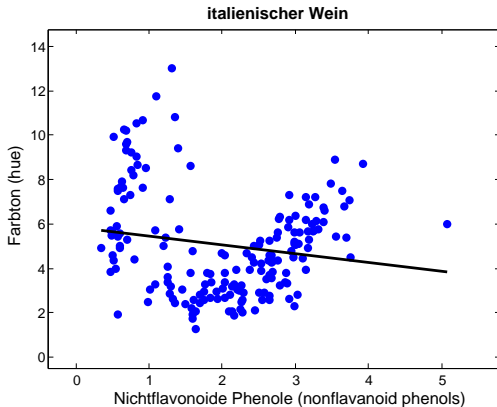


zurück





Chemische Analyse von Weinen



Pearsonkorrelation: $r = -0.17$
 Spearmankorrelation: $r_s = -0.04$
 Kendalls Tau: $\tau = 0.03$

p-Werte:

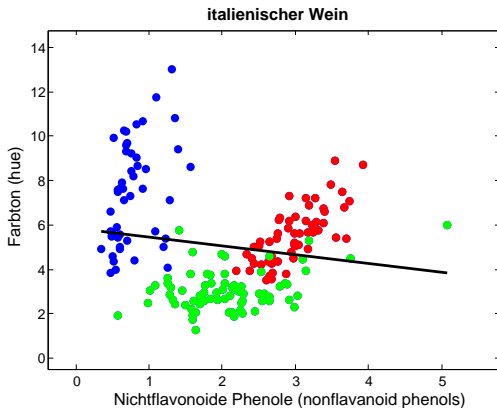
Pearson = 0.0214
 Spearman = 0.5695
 Kendall = 0.5721
 Hoeffding = 0.0000
 Bakirov = 0.0085
 LIS = 0.0056
 GRaPu = 0.0000
 GRaPf = 0.0000

zurück





Chemische Analyse von Weinen



Pearsonkorrelation: $r = -0.17$
 Spearmankorrelation: $r_s = -0.04$
 Kendalls Tau: $\tau = 0.03$

p-Werte:

Pearson = 0.0214
 Spearman = 0.5695
 Kendall = 0.5721
 Hoeffding = 0.0000
 Bakirov = 0.0085
 LIS = 0.0056
 GRaPu = 0.0000
 GRaPf = 0.0000

zurück

