Verteilungstests

Voraussetzung und verwandte Themen

Für diese Beschreibungen sind Grundlagen der Statistik vorteilhaft. Weiterführende und verwandte Themen sind:

www.versuchsmethoden.de/Normalverteilung.pdf

www.versuchsmethoden.de/Hypothesentests.pdf

Stichworte: Verteilungen – Normalverteilungstest – Hypothesentests – Anderson-Darling – Shapiro-Wilk - Chi²-Test

Einführung

Die folgenden Tests dienen dazu, ob eine Datenreihe einer vorgegebenen statistischen Verteilung entspricht. In den meisten Fällen wird auf Normalverteilung geprüft.

Ziel und Nutzen

Nur wenn man von einer bestimmten Verteilung ausgehen kann, dürfen weitere bestimmte Methoden verwendet werden. So ist z.B. die Standardberechnung auf Prozessfähigkeit nur bei Normalverteilung zulässig, ebenso der t-Test für den Mittelwerttest.

Grundlagen

Bevor man sich mit den entsprechenden Hypothesentests auseinandersetzt, ist es dringend zu empfehlen, sich die Daten visuell anzuschauen. Im Folgenden werden die wichtigsten Tests für die Normalverteilung aufgezeigt.

Übersicht der vorgestellten Tests:

▪ Shapiro-Wilk (SW)	$n < 50$, Standard im deutschsprachigem Raum, z.B. DGQ e.V.
▪ χ^2-Test (CH)	Test für $n \geq 50$, Klassierung der Daten immer nötig.
▪ Anderson-Darling (AD)	Standard in Minitab® & Six Sigma und im englisch sprachigem Raum.
▪ Kolmog,-Smirnov (KS)	Klassiker nach Literatur mit Klassierung, aber auch ohne Klassierung möglich.
▪ Cramér von Mises (CM)	relativ stark abhängig von Stichprobengröße.
▪ Epps-Pulley (EP)	weniger verbreitet, tolerant gegen Abweichungen.
▪ Jarque-Bera Test (JB)	weniger verbreitet, sehr tolerant gegen Abweichungen.

Eine Stichprobengröße von mehr als 125 ist nicht zu empfehlen. Aufgrund des immer enger werdenden Vertrauensbereiches führen kleinste Abweichungen der Daten schnell zu einer Ablehnung der getesteten Verteilung.

Bei all den Tests soll auch berücksichtigt werden, dass diese nur sinnvoll sind, wenn die Daten aus „technischer“ Sicht auch als normalverteilt anzunehmen sind. Z.B. bei Lebensdauerdaten würde man gleich auf eine Weibull-Verteilung hin testen.

Wichtig ist das Wissen um den technischen Zusammenhang und die zu erwartende Verteilung! Es ist sehr davon abzuraten Daten unbekannte Herkunft und Sachlage zu verwenden!

Es gibt Verteilungstests die eine Klassierung der Daten voraussetzen, z.B. der χ^2 Test. Dabei hängt das Ergebnis von der gewählten Klassenbreite ab, was als Nachteil angesehen wird. Der Vorteil der Klassierung ist dagegen, wenn das Mess-System eine ungenügende Auflösung besitzt und damit viele gleiche Werte vorkommen. Das bedeutet eine „Vorklassierung“, mit der die Verfahren ohne Klassierung evtl. zu einer Fehlinterpretation kommen.

Shapiro-Wilk Test

Der Shapiro-Wilk-Test prüft die Nullhypothese, dass eine Stichprobe aus einer normalverteilten Grundgesamtheit stammt. Dieser Test wurde 1965 von Samuel Shapiro und Martin Wilk veröffentlicht und wird als leistungsfähiger Standardtest zur Bestimmung der Normalverteilung angesehen.

Die Prüfgröße wird berechnet als:

$$w = \frac{b^2}{(n-1)\,s^2}$$

$$b = \sum_{i=1}^{n} a_i x_i$$

Der Test weist die Nullhypothese zurück, wenn die Prüfgröße w kleiner ist, als der kritische Wert. Ist $w < w_{krit}$ wird die Nullhypothese auf dem Signifikanzniveau α abgelehnt. Der Prüfwert und die Gewichtungen a sind u.a. *Peterson, Grundlagen der Statistik* zu entnehmen. Alternativ zum Prüfwert kann auch auf ein $T_{prüf}$ umgerechnet werden.

$$T_{prüf} = \gamma + \delta + \ln\left(\frac{w-\varepsilon}{1-w}\right)$$

Die Koeffizienten γ, δ und ε sind ebenfalls Peterson zu entnehmen. Diese Prüfgröße kann direkt mit dem u-Wert der Normalverteilung verglichen werden. Für ein Signifikanzniveau von α=0,05 ist dieser -1,645. Wenn $T_{prüf}$ < -1,645, dann ist die Nullhypothese auf Normalverteilung abzulehnen.

Kolmogorov-Smirnov Anpassungstest

Der Kolmogorov-Smirnov Anpassungstest (kurz KS-Test) prüft die Anpassung einer beobachteten an eine beliebige zu erwartende Verteilung. Besonders beim Vorliegen kleiner Stichprobenumfänge entdeckt der KS-Test eher Abweichungen von der Normalverteilung. Verteilungsirregularitäten sind im Allgemeinen besser mittels des χ^2-Tests nachzuweisen. Der KS-Test kann für stetige und diskrete Verteilungen und für Stichproben im Umfang bis 50 angewendet werden. Ab $n \geq 50$ ist der χ^2-Anpassungstest zu verwenden.
Geprüft wird die Nullhypothese: Die Stichprobe entstammt der bekannten Verteilung. Es werden für jeden Wert die relativen Summenhäufigkeiten verglichen und der maximale Differenzwert als Prüfgröße $T_{prüf}$ verwendet.

$T_{prüf} = \max|H_B - H_E|$ mit : H_B : beobachtete Häufigkeit
H_E : erwartete Häufigkeit

Bei der Bestimmung der erwarteten Summenhäufigkeiten ist bei klassierten Daten zu beachten, dass für den jeweiligen Wert x die halbe Klassenbreite dazu addiert werden muss (Betrachtung auf die jeweils obere Klassengrenze). Grundsätzlich ist die Notwendigkeit der Klassierung als Nachteil anzusehen, da die Prüfgröße von der Klassenbreite abhängt.

Diese Prüfgröße wird gegen einen kritischen Wert verglichen, der in einschlägigen statistischen Tabellen zu finden ist.

Ist $T_{prüf} > T_{kr}$ wird die Nullhypothese auf dem Signifikanzniveau α abgelehnt.

Cramér-von-Mises Test

Der Cramér-von-Mises-Test ist, wie der KS-Test auch, ein allgemeiner statistischer Test, mit dem untersucht werden kann, ob die Häufigkeitsverteilung der Daten von einer vorgegebenen Wahrscheinlichkeitsverteilung abweicht. Der Test ist benannt nach Harald Cramér und Richard von Mises, die ihn zwischen 1928 und 1930 entwickelt und veröffentlicht haben. Es wird in diesem Test die Abweichung zwischen den Häufigkeiten der Daten zur erwarteten Verteilungsfunktion betrachtet. Der Nachteil ist bei diesem Test, dass er von der Stichprobengröße relativ stark abhängig ist (siehe Vergleich der Anpassungstests am Ende).

Für den Test auf Normalverteilung gilt deshalb die Testgröße:

$$T = \frac{1}{12n} + \sum_{i=1}^{n}\left(\frac{2i-1}{2n} - H(x_i)\right)^2$$ mit $H(x)$ = Verteilungsfunktion = $F(x)$

Anderson-DarlingTest auf Normalverteilung

Der Anderson-DarlingTest prüft die Nullhypothese, dass eine Stichprobe aus einer normalverteilten Grundgesamtheit stammt. Der große Vorteil des Anderson-Darling Tests ist, dass hier keine Klassierung der Daten, wie z.B. beim KS-Test, notwendig ist.

Dieser Test ist für kleine und größere Stichproben geeignet und berücksichtigt insbesondere die Randbereiche der Daten.

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)(\ln(\Phi(u_i)) + \ln(1-\Phi(u_{n+1-i})))$$

mit

$$u_i = \frac{x_i - \bar{x}}{s}$$

bzw. $\Phi(u_i)$ für die Wahrscheinlichkeit des u-Wertes aus der Normalverteilung.

Mit A^2 wird die weitere Hilfsgröße z gebildet:

$$z = A^2\left(1 + \frac{0{,}75}{n} + \frac{2{,}25}{n^2}\right)$$

mit der der p_{value} gebildet wird.

$z \leq 0,2$	$p_{value} = 1 - \exp(-13{,}436 + 101{,}14z - 223{,}73\, z^2)$
$0,2 < z \leq 0{,}34$	$p_{value} = 1 - \exp(-8{,}318 + 42{,}796z - 59{,}938\, z^2)$
$0,34 < z \leq 0,6$	$p_{value} = \exp(0{,}9177 - 4,279z - 1{,}38\, z^2)$
$0,6 < z$	$p_{value} = \exp(1{,}2937 - 5,709z + 0{,}0186\, z^2)$

Die Nullhypothese, dass die Daten normalverteilt sind, wird verworfen, wenn $p_{value} < \alpha$ ist (Standard α = 0,05).
Eine Anwendung dieses Verteilungstests ist nur bei relativ kleinen Stichproben bis etwa n=100 sinnvoll, da bei großen Daten der Test fast immer die Nullhypothese ablehnt.

Anderson-Darling Test auf Weibull-Verteilung

Der Anderson Darling Test prüft die Nullhypothese, dass die Daten Weibull-verteilt sind. Dieser Test ist für kleine und größere Stichproben geeignet und berücksichtigt insbesondere die Randbereiche der Daten. Zunächst wird folgende Kennzahl bestimmt:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2\mathrm{i}-1)\left(\ln(H(t_\mathrm{i})) + \ln(1-H(t_{\mathrm{n+1-i}}))\right)$$

mit $H(t_i)$ für die Wahrscheinlichkeit aus der Weibull-Verteilung an der Stelle t_i. Mit A^2 wird die weitere Hilfsgröße z gebildet:

$$z = A^2 \cdot \left(1 + \frac{0{,}2}{\sqrt{n}}\right)$$

mit der der p_{value} gebildet wird.

$p_{value} = 3.98765313 \cdot \exp(-5.7849905 \cdot z)$

Die Nullhypothese, dass die Daten einer Weibull-Verteilung entsprechen, wird verworfen, wenn $p_{value} < \alpha$ (Standard α = 0,05).

Die Weibull-Verteilung ist jedoch eine universelle Verteilungsform. Die praktische Bedeutung dieses Tests ist deshalb eher gering. Bei Ablehnung der Nullhypothese kann aber von einer Mischverteilung ausgegangen werden, woraus sich evtl. unterschiedliche Ausfallursachen ableiten lassen. Eine Anwendung dieses Verteilungstests ist nur bei relativ kleinen Stichproben bis etwa n=100 sinnvoll, da bei großen Daten der Test fast immer die Nullhypothese ablehnt.

χ^2-Anpassungstest

Verglichen wird, ähnlich wie beim Kolmogorov-Smirnov-Anpassungstest, eine Stichprobe aus einer Grundgesamtheit gegen eine theoretische Verteilung. Die Prüfgröße bestimmt sich durch:

$$\chi^2 = \sum_{i=1}^{k}\frac{(H_B - H_E)^2}{H_E}$$

mit k=Anzahl der Klassen, bzw. Merkmale. Die beobachteten Häufigkeiten stehen in der Spalte 1 die erwarteten in der Spalte 2. Geprüft wird die Nullhypothese: Die beobachtete Verteilung H_B entspricht der erwarteten H_E, wobei hier die absoluten Einzelhäufigkeiten gemeint sind. Der χ^2-Anpassungstest stellt im Allgemeinen Verteilungsirregularitäten fest. Bei Vorliegen kleiner Stichprobenumfänge entdeckt der KS-Test eher Abweichungen von der Normalverteilung. Bei den erwarteten Häufigkeiten ist die Klassenbreite K zu berücksichtigen. Für die Normalverteilung gilt somit:

$$H_E = K \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-\bar{x})^2}{2s^2}}$$

Die Daten sind zur Eingabe bereits klassiert anzugeben. Diese Prüfgröße wird gegen einen kritischen Wert verglichen, der in einschlägigen statistischen Tabellen zu finden ist. Hierbei wird ein Freiheitsgrad f benötigt, der sich folgendermaßen bestimmt:

$$f = k - 1 - a$$

wobei a die Anzahl der geschätzten zusätzlichen Parameter ist. Bei der Anpassung an eine Binomial- oder Poissonverteilung ist a=1. Für eine Normalverteilung gilt: Werden $\bar{x}$ und s aus den klassierten Daten geschätzt, so ist a=2. Werden $\bar{x}$ und σ direkt aus den Originaldaten berechnet, so ist a=1 und ist μ und σ bekannt, so beträgt a=0.

Ist $\chi^2 > \chi^2_{kr}$, wird die Nullhypothese auf dem Signifikanzniveau α abgelehnt. Die alternative Vorgehensweise ist, dass H_o verworfen wird, wenn $p_{value} < \alpha$ ist (Standard $\alpha = 0{,}05$).

Einzelhäufigkeiten<1 sollten vermieden werden. Sind für bestimmte Merkmale zu kleine Einzelhäufigkeiten vorhanden (<1), so sollten diese mit anderen Werten zusammengefasst werden. Die Klassenmitte ist entsprechend anzupassen.

Siehe auch χ^2-Homogenitätstest. Für weitere Tests sei auf die Literatur verwiesen.

Epps-Pulley Test

Der Anpassungstest von Epps und Pulley (1983) hat den Vorteil, dass dieser von der Stichprobengröße weitgehend unabhängig ist, wenn nur leichte Abweichungen von der Normalverteilung vorliegen (siehe Vergleich der Anpassungstests am Ende). Der Epps-Pulley Test ist in der DIN ISO 5479 beschrieben, aber noch wenig verbreitet. Epps und Pulley schlagen folgende Gewichtungsfunktion vor:

$$EP_n(\beta) = n\frac{\sqrt{2\pi}}{\beta}\int_{-\infty}^{\infty}\left(g_{n;\beta}(x) - \frac{1}{1+h^2}\Phi\left(\frac{x}{1+h^2}\right)\right)^2 dx$$

mit dem auf die Variable z angewendeten nicht parametrischen Dichteschätzer

$$g_{n;\beta}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}\Phi\left(\frac{x-z_i}{h}\right) \quad ; \quad h = \left(\beta\sqrt{2}\right)^{-1}$$

Auf Basis umfangreicher Simulationen empfehlen Epps und Pulley für β=1.

Henze (1994, p. 311) schlägt für n≥10 folgende standardnormalverteilte Prüfgröße vor:

$$Z_n^* = \xi_4 + \xi_3 \ln\left(\frac{EP_n^* - \xi_1}{\xi_1 + \xi_2 - EP_n^*}\right)$$

$$EP_n^*(1) = \left(EP_n(1) - \frac{0{,}365}{n} - \frac{1{,}34}{n^2}\right)\left(1 - \frac{1{,}3}{n}\right)$$

$$\xi_1 = -0{,}020682 \quad \xi_2 = 2{,}26664 \quad \xi_3 = 1{,}23062 \quad \xi_4 = 3{,}55295$$

Die Hypothese, dass die Daten normalverteilt sind, wird auf dem Signifikanzniveau α abgelehnt, wenn:

$$Z_n^* > u_{1-\alpha}$$

Umgekehrt kann hieraus der p_{value} bestimmt werden (Auflösung nach α).

Jarque-Bera Test (JB)

Der Jarque-Bera-Test ist ein statistischer Test, der anhand der Schiefe und der Wölbung (Asymmetrie und Spitzigkeit der Gaußkurve) in den Daten prüft, ob eine Normalverteilung vorliegt. Der Test wurde von Carlos Jarque und Anil Bera vorgeschlagen. Dieser Test hat den Vorteil, wie der Epps-Pulley auch, dass er von der Stichprobengröße weitgehend unabhängig ist, wenn nur leichte Abweichungen von der Normalverteilung vorliegen (siehe Vergleich der Anpassungstests am Ende).

Die Schiefe S, auch als zentrales Moment 3. Ordnung bezeichnet, ist definiert durch:

$$S = \frac{\frac{1}{n}\sum (x_i - \bar{x})^3}{\left(\frac{1}{n}\sum (x_i - \bar{x})^2\right)^{3/2}}$$ bzw. für Stichproben $$\frac{n}{(n-1)(n-2)}\sum\left(\frac{x_i - \bar{x}}{s}\right)^3$$

Bei Symmetrie der Gauß-Glockenkurve ist der Wert der Schiefe = 0.

Die Wölbung, auch als Kurtosis oder zentrales Moment 4. Ordnung bezeichnet, ist definiert durch:

$$K = \frac{\frac{1}{n}\sum (x_i - \bar{x})^4}{\left(\frac{1}{n}\sum (x_i - \bar{x})^2\right)^{2}}$$ bzw. für Stichproben

$$\left(\frac{n\cdot(n+1)}{(n-1)(n-2)(n-3)}\sum\left(\frac{x_i - \bar{x}}{s}\right)^4\right) - \frac{3\cdot(n-1)^2}{(n-2)(n-3)}$$

Bei Normalverteilung ergibt sich der Wert = 3. Man verwendet anstelle der Wölbung deshalb auch den Begriff Exzess = Wölbung – 3. Bei einer Wölbung > 3, hat die Verteilung flache Enden und eine große Streuung. Die Teststatistik nach Jarque-Bera ist dann:

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4} \right)$$

Diese Kenngröße ist näherungsweise χ^2-verteilt (mit zwei Freiheitsgraden):

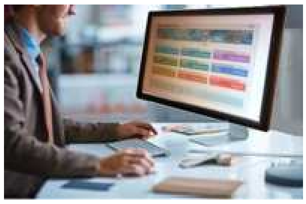
$$JB \sim \chi^2_{2;1-\alpha}$$

Bei einem Signifikanzniveau von 5% wird die Hypothese auf Normalverteilung verworfen, wenn $JB > 6$ ist. Umgekehrt kann der p_{value} durch Auflösung nach α bestimmt werden.

CRGRAPH

Software – Literatur – Consulting – Schulungen

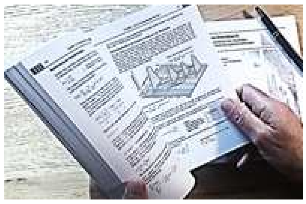
Software



Unsere Software **Visual-XSel** ist ein leistungsfähiges Tool für alle wichtigen statistischen Qualitäts- und Zuverlässigkeitsmethoden. Nicht umsonst ist diese Software in vielen großen Firmen im Einsatz – crgraph.de/Referenzen.

Weitere Informationen zum aktuellen Thema finden Sie auf den nächsten Seiten oder unter crgraph.de/Versionen

Eigene Literatur



Unser **Taschenbuch der statistischen Qualitäts- und Zuverlässigkeitsmethoden** beinhaltet weiterführende Themen, z.B. zu Systemanalysen, Weibull- und Zuverlässigkeitsmethoden, Versuchsplanung und Datenauswertung, sowie zur Mess-System-Analyse und Prozessfähigkeit.

Weitere Informationen finden Sie unter crgraph.de/Literatur

Consulting & Schulungen & Six Sigma

Bei unseren Inhouse- oder Online-Schulungen wird die praxisnahe Anwendung von statistischen Methoden vermittelt. Wir haben über 20 Jahre Erfahrung, insbesondere in der Automobilindustrie und unterstützen Sie bei Ihren Problemstellungen, führen Auswertungen für Sie durch, oder erstellen firmenspezifische Auswertevorlagen.

Weitere Informationen finden Sie unter crgraph.de/Schulungen

Hotline

Haben Sie noch Fragen, oder Anregungen? Wir stehen Ihnen gerne zur Verfügung:

Tel. +49 (0)8151-9193638

e-mail: info@crgraph.de

Besuchen Sie uns auf unserer Home-Page: www.crgraph.de

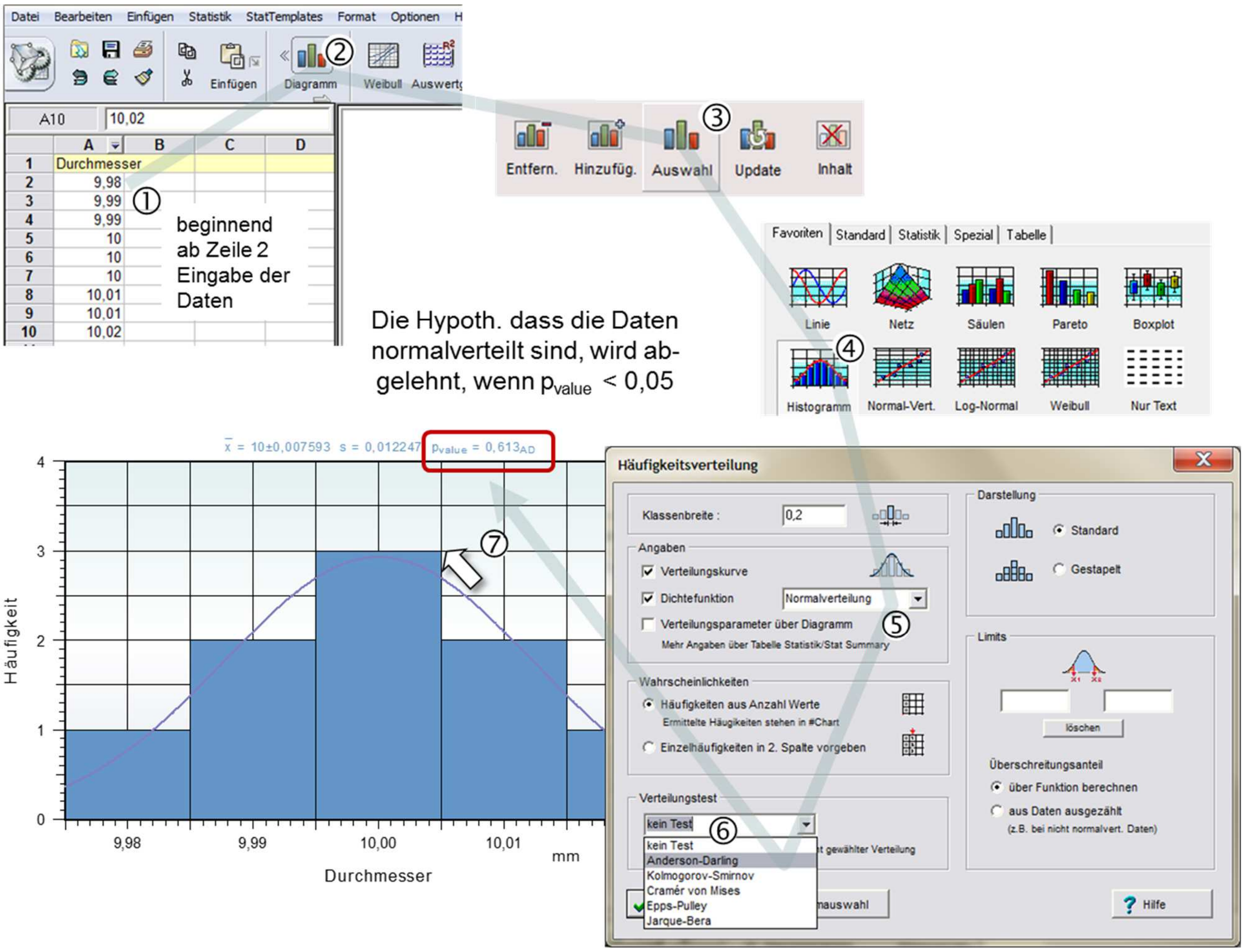
Verteilungstests

Anwendung in Visual-XSel 15.0 / 16.0

www.crgraph.de/download

In Visual-XSel gibt es 3 Möglichkeiten statistische Verteilungstests durchzuführen.

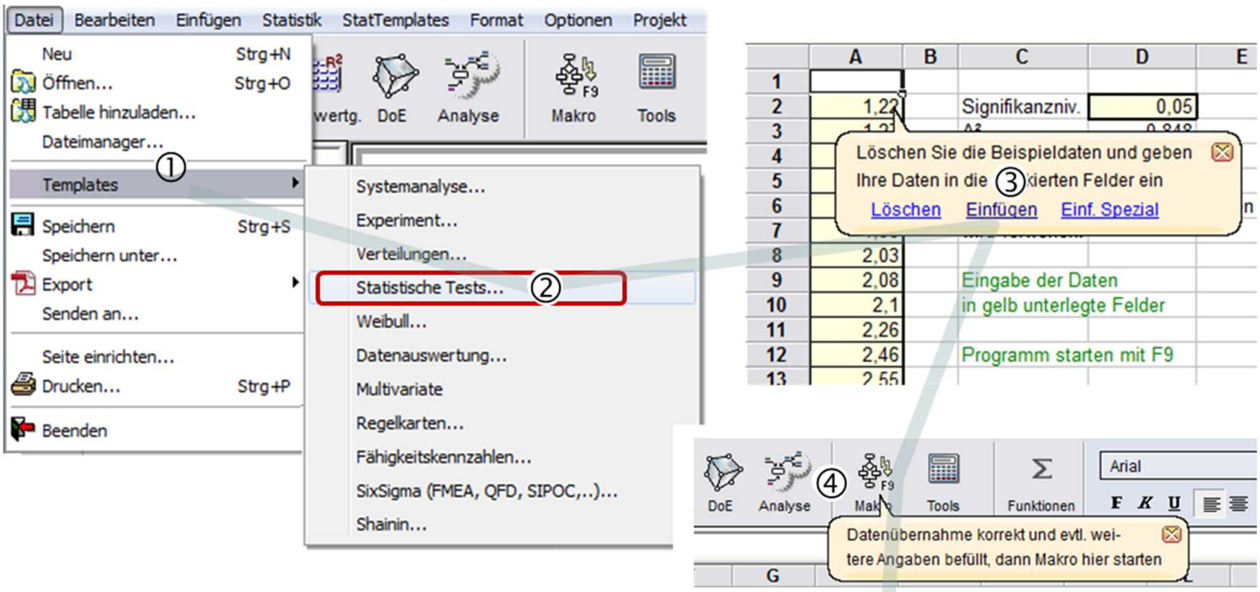
1. Einfache Grafik

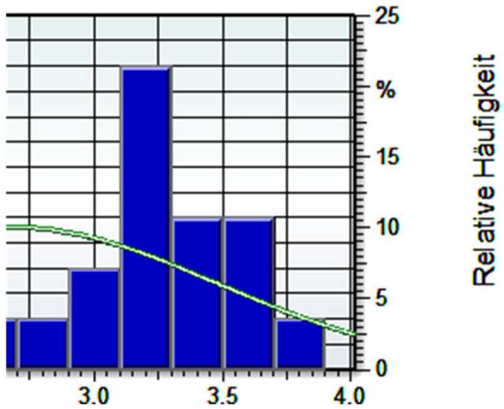


2. Verwendung von Templates

Die zu testenden Daten sollten sich in der Zwischenablage befinden.

Verteilungstests





Signifikanzniv.	0,05
A²	0,848
p-Value	0,025

⑤

Die Nullhypothese, dass die Daten normalverteilt sind, wird verworfen!

Ist p-value<SigNiveau, so ist die Nullhypothese zu verwerfen.

3. Template Verteilungstest verschiedener Verteilungen

Ab Version 16.0 befindet sich der Zugriff auf die Templates für Verteilungsvergleiche unter der Ikone Auswertung / Hypothesentests: