



AvenDATA GmbH

Auguste-Viktoria-Straße 118

14193 Berlin

www.avendata.de

Kleine IDEA Ecke

Chi Quadrat Test in IDEA

Chi Quadrat Test

Für die Aufdeckung von Unregelmäßigkeiten im Datenmaterial stellt IDEA eine Reihe von Funktionen zur Verfügung. In Veranstaltungen und Schulungen sind es dabei insbesondere Analysen von Benfords Law, die das Interesse wecken, da die Abweichungen erwarteter Häufigkeiten von Zahlen oft zu verblüffenden Ergebnissen führen. Benfords Law betrachtet die Anfangsziffern von Zahlenfeldern und belegt mit mathematischem Hintergrund eine absteigende Häufigkeitsverteilung von der Ziffer 1 bis zur 9. Entsprechen die Daten nicht den Erwartungen dieser Gesetzmäßigkeit kann dies eine Manipulation des Datenmaterials als Auslöser haben.

In diesem Zusammenhang ist auch der Schritt zu einer weiteren Methode nahe liegend, dem Chi Quadrat Test. Auch er beschäftigt sich mit der natürlichen Verteilung von Zahlen, betrachtet dabei allerdings die Ziffern vor und nach dem Komma. Vereinfacht ausgedrückt liegen die Werte bei ausreichend großen Datenmengen in einer Normalverteilung mit einer gleich hohen Wahrscheinlichkeit je Klasse vor. Die Klasse entspräche unseren Ziffern 0 bis 9, die zu jeweils 10% Anteil am Zahlenmaterial haben sollten.

Chi Quadrat in IDEA ?

Es finden sich unterschiedliche Aussagen, ob die Berechnung von Chi Quadrat Werten in IDEA geprüft werden kann. Dies liegt zum einen daran, dass das Modul AIS TaxAudit die Berechnung in seinen Funktionen zur Verfügung stellt und in der Regel mit IDEA erworben wird.

Chi-Quadrat-Test für Spalte "BETRAG_IN_HAUSWÄHRUNG_MIT_VORZEICHEN (8,2)"

Erwartete Häufigkeit je Klasse: 304,8 Signifikanzniveau: 95% => Kritischer Wert: 16,9190

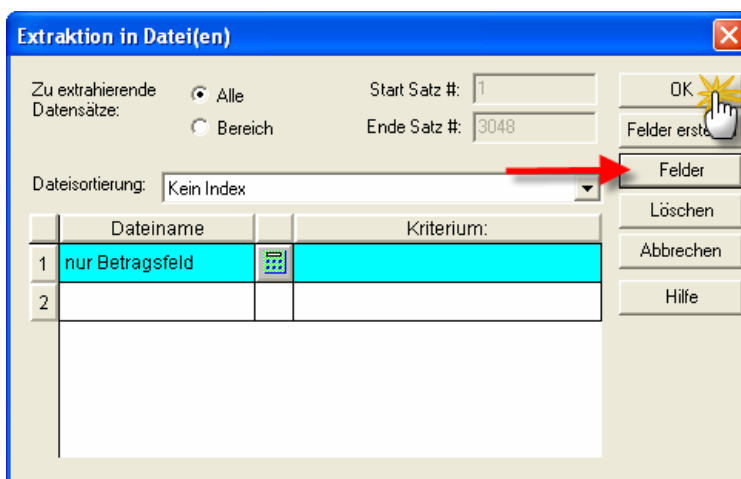
Ziffer	Zehner-Stelle		Einer-Stelle		Erste Nachkommastelle		Zweite Nachkommastelle	
	Ist-Anzahl	Differenz	Ist-Anzahl	Differenz	Ist-Anzahl	Differenz	Ist-Anzahl	Differenz
0	354	49,2	308	3,2	314	9,2	301	-3,8
1	299	-5,8	309	4,2	298	-6,8	292	-12,8
2	305	0,2	300	-4,8	287	-17,8	327	22,2
3	263	-41,8	313	8,2	301	-3,8	319	14,2
4	308	3,2	284	-20,8	278	-26,8	294	-10,8
5	304	-0,8	304	-0,8	292	-12,8	290	-14,8
6	307	2,2	319	14,2	287	-17,8	297	-7,8
7	305	0,2	294	-10,8	331	26,2	317	12,2
8	328	23,2	273	-31,8	329	24,2	305	0,2
9	275	-29,8	344	39,2	331	26,2	306	1,2
Chi ² -Wert		18,5157		11,2126		11,8753		4,6573

© 2004 Audicon GmbH

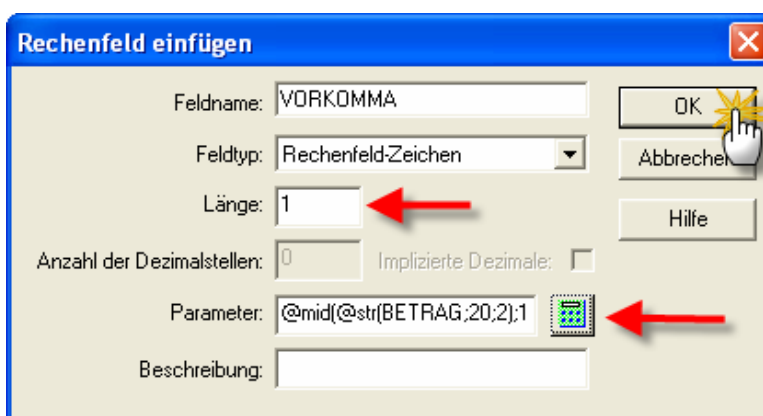
Zum anderen können die Berechnungen mit den Funktionen in IDEA realisiert werden. Um einen ungefähren Chi Quadrat Wert für ein numerisches Feldes in IDEA zu berechnen gehen

Sie wie folgt vor:

- Laden Sie eine Tabelle mit einer ausreichenden Datensatzanzahl in IDEA ein und extrahieren sie das betreffende Betragsfeld.



- Aus dem Betragsfeld extrahieren wir nun die Ziffern, der zu untersuchenden Stellen, also vor und nach dem Komma. Dafür werden berechnete Spalten genutzt. Klicken Sie mit rechts auf Ihre neue Tabelle und wählen Sie „Rechenfeld hinzufügen“.



Durch die Nutzung der Funktion @STR kann der Wert der Vorkommastelle ausgelesen werden. Die Beschreibung der Funktion in IDEA liefert Ihnen bereits die benötigte Syntax für diese Aktion.

Tragen Sie für die Berechnung der Nachkommastellen die entsprechend veränderten Werte in die Funktion ein.

Vorkomma: @mid(@str(BETRAG;20;2);17;1)

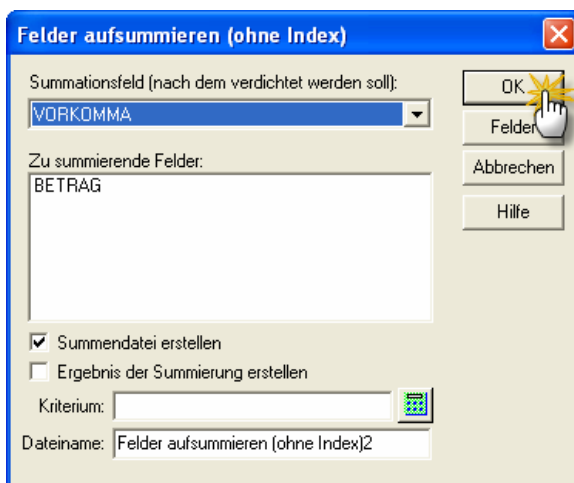
Nachkomma: @mid(@str(BETRAG;20;2);19;1)

	BETRAG	VORKOMMA	NACHKOMMA
19	63.080,43	0	4
20	22.514,17	4	1
21	10.217,83	7	8
22	26.875,54	5	5
23	22.229,37	9	3

Die Datei ist nun für die eigentliche Berechnung vorbereitet. Zuerst wird die Differenz aus der Häufigkeit des Auftretenden der einzelnen Ziffern und dem Erwartungswert gebildet. Der Chi Wert berechnet sich aus dem Quadrat der Summe dieser Differenzen, geteilt durch die erwartete Häufigkeit.

$$CHI - Wert = \frac{Summe _ Differenz^2}{erwartete _ Häufigkeit}$$

- Die Häufigkeit berechnen Sie durch den Einsatz der Summierungsfunktion.



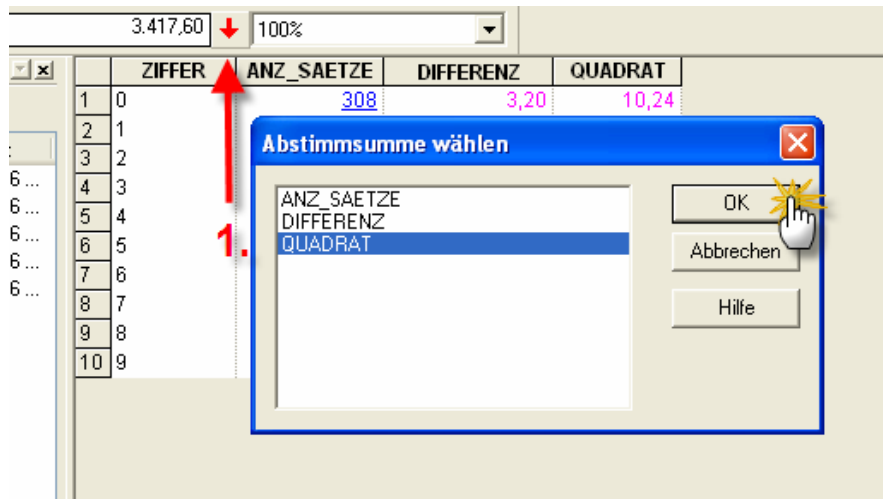
- In der neu entstandenen Summierungstabelle fügen Sie ein weiteres Rechenfeld hinzu, dass die Differenz zur erwarteten Häufigkeit (10 % der Anzahl Datensätze) berechnet.

	ZIFFER	ANZ_SAETZE	DIFFERENZ
1	0	308	3,20
2	1	309	4,20
3	2	300	-4,80
4	3	313	8,20
5	4	284	-20,80
6	5	304	-0,80
7	6	319	14,20
8	7	294	-10,80
9	8	273	-31,80
10	9	344	39,20

- Quadrieren Sie diese Differenzspalte in einem neuen Rechenfeld.

Mit Hilfe der Funktion „Abstimmsumme setzen“ können Sie sehr schnell die Summe dieser Quadrate errechnen.

- Der Wert aus der Abstimmsumme kann nun im Taschenrechner zur Berechnung des Chi Quadrat Wertes herangezogen werden.



- Berechnen Sie den Quotient aus dem Wert der Abstimmsumme und der erwarteten Häufigkeit.

$$\underline{\underline{11,21}} = \frac{3417,60}{304,80}$$

Der nun errechnete Wert von 11,21 ermöglicht eine Aussage über Wahrscheinlichkeit des Vorliegens einer natürlichen Zahlenfolge. Sind zum Zwecke der Manipulation Werte verändert worden kommt die Schwäche des Menschen zum Tragen, der durch persönliche Zahlenpräferenzen in der Regel nicht in der Lage ist diese natürliche Zahlenverteilung einzuhalten. Zur Beurteilung liegen Tabellen der mathematisch statistischen Auswertungen zugrunde, die bei einem Chi Wert von 21 bis 30 von einer hohen Wahrscheinlichkeit für eine systematische Ursache ausgehen. Bei einem Wert von 11 liegt mit höherer Wahrscheinlichkeit eine zufällige Verteilung der Zahlen abweichend vom Erwartungswert vor.

Die gezeigte Berechnung in IDEA ist nur ein sehr grober Ansatz und Berücksichtigung keine Signifikanzwerte oder Freiheitsgrade. Mit Kenntnissen in IDEA Skript lassen sich auch komplexere Berechnungen realisieren. Die Analyse kann in dieser Form lediglich Anhaltspunkte liefern oder als Ausgangspunkt für tiefer greifende Rechnungen dienen.

Besuchen Sie unsere Schulungen zum Thema IDEA und AIS TaxAudit.

Weitere Informationen zu IDEA erhalten Sie unter:

AvenDATA GmbH

Auguste-Viktoria-Straße 118

14193 Berlin

Telefon: +49-30/8 90 62-894

Telefax: +49-30/8 90 62-899

Web: www.avendata.de