

DeltaMaster clicks!

07/2009

Liebe Datenanalysten,

sicher müssen Sie nicht lange überlegen, wenn man Sie um ein Zitat oder ein geflügeltes Wort über Statistiken bittet. Auch wenn Manches davon falsch kolportiert wird: Im Durchschnitt kommt die Statistik dabei nicht gut weg. Mal sind es Zweifel an der Integrität der Daten, mal an der Integrität des Statistikers oder seines Instrumentariums, die dafür sorgen, dass mehr Schmähungen als Schmeicheleien kursieren. Und dennoch, allen Unkenrufen zum Trotz: Immer wieder brauchen wir die Statistik – „Statistik“ hier nicht verstanden als eine Berichtsform, die als reines Zahlenwerk auch ohne deutlich formulierte Botschaft einen Wert hat, sondern als eine Sammlung von mathematischen Methoden, die es uns erlaubt, eine große Menge von Objekten stellvertretend durch einige wenige Kennziffern zu beschreiben und darüber zu diskutieren. Robuste und bewährte Messgrößen wie ein Durchschnittswert, wie Standardabweichungen oder auch nur Minima und Maxima ermöglichen zusätzliche Einblicke und sind oft genug unverzichtbar. Wie Sie *DeltaMaster* unterstützt, statistische Größen in Ihre Reports zu integrieren oder, mehr noch, solche Analysewerte selbst anzulegen und damit zu arbeiten, zeigen wir Ihnen in der vorliegenden Ausgabe der *DeltaMaster clicks!*.

Herzliche Grüße

Ihr Team von
Bissantz & Company

Kunst und Datenanalyse

Dass man als Datenanalyst in der Kunst einen Quell steter Inspiration findet, haben wir schon früher einmal ausgeführt (<http://blog.bissantz.de/kunst1>). Wenn Sie in der letzten Zeit eine Veranstaltung des Bissantz Campus besucht haben, ist Ihnen dieses Werk bestimmt aufgefallen, das in 140 x 140 cm auf Alu unsere Lounge ziert. Weil einige Gäste schon danach gefragt haben: Gerne bieten wir Ihnen ein Exemplar aus einer signierten Kleinauflage an.



DeltaMaster-Matinee

16. Juli 2009, Stuttgart

Als Referenten konnten wir Achim Schalk gewinnen, Controllingleiter bei MS Motor Service International GmbH. Sein Thema: die mehrstufige Planung des weltweiten Ersatzteilgeschäfts im Rahmen eines umfassenden Vertriebscontrollings. Beides wurde natürlich mit *DeltaMaster* gelöst.
www.bissantz.de/matinee

DeltaMaster@Work

23. Juli 2009, Nürnberg

Berichte erstellen, die wirken
www.bissantz.de/dm@w

Bissantz Campus

Unser Schulungsangebot für *DeltaMaster* und Microsoft SQL Server/Analysis Services
www.bissantz-campus.de

Archiv

www.bissantz.de/clicks

Kniff des Monats Statistische Maße in Berichten und Analysen

Ein bisschen Statistik verleiht so manchem Bericht zusätzliche Würze: Das durchschnittliche Auftragsvolumen, die kleinste und größte Anzahl von Lieferscheinpositionen, die Anzahl der Kunden, die im laufenden Quartal etwas geordert haben, Standardabweichungen, Varianzen – je nach Fragestellung können solche Angaben von großem Nutzen sein. Für manche Aufgaben ist es sogar unabdingbar, eine Vielzahl von Werten in ein paar markanten Größen zusammenzufassen und diese zu diskutieren, beispielsweise in der Marktforschung oder zur Risikoanalyse. Dass *DeltaMaster* über einen umfassenden Vorrat von Methoden und Werkzeugen für die Arbeit mit statistischen, betriebswirtschaftlichen und technischen Daten verfügt, wissen Sie längst. Oft genug wirken statistische Verfahren sogar im Hintergrund, ohne dass Sie sich darüber Gedanken machen müssen. Immer wieder aber gilt es, die Statistik an die Oberfläche zu bringen, statistische Größen zu ermitteln, sie zu berichten und weitere Berechnungen damit anzustellen.

In dieser Ausgabe der *DeltaMaster clicks!* stellen wir Ihnen zwei Konzepte vor (und diese gegenüber), wie Sie mit statistischen Kennzahlen in Ihren Reports und Analysen arbeiten können.

In der Pivottabelle: Zeilen-, Spalten- und Tabellenaggregationen der Anzeigewerte

Besonders einfach fügen Sie statistische Eckdaten als zusätzliche Zeilen oder Spalten in Pivottabellen ein. Dazu dienen die *Zeilen-, Spalten- und Tabellenaggregationen* im Kontextmenü, und zwar in sämtlichen Stufen von *DeltaMaster*, vom *Offline-Reader* bis hin zum *Miner-Expert*, und selbst in einer Bildschirmpräsentation, die Sie mit der Taste *F5* gestartet haben.

Umsatz	Produkte			Summe	Mittelwert
Kunden	Luxusmodelle	Sondermodelle	Standardmodelle		
Nord 1	9.681	5.248	27.738	42.668	14.223
Nord 2	Leere und 0-Zeilen/Spalten ▶			31.329	40.894
Ost 1	Darstellung ▶			29.921	36.341
Ost 2	Zellfärbung ▶			15.774	20.081
Süd 1	Balken/Säulen ▶			230.956	682.215
Süd 2	Wasserfall ▶			13.382	23.224
West 1				388.332	389.404
West 2				40.879	44.732
Summe					1.279.559
Mittelwert					53.315

Zeilenaggregationen	Summe
Spaltenaggregationen	Anzahl
Tabellenaggregationen	Mittelwert
Speichern	Varianz
Drucken...	Standardabweichung
Standardansicht	Minimum
	Maximum
	Anteil
	Alle einblenden
	Alle ausblenden

Die Aggregationen fassen die angezeigten Werte zeilenweise oder spaltenweise zusammen, wobei verschiedene statistische Funktionen zur Auswahl stehen: die Summe, die Anzahl der Werte, der Mittelwert, die Varianz und die Standardabweichung, das Minimum und das Maximum oder der Anteil der jeweiligen Zeilen- oder Spaltensumme an der Summe aller Werte der Tabelle. Sie können eine einzelne oder auch mehrere dieser Aggregationen aktivieren. Mit der *Tabellenaggregation* blenden Sie die jeweiligen Maße sowohl in den Zeilen als auch in den Spalten ein.

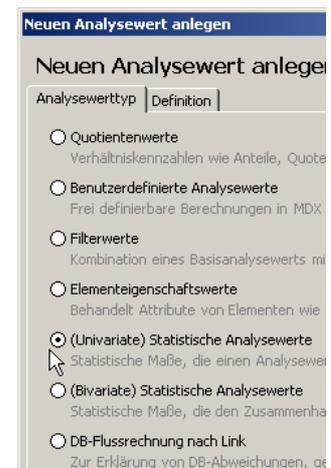
Die Berechnung bezieht sich stets auf die Werte, die in den Zeilen bzw. Spalten angezeigt werden. So ist es zum Beispiel nicht sinnvoll, die Werte in allen Zeilen einer Spalte zu summieren, wenn in den Zeilen eine „aufgeklappte“ Hierarchie dargestellt wird: Die Zwischensummen würden dann doppelt gezählt. Die Addition über die *Spaltenaggregation* ist in diesem Fall aber auch nicht nötig, denn die Summe ergibt sich ja bereits aus der Hierarchie.

Statistik mit Tiefgang: Analysewerte für statistische Berechnungen in OLAP-Modellen

Mit den oben beschriebenen Aggregationen setzen Sie die Statistik ein, um bestehende Reports mit nützlichen Zusatzinformationen anzureichern. In anderen Situationen stehen die statistischen Maße selbst im Mittelpunkt: Die „Verdichtung“ zu einigen wenigen Kenndaten muss erfolgen, bevor ein Bericht daraus entsteht.

Zu diesem Zweck können Sie mit *DeltaMaster* neue Analysewerte erzeugen, die Ihr Datenmodell ergänzen und die Sie mit allem, was *DeltaMaster* zu bieten hat, analysieren und berichten können. Diese Analysewerte wirken in der Tiefe des Datenmodells und nicht nur im aktuellen Bericht.

Um einen *neuen Analysewert anzulegen*, rufen Sie die entsprechende Funktion aus dem Menü *Modell* (Haupt-Menüleiste) oder aus dem Menü *Ich möchte* im *Analysewert-Browser* auf. Der Assistent, der daraufhin erscheint, ermöglicht das Anlegen verschiedener Arten von Analysewerten. Neben *Quotientenwerten* (Verhältniskennzahlen), *Filterwerten* und frei in MDX formulierten Berechnungen lassen sich etwa auch Kennzahlen für die Deckungsbeitragsflussrechnung erzeugen – oder eben statistische Größen.

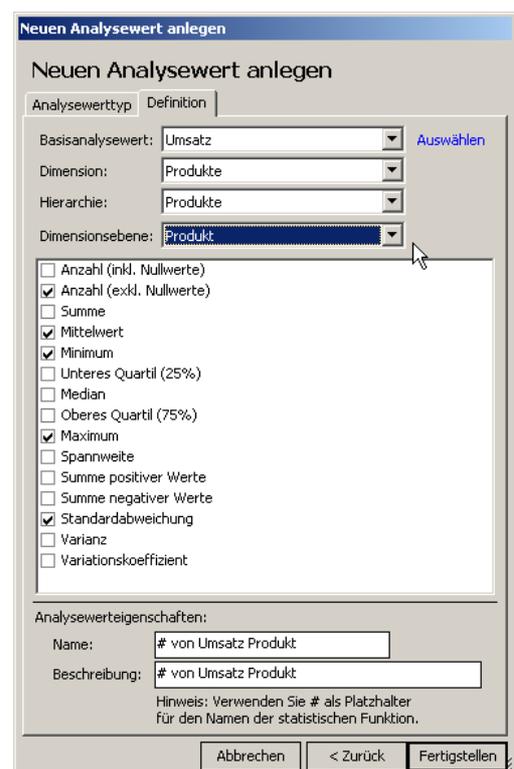


Zu unterscheiden sind univariate und bivariate Analysewerte. Die univariaten heißen so, weil sie stets genau eine Variable untersuchen, zum Beispiel den Umsatz. Die Kennzahlen, die wir hiermit erzeugen, beschreiben diesen Basisanalysewert in einer bestimmten Dimension. Beispiele sind die Anzahl der Werte, ihr Minimum, ihr Maximum oder ihr Mittelwert – jeweils berechnet für einen festzulegenden Basisanalysewert wie den Umsatz sowie für eine festzulegende Dimension, etwa die Produkte. Die bivariaten statistischen Analysewerte hingegen beschreiben nicht eine Variable, sondern zwei; genauer gesagt: den Zusammenhang zwischen den beiden Variablen, beispielsweise die Korrelation zwischen Umsatz und Rabatt.

Wir wollen uns hier auf die univariate Statistik beschränken. Die Arbeitsweisen mit bivariaten Größen sind aber identisch. Wählen Sie als Analysewerttyp die (*univariaten*) *statistischen Analysewerte* aus und wechseln Sie auf die Registerkarte *Definition*.

Der Assistent für diesen Typ ist in drei Abschnitte gegliedert. Im oberen Bereich wählen Sie aus, für welchen *Basisanalysewert* Sie statistische Maße erzeugen möchten und auf welcher *Dimensionsebene* diese gelten sollen. Das statistische Maß, das wir so erzeugen, liefert für die ausgewählte Ebene die entsprechende Zusammenfassung der Elemente. Weiter unten erläutern wir das Prinzip ausführlich.

In unserem Beispiel entscheiden wir uns für den Umsatz als *Basisanalysewert*, den wir in der *Dimension* „Produkte“ in deren gleichnamiger *Hierarchie* (zunächst) auf der *Ebene* der Produkte untersuchen wollen.



Im mittleren Bereich des Dialogs bietet *DeltaMaster* eine stattliche Menge von statistischen Funktionen an. Eine Übersicht mit weiteren Erläuterungen finden Sie am Ende dieser *clicks!*-Ausgabe. Mit den Kontrollkästchen markieren Sie alle Funktionen, die Sie in Form eines Analysewerts zugänglich machen möchten – der Assistent ist in der Lage, gleich mehrere Analysewerte in einem Arbeitsgang zu generieren. Einige sehr oft verwendete Maße hat *DeltaMaster* bereits vorausgewählt: die Elementanzahl, den Mittelwert, das Minimum und das Maximum sowie die Standardabweichung.

Unten im Dialog geben Sie den *Namen* und die *Beschreibung* für die neuen Größen ein – oder belassen es einfach bei dem Vorschlag, den *DeltaMaster* macht: In der Voreinstellung werden der Basisanalysewert sowie die gewählte Hierarchieebene einbezogen, sodass die Bezeichnungen bereits praktikabel sind. Den Platzhalter „#“ ersetzt *DeltaMaster* beim Erzeugen der Kennzahlen durch einen Kurznamen für die jeweilige Funktion (siehe Tabelle am Ende der *clicks!*). Durch den Platzhalter können Sie mehrere Analysewerte in einem Arbeitsgang erzeugen und ihnen dennoch eindeutige Namen mitgeben.

Wenn Sie die Definition *fertigstellen*, legt *DeltaMaster* alle gewünschten Kennzahlen auf einmal an. Sie finden sie anschließend im *Analysewert-Browser*, gekennzeichnet mit der Typenbezeichnung „UniStat“. Hier können Sie sie umbenennen, skalieren, formatieren usw., wie jeden anderen Analysewert auch. Und natürlich lassen sich die neuen Kennzahlen wie gewohnt zum Beispiel in einer Pivottabelle darstellen, im *Hyperbrowser* visualisieren oder mit den Analyseverfahren eingehend untersuchen.

Analysewert-Browser	
Verfügbare Analysewerte:	
Name	Typ
CountExclEmpty von Umsatz Produkt	UniStat
Avg von Umsatz Produkt	UniStat
Min von Umsatz Produkt	UniStat
Max von Umsatz Produkt	UniStat
StdDev von Umsatz Produkt	UniStat

Ebenenabhängigkeit

Um die Wirkungsweise der „UniStat“-Kennzahlen zu erläutern, werfen wir einen näheren Blick auf eine besonders einfache Funktion: die **Anzahl** („CountExclEmpty“), die wir der kompakteren Darstellung halber hier in „Anz. Umsatz Produkt“ umbenannt haben.

Der in der rechten Spalte dargestellte Analysewert ist für die Ebene Produkt definiert, wie in der Abbildung auf der vorherigen Seite angedeutet. Dementsprechend werden also Produkt-Elemente gezählt. Die einzelnen Elemente liefern jeweils „ihre“ Anzahl zurück, also 1. Nach oben hin erhalten wir die Mengenangaben der jeweils untergeordneten Ebenen.

Produkte	Umsatz	Anz. Umsatz Produkt
Alle Produkte	15.589.683	20
Luxusmodelle	1.581.042	8
Arcade	816.753	4
Arcade AE 44	433.960	1
Arcade AE 55	363.482	1
Arcade AE 66	14.558	1
Arcade AE 77	4.753	1
Discus	1.534	1
Discus Office	1.534	1
Precisio	762.755	3
Precisio JK	144.021	1
Precisio LF	609.640	1
Precisio LM	9.094	1
Sondermodelle	12.628.400	6
Standardmodelle	1.380.242	6

Wie das Ergebnis von der in den Analysewerteigenschaften ausgewählten Ebene abhängt, zeigt die folgende Abbildung. Am oberen und linken Rand sehen Sie zusätzlich die *Zellkoordinaten* der Pivottabelle, die über das Kontextmenü im Modus *Miner* eingeblendet wurden.

1	2	3	4	5		
	Kennzahlen	...				
2	Produkte	...	Anz. Umsatz Produkt	Anz. Umsatz Produktgruppe	Anz. Umsatz Produkthauptgruppe	Anz. Umsatz Produkte Top
3	Alle Produkte	20	8	3	1	
4	Luxusmodelle	8	3	1	0	
5	Arcade	4	1	0	0	
6	Arcade AE 44	1	0	0	0	
7	Arcade AE 55	1	0	0	0	
8	Arcade AE 66	1	0	0	0	
9	Arcade AE 77	1	0	0	0	
10	Discus	1	1	0	0	
11	Discus Office	1	0	0	0	
12	Precisio	3	1	0	0	
13	Precisio JK	1	0	0	0	
14	Precisio LF	1	0	0	0	
15	Precisio LM	1	0	0	0	
16	Sondermodelle	6	2	1	0	
17	Standardmodelle	6	3	1	0	

Für diese Übersicht haben wir vier Analysewerte angelegt. Alle verwenden dieselbe statistische Funktion „Anzahl (exkl. Nullwerte)“ und alle beziehen sich auf denselben Umsatz sowie dieselbe Produkthierarchie. Lediglich die Ebene innerhalb der Produkthierarchie unterscheidet sie:

- Für den Analysewert in Spalte 2 haben wir die unterste Ebene ausgewählt, das einzelne Produkt, wie in der vorherigen Abbildung.
- In Spalte 3 wurde als Ebene die Produktgruppe festgelegt. Es sollen also Produktgruppen gezählt werden. Jede Produktgruppe zählt demnach als 1, während die untergeordneten einzelnen Produkte nicht zu berücksichtigen sind und deshalb ein Ergebnis von 0 liefern. Die Produkthauptgruppe „Luxusmodelle“ enthält die Produktgruppen „Arcade“, „Discus“ und „Precisio“, sodass wir als Anzahl 3 erhalten. Zusammen mit den (nicht abgebildeten) 2 Produktgruppen der Sondermodelle und den 3 Produktgruppen in den Standardmodellen erhalten wir eine Anzahl von insgesamt 8 Produktgruppen.
- Das statistische Maß in Spalte 4 ist für die Produkthauptgruppen definiert. „Luxusmodelle“, „Sondermodelle“ und „Standardmodelle“ sind Produkthauptgruppen (Anzahl also jeweils 1). Die untergeordneten Produktgruppen und Produkte sind bzw. enthalten keine Produkthauptgruppen, sodass dort eine Anzahl von 0 resultiert.
- In Spalte 5 schließlich ist die höchste Ebene ausgewählt („Alle Produkte“). Das Ergebnis ist 1, denn es gibt genau 1 Top-Element, der Rest ist 0.

Die Wirkungsweise der Statistik Kennzahlen hängt also entscheidend von der Ebene ab, für die sie definiert sind.

Ein weiteres Beispiel soll das illustrieren. Dieses Mal nehmen wir als statistische Funktion den **Mittelwert** (das arithmetische Mittel); wir bleiben in der Produktdimension und erzeugen wieder vier Analysewerte, einen für jede der vier Ebenen dieser Hierarchie. Unsere vier Mittelwerte mit ihren unterschiedlichen Ebenen haben wir in der folgenden Abbildung nebeneinander gestellt, zusammen mit den jeweiligen absoluten Umsatzzahlen (Basisanalysewert) in der linken Spalte.

1	2	3	4	5	6	
1	Kennzahlen ...					
2	Produkte ...	Umsatz	Ø Umsatz Produkt	Ø Umsatz Produktgruppe	Ø Umsatz Produkthauptgruppe	Ø Umsatz Produkte Top
3	Alle Produkte	15.589.683	779.484	1.948.710	5.196.561	15.589.683
4	Luxusmodelle	1.581.042	197.630	527.014	1.581.042	
5	Arcade	816.753	204.188	816.753		
6	Arcade AE 44	433.960	433.960			
7	Arcade AE 55	363.482	363.482			
8	Arcade AE 66	14.558	14.558			
9	Arcade AE 77	4.753	4.753			
10	Discus	1.534	1.534	1.534		
11	Discus Office	1.534	1.534			
12	Precisio	762.755	254.252	762.755		
13	Precisio JK	144.021	144.021			
14	Precisio LF	609.640	609.640			
15	Precisio LM	9.094	9.094			
16	Sondermodelle	12.628.400	2.104.733	6.314.200	12.628.400	
17	Standardmodelle	1.380.242	230.040	460.081	1.380.242	

Das Verhalten, das wir in dieser Tabelle beobachten können, ist dasselbe wie zuvor:

- Der Analysewert in Spalte 3 bezieht sich auf Produkte – es sollen also Mittelwerte über die Elemente der Ebene „Produkte“ berechnet werden. Für jedes einzelne Produkt, zum Beispiel „Precisio JK“, stimmt der Mittelwert natürlich mit seinem direkten Umsatz überein. Für die Produktgruppe, etwa „Precisio“, erhalten wir den Mittelwert aus den Produkten, die der Gruppe zugeordnet sind, im Beispiel also den Mittelwert aus „Precisio JK“, „Precisio LF“ und „Precisio LM“. Der durchschnittliche Umsatz für die Produkthauptgruppe der Luxusmodelle wiederum errechnet sich aus dem Mittelwert aller Produkte, die der Hauptgruppe untergeordnet sind.
- Der Analysewert in Spalte 4 wurde für Produktgruppen definiert. Die einzelnen Artikel sind keine Produktgruppen und enthalten auch keine und bleiben deshalb leer. Für die Gruppen „Arcade“, „Discus“ und „Precisio“ wird ihr jeweiliger Umsatz zurückgeliefert, denn der „Mittelwert“ für nur ein Element entspricht natürlich genau dem Basisanalysewert. Erst in der Produkthauptgruppe der Luxusmodelle wird ein „echtes“ Mittel gebildet, nämlich das aus „Arcade“, „Discus“ und „Precisio“. Diese und die übrigen Produktgruppen gehen in den in Zeile 3 ausgewiesenen Mittelwert über alle Produktgruppen ein.
- In Spalte 5 sollen nur noch Produkthauptgruppen gemittelt werden. Davon gibt es 3, nämlich Luxus-, Sonder- und Standardmodelle, bei denen der Mittelwert mit dem ursprünglichen Umsatz übereinstimmt.
- Spalte 6 gibt einen Analysewert aus, in dem das Top-Element als Ebene ausgewählt wurde, das identisch mit seinem eigenen Mittelwert ist.

Funktionsvielfalt

Die folgende Liste zeigt, welche statistischen Funktionen in *DeltaMaster*-Anwendungen verfügbar sind, die auf Microsoft Analysis Services basieren. In anderen Datenbanken mag die eine oder andere Funktion fehlen, je nachdem, wie leistungsfähig ihre MDX-Schnittstelle ist. Zusätzlich ist angegeben, welche Bezeichnung *DeltaMaster* anstelle des Platzhalters „#“ in den Namen von neuen Analysewerten einfügt.

Funktion	Namensvorschlag	Anmerkung
Anzahl (inkl. Nullwerte)	CountInclEmpty	Anzahl der Elemente
Anzahl (exkl. Nullwerte)	CountExclEmpty	Anzahl der Elemente, ausgenommen diejenigen, für die der Basis-analysewert keine Werte hat („Null“)
Summe	Sum	
Mittelwert	Avg	Arithmetisches Mittel. Zwischen fehlenden Werten („Datenbank-Null“) und dem Wert Null („0“) wird unterschieden: Die „Datenbank-Null“ wird ignoriert, ein Wert von 0 hingegen wird mitgerechnet.
Minimum	Min	
Unteres Quartil (25%)	Quartile25	Die Quartile sowie der Median sind Grenzen, mit denen sich die Streuung in Verteilungen beschreiben lässt. Der Median trennt die (sortierten) Werte in zwei Hälften. Das untere Quartil teilt die 25 % kleinsten Werte ab, das obere Quartil die 25 % größten Werte. Diese Funktionen werden insbesondere für sogenannte Box-Whisker-Plots benötigt (siehe DeltaMaster deltas! 5.3.5, Punkt 11).
Median	Median	
Oberes Quartil (75%)	Quartile75	
Maximum	Max	
Spannweite	Span	Differenz zwischen Minimum und Maximum
Summe positiver Werte	SumPos	Wird beispielsweise zum Aufdecken von Kompensationseffekten bei Abweichungskennzahlen verwendet.
Summe negativer Werte	SumNeg	
Standardabweichung	StdDev	Wurzel aus der Varianz
Varianz	Var	Abweichung vom Mittelwert
Variationskoeffizient	CoV	Quotient aus Standardabweichung und Mittelwert