



Zentrale Datenstelle  
der Landesfinanzminister

# EXCEL-TUTORIAL 4

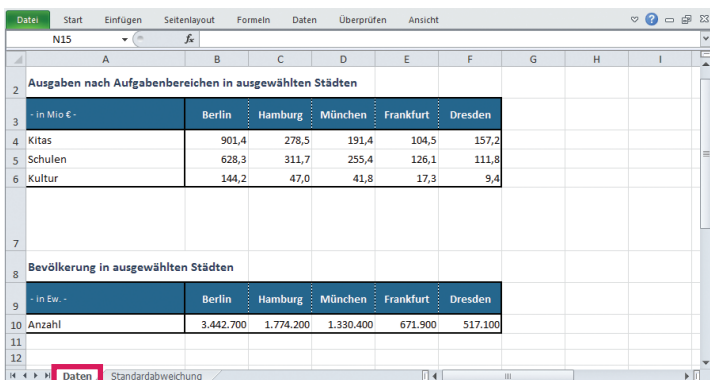
**»Auffälligkeiten mit  
Standardabweichungen ermitteln«**

## **AUFGABENSTELLUNG**

**Statistische Auffälligkeiten mit Standardabweichungen ermitteln:** Es liegen Ausgaben nach Aufgabenbereichen für verschiedene Gebietskörperschaften (hier: fünf Großstädte) vor. Ziel ist es, zu ermitteln, ob statistisch auffällige Streuungen zwischen den einzelnen Großstädten existieren, die nicht mehr mit Politikunterschieden erklärt werden können.

### Schritt 1

Die absoluten Ausgaben in den einzelnen Ausgabenbereichen schwanken stark, wie auf den ersten Blick im Tabellenblatt **Daten** zu erkennen ist. Dies ist aber dem Umstand geschuldet, dass die Städte unterschiedliche Einwohnerzahlen haben. Um die Unterschiede zu nivellieren, müssen die absoluten Beträge zunächst in eine vergleichbare Relation gebracht werden. Dazu bieten sich die Einwohnerzahlen an (siehe EXCEL-TUTORIAL 1).

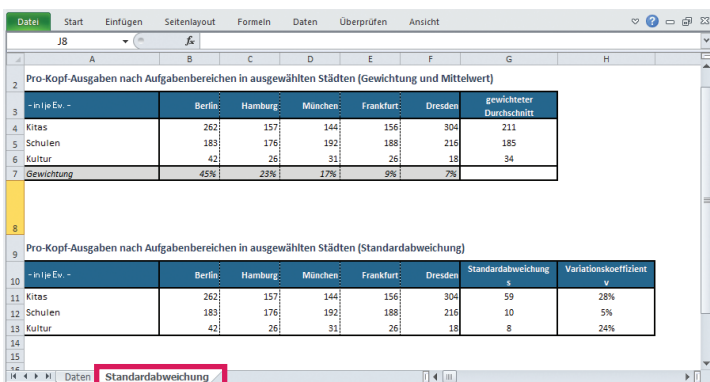


- in Mio. € -	Berlin	Hamburg	München	Frankfurt	Dresden
Kitas	901,4	278,5	191,4	104,5	157,2
Schulen	628,3	311,7	255,4	126,1	111,8
Kultur	144,2	47,0	41,8	17,3	9,4

- in Ew. -	Berlin	Hamburg	München	Frankfurt	Dresden
Anzahl	3.442.700	1.774.200	1.330.400	671.900	517.100

So werden auf dem Tabellenblatt **Standardabweichung** zunächst die Pro-Kopf-Ausgaben nach Aufgabenbereichen ermittelt.



- in je Ew. -	Berlin	Hamburg	München	Frankfurt	Dresden	gewichteter Durchschnitt
Kitas	262	157	144	156	304	211
Schulen	183	176	192	188	216	185
Kultur	42	26	31	26	18	34
Gewichtung	45%	23%	17%	9%	7%	

- in je Ew. -	Berlin	Hamburg	München	Frankfurt	Dresden	Standardabweichung s	Variationskoeffizient v
Kitas	262	157	144	156	304	59	28%
Schulen	183	176	192	188	216	10	5%
Kultur	42	26	31	26	18	8	24%



Excel screenshot showing the formula in cell B7: `=Daten!B$10/SUMME(Daten!$B$10:$F$10)`. The result is 45%.

Table 1: Pro-Kopf-Ausgaben nach Aufgabenbereichen in ausgewählten Städten (Gewichtung und Mittelwert)

- in je Ev. -	Berlin	Hamburg	München	Frankfurt	Dresden	gewichteter Durchschnitt
Kitas	262	157	144	156	304	211
Schulen	183	176	192	188	216	185
Kultur	42	26	31	26	18	34
Gewichtung	45%	23%	17%	9%	7%	

Table 2: Pro-Kopf-Ausgaben nach Aufgabenbereichen in ausgewählten Städten (Standardabweichung)

- in je Ev. -	Berlin	Hamburg	München	Frankfurt	Dresden	Standardabweichung s	Variationskoeffizient v
Kitas	262	157	144	156	304	59	28%
Schulen	183	176	192	188	216	10	5%
Kultur	42	26	31	26	18	8	24%

Tab: Standardabweichung

## Schritt 2

Nach der Berechnung der Pro-Kopf-Werte sind für die weitere statistische Analyse die Anteile der einzelnen Städte an der Grundgesamtheit von Belang, sodass zunächst die Gewichtung der einzelnen Städte ermittelt wird. Für die Gewichtung wird wiederum die Einwohnerzahl herangezogen. Dazu wollen wir den Anteil von Berlin an der Gesamtheit in der Zelle B7 ermitteln und als Prozentzahl ausdrücken. Die Formel lautet dort: `=Daten!B$10/SUMME(Daten!$B$10:$F$10)`.



Excel screenshot showing the data source for the formula in cell B7. The formula bar shows `=3442700`.

Table 1: Ausgaben nach Aufgabenbereichen in ausgewählten Städten

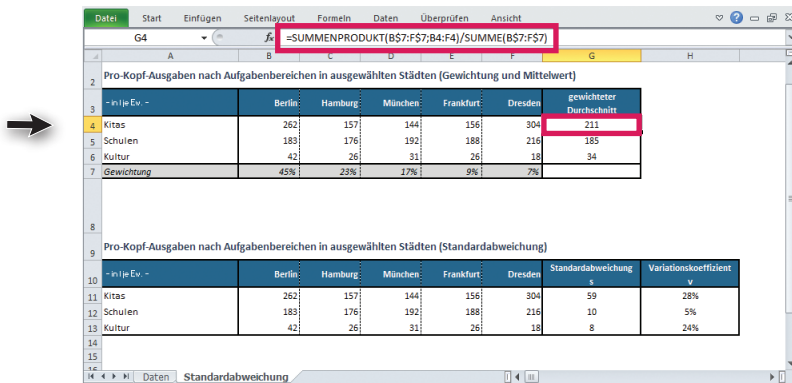
- in Mio € -	Berlin	Hamburg	München	Frankfurt	Dresden
Kitas	901,4	278,5	191,4	104,5	157,2
Schulen	628,3	311,7	255,4	126,1	111,8
Kultur	144,2	47,0	41,8	17,3	9,4

Table 2: Bevölkerung in ausgewählten Städten

- in Ew. -	Berlin	Hamburg	München	Frankfurt	Dresden
Anzahl	3.442.700	1.774.200	1.330.400	671.900	517.100

Tab: Standardabweichung

Auf dem Tabellenblatt **Daten** steht der Berliner Wert in der Zelle B10 und wird dividiert durch die Summe der Werte aller fünf Großstädte, die in den Spalten B bis F eingetragen sind. Die Referenzwerte der fünf Städte sind komplett als fester Bereich eingegeben, damit sie sich nicht verschieben. Der Zähler der Formel muss aber dynamisch sein und sich über die Spalten B bis F anpassen, daher kein **\$-Zeichen** vor dem Spaltenbuchstaben im Zähler (siehe „feste Bezüge“ in EXCEL-TUTORIAL 1). Das Ergebnis der Berechnung in B7 lautet 45 % (siehe Screenshot oben).



- in je -	Berlin	Hamburg	München	Frankfurt	Dresden	gewichteter Durchschnitt
4 Kitas	262	157	144	156	304	211
5 Schulen	183	176	192	188	216	185
6 Kultur	42	26	31	26	18	34
7 Gewichtung	45%	23%	17%	9%	7%	

- in je -	Berlin	Hamburg	München	Frankfurt	Dresden	Standardabweichung s	Variationskoeffizient v
11 Kitas	262	157	144	156	304	59	28%
12 Schulen	183	176	192	188	216	10	5%
13 Kultur	42	26	31	26	18	8	24%

### Schritt 3

Im nächsten Schritt wird der Mittelwert für die einzelnen Aufgabenbereiche ermittelt. Hierbei darf nicht die Excel-Formel **=MITTELWERT()** verwendet werden, die die Pro-Kopf-Werte der einzelnen Städte aufsummiert und durch die Anzahl der Städte dividiert. Würde man so verfahren, bekäme man den ungewichtigen Mittelwert, in dem der Pro-Kopf-Wert Berlins die gleiche Relevanz hätte wie der Wert Dresdens, obwohl Berlin mehr als sechsmal so groß ist.

Stattdessen muss man die Mittelwerte gewichten: Die Funktion **=MITTELWERT()** besitzt keine Zusatzparameter, mit denen Gewichtungen zugewiesen werden können, sodass man eine alternative Formel heranziehen muss. Sie lautet: **=SUMMENPRODUKT(Gewichtungsanteile; Einzelwert)/SUMME(Gewichtungsanteile)**. Im ersten Teil der Formel werden die Einzelwerte mit den jeweiligen Gewichtungen multipliziert. Die Summe daraus teilen Sie anschließend durch die Summe aller Gewichtungsanteile. Durch die Verwendung von absoluten und relativen Bezügen lässt sich diese Formel problemlos auf die anderen Aufgabenbereiche darunter übertragen. Sie lautet beispielsweise für den Kita-Bereich in Zelle G4:

**=SUMMENPRODUKT(B\$7:F\$7;B4:F4)/SUMME(B\$7:F\$7)**

Als Ergebnis bekommt man 211 € pro Kopf für den Kita-Bereich zurückgeben.

Zum Vergleich: der ungewichtete (in diesem Fall falsche) Wert läge bei 204 € pro Kopf. Dass der gewichtete Durchschnitt über dem ungewichteten Wert liegt, ist nachvollziehbar, da insbesondere Berlin mit seinem statistischen Gewicht von 45 % den Mittelwert nach oben zieht.

### Tipp

Die SUMMENPRODUKT-Formel ist eine Matrixfunktion, die zunächst Matrizen miteinander multipliziert und anschließend die Ergebnisse dieser Produkte aufsummiert. Die beiden Matrizen werden innerhalb der Formel durch ein Semikolon voneinander getrennt. Man kann für die einzelnen Matrizen innerhalb der Formel aber auch noch unterschiedliche Bedingungen oder Regeln hinterlegen; dann werden die Matrizen jeweils mit dem Multiplikationszeichen (\*) voneinander abgegrenzt. So kann man als alternative Formel auch **=SUMMENPRODUKT(B4:F4\*B7:F7)/SUMME(B7:F7)** verwenden.

- in je Ev. -	Berlin	Hamburg	München	Frankfurt	Dresden	gewichteter Durchschnitt
Kitas	262	157	144	156	304	211
Schulen	183	176	192	188	216	185
Kultur	42	26	31	26	18	34
Gewichtung	45%	23%	17%	9%	7%	

- in je Ev. -	Berlin	Hamburg	München	Frankfurt	Dresden	Standardabweichung s	Variationskoeffizient v
Kitas	262	157	144	156	304	59	28%
Schulen	183	176	192	188	216	10	5%
Kultur	42	26	31	26	18	8	24%

#### Schritt 4

Im nächsten Schritt wird die Standardabweichung berechnet. Dafür haben wir die Tabelle noch einmal kopiert und errechnen in Zelle G11 die Standardabweichung, die wiederum eine Gewichtung erfahren muss.

Die Formel dafür lautet:

$=\text{WURZEL}(\text{SUMMENPRODUKT}((\text{B}\$7:\text{F}\$7)*((\text{B}4:\text{F}4)^2))-(\text{SUMMENPRODUKT}(\text{B}\$7:\text{F}\$7;\text{B}4:\text{F}4)^2))$ .

Da die Formel nicht ganz einfach zu verstehen ist, wird sie in folgenden Schritt für Schritt erläutert:

(a) Vorne steht die Funktion **=WURZEL()**. Diese ist notwendig, da die Standardabweichung die Quadratwurzel der Varianz ist, die in den folgenden Teilen der Formel berechnet wird.

(b) Es folgt die erste **SUMMENPRODUKT**-Komponente der Formel: Hierbei wird der Pro-Kopf-Wert jeder Stadt quadriert  $((\text{B}4:\text{F}4)^2)$  und mit der entsprechenden Gewichtung der jeweiligen Stadt multipliziert. Anschließend werden diese einzelnen Werte summiert. Der Wert ist damit die Summe der quadrierten und gewichteten Merkmalsausprägungen.

(c) In der zweiten **SUMMENPRODUKT**-Komponente werden die Produkte aus Pro-Kopf-Ausgaben und Gewichtung gebildet, zu einer Gesamtzahl summiert und anschließend noch quadriert. Dieser Wert ist damit nichts anderes als die Potenz des gewichteten Durchschnitts.

#### Tipp

An diesem komplexem Beispiel ist die unterschiedliche Schreibweise der **SUMMENPRODUKT**-Formel zu sehen. In der ersten **SUMMENPRODUKT**-Komponente wird nur ein Teil der Matrizen quadriert, sodass man das Multiplikationszeichen verwenden muss. Im hinteren Teil werden beide Matrizen gemeinsam quadriert, sodass mit dem Semikolon gearbeitet wird.

### Schritt 5

Man erkennt, dass die Werte der Standardabweichung voneinander abweichen. Dies hat einerseits mit der Streuung der Ausprägung zu tun, andererseits aber auch mit der Skalierung der zu untersuchenden Werte. Was sagt nun eine Standardabweichung von 59 € bei einem hohen Ausgabenniveau wie bei den Kitas im Vergleich zu einer Standardabweichung von nur 8 € bei niedrigen Ausgaben im Kulturbereich?

-in je Ev. -	Berlin	Hamburg	München	Frankfurt	Dresden	gewichteter Durchschnitt
Kitas	262	157	144	156	304	211
Schulen	183	176	192	188	216	185
Kultur	42	26	31	26	18	34
Gewichtung	45%	23%	17%	9%	7%	

-in je Ev. -	Berlin	Hamburg	München	Frankfurt	Dresden	Standardabweichung s	Variationskoeffizient v
Kitas	262	157	144	156	304	59	28%
Schulen	183	176	192	188	216	10	5%
Kultur	42	26	31	26	18	8	24%

-in je Ev. -	Berlin	Hamburg	München	Frankfurt	Dresden	Standardabweichung s	Variationskoeffizient v
Kitas	262	157	144	156	304	59	28%
Schulen	183	176	192	188	216	10	5%
Kultur	42	26	31	26	18	8	24%

Um diese Frage zu beantworten, muss ein zusätzlicher Vergleichsmaßstab eingeführt werden: Dabei wird die Standardabweichung auf den Mittelwert bezogen und somit unabhängig von der Messdimension. Dieses Maß der relativen Streuung bezeichnet man als Variationskoeffizienten. Er errechnet sich aus einer einfachen Division. Der Wert für den Variationskoeffizienten z.B. der Kitas sagt aus, dass die Standardabweichung in diesem Aufgabenbereich 28 % des Kita-Durchschnitts in der Gruppe der fünf Großstädte ausmacht. Er liegt damit nur unwesentlich über dem Variationskoeffizienten des Kulturbereichs von 24 %.



### Schritt 6

Nun können die auffälligen Abweichungen weiter untersucht werden.

Die ZDL ist bei den Arbeiten zum Projekt „Vergleichbare Datengrundlagen“ von der Hypothese ausgegangen, dass bei einem Variationskoeffizienten von größer als 25 % so gravierende Unterschiede bestehen, dass sie nicht mehr durch Politikunterschiede zu erklären sind. Tatsächlich erbrachte die gemeinsam mit den Haushaltsabteilungen durchgeführte Prüfung der damals auffälligen Positionen Hinweise auf systematische Fehlbuchungen und Inkonsistenzen im Volumen von mehreren Milliarden €. Die statistischen Ämter haben die beschriebene Methode daraufhin für ihre eigenen Maßnahmen der Qualitätssicherung übernommen.

# IMPRESSUM

## Herausgeber:

Zentrale Datenstelle der Landesfinanzminister (ZDL)  
Leipziger Straße 3-4  
10117 Berlin  
Tel.: (030) 18 91 00-612  
[www.zdl-berlin.de](http://www.zdl-berlin.de)

## Satz und Layout:

junge meister\* GmbH  
Kantstraße 149  
10623 Berlin  
Tel. 030 | 437 27 - 079  
[www.junge-meister.de](http://www.junge-meister.de)

Die »Excel-Tutorials« sind urheberrechtlich geschützt (Copyright). Dies bedeutet, dass die Vervielfältigung und Verbreitung – auch auszugsweise – nur mit genauer Quellenangabe des Herausgebers gestattet ist. Die »Excel-Tutorials« sind auf die Bedürfnisse unserer Kunden zugeschnitten und verwenden daher beispielhaft Namen von Gebietskörperschaften. Die den Beispielen zugrundeliegenden Zahlen spiegeln keine realen Sachverhalte wider und dienen lediglich der Illustration. Der Auswahl der Gebietskörperschaften liegt keine politische oder sonstige Wertung zugrunde. An einem Feedback zu den »Excel-Tutorials« sind wir ausdrücklich interessiert. Wenn Sie Anregungen und Wünsche haben oder uns Hinweise auf Fehler geben können, sind wir für eine Nachricht dankbar.

Erschienen: November 2012