

Erstellung von großen Datentabellen mit Hilfe von Zufallszahlen

Zufallszahlen

Zentraler Inhalt jedes Datenbankunterrichts ist die Abfrage. Diese macht aber erst dann richtig Sinn und Spaß, wenn man sie in sehr große Datentabellen durchführt, wobei groß hier eine große Anzahl von Datensätzen meint, mindestens 1000, besser mehr. Woher soll man nun aber diese Tabellen nehmen, wenn man nicht Tausende Datensätze von Hand eintippen möchte? Eine Hilfe bietet der Zufallszahlengenerator in Excel, der im Programm als Funktion integriert ist. Diese Funktion ZUFALLSZAHL() liefert zunächst eine beliebige Zahl zwischen 0 und 1. Möchte man größere Zahlen erzeugen, zum Beispiel eine zufällige Zahl zwischen 0 und 10, so muss man einfach mit 10 multiplizieren: ZUFALLSZAHL() * 10. Dazu einige Beispiele:

	A1	=	=ZUFALLSZAHL()*10
	A	B	C
1	1,14159389010742		
2	4,62474332033494		
3	1,71741900943222		
4	8,88536163380570		
5	5,24987560573161		
6	8,14578567496723		
7	1,87438385320666		
8	0,43700334063305		

Zufallszahlen zwischen 0 und 10

Man sieht, dass die Zufallszahlen unendliche Dezimalzahlen sind, was für unseren Zweck oft ungünstig ist. Daher schneiden wir den Nachkommaanteil der Zahlen mit der Funktion GANZZAHL() ab. Die Formel für eine ganzzahlige

	A1	=	=GANZZAHL(ZUFALLSZAHL()*20)	
	A	B	C	D
1	8			
2	17			
3	10			
4	14			
5	2			
6	9			
7	1			
8	3			

Zufallszahlen zwischen 0 und 10 als Ganzzahlen

Zufallszahl zwischen 0 und 20 lautet somit:

=GANZZAHL(ZUFALLSZAHL()*20)

Die Zahlen werden ohne Dezimalstellen angezeigt. Nach diesem Verfahren kann man auch beliebige Zahlen in einem Bereich zwischen einer unteren

(MIN) und einer oberen Grenze (MAX) erzeugen. Die Formel dafür lautet:

=GANZZAHL(ZUFALLSZAHL() * (MAX - MIN) + MIN)

Folgende Anweisung erzeugt eine ganzzahlige Zufallszahl zwischen 20 und 70:

=GANZZAHL(ZUFALLSZAHL() * (70 - 20) + 20)

Tabellen

Eine gute Grundlage für Personentabellen liefern die Schülerdaten aus der Schulverwaltung, an unserer Schule sind darin derzeit ca. 1200 Datensätze. Dieser Datei entnimmt man folgende Datenfelder samt Inhalten: Name, Vorname, Geschlecht, Straße und eventuell noch die PLZ und den Wohnort. Die beiden Felder PLZ und Wohnort kann man auch weglassen, da diese auch leicht neu erzeugt werden können, während dies bei Name, Vorname (verknüpft mit dem Geschlecht) und Straße schlecht möglich ist. Wird nun diese Tabelle gründlich durchgewürfelt, steht zum Schluss kein Vorname mehr beim ehemaligen Namen und kein Name mehr bei der ehemaligen Straße, nur Vorname und Geschlecht sollte man nie trennen. In der neuen Tabelle wird sich bestimmt keine Schülerin und kein Schüler wieder finden, der Datenschutz ist also erfüllt.

	A	B	C	D
	Name	Vorname	Geschlecht	Straße
1	Pfisterer	Tobias	m	Kirchstraße 7
2	Schlör	Yvonne	w	Apfelweg 4
3	Eisele	Martina	w	Lautenstraße 27
4	Schmid	Sandra	w	Kellergasse 34
5	Begic	Richard	m	Rathausstr. 28
6	Katz	Martina	w	Steinhölzle 13
7	Frank	Martina	w	Gartenstr.27
8	Wittlinger	Deborah	w	Waldweg 15
9	Schlien	Renate	w	Zum Spornplatz 12

Die neue anonymisierte Tabelle

Würfeln

1. Schritt: Trennung von Name und Vorname und der Straßen

Einfache Methode: Die Spalte mit dem Namen wird einzeln alphabetisch sortiert. Der Nachteil dabei ist, dass die Daten dann schon sortiert vorliegen.

Zweite Methode: Vor der Spalte mit den Namen fügt man eine leere Spalte ein. Diese wird mit Zufallszahlen zwischen 0 und Anzahl der Datensätze (hier 1200) aufgefüllt: (Abb. auf der nächsten Seite)

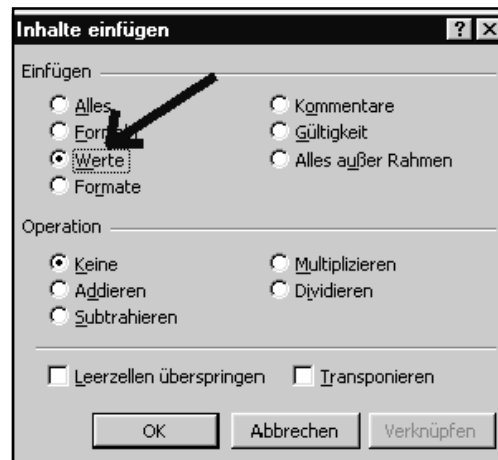
	A	B	C	D	E
	=GANZZAHL(ZUFALLSZAHL()*1200)				
1	Hilfsspalte	Name	Vorname	Geschlecht	Strasse
2	691	Pfisterer	Tobias	m	Kirchstraße 7
3	97	Schlor	Yvonne	w	Apfelweg 4
4	1164	Eisele	Martina	w	Lauternstraße 27
5	547	Schmid	Sandra	w	Kellergasse 34
6	1028	Begic	Richard	m	Rathausstr. 28
7	550	Katz	Martina	w	Steinhölzle 13
8	439	Frank	Martina	w	Gartenstr.27
9	704	Wittlinger	Deborah	w	Waldweg 15
10	850	Schlien	Benate	w	Zum Sportplatz 12

Eingefügte Spalte mit Zufallszahlen

Da sich die Felder mit den Formeln nicht sortieren lassen, fügt man zwischen der Hilfsspalte und der Namensspalte eine weitere Spalte „Zufallszahlen“ ein. In diese neue Spalte kopiert man die Inhalte der Hilfsspalte mit dem Befehl Bearbeiten / Inhalte einfügen ... und zwar nur die Werte: (siehe Abb. rechts)

Nun sortiert man die Spalte B und die Spalte C zusammen - aufsteigend nach Spalte B, das heißt nur die beiden Spalten B und C dürfen markiert sein, die Daten in Spalte D, E, F, ... bleiben stehen (siehe Abb. unten).

Das Ergebnis sieht dann folgendermaßen aus: (siehe Abb. auf der nächsten Seite)



So fügt man richtig nur die Inhalte der Hilfsspalte ein

	B	C	D	E	F
24	Zufallszahl	Name	Vorname	Geschlecht	Strasse
24	705	Pfisterer	Tobias	m	Kirchstraße 7
31	299	Schlor	Yvo		
31	491	Eisele	Mar		
16	1179	Schmid	San		
27	925	Begic	Rich		
92	25	Katz	Mar		
44	1052	Frank	Mar		
80	403	Wittlinger	Deb		
69	27	Schliep	Rer		
71	507	Ley	Julia		
46	37	Wiedenmann	San		
41	645	Kellner	Cinc		
33	142	Deutschländer	Bett		
81	870	Korn	Anja		
62	647	Steinle	Slav		
52	654	Diem	Nico		
65	946	Aubele	Silv		
05	1130	Großkinsky	Oliver	w	Lembchenstr. 43

Sortieren der Spalte B und C zusammen nach der Spalte B

B	C	D	E	F
Zufallszahl	Name	Vorname	Geschlecht	Strasse
1	Sax	Tobias	m	Kirchstraße 7
1	Röschl	Yvonne	w	Apfelweg 4
2	Wegner	Martina	w	Lauternstraße 27
2	Korkaric	Sandra	w	Kellergasse 34
4	Ledermann	Richard	m	Rathausstr. 28
5	Dimoti	Martina	w	Steinhölzle 13
6	Kurt	Martina	w	Gartenstr.27
6	Münst	Deborah	w	Waldweg 15
6	Kulik	Renate	w	Zum Sportplatz 12
7	Spann	Julia Elisabeth	w	Seidlheck 9
7	K...	R...	w	M...

Eingefügte Spalte mit Zufallszahlen

Dasselbe Verfahren wendet man an um auch noch die Straßen neu an die Personen zu verteilen (siehe Abb. oben). Dabei kann man die selbe Hilfsspalte verwenden, da bei jeder Aktion des Programms die Zufallszahlen in den einzelnen Zellen neu berechnet werden. (Daher lassen sich auch die Felder mit den Formeln nicht sortieren.)

2. Schritt: Neues Geburtsdatum

Unsere Personen benötigen nun ein Geburtsdatum. Dazu muss man wissen, dass in Excel jedes Datum als Datumszahl (Anzahl der Tage seit dem 1.1.1900) dargestellt ist. Möchte man zum Beispiel Daten zwischen dem 1.1.1950 und dem 1.1.1960 erzeugen, so muss man Zufallszahlen zwischen diesen beiden Datumszahlen erzeugen: (Abb. rechts)

*Geburtsdaten
z w i s c h e n
1. Januar 1950
und 1. Januar
1960*

Datum	Datumszahl
1. Januar 1950	18264
1. Januar 1960	21916

Die Formel dazu lautet:

$$=\text{GANZZAHL}(\text{ZUFALLSZAHL}()*(21916-18264))+18264)$$

Wichtig ist auch hier, dass man zunächst mit einer Hilfsspalte arbeitet und dann nur die Werte in die Geburtsdatumsspalte einfügt, sonst würden unsere Personen laufend ihr Geburtsdatum verändern. (vergl. Abb. unten) Zum Schluss nur noch im Datumsformat formatieren.

A2	=GANZZAHL(ZUFALLSZAHL()*(21916-18264))+18264)					
	A	B	C	D	E	
1	Hilfsspalte	Name	Vorname	Gebdat	Geschlecht	St
2	20371	Sax	Tobias	19.01.55	m	Kir
3	20059	Röschl	Yvonne	17.04.55	w	Ap
4	20707	Wegner	Martina	15.01.58	w	La
5	20755	Korkaric	Sandra	05.02.56	w	Ke
6	20831	Ledermann	Richard	05.10.59	m	Ra
7	18517	Dimoti	Martina	13.04.53	w	St
8	20712	Kurt	Martina	02.04.59	w	Ge
9	20844	Münst	Deborah	24.08.59	w	W
10	21891	Kulik	Renate	29.08.52	w	Zu
11	19722	Spann	Julia Elisabeth	06.01.51	w	Se

Tabelle mit Hilfsspalte zur Erstellung von Geburtsdaten

Neue Datenfelder anfügen

T	U
0	Tanzen
1	Wandern
2	Schilaufen
3	Schwimmen
4	Reisen
5	Tennis
6	Reiten
7	Malen
8	Volleyball
9	Lesen
10	Briefmarken
11	Tauchen
12	Klettern
13	Basteln
14	Musik
15	Theater
16	Garten

Datenfeld: Hobby

2. Schritt:

In der Tabelle eine Spalte einfügen. In diese Spalte folgende Formel eingeben

=GANZZAHL(ZUFALLSZAHL()*16), wobei hier mit der höchsten Zahl der Tabelle multipliziert wird, hier mit 16.

=GANZZAHL(ZUFALLSZAHL()*(Anzahl Feldinhalte-1))

Diese Spalte ist die Spalte R! In dieser Spalte erscheinen nun zufällig verteilt die Zahlen von 0 bis 16.

3. Schritt:

Vor dieser Spalte (Spalte R) nun noch eine leere Spalte (dies ergibt die Spalte Q) in die Tabelle einfügen. In diese Spalte folgende Formel eingeben:

=SVERWEIS(R3;T\$1:U\$17;2)

Q	R	S	T	U
Hobby				0 Tanzen
Wandern	1			1 Wandern
Basteln	13			2 Schilaufen
Tanzen	0			3 Schwimmen
Basteln	13			4 Reisen
Lesen	9			5 Tennis
Musik	14			6 Reiten
Reisen	4			7 Malen
Briefmarken	10			8 Volleyball
Wandern	1			9 Lesen
Klettern	12			10 Briefmarken

Die neu verteilten Hobbys

Unsere Personen bekommen ein Hobby zugeteilt.

1. Schritt:

Für gewünschte Feldinhalte in einem neuen Tabellenblatt oder wie hier im rechten, freien Tabellenbereich eine Hilfstabelle erstellen, die in der ersten Spalte die Zahlen von Null aufsteigend und in der zweiten Spalte die gewünschten Feldinhalte enthält. Hier durchlaufen die Zahlen in der ersten Spalte bei 17 verschiedenen Feldinhalten die Zahlen von 0 bis (Anzahl-1), also 16. (vergl. Abb. links)

Dadurch erscheinen nun in dieser Spalte Q die gewünschten Feldinhalte zufällig verteilt: (siehe Abb. unten)

Da nun aber in der Spalte Q die Inhalte durch eine Formel entstehen, kann diese Tabelle noch nicht in die Datenbank übernommen werden, man benötigt noch einen weiteren Schritt.

4. Schritt:

Vor dieser Spalte (Spalte Q) nun noch eine leere Spalte (dies ergibt die Spalte R) in die Tabelle einfügen. Diese Spalte wird später in die Datenbank übernommen. Alle Daten der Spalte Q werden markiert und in die Zwischenablage kopiert. Diese Daten werden dann in die Spalte R eingefügt, aber nur die Inhalte der Felder und nicht die Formeln. Dies geschieht wieder mit dem Befehl:

Bearbeiten / Inhalte einfügen ... / Werte

Nun können alle Hilfsspalten samt der Hilfstabelle gelöscht werden (Spalten Q bis U).

Auf diese Weise können alle erdenklichen Datenfelder (Augenfarbe, Kontostand, Bildung, Familienstand, offenen Rechnungsbeträge, usw) mit entsprechenden Inhalten angefügt werden, zum Beispiel kann man sich eine Hilfstabelle mit Postleitzahlen und Wohnorten erstellen, falls man diese Daten nicht aus der Ursprungsdatei übernommen hat. Wichtig ist dabei, dass man keine Felder mit Formeln in die Datentabelle übernimmt, sondern stets nur die Werte einkopiert. Bei Geldbeträgen sollte man darauf achten, dass die Zahlen nur zwei Nachkommastellen haben. Die erreicht man zum Beispiel für beliebige Beträge zwischen 150 DM und 500 DM durch folgende Formel:

=GANZZAHL((ZUFALLSZAHL()*(500-150) + 150) * 100) / 100.

Formatieren der Zahlen auf zwei Nachkommastellen würde nicht ausreichen, da sich dies nur auf die Anzeige, nicht aber auf die Feldinhalte auswirkt.

Und nun viel Spaß und gute Ideen!

Ulrich Stark (ZPG)

Und der Spruch zum Schluss:

"Ich glaube an das papierlose Büro genauso wenig wie an das papierlose Klo."

Der stellvertretende Landtagspräsident Frieder Birzele in der Debatte über den Etat und die Ausstattung des Parlaments in Stuttgart mit neuen Computersystemen