

# Zonk, „Wetten, dass“ und der t-Test

**Jörg Sellmann**

*joerg.sellmann@julius-kuehn.de*

Tor 1



Tor 2



Tor 3



# Zentrale Fragen:



Soll man wechseln?

Wie hoch sind die Wahrscheinlichkeiten,  
50:50 oder anders?

## Antworten:

Ja, man(n)/frau sollte wechseln.

Die Fragestellung ist theoretisch gelöst, die  
Wahrscheinlichkeiten liegen bei  $\frac{2}{3} : \frac{1}{3}$

Aber:



Es handelt sich um Wahrscheinlichkeiten.

Im konkreten Fall kann es falsch sein.

# Lösungsansätze:

1. Basic
2. SAS<sup>®</sup>
3. Python
4. Gesunder Menschenverstand

Basic (2008):



```

1  DEFINT A-Z
2  RANDOMIZE TIMER                                'Der Zufall wird vorbereitet
3  PRINT "Moderator darf weder Auto- noch Wahltür öffnen"
4  PRINT ""
5  FOR I = 1 TO 10                                'Zehn Versuche
6      R = 0                                       'Richtig <-- 0
7      F = 0                                       'Falsche <-- 0
8      FOR J = 1 TO 1000                          '1000 Wiederholungen
9          A = INT (3 * RND(1) + 1)                'Zufallswahl der Autotür A
10         W1 = INT (3 * RND(1) + 1)               'Zufällige Erstwahl W1
11         DO
12             M = INT (3 * RND(1) + 1)            'Moderator will Tür M öffnen
13             LOOP UNTIL (M <> A) AND (M <> W1)    'darf weder Auto noch W1 sein
14             W2 = 6 - M - W1                     'W2 wäre die Tür, zu der nun gewechselt werden kann
15             IF W2 = A THEN
16                 R = R + 1                        'Auto wurde getroffen
17             END IF
18             IF W1 = A THEN
19                 F = F + 1                        'Auto wurde verfehlt
20             END IF
21         NEXT J
22     PRINT "Wechseln richtig: "; R ; " Wechseln falsch: "; F
23 NEXT I
24

```



# Basic (2008):

Console

Moderator darf weder Auto- noch Wahltür öffnen

Wechseln richtig: 637	Wechseln falsch: 363
Wechseln richtig: 680	Wechseln falsch: 320
Wechseln richtig: 676	Wechseln falsch: 324
Wechseln richtig: 657	Wechseln falsch: 343
Wechseln richtig: 646	Wechseln falsch: 354
Wechseln richtig: 648	Wechseln falsch: 352
Wechseln richtig: 653	Wechseln falsch: 347
Wechseln richtig: 656	Wechseln falsch: 344
Wechseln richtig: 660	Wechseln falsch: 340
Wechseln richtig: 687	Wechseln falsch: 313

SAS® (2015):



```

data zonk;
  do j=1 to 100;
    R=0;
    F=0;

    do i=1 to 1000;
      Auto=int(3*rand('UNIFORM')+1);
      Wahl1=int(3*rand('UNIFORM')+1);
      Moderator=0;
      do until (Moderator ne Wahl1);
        do until (Moderator ne Auto);
          Moderator=int(3*rand('UNIFORM')+1);
        end;
      end;
      Wahl2=6 - Moderator - Wahl1;
      if (Wahl2 = Auto)
        then R=R+1;
        else F=F+1;
      end;

      output;
    end;

    keep R F;
  run;

```

\* 100 Simulationen ;  
 \* summiert die Richtigen --> Auto ;  
 \* summiert die Falschen --> Zonk ;

\* 1000 Wiederholungen pro Simulation ;  
 \* Tor (1,2,3) vor der Sendung für das Auto wählen ;  
 \* Kandidat wählt in der Sendung ein Tor (1,2,3);

\* Moderator darf nicht Kandidatentor öffnen ;  
 \* Moderator darf nicht Tor mit dem Auto öffnen ;

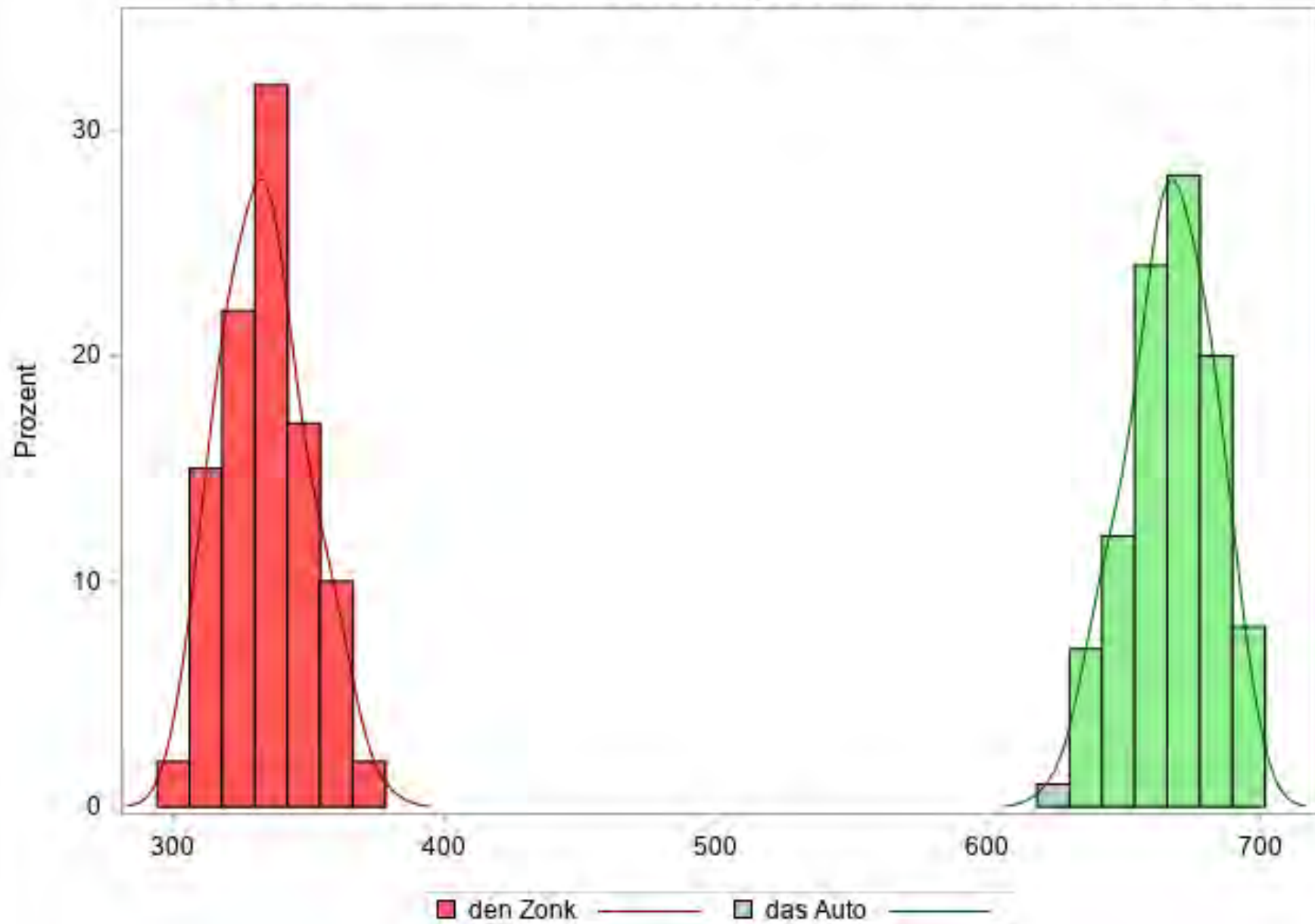
\* Kandidat wechselt das Tor ;

\* Auto wurde gewählt ;  
 \* Zonk wurde gewählt ;

\* Simulationsergebnis speichern ;

\* nur die Summen interessieren ;

### Wenn der Kandidat 1000-mal das Tor gewechselt hätte, dann hätte er ...



# Python (2023):



```
1  #!/usr/bin/python
2
3  import random
4
5  print("Moderator darf weder Auto- noch Wahluer oeffnen")
6  print("")
7
8  for i in range(10):
9      R=0
10     F=0
11
12     for j in range(1000):
13         A= random.randint(1,3)      # Zufallswahl der Autotuer A - zufaellige Zahl zwischen 1 und 3
14         W1=random.randint(1,3)     # Zufaellige Erstwahl W1
15
16         while 1:                    # while TRUE:
17             M=random.randint(1,3)  # Moderator will Tuer M oeffnen
18             if (M != A) and (M != W1):
19                 break
20
21         W2 = 6-M-W1                 # W2 waere die Tuer, zu der nun gewechselt werden kann
22
23         if W2 == A:
24             R = R + 1                # Auto wurde getroffen
25         if W1 == A:
26             F = F + 1                # Auto wurde nicht getroffen
27
28     print("Runde: ", '{:3d}'.format(i+1), "    Wechseln richtig: ", R , "    Wechseln falsch: ", F)
```

# Python (2023):



```
192.168.237.131 - PuTTY
sellmann@ubuntu:~$ ./zonk.py
Moderator darf weder Auto- noch Wahluer oeffnen

Runde:      1      Wechseln richtig:  687      Wechseln falsch:  313
Runde:      2      Wechseln richtig:  672      Wechseln falsch:  328
Runde:      3      Wechseln richtig:  654      Wechseln falsch:  346
Runde:      4      Wechseln richtig:  642      Wechseln falsch:  358
Runde:      5      Wechseln richtig:  667      Wechseln falsch:  333
Runde:      6      Wechseln richtig:  681      Wechseln falsch:  319
Runde:      7      Wechseln richtig:  672      Wechseln falsch:  328
Runde:      8      Wechseln richtig:  679      Wechseln falsch:  321
Runde:      9      Wechseln richtig:  649      Wechseln falsch:  351
Runde:     10      Wechseln richtig:  653      Wechseln falsch:  347
sellmann@ubuntu:~$ █
```

## Gesunder Menschenverstand (zeitlos 😊)



Die Wahrscheinlichkeit, dass der Wagen hinter der erstgewählten Tür ist, beträgt  $\frac{1}{3}$ .

Die Wahrscheinlichkeit, dass er hinter einer der beiden anderen Türen ist, beträgt somit  $\frac{2}{3}$ .

Wenn ich nun erfahre, hinter welcher der beiden anderen Türen er nicht ist, weiß ich sofort die Tür, hinter der er mit einer Wahrscheinlichkeit von  $\frac{2}{3}$  ist.





$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{2}{n}}}$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Zentrale Frage:



Was macht eigentlich der t-Test?

Antwort:



Eine Aussage über einen Lageparameter einer nicht vollständig untersuchbaren Grundgesamtheit anhand einer Stichprobe treffen.

Aber:



Wie „gut“ sind die Stichproben?

Kann man von der Stichprobe auf  
die Grundgesamtheit schließen?

Drei Ziele:

Arbeiten mit Arrays in SAS<sup>®</sup>

Dynamisches Bilden von Mittelwerten

Simulieren des Stichproben-Ziehens

# Arbeiten mit Arrays in SAS®:

```
%LET wdh      = 20;          /* Größe der Grundgesamtheiten */
%LET sample  =  5;          /* Größe der Stichprobe */
%LET start_m =  5;          /* Startwert M */
%LET start_w =  1;          /* Startwert W */

%LET dim = %sysfunc(comb(&wdh., &sample.));
                                     /* Anzahl Permutationen */
```

dim -> 15504

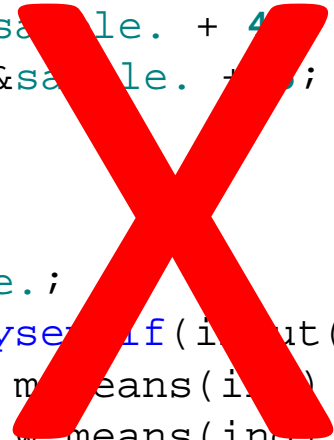
# Arbeiten mit Arrays in SAS®:

```
/* Daten erzeugen mittels Array */  
data wiesbaden2 (drop=i);  
  call streaminit(123);  
  array m{&wdh.} M01-M&wdh.;  
  array w{&wdh.} W01-W&wdh.;  
  do i=1 to &wdh.;  
    m{i}=&start_m. + i + rand("Normal",0,1);  
    w{i}=&start_w. + i + rand("Normal",0,1);  
  end;  
run;
```

# Dynamisches Bilden von Mittelwerten:



```
ind=0;
do i1=1 to &wdh. - &sample. + 1;
  do i2=i1+1 to &wdh. - &sample. + 2;
    do i3=i2+1 to &wdh. - &sample. + 3;
      do i4=i3+1 to &wdh. - &sample. + 4;
        do i5=i4+1 to &wdh. - &sample. + 5;
          ind+1;
          m_means(ind)=0;
          w_means(ind)=0;
          do index=1 to &sample.;
            %LET mindex=%sysfunc(which_max(m(i&index.),best.));
            m_means(ind) = m_means(ind) + m(i&mindex.);
            w_means(ind) = w_means(ind) + w(i&mindex.);
          end;
          m_means(ind)=m_means(ind)/&sample.;
          w_means(ind)=w_means(ind)/&sample.;
        end;
      end;
    end;
  end;
end;
end;
```





```
%macro mittelwerte(wdh=, sample=);
```

```
data mittelwerte(keep=QM0001-QM&dim. QW0001-QW&dim.);
```

```
set wiesbaden2;
```

```
array _vektor(1:&wdh.) a1-a&wdh.;
```

```
array m_means{1:&dim.} QM0001-QM&dim.;
```

```
array w_means{1:&dim.} QW0001-QW&dim.;
```

```
array m{1:&wdh.} M01-M&wdh.;
```

```
array w{1:&wdh.} W01-W&wdh.;
```

```
index=0;
```

```
%do i = 1 %to &wdh.;
```

```
do x&i.= 0 to 1;
```

```
a&i.=x&i.;
```

```
%end;
```

```
if sum(of a1-a&wdh.)=&sample. then do; /* Nächste Folie */ end;
```

```
%do i = 1 %to &wdh.; end; %end;
```

```
run;
```

```
%mend mittelwerte;
```



	a01	a02	a03	a04	a05	a06	a07	a08	a09	a10	a11	a12	a13	a14	a15	a16	a17	a18	a19	a20
1											1	1	1	1	1	1	1	1	1	1
2										1		1	1	1	1	1	1	1	1	1
3										1	1		1	1	1	1	1	1	1	1
4										1	1	1		1	1	1	1	1	1	1
5										1	1	1	1		1	1	1	1	1	1
6										1	1	1	1	1		1	1	1	1	1
7										1	1	1	1	1	1		1	1	1	1
8										1	1	1	1	1	1	1		1	1	1
9										1	1	1	1	1	1	1	1		1	1
10										1	1	1	1	1	1	1	1	1		1
...																				
184747	1	1	1	1	1	1	1	1	1											1
184748	1	1	1	1	1	1	1	1	1									1		
184749	1	1	1	1	1	1	1	1	1								1			
184750	1	1	1	1	1	1	1	1	1							1				
184751	1	1	1	1	1	1	1	1	1						1					
184752	1	1	1	1	1	1	1	1	1					1						
184753	1	1	1	1	1	1	1	1	1				1							
184754	1	1	1	1	1	1	1	1	1			1								
184755	1	1	1	1	1	1	1	1	1		1									
184756	1	1	1	1	1	1	1	1	1	1										



Original:  
Permutationsmatrix  
10 aus 20

```
%macro mittelwerte(wdh=, sample=);  
  
...  
  
if sum(of a1-a&wdh.)=&sample. then do;  
    index+1;  
    mwm=0; mww=0;  
    do j=1 to &wdh.;  
        mwm = mwm + m(j)*_vektor(j);  
        mww = mww + w(j)*_vektor(j);  
    end;  
    m_means(index)= mwm/&sample.;  
    w_means(index)= mww/&sample.;  
  
end;  
  
...  
  
%mend mittelwerte;  
  
%mittelwerte(wdh=&wdh., sample=&sample.);
```

# Simulieren des Stichproben-Ziehens:



Jedes M mit jedem W:

M=15.63

W=11.65

m_greater	w_greater	vergleiche1	vergleiche2	proz_m	proz_w
210.341.481	30.032.535	240.374.016	240.374.016	87.5059	12.4941

Zufällige Auswahl von M als auch W:

anzahl	m_greater	w_greater	proz_m	proz_w
1.000	871	129	87.1000	12.9000
10.000	8.685	1.315	86.8500	13.1500
100.000	87.482	12.518	87.4820	12.5180
1.000.000	875.333	124.667	87.5333	12.4667
10.000.000	8.750.951	1249.049	87.5095	12.4905

# Ergebnis t-Test:



Method	Variances	DF	t Value	Pr >  t
Pooled	Equal	38	2.01	0.0514
Satterthwaite	Unequal	37.956	2.01	0.0514

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	19	19	1.07	0.8836



Wie viele Personen sollten im Raum sein, damit ich eine realistische Chance habe die Wette zu gewinnen, dass wenigstens 2 Personen **am gleichen Tag Geburtstag** haben?

183 ?    Geburtstagsparadoxon

## Zentrale Fragen:

Was heißt „realistische Chance“ ?

Sind es mehr oder weniger als 183?

Kann SAS<sup>®</sup> mir helfen?

Antworten:

So um die 90 Prozent.

Weniger, aber wie viele genau?

Ja.



# Theoretische Lösung:

$$W_n = \frac{365 * 364 * \dots * (365 - n + 1)}{365^n}$$

$$W_n = \frac{365!}{(365 - n)! * 365^n}$$

$$W_n \approx \left(1 - \frac{n}{365}\right)^{n-365,5} * e^{-n}$$

$W_n$  = Wahrscheinlichkeit,  
dass von  $n$  Ereignissen  
**nicht** zwei am selben Tag  
stattfinden

Gilde/Altrichter: Mehr Spaß mit  
dem **Taschenrechner**, Leipzig, **1989**

# SAS® Studio:

$$W_n = \frac{365 * 364 * \dots * (365 - n + 1)}{365^n}$$



Anzahl Personen in einem Raum	Wahrscheinlichkeit
1	0,000 %
2	0,300 %
3	0,800 %
4	1,600 %
5	2,700 %
6	4,000 %
7	5,600 %
8	7,400 %
9	9,500 %
10	11,700 %
11	14,100 %
12	16,700 %
13	19,400 %
14	22,300 %
15	25,300 %
16	28,400 %
17	31,500 %
18	34,700 %
19	37,900 %
20	41,100 %

Anzahl Personen in einem Raum	Wahrscheinlichkeit
21	44,400 %
22	47,600 %
23	50,700 %
24	53,800 %
25	56,900 %
26	59,800 %
27	62,700 %
28	65,400 %
29	68,100 %
30	70,600 %
31	73,000 %
32	75,300 %
33	77,500 %
34	79,500 %
35	81,400 %
36	83,200 %
37	84,900 %
38	86,400 %
39	87,800 %
40	89,100 %

SAS® (per Hand 🤖):

Erzeugen von 40 Geburtstagen

Oder einfacher:

Erzeugen von 40 Zahlen zwischen 1 und 365

Herausfinden, ob eine Zahl doppelt vorkommt mittels  
PROC sql;

Das Ganze oft genug wiederholen.

```
%macro doppelt_dataset (personen);
```

```
options nonotes;
```

```
data geburtstage;
```

```
  do n=1 to &personen.;
```

```
    tag = %RandBetween(1, 365);
```

```
    output;
```

```
  end;
```

```
  drop n;
```

```
run;
```

```
proc sql noprint;
```

```
  insert into doppelte_tage select count(*) from
```

```
(select distinct tag from geburtstage group by tag having count(*)>1);
```

```
quit;
```

```
%mend;
```



```
%macro doppelt_array (personen);
```

```
options nonotes;
```

```
data geburtstage;
```

```
array tag{&personen.} T01-T&personen.;
```

```
do n=1 to &personen.;
```

```
    tag(n) = %RandBetween(1, 365);
```

```
end;
```

```
drop n;
```

```
run;
```

```
proc transpose data=geburtstage
```

```
out=tgeburtstage (rename=(coll=tag));
```

```
run;
```

```
proc sql noprint; ... ;
```

```
%mend;
```



```
/* Leere Ergebnismenge */  
data doppelte_tage;  
input anzahl;  
cards;  
;  
run;  
  
%macro RandBetween(min, max);  
    (&min + floor((1+&max-&min)*rand("uniform")))  
%mend;  
  
%macro simulation (makro, wdh, personen);  
  
    data _null_;  
        %do i=1 %to &wdh.;  
            %&makro.(&personen.);  
        %end;  
    run;  
  
%mend;
```

```

%let wdh=1000;           /* Anzahl Simulationen */
%let personen=40;      /* Wieviel Personen werden befragt */

%let makro1=doppelt_dataset;
%let makro2=doppelt_array;
%let auswahl=2;        /* Welches Makro wird genutzt */
%let makro=&&makro&auswahl;

%let launchTime = %sysfunc(time());           /* Startzeit */
%simulation(&makro., &wdh., &personen.);
%let landTime = %sysfunc(time());            /* Endezeit */

/* Writes time values using the ISO 8601 extended notation hh:mm:ss.ffffff. */
%let timeTaken = %sysevalf(&landTime.-&launchTime.); /* Dauer */
%let timeTakenFmt = %sysfunc(putn(&timeTaken.,e8601tm12.0));

proc sql;
select "&Makro." as Makro, &wdh. as Wiederholungen, "&timeTakenFmt." as Dauer,
100-100*keine/&wdh. as Erfolgsrate from
(select count(*) as KEINE from doppelte_tage where anzahl=0);
delete from doppelte_tage;
quit;

```



# Ergebnis:



Makro : 40	Wiederholungen	Dauer	Erfolgsrate
doppelt_array	1000	00:00:37	88.9

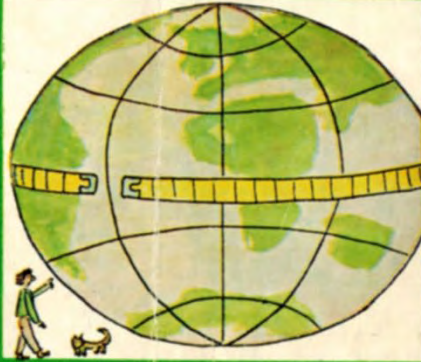
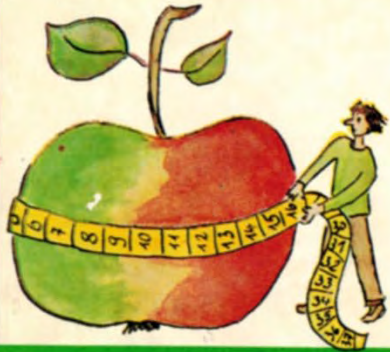
Makro : 40	Wiederholungen	Dauer	Erfolgsrate
doppelt_dataset	1000	00:00:21	88.6

Makro : 22	Wiederholungen	Dauer	Erfolgsrate
doppelt_dataset	1000	00:00:21	49.9



# Mehr Spaß mit dem Taschenrechner

Gilde / Altrichter



Gero von Randow

Das Ziegenproblem

Denken in

Wahrscheinlichkeiten

science

ro  
ro  
ro



Falls Sie mit dem erworbenen Wissen zu Reichtum gelangen sollten, meine Mailadresse haben Sie 😊

Programme unter:  
<https://sf.julius-kuehn.de/sas/2023>

[www.julius-kuehn.de](http://www.julius-kuehn.de)