Tillämpningar på logaritmer kan avslöja fuskare

Benfords lag har fascinerat fackfolk och allmänheten i över ett sekel. 1881 upptäckte den amerikanske astronomen Simon Newcomb att böcker med logaritmtabeller (detta var innan vi fick räknestickor och långt senare miniräknare) alltid verkade mer slitna på de första sidorna med logaritmer för tal som börjar med 1, 2 osv och mindre slitna på de sidor där det fanns logaritmer av tal som började med siffror som 8 och 9.

Vad gjorde logaritmtabeller för nytta. Jo, det var en metod att förenkla krångliga och tidsödande multiplikationer och divisioner med många siffror och ersätta dem med betydligt enklare additioner och subtraktioner. Av någon anledning verkade det som folk slog upp logaritmen för tal som började med siffrorna 1 och 2 mycket oftare än de slog upp logaritmen för tal som började med 8 och 9. Det verkade faktiskt som siffror som började med 1 och 2 faktiskt förekom oftare i beräkningar än siffror som började med 8 och 9.

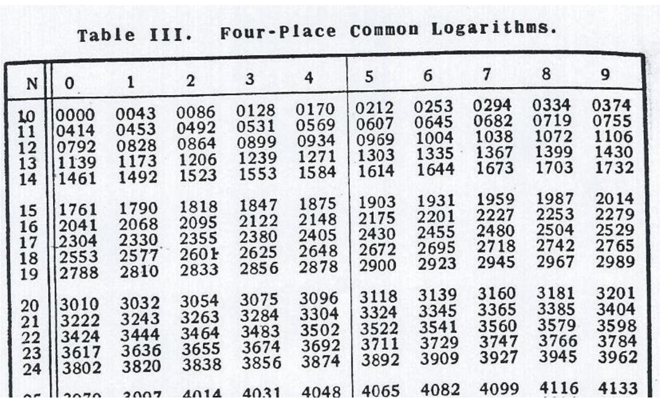


Table III. Four-Place Common Logarithms.

N	0	1	2	3	4	5	6	7	8	9
10	0000	0043	0086	0128	0170	0212	0253	0294	0334	0374
11	0414	0453	0492	0531	0569	0607	0645	0682	0719	0755
12	0792	0828	0864	0899	0934	0969	1004	1038	1072	1106
13	1139	1173	1206	1239	1271	1303	1335	1367	1399	1430
14	1461	1492	1523	1553	1584	1614	1644	1673	1703	1732
15	1761	1790	1818	1847	1875	1903	1931	1959	1987	2014
16	2041	2068	2095	2122	2148	2175	2201	2227	2253	2279
17	2304	2330	2355	2380	2405	2430	2455	2480	2504	2529
18	2553	2577	2601	2625	2648	2672	2695	2718	2742	2765
19	2788	2810	2833	2856	2878	2900	2923	2945	2967	2989
20	3010	3032	3054	3075	3096	3118	3139	3160	3181	3201
21	3222	3243	3263	3284	3304	3324	3345	3365	3385	3404
22	3424	3444	3464	3483	3502	3522	3541	3560	3579	3598
23	3617	3636	3655	3674	3692	3711	3729	3747	3766	3784
24	3802	3820	3838	3856	3874	3892	3909	3927	3945	3962
[illegible]	[illegible]	[illegible]	4014	4031	4048	4065	4082	4099	4116	4133

Sid 2

Logaritmlagen $\log(a)-\log(b) = \log\left(\frac{a}{b}\right)$ ger då att

$$\log(n+1)-\log(n) = \log\left(\frac{n+1}{n}\right) = \log\left(1+\frac{1}{n}\right)$$

Vi kan beräkna totala andelen genom summering:

$$\sum_{n=1}^{9}\left(\log_{10}\left(1+\frac{1}{n}\right)\right) \blacktriangleright 1$$

Det stämmer! Vi summerar utan att använda TI-Nspire´s analysverktyg och skriver termerna på formen $\log\left(\frac{n+1}{n}\right)$

$\log(2)+\log(3/2)+\log(4/3)+\log(5/4)+\log(6/5)+\log(7/6)+\log(8/7)$
$+\log(9/8)+\log(10/9)=\log(2\cdot 3/2\cdot 4/3\cdot 5/4\cdot 6/5\cdot 7/6\cdot 8/7\cdot 9/8\cdot 10/9)$
$=\log(10)=1$

Här använder vi andra logaritmlagen och visar analysverktyget summa kan användas som ett förkortat skrivsätt när man ska summera ett antal termer.

Sid 3-4

Benfords lag kan exempelvis användas för att avslöja fiffel med skatteavdrag. Fejkade siffror följer sällan Benfords lag. Lagen fungerar inte inom snäva områden som exemelvis lägenhetshyror för trerumslägenheter eller kroppslängder. Data ska också gå över stora intervall, dvs. det ska finnas värden t.ex. mellan 0 till 10, 10 till 100, 100 till 1000. Vi ska försöka ge en enkel förklaring:

Tänk dig att en vara kostar 1000 kr. Så höjs priset med 20 %. Den kostar då 1200 kr. Så höjs priset igen med 20 %. Då kostar den 1440 kr. Om det håller på så *fyra gånger* med prishöjningar så passerar priset 2000 kr och vi får en ny förstasiffra. Det tar alltså 4 år att gå från en etta till en tvåa som första siffra.

Nu tänker vi oss att vi börjar med 2000 kr istället. Efter *tre höjningar* med 20 % har vi passerat 3000 kr. Vi gör samma sak med 3000 kr som starvärde. Efter *två höjningar* så har vi passerat 4000 kr.

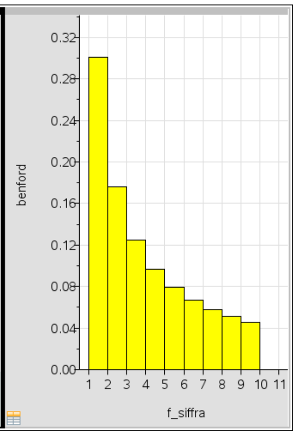
På sid 3 och 4 ger vi en enkel förklaring till varför ”låga” siffror är vanligare i vissa typer av data. På nätet finns också många sidor som kan ge förklaringar till varför det är så. Sök på Benford.

Sid 5

Observera att **log (4/3)** t.ex. skrivs som **-log (3/4)**. Se till att eleverna förstår den omskrivningen.

	A f_siffra	B benford	C
=		=log(1+1/'f_siffra)	=log(1.+1/
1	1	log(2)	0.30102…
2	2	log(3/2)	0.17609…
3	3	-log(3/4)	0.12493…
4	4	log(5/4)	0.09691…
5	5	-log(5/6)	0.07918…
6	6	log(7/6)	0.06694…
7	7	-log(7/8)	0.05799…
8	8	log(9/8)	0.05115…
9	9	log(10/9)	0.04575…
10			

B3 $=-\log_{10}\left(\frac{3}{4}\right)$



Sid 7

	A	B	C	D	E fsiffra...	F
=						
1	China	1330044544	1330044...	1	1	
2	India	1147995904	1147995...	1	1	
3	United States of America	303824640	3038246...	3	3	
4	Indonesia	237512352	2375123...	2	2	
5	Brazil	196342592	1963425...	1	1	
6	Pakistan	172800048	1728000...	1	1	
7	Bangladesh	153546896	1535468...	1	1	
8	Nigeria	146255312	1462553...	1	1	
9	Russia	140702096	1407020...	1	1	
10	Japan	127288416	1272884...	1	1	
11	Mexico	109955400	1099554...	1	1	

Hör börjar nu olika exempel på datamaterial som mer eller mindre följer Benfords lag. Exemplet ovan är folkmängden i världens länder (239 st). För att komma fram till förstasiffran så gör vi en del manipulationer med våra data, som från början är i numeriskt format.

Vi visar här hur det går till:

string(234) ▸ 234 string(349.6) ▸ 349.6

Förenklar det numeriska uttrycket inom parentesen och ger resultatet som en *teckensträng*.

left("234",1) ▸ 2 left("349.6",1) ▸ 3

Tar fram det numeriska tecknet längst till vänster (position 1) i teckensträngen

expr("2") ▸ 2 expr("3") ▸ 3

Gör om teckensträngen till numeriskt format

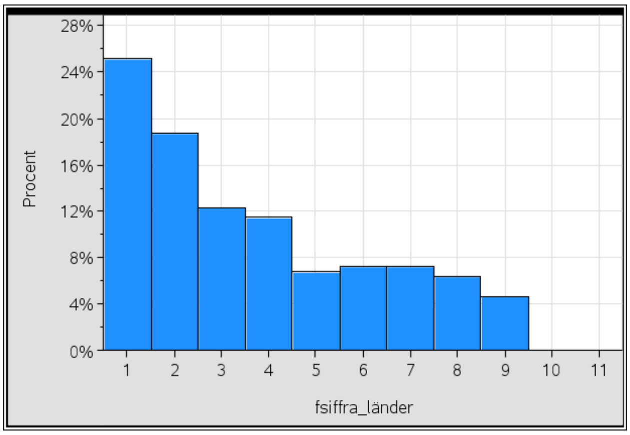
Dessa manipulationer görs alltså i kolumnerna C, D och E. Observera alltså att numeriska data visas högerställt i kalkylbladets celler och textsträngar positioneras till vänster

Vi får alltså resultatet i kolumn E som vi här markerat med gul färg.

	B	C	D	E fsiffra...	F
=					
1	1330044544	1330044...	1	1	
2	1147995904	1147995...	1	1	
3	303824640	3038246...	3	3	
4	237512352	2375123...	2	2	
5	196342592	1963425...	1	1	
6	172800048	1728000...	1	1	

E1 =expr(d1)

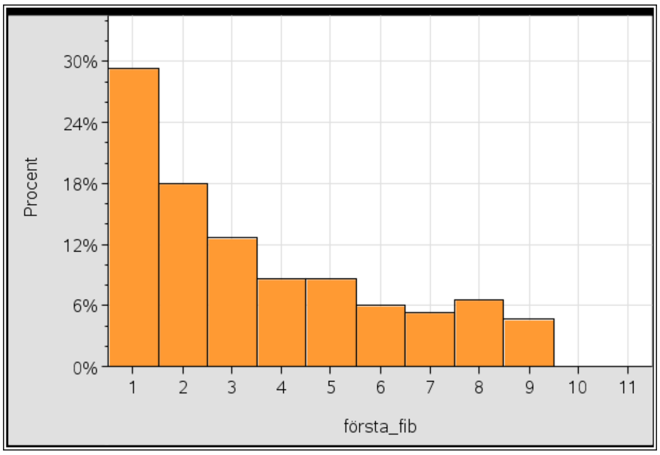
Sid 8



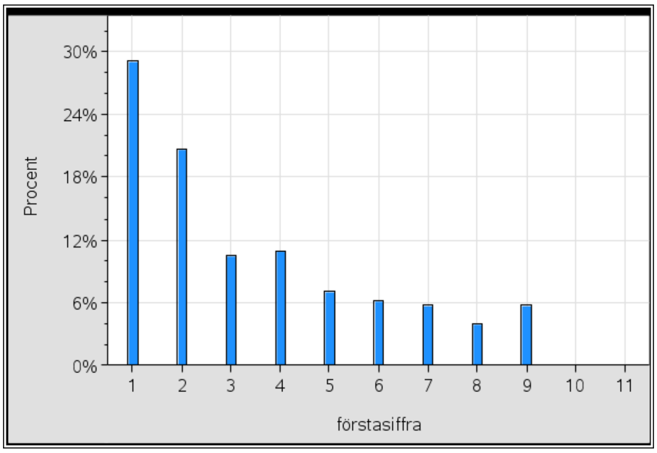
Sär ser fördelningen av förstasiffra ut. Skalan på den lodräta axeln är *relativ* frekvens som markeras som ”procent”. Man kan också ha absoluta värden.

På de följande sidorna finns några fler exempel. Låt eleverna själva ta fram data som de kan testa. Man kan oftast kopiera data direkt och klistra in direkt i kalkylarket. Man kan få pröva sig fram.

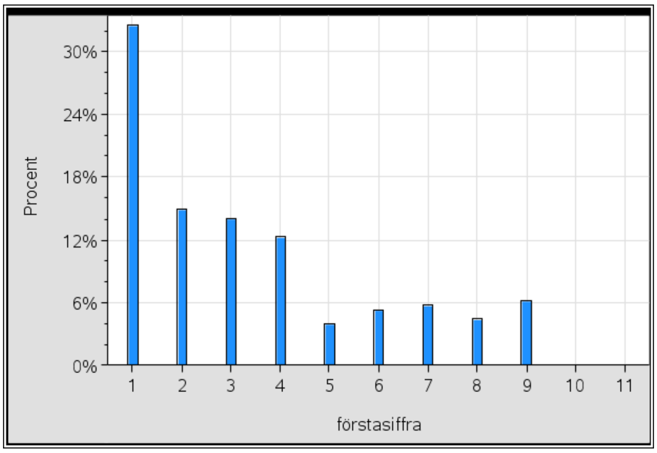
Det visar sig att även de så kallade *Fibonaccitalen* följer Benfords lag ganska bra. Förklara för eleverna hur denna talföljd är uppbyggd.



En intressant egenskap hos data som följer Benfords lag är att data ska vara *skalinvarianta*. Nedan har vi fördelningen av förstasiffra för arealen för jordens länder. Data finns med i ett kalkylblad i TI-Nspire-filen. Arealen mäts här i kvadratkilometer (km^2). Om vi om-vandlar till *square miles* så får vi en fördelning som också liknar den teoretiska Benfordfördelningen. Se diagrammen på nästa sida.


Procent
30%
24%
18%
12%
6%
0%
1
2
3
4
5
6
7
8
9
10
11
förstasiffra

km^2


Procent
30%
24%
18%
12%
6%
0%
1
2
3
4
5
6
7
8
9
10
11
förstasiffra

sq mi