Pinocchio testing in the forensic analysis of waiting lists: using public waiting-list data from Finland and Spain for testing Newcomb-Benford's law.

Jaime Pinilla, Beatriz G López-Valcárcel, Christian González-Martel, Salvador Peiró.

ANNEX 1

Are waiting lists candidates for Benford's test?

Newcomb-Benford's Law (NBL) does not apply to datasets of truly random numbers (e.g. lottery), sequential numbers (calendar years), assigned numbers (e.g. zip codes) and numerical series with some restrictions (e.g., with data between 800 and 999).

Some authors have proposed some preliminary tests that reveal whether or not Benford's law applies to a particular data set.

Wallace (2002) suggests that if the mean of a set of numbers is larger than the median and the skewness value is positive, the data set likely follows a Benford distribution. The rational is that the larger the ratio of the mean divided by the median, the more closely the set will follow Benford's law. The table 1 of this Annex show the summary statistics for the datasets used in our study. Since both of our datasets has mean greater than median and is positively skewed, we can consider both datasets as appropriate candidates for NBL analysis.

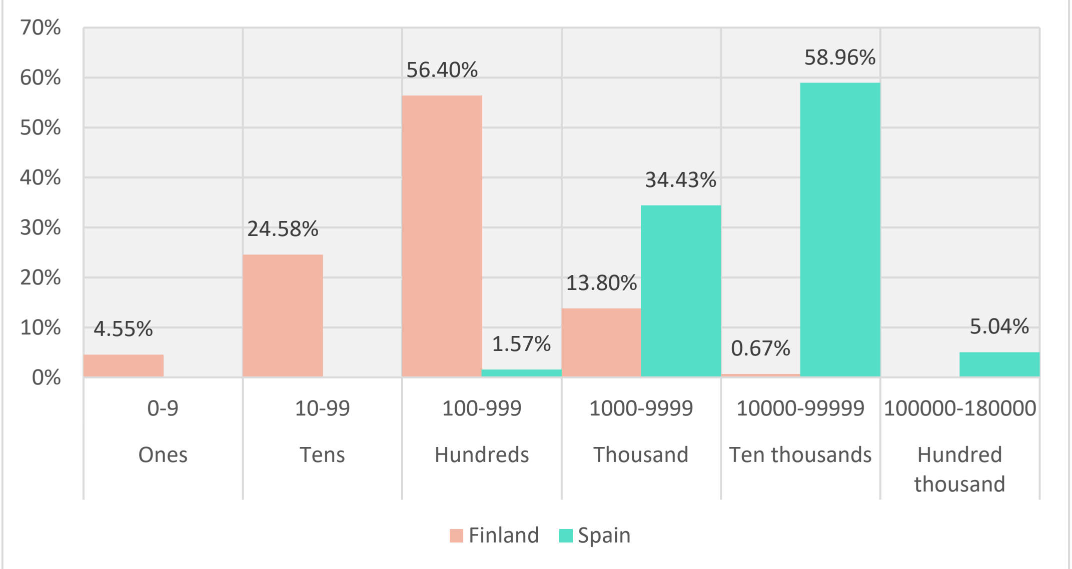
Annex 1. Table 1. Summary statistics for Finish and Spanish data set

	Finish data set	Spanish data set	First digit Benford distribution
Mean	720.82	19117.42	3.44
Median	266	9678	3
Skewness	4.67	2.44	0.8
Mean/Skewness	2.71	1.98	1.15

Benford's law tends to apply most accurately to data that are distributed across several orders of magnitude. As a rule of thumb, the more orders of magnitude that the data covers, the more accurately Benford's law applies. The so-called spread hypothesis (Miller SJ, 2015) states that if a data set is distributed over several orders of magnitude (as our datasets), then the leading digits will approximately follow Benford's Law.

Although we must bear in mind that Finland's population is about 5.5 million while Spain's population is 46.5 million, the spread out of the two data sets used in our study (see figure 1 in this Annex) are between 1 and 13,533 for Finnish dataset (excluding 14 registers with 0 persons in the WL) and between 871 and 173,006 for Spanish data set, both long enough to approximate the NBL distribution.

Annex 1. Figure 1. Value chart distribution for Finnish and Spanish respectively waiting list data set



References

Miller SJ, ed. Benford's Law: Theory and Applications. Princeton, NJ: Princeton University Press; 2015.

Wallace WA. Assessing the quality of data used for benchmarking and decision making. J Gov Financ Manag. 2002;51(3):16–22.

Does seasonal variability influence the NBL test?

Waiting lists can have important seasonal variations, with certain accumulations in summer or during holiday periods due to the reduction in activity associated with vacations. Since waiting lists are reported differently in Spain and Finland (two or three times a year depending on the country). There is a possibility that this variability could influence the results of the NBL analysis. Therefore, we repeated our analyses but using only the data reported in the month of December (common for both countries).

The results (see table 2 and 3, Annex 1) are completely superimposable to the overall results presented in the main manuscript, with data from Finland adjusting perfectly to the NBL, while those from Spain being significantly different from this distribution. This suggests that the different behaviour regarding the adjustment to the NBL distribution is not related to the time of WL reporting.

Annex 1. Table 2. Test statistics for the first digits of Finnish data (December)

Value	Count	Frequency observed	Frequency expected (Benford's Law)	Diff. (MAD)	p-value
1	67	0.30317	0.30103	0.00214	0.9416
2	46	0.20814	0.17609	0.03205	0.2161
3	23	0.10407	0.12494	-0.02087	0.4152
4	18	0.08145	0.09691	-0.01546	0.4960
5	21	0.09602	0.07918	0.01584	0.3818
6	12	0.05430	0.06695	-0.01265	0.5888
7	16	0.07240	0.05799	0.01441	0.3851
8	11	0.04977	0.05115	-0.00138	1.0000
9	7	0.03167	0.04576	-0.01408	0.4188
Total	221				
Pearson's $\chi^{2\text{-test}}$		5.5937 (p-value 0.6926)			
Mean test (absolute value)		1.4319			
Kuiper test		0.5463			

*Significant test value on the 5% level; ** significant test value on the 1% level. The p-value of the Z-test statistics for each digit are presented in the last column. The critical test values for the respective significance levels are: Pearson's χ^2 test (8df): 15.51 and 20.09; Mean test: 1.96 and 2.58; Kuiper test: 1.75 and 2.00.

Annex 1. Table 3. Test statistics for the first digits of Spanish data (December)

Value	Count	Frequency observed	Frequency expected (Benford's Law)	Diff. (MAD)	p-value
1	167	0.39397	0.30103	0.09284	0.0000**
2	65	0.15330	0.17609	-0.02279	0.2508
3	27	0.06368	0.12494	-0.06126	0.0000**
4	23	0.05425	0.09691	-0.04266	0.0018**
5	28	0.06604	0.07918	-0.01314	0.3679
6	17	0.04009	0.06695	-0.02685	0.0250**
7	31	0.07311	0.05799	0.01512	0.1768
8	25	0.05896	0.05115	0.00781	0.4400
9	41	0.09670	0.04576	0.05094	0.0000**
Total	424				
Pearson's $\chi^{2\text{-test}}$		65.8043 (p-value 0.000**)			
Mean test (absolute value)		3.704**			
Kuiper test		3.4606**			

*Significant test value on the 5% level; ** significant test value on the 1% level. The p-value of the Z-test statistics for each digit are presented in the last column. The critical test values for the respective significance levels are: Pearson's χ^2 test (8df): 15.51 and 20.09; Mean test: 1.96 and 2.58; Kuiper test: 1.75 and 2.00.