Is the 'n = 30 rule of thumb' of ecological field studies reliable? A call for greater attention to the variability in our data

A. Martínez–Abraín

Martínez–Abraín, A., 2014. Is the 'n = 30 rule of thumb' of ecological field studies reliable? A call for greater attention to the variability in our data. *Animal Biodiversity and Conservation*, 37.1: 95–100, Doi: https://doi.org/10.32800/abc.2014.37.0095

Abstract

Is the 'n = 30 rule of thumb' of ecological field studies reliable? A call for greater attention to the variability in our data.— A common practice of experimental design in field ecology, which relies on the Central Limit Theorem, is the use of the 'n = 30 rule of thumb'. I show here that papers published in *Animal Biodiversity and Conservation* during the period 2010–2013 adjust to this rule. Samples collected around this relatively small size have the advantage of coupling statistically–significant results with large effect sizes, which is positive because field researchers are commonly interested in large ecological effects. However, the power to detect a large effect size depends not only on sample size but, importantly, also on between–population variability. By means of a hypothetical example, I show here that the statistical power is little affected by small–medium variance changes between populations. However, power decreases abruptly beyond a certain threshold, which I identify roughly around a five–fold difference in variance between populations. Hence, researchers should explore variance profiles of their study populations to make sure beforehand that their study populations lies within the safe zone to use the 'n = 30 rule of thumb'. Otherwise, sample size should be increased beyond 30, even to detect large effect sizes.

Key words: Sample size, Variance, Statistical power, Effect size, Field ecology, Reliability

Resumen

¿Es fiable la regla de oro de n = 30 de los estudios ecológicos de campo? Se debe prestar más atención a la variabilidad de nuestros datos.— La utilización de la regla de oro de n = 30 es una práctica común del diseño experimental en ecología de campo que se fundamenta en el teorema del límite central. A continuación se muestra que los artículos publicados en A*nimal Biodiversity and Conservation* durante el período comprendido entre los años 2010 y 2013 se ajustan a esta regla. Las muestras recogidas cuyo tamaño se aproxima a esta cifra relativamente pequeña tienen la ventaja de relacionar resultados estadísticamente significativos con efectos de gran magnitud, lo cual es positivo porque por lo general los investigadores de campo están interesados en los efectos ecológicos de gran magnitud. No obstante, la posibilidad de detectar un efecto de gran magnitud no solo depende del tamaño de la muestra, sino también en gran medida de la variabilidad existente entre las poblaciones. Mediante un ejemplo hipotético, a continuación se muestra que la potencia estadística se ve poco afectada por los cambios pequeños o medios de varianza que pueda haber entre las poblaciones. Sin embargo, la potencia se reduce bruscamente a partir de un determinado límite, que nosotros establecemos aproximadamente en una diferencia de cinco veces en la varianza entre poblaciones. Por consiguiente, los investigadores deberían analizar los perfiles de varianza de sus poblaciones de estudio con el fin de asegurarse de antemano de que sus poblaciones en estudio se encuentran en la zona de seguridad en que puede emplearse la regla de oro de n = 30. De lo contrario, será necesario aumentar el tamaño de la muestra a más de 30, incluso para detectar efectos de gran magnitud.

Palabras clave: Tamaño de la muestra, Varianza, Potencia estadística, Magnitud del efecto, Ecología de campo, Fiabilidad

Received: 1 VIII 13; Conditional acceptance: 11 XI 13; Final acceptance: 20 XII 13

ISSN: 1578–665 X
eISSN: 2014–928 X

Alejandro Martínez–Abraín, Depto. de Bioloxía Animal, Bioloxía Vexetal e Ecoloxía, Univ. da Coruña, Fac. de Ciencias, Campus da Zapateira s/n., 15071 A Coruña, España (Spain); Population Ecology Group, IMEDEA (CSIC–UIB), Miquel Marquès 21, 07190 Esporles, Mallorca, España (Spain).

E–mail: amartinez@imedea.uib–csic.es

It is common to read in introductory books on biostatistics that working with a sample size of at least 30 is safe for the design of field studies (*e.g.* pg. 43 in Cohen & Cohen, 1995). This recommendation relies on the Central Limit Theorem, according to which if random samples of size n are drawn from a normal population the means of these samples will conform to a normal distribution (Zar, 1999). Supposedly random samples of a minimum size = 30 allow recovery of a normal distribution of the mean even if samples are non–normal. A common consequence is that field researchers tend to use samples adjusted to this minimum. To verify this tendency, I analyzed the sample size used in the papers published in Animal Biodiversity and Conservation from 2010 to 2013 (n = 4 years), and the results suggest that this rule holds for ecological and conservation field studies, since the arithmetic mean of both the medians and averages of sample sizes used in each paper was very close to 30 (see table 1).

This is in contrast with experimental design recommendations, where sample size is known to be directly dependent on the variance in the population (as estimated by the sample standard deviation), and indirectly dependent on the maximum allowable absolute difference between the estimated population parameter and the true population parameter (d) from equation 1,

$$n \geqslant \frac{z^2\,\sigma^2}{d^2}$$

where z is 1.96, the value for a 95% confidence interval from a normal distribution (Quinn & Keough, 2002). That is, the minimum required sample size to accurately estimate a parameter will be higher if a) population variance is high, in order to minimize the risk of our sample not being representative of the statistical population, and b) if the desired accuracy is high, for a given confidence level, since increasing the confidence level also increases the required sample size. Of course, deciding the value of d means that we have previous knowledge about the true magnitude of the study parameter, which is not usually the case in field ecological studies (*e.g.* Martínez–Abraín, 2008, 2013).

Sample size and null hypothesis testing

This is not only the theoretical basis of experimental design for parameter estimation, but also for a priori or prospective power tests within the framework of null hypothesis statistical testing (Zar, 1999; Schneir & Gurevitch, 2001). The required sample size to couple biological and statistical significance (and hence to make sense of statistically non–significant results) is determined after providing alpha (*i.e.* the type–I error rate, typically fixed at 5%), power (*i.e.* 1–the type II error rate or probability of correctly failing to reject a null hypothesis, typically fixed at 80%), and effect size (the minimum magnitude of the difference between two populations that is considered to be biologically relevant if dealing with mean comparison, or the amount of variance of each variable that is explained by other variables considered, if dealing with regression problems). Again, this implies that we have some a priori knowledge of what represents a biologically–relevant effect in our study system, which unfortunately is seldom the case in field ecological studies.

Empirically, sample size could be obtained by plotting the standard deviation against sample size until a plateau is reached, provided that our sample correctly represents the variability in the population. However, this is mostly viable for experimental studies (preferentially lab studies, although also some field studies), where sample size can be modified along a large range of possible values. A better–suited option for field studies could be to perform this plotting by applying resampling with repetition (bootstrap) to our data, if our sampling protocol includes samples of different size. Moreover, this exercise would be necessary for each variable under study, because each variable has its own profile. This means that the usual procedure of measuring many variables from the same sample of individuals —taking advantage of having captured them, for example— does not respect the prerequisite of accounting for variance to determine the right sample size, because variances for each trait of an individual do not necessarily co–vary in a strong way. Different sample sizes would thus be necessary to study different traits, something that seems impracticable in field studies where information

Table 1. Sample size extracted from n = 76 papers published in *Animal Biodiversity and Conservation* from 2010 to 2013: V. Journal volume; I. Journal isue; i. Initial page; f. Final page; n. Sample size; M1. Median sample size; M2. Average sample size; E. Cause of exclusion (1. Species description; 2. Ring recoveries; 3. Essay; 4. Survey, hunting–bag data, bioacoustics/RADAR data; 5. Species atlas or similar; 6. Review papers; 7. Genetics; 8. Museum collections; 9. Simulated data).

Tabla 1. Tamaño de la muestra extraída de n = 76 artículos publicados en Animal Biodiversity and Conservation *entre 2010 y 2013: V. Volumen de la revista; I. Número de la revista; i. Página inicial; f. Página final; n. Tamaño de la muestra; M1. Mediana del tamaño de la muestra; M2. Media del tamaño de la muestra; E. Motivo de exclusión (1. Descripción de especies; 2. Recuperación de anillas; 3. Ensayo; 4. Encuesta, datos de caza, datos de bioacústica y radar; 5. Atlas de especies o similar; 6. Artículos de revisión; 7. Genética; 8. Colecciones museísticas; 9. Datos simulados).*

Year	V	I	i	f	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	M1	M2	E
2010	33	1	1	13	22	19	11														13	13.2	
2010	33	1	15	18																			3
2010	33	1	19	29	18	4	17	39	23	26	18	27	24	21	11	27					22	21.6	
2010	33	1	31	45	5	10	20	30	53												30	27.7	
2010	33	1	47	51																			1
2010	33	1	53	61	7																53	40.3	
2010	33	1	63	87	13	16	11	20	11	9	10	10	12	14	10	19					12.5	21.8	
2010	33	1	89	115																			2
2010	33	1	117	117																			3
2010	33	2	119	129	50	50	50	50	50												50	50.0	
2010	33	2	131	142																			1
2010	33	2	143	150	16	16	21	25	23												21	20.2	
2010	33	2	151	185																			6
2010	33	2	187	194	30	30															30	30.0	
2010	33	2	195	203																			1
2010	33	2	205	208																			3
2011	34	1	1	10																			1
2011	34	1	11	21	39	41	19	31	20	10											25.5	26.7	
2011	34	1	23	29	60	60	60	60													60	60.0	
2011	34	1	31	34	8																8	8.0	4
2011	34	1	35	45	20	14	36	24	16	11	16	36									18	21.6	
2011	34	1	47	66																			1
2011	34	2	229	247	17	3															10	10.0	5
2011	34	2	249	256	13	132															72.5	72.5	
2011	34	2	257	264																			1
2011	34	2	265	272																			1
2011	34	2	273	285																			8
2011	34	2	287	294	4	17															10.5	10.5	
2011	34	2	295	308																			1
2011	34	2	309	317																			6
2011	34	2	319	330	10																10	10.0	
2011	34	2	331	340	8																8	8.0	
2011	34	2	341	353																			3
2011	34	2	355	361	37	25	22														25	28.0	
2012	35	1	1	11																			5

Table 1. (Cont.)

Year	V	I	i	f	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	M1	M2	E
2012	35	1	13	21																			1
2012	35	1	23	26																			7
2012	35	1	27	50																			1
2012	35	1	51	58																			8
2012	35	1	59	69																			7
2012	35	1	71	94																			1
2012	35	1	107	117	73	51															62	62.0	
2012	35	1	119	124	29	26	39	35	36	33											34	33.0	
2012	35	1	125	139	30																30	30.0	
2012	35	1	141	150	10	4	2	14													7	7.5	
2012	35	2	151																				3
2012	35	2	153	154																			3
2012	35	2	155																				3
2012	35	2	159	161																			3
2012	35	2	163	170																			4
2012	35	2	171	174																			3
2012	35	2	175	188																			4
2012	35	2	189	196	18	10	13	15	54	44	33	28									23	26.9	
2012	35	2	197	207	27	64	60	6	48	12	5										27	31.7	
2012	35	2	209	217																			
2012	35	2	219	220																			3
2012	35	2	221	233	27																27	27.0	
2012	35	2	235	246	5	29	42	9	21	15											18	20.2	
2012	35	2	247	252																			5
2012	35	2	253	265	5	20	49	158	42												42	54.8	
2012	35	2	267	275																			7
2012	35	2	277	283																			5
2012	35	2	285	293	48	28	16	16	16	16	16	16	16	4							16	19.2	
2012	35	2	295	306																			3
2013	36	1	1	11	12	12	29	31	32	3											20.5	19.8	
2013	36	1	13	31	25	9															17	17.0	
2013	36	1	33	36																			3
2013	36	1	37	46																			9
2013	36	1	47	57																			4
2013	36	1	59	67																			1
2013	36	1	69	78																			2
2013	36	1	79	88																			4
2013	36	1	89	99	50	51	41	50	50	50	45	49	59	22	51	30	30	21	30	6	47	39.7	
2013	36	1	101	111																			1
2013	36	1	113	121	3	30	14	91	30	21	45	11	13	20	30	28					24.5	28.0	
2013	36	1	123	139																			
Mean																					27.3	28.0	
SD																					17.2	16.7	

Table 2. Change in statistical power as between–population variance increases, using the n = 30 rule of thumb of field ecological statistics. Exercise using fictitious data in open software G*Power 3.1.3: n. Sample size; M1, M2. Means of both populations; Sd1, Sd2. Standard deviations of both populations; Ratio. Ratio Sd1/Sd2; Es. Effect size (Cohen's d); Po. Statistical power.

*Tabla 2. Cambio en la potencia estadística a medida que aumenta la varianza entre poblaciones, utilizando la regla de oro de n = 30 de los estadísticos de la ecología de campo. Ejercicio que utiliza datos ficticios en el programa informático abierto G*Power 3.1.3: n. Tamaño de la muestra; M1, M2. Medias de ambas poblaciones; Sd1, Sd2. Desviaciones estándar de ambas poblaciones; Ratio. Coeficiente Sd1/Sd2; Es. Magnitud del efecto (d de Cohen); Po. Potencia estadística.*

ID	n	M1	M2	Sd1	Sd2	Ratio	Es	Po
1	30	1.5	1.0	0.2	0.2	1.0	2.50	1.00
2	30	1.5	1.0	0.3	0.2	1.5	1.96	1.00
3	30	1.5	1.0	0.4	0.2	2.0	1.58	1.00
4	30	1.5	1.0	0.5	0.2	2.5	1.31	1.00
5	30	1.5	1.0	0.6	0.2	3.0	1.12	0.99
6	30	1.5	1.0	0.7	0.2	3.5	0.97	0.96
7	30	1.5	1.0	0.8	0.2	4.0	0.86	0.90
8	30	1.5	1.0	0.9	0.2	4.5	0.77	0.83
9	30	1.5	1.0	1.0	0.2	5.0	0.69	0.75
10	30	1.5	1.0	1.1	0.2	5.5	0.63	0.67
11	30	1.5	1.0	1.2	0.2	6.0	0.58	0.60
12	30	1.5	1.0	1.3	0.2	6.5	0.54	0.53
13	30	1.5	1.0	1.4	0.2	7.0	0.50	0.48
14	30	1.5	1.0	1.5	0.2	7.5	0.47	0.43

for each individual is exploited as much as possible given the difficulty in obtaining it.

The 'rule of thumb of n = 30' in field ecological studies comes from the fact that we are usually interested in 'large' effect sizes (of unknown absolute magnitude). The reasoning follows that if we are able to obtain a statistically–significant result (this only meaning that the properties of our sample can be applied to the whole statistical population, and hence that the desired inference from particular to general can be done) with a small sample size such as 30, the effect we are dealing with is probably large, and hence, most likely a biologically–relevant effect (Martínez–Abraín, 2007).

However, this approach of reasoning around n = 30 in relation to the magnitude of the effects (in the denominator of equation 1) is influenced by variance (in the numerator of equation 1). Power decreases with increasing between–population variance, when sample size is kept constant at n = 30 (table 2, fig. 1). However, this decrease proceeds in a non–linear fashion, indicating a strong resilience of statistical power to small–to–medium changes in between–population variance. Only when the change in between–populations variance is large (around a five–fold difference in the variance between groups, which corresponds to our study case #9) does power drop abruptly below the desired minimum value of 0.8 (fig. 1). In this hypothetical example, it is necessary to increase our sample size to n = 34 in case study #9, and to n = 73, in case study #14, to allow the recovery of a 0.80 power.

The n = 30 rule of thumb also overlooks the possibility that small or medium effect sizes can be biologically relevant in some cases (Igual et al., 2005). Since we typically do not know when that is the case, we are forced or limited to work with large effect sizes. On the contrary, working with too large a sample size (as is commonplace among theoretical ecologists) could even be counter–productive at times because we could be focusing on small effects which could be biologically irrelevant.

Hence, it seems reasonable to use the n = 30 rule in ecological field studies to make inferences on parameter values or to test null hypotheses, owing to our usual lack of knowledge on *d* or effect size of interest, but we should make an effort to explore beforehand the variance profile of our study populations in order to be able to detect large effects with that sample size and with a high power. Populations with

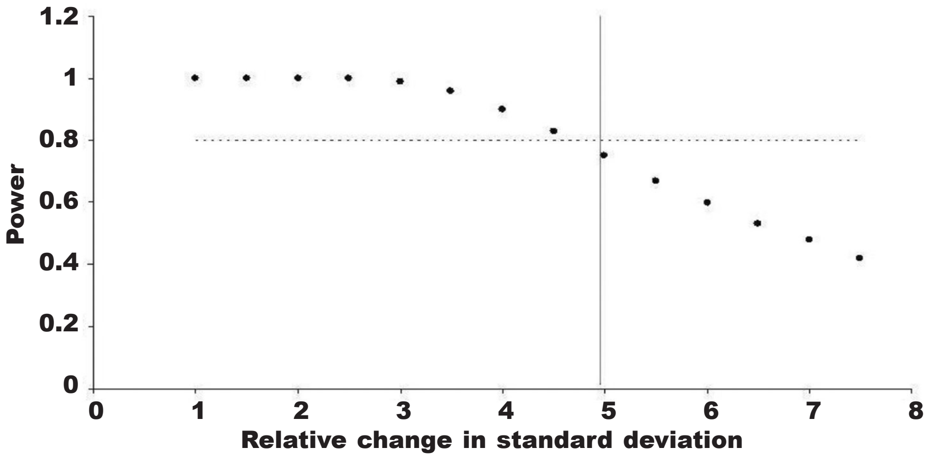

Fig. 1. Non–linear decrease of statistical power with increasing variance between two populations (ratio of standard deviations between two groups) while keeping means and sample sizes constant (n = 30), for data in table 2. The reference line for a desired power of 0.80 is shown as well as the line for a five–fold difference in between–populations variance.

Fig. 1. Para los datos que figuran en la tabla 2, la potencia estadística disminuye de forma no lineal a medida que aumenta la varianza entre dos poblaciones (coeficiente de las desviaciones estándar entre dos grupos) y con las medias y los tamaños de las muestras (n = 30) constantes. Se muestran también la línea de referencia de una potencia deseada de 0,80 y la línea que representa una diferencia de cinco veces en la varianza entre poblaciones.

high differences between them in variance will require larger sample sizes (compared to 30) to detect even large differences.

Rules of thumb exist for a reason, but they should be used with great caution and as an approximation. The exercise of thinking, typical of the scientific enterprise, cannot and should not be set aside during the process of experimental design, despite the added complexities of ecological field studies compared to lab studies. Reduced power, such as increases in between–populations variance, can lead to increased prevalences of Type II errors (*i.e.* incorrectly failing to reject a null hypothesis of equality to zero), resulting in serious problems regarding decision–making in conservation, when evaluating the effect of human activities. We may conclude that nothing happens when indeed it does. Let's hence give more attention to the variability of our field data for the benefit of proper knowledge acquisition and correct decision making.

Acknowledgements

I would like to thank the editor and two anonymous referees who helped improve the manuscript with their critical comments. Daniel Oro commented on an early version of the manuscript.

References

Cohen, J. & Cohen, L., 1995. *Statistics for ornithologists*. BTO Guide 22.

Igual, J. M., Forero, M. G., Tavecchia, G., González–Solís, J., Martínez–Abraín, A., Hobson, K. A., Ruiz, X., Oro, D., 2005. Short–term effects of data–loggers on Cory's shearwater (*Calonectris diomedea*). *Marine Biology*, 146: 619–624.

Martínez–Abraín, A., 2007. Are there any differences? A non–sensical question in ecology. *Acta Oecologica,* 32: 203–206.

– 2008. Statistical significance and biological relevance: A call for a more cautious interpretation of results in ecology. *Acta Oecologica*, 34: 9–11.

– 2013. Why do ecologists aim to get positive results? Once again, negative results are necessary for better knowledge accumulation. *Animal Biodiversity and Conservation,* 36.1: 33–36.

Quinn, G. P. & Keough, M. J., 2002. Experimental design and data analysis for biologists. Cambridge University Press, Cambridge.

Schneir, S. M. & Gurevitch, J. (Eds.), 2001. *Design and analysis of ecological experiments*. Oxford University Press, Oxford.

Zar, J. H., 1999. *Biostatistical analysis*. Prentice Hall, Upper Saddle River, New Jersey.