© 2021 The Authors. *The Bulletin* published by Wiley Periodicals LLC on behalf of Ecological Society of America.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



Contributions

Relative Inference for Paired Data: More than Meets the Eye

Kenneth Gerow[1], David R. Stewart[2], and Calvin Farris[3]
[1]Department of Mathematics and Statistics, University of Wyoming, Laramie, Wyoming 82071, USA
[2]Division of Biological Sciences, US Fish and Wildlife Service, Albuquerque, New Mexico 87103, USA
[3]Division of Fire and Aviation Management, National Park Service, Klamath Falls, Oregon 97603, USA

Introduction

A classical analysis for measurements from a paired design is the paired *t* tool, useful for both estimation via confidence intervals and testing. It is usual to assume the pairs are independent of one another. There is, however, correlation between data points *within* pairs; indeed, it is this correlation that gives the design its power as a statistical tool. This tool operates on *increments of change* (i.e., change measured by subtraction) within each pair by taking the differences, and analyzing the resulting single sample thereof, or by taking the difference in the two means.

On the other hand, inference on *relative change* is often of interest with biological data. For example, the National Park Service implements controlled burns with management goals that are often stated on a percentage basis (e.g., aiming to reduce fuel load by 50%, or to limit postfire tree mortality to <20%). Relative change objectives are useful when a trend over time is more important than the specific current or future state (USDI National Park Service 2003). For example, a reduction in tree density of 40% may be more useful to resource managers than a decline to 500 individuals per hectare across multiple burn units (USDI National Park Service 2003).

Gerow, K., D. R. Stewart, and C. Farris. 2022. Relative Inference for Paired Data: More than Meets the Eye. Bull Ecol Soc Am 103(2):e01960. https://doi.org/10.1002/bes2.1960

In monitoring a fish population from one time point to another, the relative increase (or decrease) in the population size may be of more interest than the increment of change. For example, many small-bodied fish may produce multiple cohorts of larvae in a summer, with some of the progeny reaching maturity by the end of their first summer. Thus, and depending on environmental conditions and prey availability both owing to variable levels of recruitment and natural mortality, population sizes from one year to the next can be highly variable (Herwig and Zimmer 2007). Consequently, relative increases across a longer time period are useful to draw inference to size and status of populations.

In what follows, we will show that the ratio of means is often superior to a more traditional approach (mean of ratios), examine some theory behind the approach, and exemplify the method with several real examples.

Background

We note that when using a classical paired *t*, the order of subtraction is a choice made for ease in telling the story; for simplicity, we will simply denote the two settings as condition "a" and "b" and use an arbitrary order. Done properly (i.e., taking care with the standard error calculation in the second case), the methods are the same. This is because, given differences $d_i = Y_{ai} - Y_{bi}$, the standard error is

$$\widehat{\mathrm{SE}}\left(\overline{d}\right) = s_d/\sqrt{n} = \widehat{\mathrm{SE}}\left(\overline{Y}_a - \overline{Y}_b\right) = \sqrt{\left(\widehat{\mathrm{SE}}\left(\overline{Y}_a\right)\right)^2 - 2\widehat{\mathrm{Cov}}\left(\overline{Y}_a, \overline{Y}_b\right) + \left(\widehat{\mathrm{SE}}\left(\overline{Y}_b\right)\right)^2},$$

where s_d is the standard deviation among the sample of differences, and $\widehat{\mathrm{Cov}}\left(\overline{Y}_a, \overline{Y}_b\right)$ is the estimated covariance between the two means.

For purpose of studying relative change, we will here consider the ratio of condition *b* to condition *a*, without prescribing which order is best for a given real situation; that is, in fact, a story-telling choice on the part of the researcher. The two possible analogs to the aforementioned approach to incremental change are the mean of the individual ratios (MoR) and the ratio of the two means (RoM).

For relative change, it *does* matter how one approaches the estimation. Unlike incremental change where the mean of differences is equal to the difference of means, the ratio of the means and the mean of the ratios will usually not yield the same answers, and these differences are in fact meaningful. Table 1 illustrates these points using a simple example between two hypothetical conditions *a* and *b* ($n = 2$). If asked, "On average, by how much did these individuals change?" the answer would be fourfold. On the other hand, the relative change over both of the sampling units is only a doubling (twofold).

Manley et al. (2002) discuss choosing between RoM and MoR in their resource selection book. They note that the two statistics estimate different population parameters; RoM estimates $R = \mu_b/\mu_a$, whereas the average of the ratios $\overline{r} = (\sum_{i=1}^{n} Y_{b,i}/Y_{a,i})/n$ estimates μ_r, the population mean of all such ratios. However, and as illustrated in Table 1, and explained more fully below, the two in fact are not identical.

The ratio of means appears in the sampling literature, albeit not by that name. Ratio estimation is one of several regression-based model-assisted methods for estimating the mean (or, if rescaled) total of

some variable in a population. These methods require information, ideally for the whole population, of an auxiliary variable that is correlated with the variable of primary interest. See, for example, Section 6.2 of Schaffer et al. 2012.

Lohr (1999) develops the RoM more fully, showing its use as a regression tool in its own right. Suppose the response variable (Y) is proportional to a predictor (X). then it is reasonable that (1) Y and X can only take non-negative values; (2) that a straight line relationship through the origin would be a good fit, and (3) that variation in Y also grows with increasing X values. Under those conditions, the RoM $\left(\overline{Y}/\overline{X}\right)$ is a natural estimator for the slope. The basic ideas are presented in Section 3.1 (pages 60–73) of Lohr (1999), with model formalities addressed in Section 3.4 (pages 81–88). The sample RoM is slightly (downward) biased, although the bias is usually small and shrinks with increasing sample size. An estimator of the bias is presented on page 67.

Investigation of the properties of the RoM of two normally distributed variables (with sufficient sample sizes, the two means will be so) is a classical problem in statistics, known as the Fieller-Creasy Problem; various modern approaches to that problem include Bayesian methods (e.g., Mendoza and Guitierrez-Pena 1999 and Kim et al. 2018) and constrained normal approximations (see Diaz-Frances and Sprott (2004) and Diaz-Frances and Rubio (2013). (The latter studies the ratio of two independent variables; in the current setting, they will be correlated.) The novelty in the current paper is our proposal that has a useful and largely unexploited role in estimation for paired data.

Conceptual Development

It is well-known in the statistics literature that the RoM is a weighted form of MoR, which we show here for convenience. First, $\text{MoR} = \sum_{i=1}^{n}\left(\frac{1}{n}\right)\frac{Y_{i,b}}{Y_{i,a}}$. MoR is an equally weighted sum of the ratios. Now,

$$\text{RoM} = \left(\sum Y_{i,b}/n\right)/\left(\sum Y_{i,a}/n\right) = \left(\sum Y_{i,b}\right)/\left(\sum Y_{i,a}\right) = \sum_{i=1}^{n}\left(Y_{i,b}/\sum Y_{i,a}\right).$$

Next, we deploy a standard mathematical trick of multiplying by 1, which changes nothing numerically, but with "1" presented in an algebraic form $\left(Y_{i,a}/Y_{i,a}\right)$ we can exploit.

$$\text{RoM} = \sum_{i=1}^{n}\left(Y_{i,a}/Y_{i,a}\right)\left(Y_{i,b}/\sum Y_{i,a}\right) = \sum_{i=1}^{n}\left(Y_{i,a}/\sum Y_{i,a}\right)\left(Y_{i,b}/Y_{i,a}\right) = \sum_{i=1}^{n}\left(Y_{i,a}/\sum Y_{i,a}\right)\left(r_i\right).$$

Table 1. Illustration of RoM versus MoR.

Pair	Condition *a*	Condition *b*	Differences	Ratios
1	2	14	12	7
2	10	10	0	1
Means	6	12	6	4.0

The leading terms $\left(Y_{i,a} / \sum Y_{i,a}\right)$ sum to unity, and so they qualify as weights in a weighted average. Thus, RoM can be seen as a weighted form of MoR.

MoR will always be greater than or equal to RoM. They will be equal when the denominator terms are the same. It sometimes happens that the values in the denominator have small variation; in that case, MoR and RoM will yield similar answers numerically. By way of analogy, when distributions are approximately symmetric, the mean and the median may be similar numerically, but they are still conceptually meaningfully different. Here, we argue that the choice between MoR and RoM is similarly germane. Using the individual ratios is appropriate when change (on average) at the level of individuals is of interest; a ratio of the means wins when change at the population level is of interest. When the question of interest in many studies is with regard to change over the population, we argue that RoM will be the better choice. We illustrate the process of choosing among the classical paired *t* tool, MoR and RoM in Figure 1.

When using individual ratios (MoR), a value of 0 in the denominator will yield an undefined ratio (or infinite). One can choose to replace the 0 with a judiciously chosen (the choice will affect the resulting estimate) small value (akin to what is sometimes done to accommodate a 0 when using logarithmic transformations) or drop that data point.

Some studies do indeed focus on individual responses. For instance, in studies of changes in oxygen consumption rates in seals while diving as compared to resting (e.g., Webb et al. 1998, Green et al. 2007) measuring the change on a per-seal basis makes sense, and if relative change is of interest, then the mean of individual ratios is attractive. In studies of changes in percent body fat of hibernating animals (e.g., Buck and Barnes 1998, Kunz et al. 1998), answering the question of change on average of individuals also makes sense. When using the MoR, the central limit theorem applies, since the statistic being considered is the mean of individual values (ratios). For studies like these, the RoM does not answer a relevant question (e.g., the total sum of body fat across the entire bear population).

On the other hand, when interest is in change at the population (rather than individual), the RoM comes into its own. A simple and elegant statistic, it comes with a few twists, as we now investigate.

Bootstrapping for Confidence Intervals of the Ratio of Means

Analysis of the mean of individual ratios can proceed in the usual one-sample way; the *t*-distribution is viable given that the sample size is sufficiently large so that the distribution of the mean is approximately normal. However, those conditions do not hold for the RoM, for two reasons. First, the central limit theorem does not apply to the ratio of two means, except under certain constraints (e.g., Diaz-Frances and Rubio 2013). Second, the standard error (SE) formula does not always work well.

The SE of a ratio of two means appears in two equivalent general forms. The first is written as,

$$\widehat{\mathrm{SE}}\left(\overline{Y}_b/\overline{Y}_a\right) = \left(\overline{Y}_b/\overline{Y}_a\right)\sqrt{\left(\widehat{\mathrm{SE}}\left(\overline{Y}_b\right)/\overline{Y}_b\right)^2 - 2\left(\widehat{\mathrm{Cov}}\left(\overline{Y}_a,\overline{Y}_b\right)\right)/\left(\overline{Y}_a\overline{Y}_b\right) + \left(\widehat{\mathrm{SE}}\left(\overline{Y}_a\right)/\overline{Y}_a\right)^2}.$$

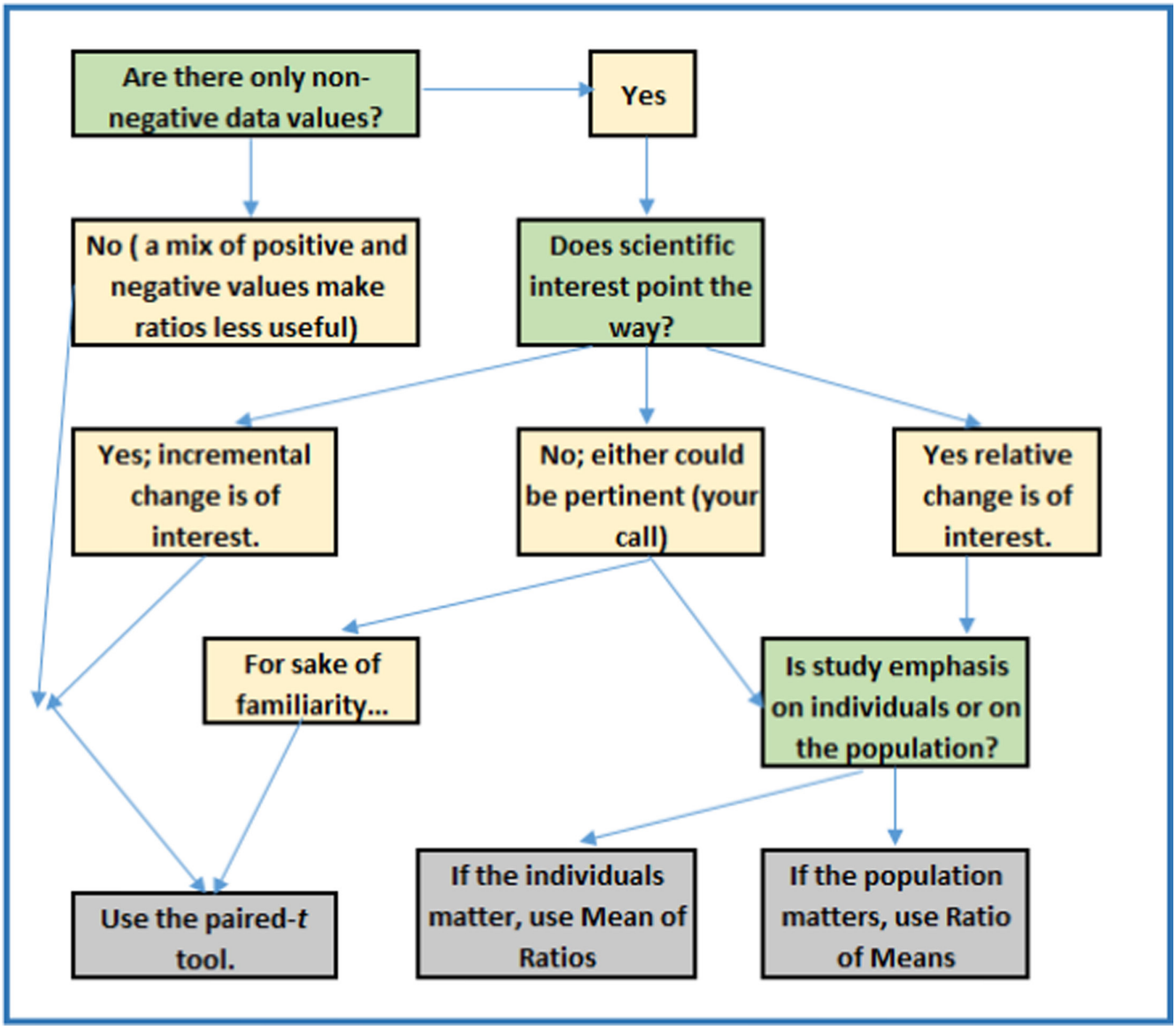


Fig. 1. Flow chart showing decision process for paired data.

The second form is adapted from the finite population sampling literature, where the ratio of two means appears in so-called ratio estimation. In that context, the ratio of two means is the estimator for the slope in a regression model that arises from assuming that the response Y_b is proportional to Y_a, and thus, our formula appears as.

$$\widehat{\mathrm{SE}}\left(\overline{Y}_b/\overline{Y}_a\right) = \left(1/\overline{Y}_a\right)\sqrt{\left(\sum_{i=1}^{n}\left(Y_{bi}-\left(\overline{Y}_b/\overline{Y}_a\right)Y_{ai}\right)^2/(n-1)\right)}.$$

The conventional advice is that the estimator works well when $\overline{Y}_a$ is estimated precisely (necessary since the derivation of the estimator treats it as essentially constant), typically described as having a coefficient of variation $\left(\widehat{\mathrm{SE}}\left(\overline{Y}_a\right)/\overline{Y}_a\right)$ less than 0.1. Ironically, it is precisely when this is *not* true that

choosing between RoM and MoR is critical numerically. Aside from the sometimes unstable behavior of the SE formula, the matter is further complicated by the fact that the central limit theorem does not routinely apply to the distribution of the ratio $\left(\overline{Y}_b/\overline{Y}_a\right)$; we cannot argue that it is approximately normal, hence the need for bootstrapping.

As a nonparametric method for creating confidence intervals, bootstrapping was the creation of Efron (1979), but see Efron and Tibshirani (1993) for a thorough introduction. The simplest way to form bootstrap confidence intervals is the empirical method. Suppose one uses $B = 1000$ bootstrap replicates with intent to make a 95% confidence interval. Sort the 1000 bootstrapped estimates from smallest to largest. Then, 2.5% of the simulated distribution will be less than the 26th smallest value and 2.5% will be larger than the 97.5th smallest. Those two values are the endpoints of the interval.

Currently the *bca*-bootstrap procedure is commonly used; *bca* stands for "bias-corrected and accelerated." Many statistics are biased estimators of their relevant parameter (the usual Pearson product-moment correlation coefficient and the standard deviation are two oft-used examples). The bootstrap estimate of bias in a statistic is measured by the proportion of simulated estimates that are less than that in the original sample. An acceleration parameter is calculated that is proportional to the skewness in the bootstrap distribution. These two values adjust the endpoints of the interval. For instance, if the simulated distribution is skewed right, the CI is adjusted to the right.

To illustrate the differences among the three statistical methodologies, we present three different examples below.

Three Case Studies

Example 1

In southern Arizona (USA), seven fish species are endemic to the Río Yaqui watershed and four remain at and around San Bernardino and Leslie Canyon National Wildlife Refuges (NWRs). The Yaqui catfish *Ictalurus pricei* is functionally extinct (Stewart et al. 2017a), the beautiful shiner *Cyprinella formosa* and Yaqui chub *Gila purpurae* are federally listed as threatened, and the Yaqui topminnow *Poeciliopsis occidentalis sonoriensis* is federally listed as endangered (USFWS 1994). Our focus is on the Yaqui chub. This species occupies more than 35 wetland ponds. From 2007 to 2012, Federal, State, and multiple conservation organizations annually monitored populations of these fish to insure their persistence and evaluate their recover (Stewart et al. 2017b, 2019). Of the 35 wetland ponds, 10 were selected and ranged in size from 28 to 8813 m^2, and their depths ranged from 0.4 to ~4 m. Most ponds are clear and dominated by submergent aquatic vegetation (e.g., spiny naiad *Najas minor*, sago pondweed *Potamogeton pectinatus*, coontail *Ceratopyllum demersum*, and horned pondweed *Zannichellia palustris*). Typically, the number of traps used to survey each wetland pond ranged from 8 to 30 and was set following a proportional stratified random sampling design based on vegetation and depth surveys conducted 1 wk prior to sampling (Stewart et al. 2017b). For example, if 50% of the pond was dominated by a certain submergent aquatic vegetation type, then 50% of the traps were randomly set in vegetation and 50% randomly outside the vegetation (Stewart et al. 2017*b*). We calculated the total number of fish trapped from the historical survey data set for Yaqui chub by pond and sample year, to determine population trajectories from 2007 to 2012, on a relative basis (by choice), using the MoR and RoM statistics (Table 2).

Based on the Yaqui chub data, if we used the mean of individual ratios, data from pond 3 must be excluded due to the observed 0 in 2007. After discarding pond 3, and based on $n = 9$, the MoR is 56.8, and a *t*-based 95% CI is (−40.8, 154.4). The Yaqui chub data are so heavily skewed that the assumption of approximate normality is likely invalid; that and the fact that the resulting interval has a negative value for the lower endpoint also makes it clear that the *t*-based inference here is not valid. A *bca*-bootstrapped interval for these data is 2.0, 125. If inference is to be made to the population of ponds rather than average behavior of individual ponds, then the ratio of the means is the more appropriate statistic. Here, we use data for all 10 ponds. The RoM is 2.5, and a 95% *bca*-bootstrapped CI is 1.2, 6.9. This example illustrates how different the two estimates can be. One suggests an almost 60-fold increase in fish abundance. The other more realistically suggests a 2.5-fold increase. The reason for the dramatic difference can be traced to the ponds with only a very small number of fish detected in 2007 (Table 2). Thus, if the objective is to infer overall species population status and trend through time at the species and NWR level, then the RoM is the more appropriate statistic.

Example 2

The National Park Service (NPS) fire management program conducts systematic monitoring of prescribed burns to determine whether management goals are being achieved. Within each park, a series of fire management objectives are established for each major vegetation community, referred to hereafter as monitoring types. Fire ecologists typically use random sampling within each monitoring type to collect pre- and postburn data at regular time intervals.

In this area of work, the choice to focus on relative change often rises to the level of routine policy. For example, an important metric used to predict fire intensity and potential fire impacts is the surface fuel loading, defined as the total amount of dead biomass (wood, duff, and litter) on the forest floor available for consumption (measured in dry weight as tons per acre). A common management objective listed in Fire Management Plans is to reduce the total surface fuel load in a given monitoring type by some designated percentage. The data for Example 2 show surface fuel loading values for 11 random plots established in the mixed conifer monitoring type before (pre) and 1-yr after (post) a prescribed burn in 2008 in Lassen Volcanic National Park (northern California, USA) (Table 3). The park's programmatic management objective for this monitoring type is to reduce surface fuel load by 50%–65% one year after burning.

Since inference is to be made to the entire population in the burn unit as opposed to estimating how, on average, individual plots behaved, estimation using the means is preferred. The RoM is 0.42 with a 95% *bca*-bootstrapped CI of (0.27, 0.55), indicating that the postfire fuel load was 42% of the prefire fuel load. To interpret the RoM in the context of the management objective, we subtract the RoM from 1 (and rescale) to show the percent reduction in fuel loading from prefire levels: $100\% \times \left(1 - \overline{Y}_{\text{post}}/\overline{Y}_{\text{pre}}\right) = 58\%$, with a 95% CI of (45%, 73%). Thus, we can conclude that the burn reduced fuel loading by 58%, and the management objective of reducing fuel load by 50%–65% was achieved.

For sake of comparison, the analogous result using the MoR shows an estimated fuel load reduction of 66%, CI of (50%, 78%), slightly greater than the upper management objective range (although not statistically significant). The MoR suggests a greater overall reduction than the RoM because plots with

Table 2. Total number of Yaqui chub *Gila purpurea* captured from 10 wetland ponds in southern Arizona.

Year	Pond									
	1	2	3	4	5	6	7	8	9	10
2007	10	38	0	498	1	3	165	145	100	633
2012	1	593	350	350	300	552	474	154	597	604
’12/‘07	0.10	15.61	NA	0.70	300	184	2.87	1.06	5.97	0.95

relatively low preburn fuel loads experienced a larger percent reduction than plots with higher fuel loads; these plots receive a lower weight in the RoM because they contribute less to the total loading of the burn unit (population). This is not an unusual occurrence in spring burns because areas with open forest canopies (and often lower corresponding surface fuel loads) can dry quicker than areas with more dense canopies (and often higher corresponding fuel loads). Thus, the burn may consume a greater percentage of biomass in those open forests, even though more total tons per acre of fuels are being consumed in the denser forests. The RoM better reflects the population fuel load across the unit.

Example 3

This example illustrates that the two methods not only usually produce different estimates numerically, but that they can actually point in opposite directions! Here, we have native grass percent cover from 42 paired plots, preburn and five years postburn, in the sagebrush-steppe monitoring type in Lava Beds National Monument (northern California, USA) (Figure 2; Table 4). These data are intended to measure park-wide trends across the sagebrush monitoring type, so the RoM is the appropriate metric here. The MoR is 1.17, suggesting a slight (17%) increase in native grass cover five years after burns, whereas the RoM (the correct metric) yielded an 8% *decrease*. Here, the MoR approach based on individual ratios yields a 95% CI of 0.96–1.53 strongly suggestive of an increase, whereas the RoM yields a 95% CI of 0.80, 1.07, equally strongly suggestive of a decrease! *Caveat emptor*!

The reason for the discrepancy between the two metrics is a small handful plots with very low prefire native grass cover that exert high leverage on the MoR due to equal weighting (Figure 2). For example, grass cover in one plot increased from 1% to 5% after the burn, which is a fivefold increase even though real cover increased by only 4 percentage points! In fact, just four plots with prefire cover of ≤4% (no other plot in the park had <12%) have an MoR of 3.2, compared to 0.95 for the other 38 plots (more in line with the RoM of 0.91). These plots contribute very little to overall grass cover across the monitoring type in the park, which is what the monitoring was intended to measure.

Table 3. Surface fuel loading values for 11 random plots established in the mixed conifer monitoring type before (pre) and 1-yr after (post) a prescribed burn in 2008 in Lassen Volcanic National Park (northern California, USA).

	Plot										
	1	2	3	4	5	6	7	8	9	10	11
Pre	22	60	95	18	39	45	47	24	47	18	57
Post	1	29	50	4	6	27	24	5	6	2	45
1 − Post/Pre (%)	95	52	48	76	84	39	49	80	88	89	21

Table 4. Native grass percent cover from 42 paired plots, preburn and five years postburn, in the sagebrush-steppe monitoring type in Lava Beds National Monument (northern California, USA).

Plot	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Pre	23	28	18	14	46	32	18	3	25	50	38	49	38	42
Post	24	24	8	7	27	15	12	8	34	33	39	28	34	29
Plot	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**
Pre	69	39	42	40	62	18	31	22	27	54	55	39	4	1
Post	51	27	36	23	39	21	24	14	43	62	38	49	13	2
Plot	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**
Pre	26	12	5	25	1	21	29	17	31	31	30	46	37	62
Post	17	5	7	23	5	20	47	44	35	39	65	73	15	28

In this example, the preburn percent cover of grasses was approximately 30%. If a specific management goal had been declared (it was not here) in relative terms, it might have been, say, "a 50% increase." Since the measurement units are percentage, this statement is unclear. A 50% *relative* change would be 15 percentage points (an increase to 45% cover), whereas a 50 percentage point increase would yield 80% cover after the burn.

Additional points

In our examples, data were paired across time. This is a common circumstance, but not exclusively so. For example, pairing can have a spatial basis. Long et al. (2017) employed paired plots in a study examining the effect of removing feral pigs on soil and nutrient availability. One of each pair was inside an exclosure, the other out. Additionally, paired sampling is often done in ecology to measure the effects of natural or intentional experimental treatments (e.g., burned vs unburned, or sprayed vs unsprayed, etc.) for practical reasons, yet the overall inference is still on the population rather than individual pairs. The same theory and application of RoM vs MoR would apply.

RoM and MoR can yield wildly different results driven by, in particular, high variation in the denominator terms. In our first example, the conclusions were approximately an order of magnitude different between the two methods. In the third example, the two estimators pointed in opposite directions (MoR suggested an increase, while MoR pointed to a decrease)!

Just as it is possible that one might, for different purposes, use both the mean and median in turn (applicable if the data are skewed), so it is also possible that one might, for different ends, use both the RoM and the MoR in a given study.

The mean of ratios can be handled with standard statistical tools, since summarized data are simply a single sample of values, which happen to be ratios. The ratio of means, however, is better served by use of bootstrapping. Gerow has created an interactive Excel app for confidence intervals and will gladly send such to interested readers.

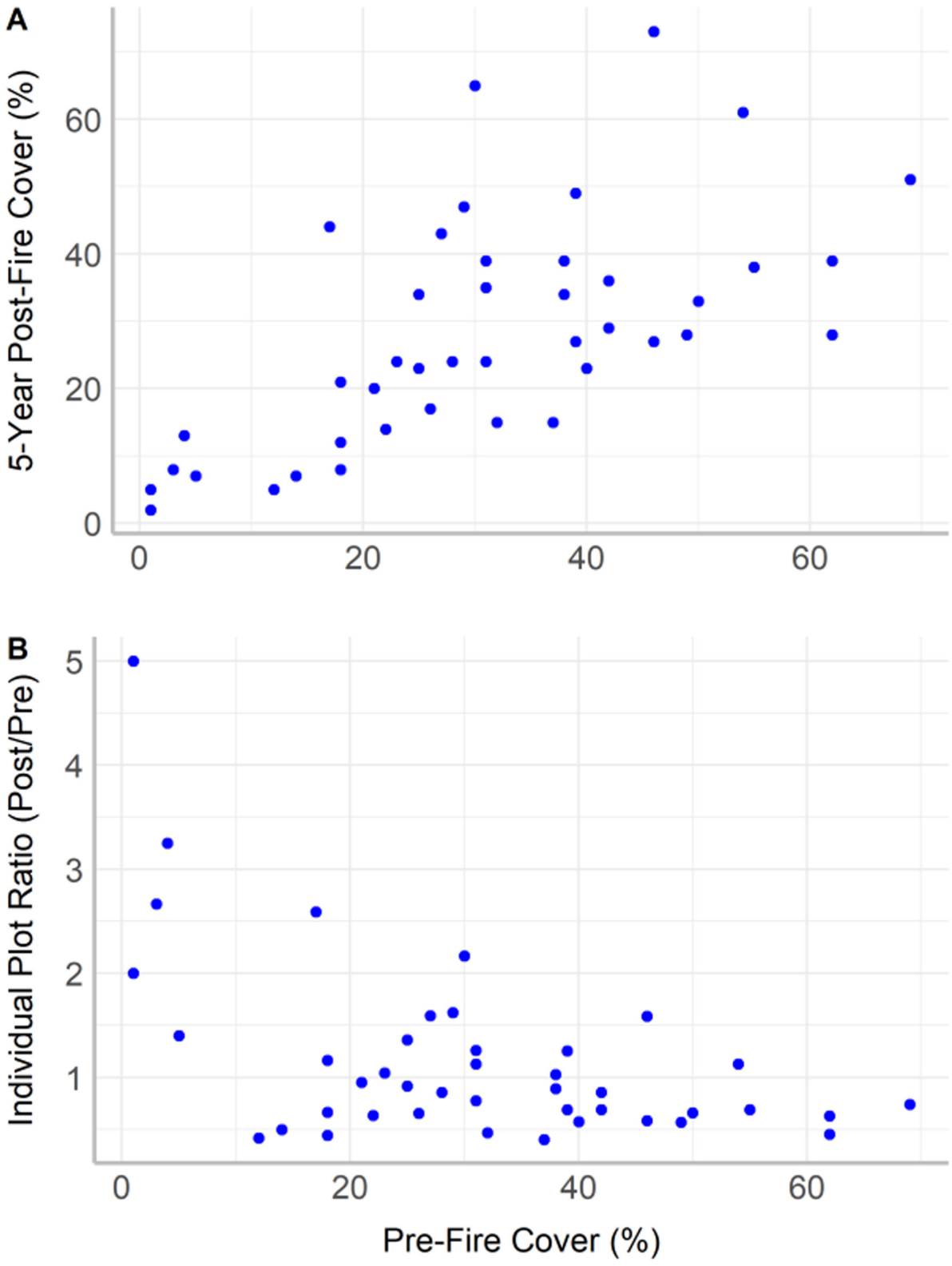


Fig. 2. Pre- and postfire (5-yr) cover (A) and individual plot ratio (Post/Pre) data (B) from sagebrush-steppe monitoring type in Lava Beds National Monument (northern California, USA).

Summary

The intention when using paired data in ecological studies is often to make inference at the population level: how much change has there been at the landscape or watershed level. In that case, the ratio of the two means is estimating the parameter of interest, whereas the mean of ratios is useful if, on average, individual change is of interest. Our examples showed that the two approaches (RoM or MoR) can produce markedly different results, so careful attention to choosing between them is paramount.

For the MoR, any standard single sample procedure will work (e.g., use a one-sample *t* if sample sizes are large enough). Bootstrapping is useful for confidence intervals when using the RoM; we recommend that *bca*-bootstrap procedure in that case. If interest happens to be at the individual level, in which case the mean of pairwise ratios is suitable, we note that zero values in the denominator require either dropping that data point or replacing the zero with some arbitrary small value (much as we deal with zero values when deploying logarithmic transformation).

Literature Cited

Buck, C. L., and B. M. Barnes. 1998. Annual cycle of body composition and hibernation in free-living arctic ground squirrels. Journal of Mammology 80:430–442.

Díaz-Francés, E., and F. J. Rubio. 2013. On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables. Statistical Papers 54:309–323.

Díaz-Francés, E., and D. A. Sprott. 2004. Inference for the ratio of two normal means with unspecified variances. Biometrical Journal: Journal of Mathematical Methods in Biosciences 46:83–89.

Efron, B. 1979. Bootstrap methods: another look at the jackknife. Annals of Statistics 7:1–26.

Efron, B., and R. Tibshirani. 1993. An introduction to the bootstrap. Chapman & Hall, New York, New York, USA.

Green, J. A., L. G. Halsey, P. J. Butler, and R. L. Holder. 2007. Estimating the rate of oxygen consumption during submersion from the heart rate of diving animals. Comparative and Evolutionary Physiology 292:2028–2038.

Herwig, B. R., and K. D. Zimmer. 2007. Population ecology and prey consumption by fathead minnows in prairie wetlands: importance of detritus and larval fish. Ecology of Freshwater Fish 16:282–294.

Kim, H. K., W. D. Lee, S. G. Kang, and Y. Kim. 2018. Objective Bayesian tests for Fieller-Creasy problem. Computational Statistics 34:1159–1182.

Kunz, T. H., J. A. Wrazen, and C. D. Burnett. 1998. Changes in body mass and fat reserves in prehibernating little brown bats (*Myotis lucifugus*). Ecoscience 5:8–17.

Lohr, S. 1999. Sampling: design and analysis. Pages 494. Duxbury Press, Pacific Grove, California, USA.

Long, M. L., C. M. Litton, C. P. Giardina, J. Deenik, R. J. Cole, and J. P. Sparks. 2017. Impact of nonnative feral pig removal on soil structure and nutrient availability in Hawaiian tropical montane wet forests. Biological Invasion 19:749–763.

Manley, B., L. McDonald, D. Thomas, T. McDonald, and W. Erickson. 2002. Resource selection by animals: statistical design and analysis for field studies. Second edition. Pages 231. Kluwer Academic Publishers, New York, New York, USA.

Mendoza, M., and E. Gutierrez-Pena. 1999. Bayesian inference for the ratio of the means of two normal populations. Biometrical Journal: Journal of Mathematical Methods in Biosciences 41:133–147.

Scheaffer, R., W. Mendenhall, L. Ott, and K. Gerow. 2012. Elementary survey sampling. Seventh edition. Pages 436. Brooks/Cole, Boston, Massachusetts, USA.

Stewart, D. R., M. J. Butler, G. M. Harris, L. A. Johnson, and W. R. Radke. 2017b. Estimating abundance of endangered fish by eliminating bias from non-constant detectability. Endangered Species Research 32:187–201.

Stewart, D. R., M. J. Butler, G. M. Harris, and W. R. Radke. 2017a. Mark-recapture models identify imminent extinction of Yaqui catfish *Ictalurus pricei* in the United States. Biological Conservation 209:45–53.

Stewart, D. R., M. J. Butler, L. A. Johnson, A. Cajero, A. N. Young, and G. M. Harris. 2019. Efficacy of depletion models for estimating abundance of endangered fishes in streams. Fisheries Research 209:208–2017.

USDI National Park Service. 2003. Fire Monitoring Handbook. Pages 274. Fire Manage Program Center, National Interagency Fire Center, Boise, Idaho, USA.

USFWS (US Fish and Wildlife Service). 1994. Yaqui fishes recovery plan. US Fish and Wildlife Service, Albuquerque, New Mexico, USA.

Webb, P. M., R. D. Andrews, D. P. Costa, and B. J. Le Boeuf. 1998. Heart rate and oxygen consumption of northern seals during diving in the laboratory. Physical Zoology 71:116–125.