

Neglected biological patterns in the residuals

A behavioural ecologist's guide to co-operating with heteroscedasticity

Ian R. Cleasby · Shinichi Nakagawa

Received: 25 May 2011 / Revised: 17 August 2011 / Accepted: 24 August 2011 / Published online: 11 October 2011
© Springer-Verlag 2011

Abstract One of the fundamental assumptions underlying linear regression models is that the errors have a constant variance (i.e., homoscedastic). When this assumption is violated, standard errors from a regression can be biased and inconsistent, meaning that the associated p values and 95% confidence intervals cannot be trusted. The assumption of homoscedasticity is made for statistical reasons rather than biological reasons; in most real datasets, some form of heteroscedasticity is likely to exist. However, a survey of the behavioural ecology literature showed that only about 5% of articles explicitly mentioned heteroscedasticity, leaving 95% of articles in which heteroscedasticity was apparently absent. These results strongly indicate that the prevalence of heteroscedasticity is widely under-reported within behavioural ecology. The aim of this article is to raise awareness of heteroscedasticity amongst behavioural ecologists. Using topical examples from fields in behavioural ecology such as sexual dimorphism and animal personality, we highlight the biological importance of considering heteroscedasticity. We also emphasize that researchers should pay closer attention to the variance in

their data and consider what factors could cause heteroscedasticity. In addition, we introduce some simple methods of dealing with heteroscedasticity. The two methods we focus on are: (1) incorporating variance functions within a generalised least squares (GLS) framework to model the functional form of heteroscedasticity and; (2) heteroscedasticity-consistent standard error (HCSE) estimators, which can be used when the functional form of heteroscedasticity is unknown. Using case studies, we show how both methods can influence the output from linear regression models. Finally, we hope that more researchers will consider heteroscedasticity as an important source of additional information about the particular biological process being studied, rather than an impediment to statistical analysis.

Keywords Heteroscedasticity · Homoscedasticity · Linear regression · Residuals · Standard errors · Variance · Within-group errors

Introduction

For most scientists, a fundamental part of research involves applying statistical models to their data in order to make valid conclusions. While there is a vast array of different techniques to choose from, there is little doubt that the humble linear regression represents the foundation upon which more complex models can be constructed (Hayes and Cai 2007). Whenever a researcher performs a traditional ORDINARY LEAST SQUARES (OLS) linear regression, they implicitly make a number of assumptions about their data (see Table 1 for a glossary of all terms written in uppercase in their first appearance). Most of these assumptions will be familiar to any reader who has taken part in even basic

Communicated by L. Z. Garamszegi

Electronic supplementary material The online version of this article (doi:10.1007/s00265-011-1254-7) contains supplementary material, which is available to authorized users.

I. R. Cleasby (✉)
Department of Animal and Plant Sciences,
University of Sheffield,
Western Bank,
Sheffield S10 2TN, UK
e-mail: i.r.cleasby@gmail.com

S. Nakagawa
Department of Zoology, University of Otago,
340 Great King Street, PO Box 56, Dunedin, New Zealand

Table 1 Glossary of statistical terms used in the text

Term	Definition
AIC	The Akaike Information Criterion is a measure of the relative goodness-of-fit of a statistical model based upon the information theories. When comparing two models the one with the lower AIC score is judged to fit the data better. In most cases, using a corrected version of the AIC score known as AICc is recommended.
Covariance	A measure of how much two variables change (covary) together.
Error variance	The variance in the statistical errors of a given model. Statistical errors are the difference between a sample and its true function value. Because this true value is unobservable we cannot calculate the true model errors and have to rely upon the model residuals as an estimate instead.
Fixed effect	Terms in a statistical model for which regression coefficients are estimated. The exact definition of fixed effects is controversial (Gelman 2005).
Generalised least squares (GLS)	An extension of the traditional linear model. Can be used when data are heteroscedastic, or when observations are correlated. Involves altering the basic variance–covariance matrix.
Heteroscedasticity	Occurs when error terms <i>do not have</i> a constant variance.
Heteroscedasticity consistent standard error estimator (HCSE)	A type of sandwich estimator that can generate more reliable standard errors when data are heteroscedastic. Useful when the form of heteroscedasticity is unknown.
Homoscedasticity	Occurs when errors terms have a constant variance.
Leverage	A measure of the influence that a data point has on the results of a fitted regression model. Data points with high leverage values will tend to draw the fitted regression curve towards them, potentially skewing the results.
Mixed-effect model	A model that includes both fixed effects and random effects.
Moment	In statistics a moment is quantitative measure of the distribution, or shape, of a set of points. The first moment of a distribution is usually the mean, with the second moment being the variance. Further, higher-order moments include skewedness and kurtosis.
Nested models	Models that are subsets of a more complex model, often with one parameter removed, or set at a value of 0.
Ordinary least squares (OLS)	Traditional method of fitting a linear regression. Minimizes the sum of the squared distances between observed responses and predicted responses. Makes a number of assumptions about the data.
Random effects	Factors whose levels are sampled from a larger population, or effects where interest lies in the variation among levels rather than on the specific effects of each level. The exact definition of random effects is controversial (Gelman and Hill 2007).
Residuals	An <i>estimate</i> of the unobservable statistical error. Typically, the difference between a sample and its estimated value from the data.
Variance	A measure of the dispersion of a set of data points around their average value.
Variance–covariance matrix	A matrix where the variance is presented on the diagonal elements and the covariance is presented on the off-diagonal elements.
Variance function	Function allowing the variance to depend on the fixed effects and random effects in a specified way.

statistical courses and they include: normality of the RESIDUALS, independence of data points and HOMOSCEDASTICITY. For the rest of the paper, our focus will be upon the assumption of homoscedasticity and methods of modelling data when this assumption is violated and our data are HETEROSCEDASTIC.

What are homoscedasticity and heteroscedasticity?

The terms homoscedasticity and heteroscedasticity are both concerned with the ERROR VARIANCE in data. Homoscedasticity occurs when the error terms in a model have a constant variance. In contrast, heteroscedasticity occurs when this assumption is broken. In certain situations heteroscedasticity may also be referred to as variance heterogeneity (Congdon 2010), but for the sake of simplicity we will only

use the term heteroscedasticity here. The assumption of homoscedasticity can be made explicit if we consider the equation for a simple linear regression:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \text{ where } \varepsilon_i \sim N(0, \sigma^2), \quad (1)$$

where Y_i is the response variable, X_i is the predictor variable, parameters β_0 and β_1 are the intercept and slope of the model, respectively, and ε_i represents the model error. Equation 1 specifies that the errors from the model are normally distributed with a mean of 0 and, crucially, that the variance of the error terms is σ^2 . Assuming homoscedasticity also implies that the variance of error terms is not related to any predictor, or linear combination of predictors (Hayes and Cai 2007). In some instances, homoscedasticity may be violated even if the magnitude of the variance in the error terms is not a function of any of the predictors; this is termed

heteroscedasticity of unknown form (White 1980). The model error terms, ε_i , represent the difference between the i th observation and the unobserved true population mean. However, because the true population mean is unknown we cannot actually estimate the model errors. Instead, the model residuals ($\hat{\varepsilon}_i$) are used as an estimate of the model errors. The residuals from a model represent the difference between an observation and the predicted value (estimated via regression) and can therefore be calculated from the data. Figure 1a and b shows plots of the residuals from two different linear regressions. In Fig. 1a, the assumption of homoscedasticity was valid. In Fig. 1b, however, it is clear that the variance in the residuals increases and the assumption of homoscedasticity was violated. In many cases, heteroscedasticity will appear as an increasing dispersion in model residuals at higher fitted/predictor values (Congdon 2010), but it can take many forms. For instance, the variance in residuals could decrease linearly as a function of one or more predictor, or be greatest at moderate values of a particular predictor.

When heteroscedasticity occurs, it does not usually cause any bias in the estimated model coefficients themselves (White 1980). However, it will influence the standard errors around these coefficient estimates, making any statistical inferences and predictions based on the model unreliable. Heteroscedasticity can cause serious problems when using a traditional hypothesis testing approach because when errors are heteroscedastic then the standard error estimates from an OLS regression are both biased and inconsistent, making model p values and confidence

intervals unreliable (Hayes and Cai 2007). As a result, heteroscedasticity can reduce the statistical power of standard ANOVA and linear regression techniques (Wilcox 1998).

To demonstrate precisely how the estimated standard error of model coefficients can influence our p values, consider that most statistical packages provide a t value (sometimes a z value) for each coefficient upon which the p value is based. In many cases this test statistic is calculated using the estimated standard errors as:

$$\frac{\hat{\beta}}{\text{s.e.}(\hat{\beta})}, \quad (2)$$

where $\hat{\beta}$ represents the estimated regression coefficients for the model in question and $\text{s.e.}(\hat{\beta})$ represents the standard error of this estimate. Clearly, any changes to the denominator in the above equation will influence the test statistic and, hence the p values, that we obtain.

The extent to which heteroscedasticity will affect our results depends on both its severity and form. When there is only weak heteroscedasticity it is unlikely that the outcome of the analysis will be drastically changed. However, even weak heteroscedasticity can have important effects when sample size is large. When the spread of residuals is smaller around the extreme values of one of the predictors in the model then the standard errors from the model tend to be overestimated and the p values are too large. On the other hand, if the spread of residuals is greater at either the higher or lower values of a predictor variable then the standard errors tend to be underestimated and the p values will be too small (Hayes and Cai 2007). Unfortunately, simply increasing the sample size will not necessarily reduce this problem, and could even make things worse (Long and Ervin 2000). In addition, Venables (2000) points out that when two samples show a large difference in their variance, this difference will often be more important than any difference in their means.

As well as influencing the size of the standard errors around our estimated regression coefficients, heteroscedasticity can also influence the reliability of the variance components when using a MIXED-EFFECTS MODEL approach. Within the field of animal breeding the importance of accounting for heteroscedasticity when seeking accurate genetic evaluations of livestock has long been recognised (Hill 1984). For animal breeders, heteroscedasticity is particularly important because phenotypic variance will influence the response to selection and the uniformity of the end product (e.g., milk yield). For an example of the effect that different models of heteroscedasticity can have on the variance components from a mixed model, we recommend reading Rönnegård et al. (2010).

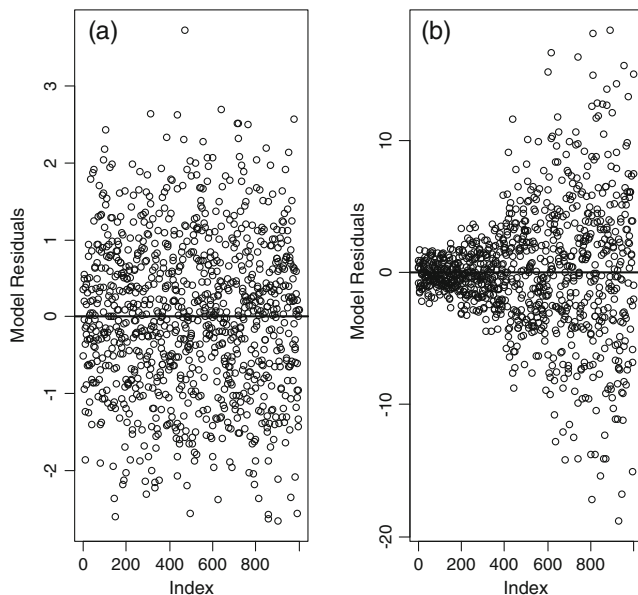


Fig. 1 Plots of the residuals from two linear regressions one in which the assumption of homoscedasticity was valid (**a**) and one in which it was invalid (**b**). Here, the x -axis index represents the identity of each observation in the model from 1 through to 1,000

How often is heteroscedasticity mentioned?

We would imagine that almost all behavioural ecologists are aware that examining model assumptions is essential. However, to see how frequently heteroscedasticity is referred to in the literature we performed an advanced search of three different behavioural ecology journals (*Behavioural Ecology*, *Behavioural Ecology & Sociobiology* and *Ethology*) over the period from January 2007 until March 2011. We performed advanced searches using Boolean logic to look for the words ‘heteroscedasticity’ AND ‘residuals’, or ‘heterogeneity’ AND ‘residuals’ anywhere in the text of an article. All searches were conducted via the search engines found on the home page of each journal. We then sorted through the articles found and recorded how many of them referred to heteroscedasticity within a statistical context. We also performed another separate search in which were looked for the words ‘regression’ AND (‘linear’ OR ‘multiple’) anywhere within the text in order to estimate of the total number of articles in which heteroscedasticity could have been mentioned. We did this to try and exclude articles in which checking for heteroscedasticity was irrelevant (e.g., theoretical papers). We focused only upon original articles; review papers, methods papers and short communications were not included within our searches. The results from these investigations are provided in Table 2. In the behavioural ecology journals examined heteroscedasticity was rarely mentioned explicitly and in the vast majority of cases was, presumably, absent. Of course, in some cases the focus of the article or the approach used may have made consideration of heteroscedasticity redundant. In other cases, the assumption of homoscedasticity may have been validated, but went unreported. Nevertheless, our quick survey would suggest that heteroscedasticity is a problem only rarely encountered in behavioural ecology datasets. Similarly, Ruxton (2006) reported that the unequal variances *t*-test has been neglected by behavioural ecologists in the past. However, there is usually no underlying biological reason why homoscedasticity is assumed and one could argue that in many cases we might expect heteroscedasticity (see below). Indeed, Erceg-Hurn and Mirosevich (2008) point out that the assumption of homoscedasticity is rarely met when analysing real data. Moreover, biological research has been identified as an area in which unequal variances are a common problem and

may often be hard to avoid (Hill and Zhang 2004; Herberich et al. 2010). Given this, the paucity of articles in the behavioural ecology literature that report heteroscedasticity seems somewhat puzzling. After all, this is an area in which we could argue that, not only should we expect heteroscedasticity, but also that it would actually be biologically interesting. In fact, we would suggest that heteroscedasticity is under-reported by behavioural ecologists who may be unaware of some of the different techniques available for dealing with heteroscedastic data when the need arises.

When might we see heteroscedasticity in behavioural ecology?

In the following section the goal is to try and give some examples of when heteroscedasticity might occur. The list is far from exhaustive and we hope that researchers reading this article will be able to come up with further examples from their own work.

Any study in which measurements from males and females are included within the same analysis is a candidate for heteroscedasticity. The problem may be particularly severe in species with a marked sexual size dimorphism, because the error variance will tend to increase as body size increases. A similar problem is also encountered in allometric studies when comparing animals, or species, of different sizes. Even in species with less obvious sexual dimorphism males and females may behave or respond differently, potentially leading to heteroscedasticity. As a specific example, studies looking at the parental care behaviour of the house sparrow (*Passer domesticus*) have shown that there is more variation in the provisioning rate of females than of males (Schwagmeyer and Mock 2003; Nakagawa et al. 2007a), which could cause heteroscedasticity in the model errors. In addition, there are a multitude of studies reporting that one sex is more sensitive to poor rearing environment (Jones et al. 2009). In such instances, we might expect the variation expressed in certain traits to differ between males and females depending on the environment experienced during development. Such sex-specific variation will play an important role in the evolution of sexual dimorphism (Badyaev 2002). Indeed, any study in which a relationship between environmental

Table 2 The frequency with which heteroscedasticity is reported within three of the major behavioural ecology journals

Journal	No. articles	No. articles mentioning heteroscedasticity	% articles mentioning heteroscedasticity
<i>Behavioural Ecology</i>	354	16	4.52%
<i>Behavioural Ecology & Sociobiology</i>	265	6	2.26%
<i>Ethology</i>	195	6	3.08%
Total	814	28	3.44%

quality and phenotypic variation is expected (Teder et al. 2008) is a candidate for heteroscedastic errors.

In general, an argument for heteroscedasticity may be plausible whenever there is a distinct stratification in the data (Keppel and Wickens 2004). For example, Clayton et al. (1957) reported that selected lines of the fruitfly (*Drosophila melanogaster*) showed a higher degree of phenotypic variance than non-selected lines. Similarly, other studies have shown that heteroscedasticity can result from differences between control and treatment groups (Honkanen and Jormalainen 2005; Erceg-Hurn and Mirosevich 2008), different herds/populations (Brotherstone and Hill 1986), different breed groups (Cardoso et al. 2005), or when a genotype-by-environment interaction is suspected (Dutilleul and Potvin 1995). In certain cases, we may also find the variance in the model residuals is related to a particular continuous predictor (Zuur et al. 2009). In studies of growth, the residual variance tends to increase as individuals' age and grow larger (Cleasby 2010). As another example, in the blue tit (*Cyanistes caeruleus*) the variance in mean residual nestling mass appears to increase as the distance from the nest to the nearest oak tree increases (Wilkin et al. 2009). Heteroscedasticity can also be an issue whenever data has either a temporal or spatial component (Bivand et al. 2008; Cryer and Chan 2008). Both Grissom and Kim (2005) and Erceg-Hurn and Mirosevich (2008) provide a more detailed discussion about why heteroscedasticity occurs in real data sets. In most cases, we would argue that the existence of heteroscedasticity fits nicely with our theoretical and practical knowledge of biology; thus, discovering that our data are heteroscedastic could be seen as a positive rather than a cause for concern.

The recent increase in animal personality studies provides a specific example of how consideration of heteroscedasticity and the variance in our data will become increasingly important to behavioural ecologists. Broadly, animal personality can be defined as consistent, individual differences in the behaviour of animals across time and different contexts (Dingemanse et al. 2010). As a result, consideration of the variance present in data is a crucial component of animal personality studies. Typically in animal personality studies, there has been a focus upon between-individual differences in the average level of a particular behaviour (Dingemanse et al. 2010). However, there is now a growing realisation that we need to pay more attention to the variation around the average behaviour of individuals. In particular, Reale and Dingemanse (2010) suggest that by focussing on the within-individual variance in animal behaviour, we may be able to identify individuals as being generalists or specialists. In this context, specialists are individuals that show a low degree of variation in their behaviour, whereas generalists show a high degree of variation. Alternatively, one could focus on individual changes in behavioural consistency over time to

investigate how personality develops (e.g., Sinn et al. 2008). To some extent, this increased focus on variance mirrors what has already occurred in other branches of biology, where understanding and modelling variance has become increasingly important (e.g., Hill 1984). In a sense, all we are really doing is shifting our attention from one MOMENT of a distribution (the mean) to another moment (the variance). However, it is important to realise that if individuals do differ in the level of consistency they exhibit then the data will not be homoscedastic. Incorporating heteroscedasticity into statistical models provides one way in which we can begin to describe the extent of within-individual variation in a particular behaviour. For example, by allowing the residual variance to differ by individual we can assess whether the fit of a model is improved and describe the pattern of within-individual variation across the population. Moreover, it should be possible to incorporate heteroscedasticity within the behavioural reaction norm approach outlined by Dingemanse et al. (2010) in order to represent within-individual variation, alongside between-individual variation in intercepts and slopes. When one performs a regression the focus usually rests upon population averages, which is what regression coefficients typically represent. However, animal personality studies provide a good example of how behavioural ecologists could gain additional insights by paying more attention to the variance in their data. The growing use of mixed effects models within behavioural ecology suggests that researchers are beginning to scrutinize the variance in their data more closely, and a greater consideration of heteroscedasticity should become more common in the future as a result. Ultimately, this could aid our understanding of evolution in general. After all, as Hill and Zhang (2004) point out; selection must act upon variance per se, not just upon variability in means.

Detecting heteroscedasticity

The simplest way for researchers to detect heteroscedasticity is to plot their data and look for visual patterns. Plotting residuals against fitted values from the model in question provides one means of detecting patterns in the residuals. One can also plot model residuals against various predictor variables in order to identify if any predictors contribute to heteroscedasticity; this procedure may provide some insight into the functional form of any heteroscedasticity (Zuur et al. 2009). The assumption of homoscedasticity is validated if the dispersion of residuals is the same across all fitted values, or predictor values. Figure 2a and b shows a simulated dataset in which an initial plot of the raw data (Fig. 2a) suggested heteroscedasticity would be a problem; this was then supported by a residual plot from the resulting linear regression (Fig. 2b). One potential

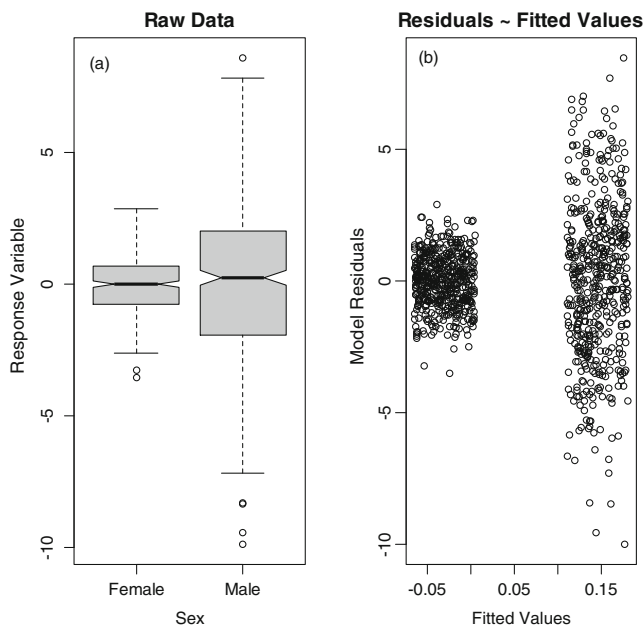


Fig. 2 A plot of simulated raw data (**a**) suggesting that heteroscedasticity might be a problem prior to running a model a plot of the residuals from the model (**b**) also showed evidence of heteroscedasticity. **a** The bold black line represents the median of the response variable, the hinges of the box represent the 25% and 75% quantiles, respectively, and the whiskers extend to 1.5 times the length of the box. The notches displayed on the box extend to $\pm 1.58 \times \text{inter} - \text{quartile range} / \sqrt{n}$. In **b**, the x-axis was jittered to display the dispersion in the residuals more clearly

drawback with this visual method is its subjectivity, obvious patterns in the data are easy to detect, but assessing heteroscedasticity becomes difficult with more subtle patterns (Darlington 1990). To combat this, there are a variety of different statistical tests for heteroscedasticity available. In the case of regression the White test (White 1980), the Breusch–Pagan test (Breusch and Pagan 1979) and the Goldfeld–Quandt test (Goldfeld and Quandt 1965) are some of the more common tests of heteroscedasticity. However, like all statistical tests the various tests make certain assumptions about the data and the functional form of heteroscedasticity under consideration. Therefore, one should be aware of these assumptions before using them. As an alternative, we can compare a model in which heteroscedasticity has been modelled in some way to a simpler model in which homoscedasticity is assumed. Such models are not NESTED, but model AIC scores can be compared in order to choose the best model (Zuur et al. 2009).

Methods of dealing with heteroscedasticity

Assuming that one finds evidence of heteroscedasticity in the residuals the next problem is how to deal with it. Probably the most common method of dealing with heteroscedasticity is to transform the data (the log transformation, in particular, appears

to be widely used). While it is not necessarily a bad thing to transform data in this way the procedure does have some drawbacks: (1) it might not remove heteroscedasticity; (2) results on the transformed-scale might be more difficult to interpret; (3) it can be difficult to know which kind of transformation is best without some understanding about the functional form of the heteroscedasticity; and (4) the existence of heteroscedasticity is not simply an impediment to an analysis; it may represent important information (Carroll 2003; Hayes and Cai 2007; Zuur et al. 2009). In the following sections, we focus on two different methods of dealing with heteroscedastic data that do not involve data transformation, the first method involves using GENERALISED LEAST SQUARES (GLS) regression and the second method involves using HETEROSCEDASTICITY-CONSISTENT STANDARD ERROR (HCSE) estimators. We claim no originality for either of these statistical techniques, both of which have existed for many years (White 1980; Pinheiro and Bates 2000). However, we present them as often overlooked methods that are extremely relevant to behavioural ecologists, and in other areas of ecology as well.

Modelling heteroscedasticity using GLS

When constructing a model in which heteroscedasticity is incorporated in some way it is helpful to consider that heteroscedasticity can be modelled within the random part of a model. The random part of a model is concerned with variance in the data and also includes random effects, as well as both temporal and spatial-autocorrelation structures (Zuur et al. 2009). Hence, including heteroscedasticity in a model will involve altering the basic VARIANCE–COVARIANCE MATRIX. Essentially, a GLS is a form of weighted regression and can be very useful provided that we know something about the form of the heteroscedasticity in our model (Mardia et al. 1979). In a GLS model we can use a VARIANCE FUNCTION to specify the form of the within-group errors within our data (Pinheiro and Bates 2000). The basic variance function formula is very flexible and it allows the error variance to depend on the FIXED EFFECTS (and also the RANDOM EFFECTS in the case of a mixed model) in a defined manner. Thus, incorporating a variance function changes the basic linear regression equation. For instance, if we wanted to include a simple variance function that allowed the variance in our errors to increase as the value of a predictor, X , increased we could simply write the linear regression equation in Eq. 1 as:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \text{ where } \varepsilon_i \sim N(0, \sigma^2 \times X_i) \quad (3)$$

The crucial difference is in how we have defined the variance in the model errors.

To describe the basics of GLS, we present a simple case study using data collected as part of the Lundy house sparrow project (for more details, see Griffiths et al. 1999; Nakagawa et al. 2007a). Our GLS models were conducted in the R environment (R version 2.12.1, R Development Core Team 2010) using the package nlme (Pinheiro et al. 2010). We believe this package provides a good introduction to fitting GLS models and it includes a number of ready-made variance functions. Excellent descriptions of how to use variance functions in nlme can be found in the work of Pinheiro and Bates (2000) and Zuur et al. (2009); the R code and datasets used here can be found within the electronic supplement (ESM). As well as R, the programs SAS, SPSS and STATA allow the use of weighted least squares regression, which can be used to model heteroscedasticity.

In our example, the tarsus length of adult house sparrows was modelled using an individual's sex, and whether or not they received food supplements when they were a chick, as two predictor variables. Initial plots of the data suggested that the variance observed in adult tarsus length differed between males and females depending upon whether or not they received supplemental food during growth (Fig. 3). We present two analyses of this dataset, one in which the traditional assumption of homoscedasticity was made and another where we used a variance function to specify that the variance in adult tarsus length was allowed to differ

between the strata (i.e., sex \times treatment) in the data. The variance function used in the latter model can be written as:

$$\varepsilon_{ijk} \sim N(0, \sigma_{jk}^2), \text{ where } j = \text{sex}; k = \text{control or treatment group} \quad (4)$$

Equation 5 specifies that there are four distinct strata in our data (two sexes \times two treatment groups), each with their own particular error variance. Including this variance function in our model influenced our results, in particular the effect of the food treatment on adult tarsus length, which went from being non-significant to significant (Table 3). Comparing the two models on the basis of their AIC score suggests that the heteroscedastic model should be preferred, although only just. Plots of the normalized residuals from both models also suggest that using the heteroscedastic model resulted in a slightly more even dispersion of the residuals (Fig. 4a and b).

Keen observers will have noticed that the heteroscedastic model used three more degrees of freedom than the simpler homoscedastic model. To understand why, remember that we have four strata in the data and we need to estimate the variance for each of these strata. In nlme one of these strata is used as a reference point (by default, the reference strata is set to 1) and then the ratios between the standard deviation of each remaining stratum and the reference stratum are used as parameters within the variance function (Pinheiro and Bates 2000). Hence, with four strata we have to estimate three different ratios, using up three degrees of freedom in the process. The fact that incorporating variance functions into a model uses up degrees of freedom highlights one of the costs of using variance functions. In certain circumstances, particularly when dealing with small data sets, adding variance functions may not improve model fit, even when the data appear to be heteroscedastic.

The GLS models introduced here provide a relatively straightforward introduction to the modelling of variance structures. The basic GLS methods in R can be extended further into both generalised mixed-effects models and generalised additive models. Some care must be taken when including variance functions and random effects within the same model as they can compete with each other to some extent (Pinheiro and Bates 2000). Bayesian methods also provide a very flexible means of modelling heteroscedasticity. In the ESM, we provide information on how the analysis above could be run in the R package MCMCglmm (Hadfield 2010) using Bayesian techniques.

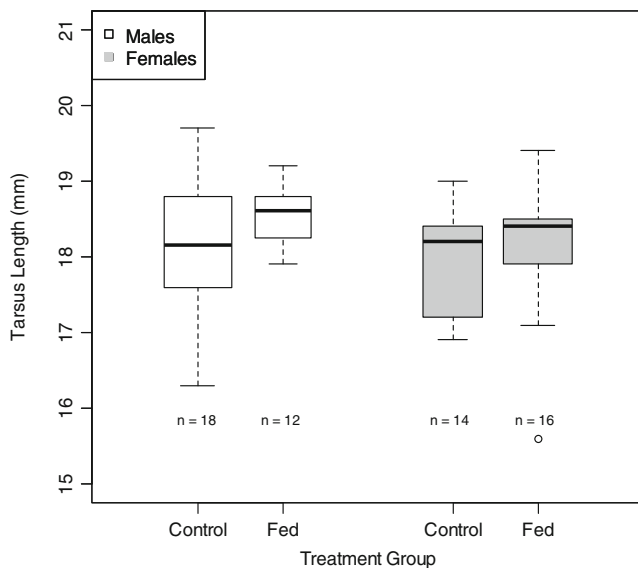


Fig. 3 The tarsus length of male and female house sparrows in the control and treatment groups. The **bold black line** represents the median of the response variable, the **hinges of the box** represent the 25% and 75% quantiles, respectively, and the whiskers extend to 1.5 times the length of the box

Table 3 The results from two different models of adult tarsus length in the house sparrow

The homoscedastic model represents the standard OLS regression, whereas the heteroscedastic model includes a variance function allowing the variance to differ between the different strata in the data, $n=64$

	Model					
	Homoscedastic			Heteroscedastic		
	β	SE	p	β	SE	p
Intercept	17.894	0.166	<0.001	17.891	0.156	<0.001
Sex	0.296	0.185	0.115	0.307	0.173	0.082
Treatment	0.311	0.186	0.099	0.351	0.163	0.036
	AIC=152			AIC=150		
	$df=4$			$df=7$		

Dealing with heteroscedasticity using HCSE estimators

Using GLS can be a very convenient way of dealing with heteroscedasticity. However, using this method implies that the researcher has some underlying knowledge about the mechanism generating the heteroscedasticity. In certain cases, such knowledge may not be available. For instance, imagine that the heteroscedasticity was being generated by a variable that we had not thought to measure. In such situations, an alternative method of reducing the effects of heteroscedasticity in our model is to use a HCSE estimator (White 1980; Long and Ervin 2000). The basic HCSE estimator is relatively straightforward. The regression model is estimated as usual using OLS, but then an alternative method of estimating standard errors that does not assume homoscedasticity is used. The advantage of this

technique is that it does not require any knowledge of the mechanism generating heteroscedasticity. Therefore, we no longer have to make any assumption about the structure of the errors within our data, nor do we have to use any degrees of freedom to estimate extra parameters (Hayes and Cai 2007). HCSE estimators appear to be widely used within econometrics where they have been quite successful in allowing researchers to report more reliable statistics (Wooldridge 2000). However, to the best of our knowledge, HCSE estimators have never been used in behavioural ecology. HCSE estimators are relatively simple, but in order to understand how they work it is best to rewrite the linear regression equation using matrix notation, hence Eq. 1 becomes:

$$y = X\beta + \varepsilon, \quad (5)$$

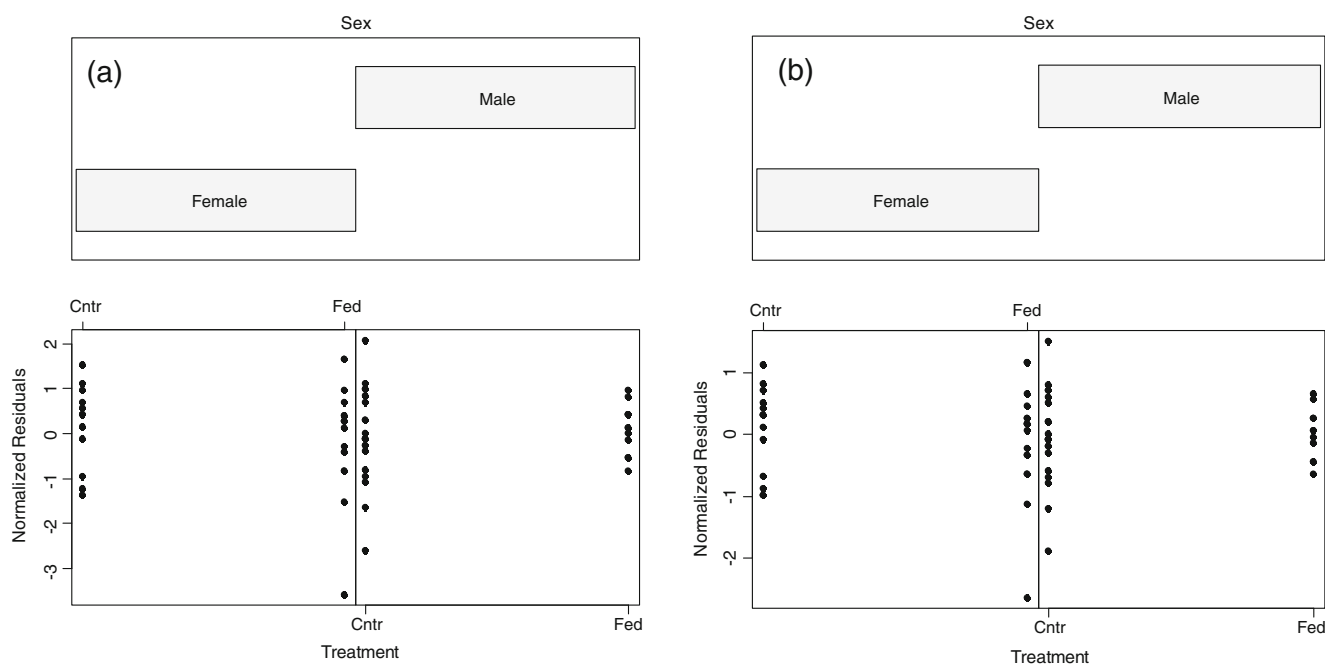


Fig. 4 Plots of the normalized residuals versus predictors for (a) the homoscedastic model and (b) the heteroscedastic model. In the bottom left of each figure the standardized residuals for females in the control (0) and treatment (1) groups are given. In the bottom right of each

figure the standardized residuals for males in the control (0) and treatment (1) groups are given. Note that the spread of the residuals is more even across all groups in b, in which the heteroscedasticity was modelled with a variance function

where \mathbf{y} is an $n \times 1$ vector denoting the response variable, \mathbf{X} is an $n \times (p+1)$ matrix of predictor variables (which includes a column of 1's representing the regression constant), $\boldsymbol{\beta}$ is a vector of length p containing the model coefficients, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of errors, and where n is the sample size and p is the number of predictors. In a traditional regression, the homoscedasticity of errors is assumed and the variance–covariance matrix of the errors, $\boldsymbol{\phi}$, is taken to be $\sigma^2 \mathbf{I}$, where \mathbf{I} represents an identity matrix of appropriate dimensions, and σ^2 represents the variance of model residuals (Hayes and Cai 2007). Thus, we have a matrix with the estimated residual variance on the diagonal and all off-diagonal elements of the matrix set at zero. Furthermore, because homoscedasticity is assumed all diagonal elements on the matrix will take the same value. The estimated variance–covariance matrix of the regression coefficients, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}$, can then be written using matrix notation as:

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\phi} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \quad (6)$$

where the superscript ‘T’ denotes the transpose of a matrix, and the superscript $^{-1}$ denotes matrix inversion. The squares of the estimated standard errors for each of the regression coefficients are found on the main diagonal of $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}$, and the standard errors of each regression coefficient can be found by taking the square root of each diagonal element (Hayes and Cai 2007). However, when the error variance is heteroscedastic, this approach is unreliable.

In order to create a heteroscedasticity-consistent estimate of the variance–covariance matrix, White (1980) suggested putting the i th squared error into the i th row of the diagonal of the $\boldsymbol{\phi}$ matrix, using the residuals from an OLS regression as an estimate of error. Thus, the diagonal elements of the $\boldsymbol{\phi}$ matrix would be the squared residuals from the model in question and all off-diagonal elements of the matrix would be zero, this new matrix is denoted as $\widehat{\boldsymbol{\phi}}$ to represent the fact that the residuals are

only estimates of the statistical errors. This approach is often referred to as the HC0 estimator of the variance matrix and can be written as:

$$\text{HC0} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{diag}[\varepsilon_i^2] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \quad (7)$$

The entries on the diagonal of the HC0 matrix are estimates of the squared standard error of the regression coefficients (see Kauermann and Carroll 2001; Wilcox 2005). The HC0 estimator and other estimators in its family are commonly referred to as sandwich estimators. To see why, imagine that the $\mathbf{X}^T \text{diag}[\varepsilon_i^2] \mathbf{X}$ represents the filling of a sandwich (e.g., cheese and pickle) and that the $(\mathbf{X}^T \mathbf{X})^{-1}$ on either side of it in the equation represent two slices of bread. A worked example of the HC0 estimator is given in electronic supplementary materials so that readers can go through the matrix operations required step-by-step ESM.

The HC0 estimator is a consistent estimator when errors are heteroscedastic, but for small sample sizes the standard errors estimated are often biased downwards; although they improve as sample size increases. Since the initial work of White (1980), four alternative estimators, based upon the original HC0 estimator, have been developed HC1, HC2, HC3 and HC4, all involve modifications to the general HC0 estimator, but perform better when the sample size is small (Table 4). The general recommendation from the literature appears to be to use either HC3 or HC4 estimators. HC3 has been shown to perform well in most circumstances (Long and Ervin 2000; Cribari-Neto et al. 2005) and it is the default HC estimator in most R packages that use such estimators. However, the HC4 method performs better when there are certain points with very high LEVERAGES (Cribari-Neto 2004). Apart from the HC estimators mentioned above there are a variety of other estimators available and new estimators are still being developed (see Furno 1996; Cribari-Neto et al. 2000; Qian and Wang 2001; Cribari-Neto et al. 2007; Cribari-Neto and da Silva 2011).

Table 4 A list of the different HC sandwich estimates

HC Estimator	Formula	Details
OLS	$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\phi} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$	Homoscedasticity assumed
HC0	$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{diag}[\varepsilon_i^2] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$	Uses squared OLS residuals on the diagonal of the variance–covariance matrix
HC1	$\frac{n}{n-p-1} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{diag}[\varepsilon_i^2] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$	Uses a degree of freedom correction
HC2	$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{diag}\left[\frac{\varepsilon_i^2}{1-h_{ii}}\right] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$	h_{ii} 's are leverage values obtained as the diagonal elements from the ‘hat’ matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$
HC3	$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{diag}\left[\frac{\varepsilon_i^2}{(1-h_{ii})^2}\right] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$	Probably the most widely used HCSE estimator
HC4	$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{diag}\left[\frac{\varepsilon_i^2}{(1-h_{ii})^{\delta_i}}\right] \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$ where $\delta_i = \min\left\{4, \frac{nh_{ii}}{p+1}\right\}$	Designed to take points with large leverages into account. The parameter δ controls the level of discounting for the i th observation.

HCSE estimators are relatively easy to implement in R using any one of a number packages, including *car* (Fox and Weisberg 2010), *lme4* (Zeileis and Hothorn 2002) and *sandwich* (Zeileis 2004). HCSE estimators can be used to account for heteroscedasticity in both linear regressions and ANOVA, however, at present we know of no simple way of using them in combination with mixed models. In the following example, we will analyse another dataset from the Lundy house sparrow project using the R packages *lme4* and *sandwich*. HCSE estimators can also be used in the packages SAS and SPSS (see Hayes and Cai 2007 for the code) and in STATA (using the robust variance estimators). The analysis is concerned with modelling the size of the black throat patch of male house sparrows as a function of four predictors, age, mass, wing length, and the size of the brood in which the focal individual hatched. The black throat patch of male house sparrows is an example of a ‘badge of status’, and has a potential role in both social signalling and sexual selection (Nakagawa et al. 2007b). A residual scatterplot of the dataset suggested there was evidence of heteroscedasticity, with the variance in model residuals increasing with higher predicted badge size values. The Breusch–Pagan test confirmed the presence of heteroscedasticity in the data ($\chi^2=23.37$, $df=4$, $p<0.001$). The dataset contains repeated measures of some individuals, which violates the assumption that our data points are fully independent. However, we will ignore this in the following example because the aim is simply to demonstrate how using HCSE estimators can influence model standard errors. The coefficient estimates from a standard OLS regression that assumes homoscedasticity are compared with the results obtained using both an HC3 and an HC4 estimator (Table 5). What is clear from Table 5 is that the standard errors for the OLS regression are quite different from those calculated using either the HC3 or HC4 estimator. Moreover, if we based our inferences solely on p values then body mass would be identified as a statistically significant predictor of badge size when using the OLS model, but not when using either of the HC estimators. Of course, this also highlights why using p

values as a basis for statistical inference is questionable at best (Cohen 1990; Nakagawa and Cuthill 2007). Unlike with the previous GLS models, using an HCSE estimator has no effect upon model residuals, thus they will still appear heteroscedastic when plotted. Incidentally, it should also be pointed out that HCSE estimators can also be used for ANOVA tests as well as linear regression. Long and Ervin (2000) recommended that investigators should employ a HCSE estimator as a matter of course and the HC estimators introduced here provide a simple way for researchers in behavioural ecology to check that their results are not being influenced by heteroscedasticity. We should make it clear that HCSE estimators are not a magical cure for all the problems that heteroscedasticity can create. In general, estimated standard errors tend to be larger when using HCSE estimators than they do when using a GLS approach in which the form of the heteroscedasticity is correctly specified. However, if we are not confident that we can model heteroscedasticity correctly HCSE estimators provide an alternative that make less assumptions about the errors in our data.

Conclusions

Within the behavioural ecology literature heteroscedasticity is rarely mentioned. However, the assumption that data are homoscedastic is not made on biological grounds. Given the complexity we see in nature, there are a variety of situations in which heteroscedasticity could be expected. Moreover, careful consideration of the variance present in data could prove useful. An increasing focus upon variance will naturally lead behavioural ecologists to consider whether their data are heteroscedastic in more detail. In particular, we identify animal personality studies as one area in which heteroscedasticity could have an interesting biological interpretation. In addition, not accounting for heteroscedasticity when it is present will make the standard errors and the p values from a regression model unreliable. Simple data transformations may solve the problem of

Table 5 Modelling male badge size in the house sparrow using different standard error estimators ($n=60$ adults)

	β	Standard error estimator					
		OLS		HC3		HC4	
		SE	p	SE	p	SE	p
Intercept	28.311	16.404	0.086	17.616	0.108	17.767	0.111
Age	1.893	0.394	<0.001	0.449	<0.001	0.464	<0.001
Body mass	0.351	0.176	0.048	0.197	0.075	0.197	0.075
Wing length	−0.014	0.226	0.950	0.234	0.952	0.237	0.952
Brood size	−0.366	0.338	0.280	0.421	0.385	0.425	0.389

heteroscedasticity in some cases and we are not seeking to criticize this approach, but there will be situations when it is not the best option. Here, we introduced two simple methods of dealing with heteroscedasticity, GLS regression and HCSE estimators, which we hope will prove useful to behavioural ecologists the next time they encounter unusual patterns in their residuals. Beyond GLS and HCSE there are a variety of more complex techniques such as Bayesian or hierarchical GLMs (Gelman and Hill 2007; Rönnegård et al. 2010). For more simple analyses, the robust estimation methods covered by Wilcox (2005) may also prove useful. We emphasize that this article is designed to provide an introduction to some of the various techniques available for dealing with heteroscedasticity; and we hope that individual researchers will explore further. To end, we reiterate the point made earlier; heteroscedasticity should not just be considered as a nuisance when performing an analysis, instead it may be able tell us something about the biological process being studied, which should be of interest to all behavioural ecologists.

Acknowledgements We thank Barbara Morrissey, Eduardo Santos and Alistair Senior for commenting upon earlier versions of this manuscript. We are grateful to three anonymous reviewers for comments which improved the manuscript. We would also like to thank Terry Burke for his encouragement during the writing of this manuscript. S.N. is supported by the Marsden Fund.

References

- Badyaev AV (2002) Growing apart: an ontogenetic perspective on the evolution of sexual size dimorphism. *Trends Ecol Evol* 17:369–378
- Bivand RS, Pebesma EJ, Gómez-Rubio V (2008) *Applied spatial data analysis with R*. Springer, New York
- Breusch TS, Pagan AR (1979) A simple test for heteroskedasticity and random coefficient variation. *Econometrica* 47:1287–1294
- Brotherstone S, Hill WG (1986) Heterogeneity of variance amongst herds for milk production. *Anim Prod* 42:297
- Cardoso FF, Rosa GJM, Tempelman RJ (2005) Multiple-breed genetic inference using heavy-tailed structural models for heterogeneous residual variances. *J Anim Sci* 83:1766–1779
- Carroll RJ (2003) Variances are not always nuisance parameters. *Biometrics* 59:211–220
- Clayton GA, Morris JA, Robertson A (1957) An experimental check on quantitative genetical theory: I. Short term responses to selection. *J Genet* 55:131–151
- Cleasby IR (2010) The influence of early environment and parental care on offspring growth and survival in the house sparrow. PhD thesis, University of Sheffield, UK
- Cohen J (1990) Things I have learned (so far). *Am Psychol* 45:1304–1312
- Congdon PD (2010) *Applied Bayesian hierarchical methods*. CRC Press, Florida
- Cribari-Neto F (2004) Asymptotic inference under heteroskedasticity of unknown form. *Comput Stats Data Anal* 45:215–233
- Cribari-Neto F, da Silva WB (2011) A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model. *Adv Stat Anal* 95:129–146
- Cribari-Neto F, Ferrari SLP, Cordeiro GM (2000) Improved heteroscedasticity-consistent covariance matrix estimators. *Biometrika* 87:907–918
- Cribari-Neto F, Ferrari SLP, Oliveira WASC (2005) Numerical evaluation of tests based on different heteroskedasticity consistent covariance matrix estimators. *J Stat Comput Simul* 75:611–628
- Cribari-Neto F, Souza TC, Vasconcellos KLP (2007) Inference under heteroskedasticity and leveraged data. *Commun Stat Theory Methods* 36:1877–1888, Errata 37:3329–3330, 2008
- Cryer JD, Chan K (2008) *Time-series analysis with applications in R*, 2nd edn. Springer, New York
- Darlington RB (1990) *Regression and linear models*. McGraw-Hill, New York
- Dingemanse NJ, Kazem AJN, Réale D, Wright J (2010) Behavioural reaction norms: animal personality meets individual plasticity. *Trends Ecol Evol* 25:81–89
- Dutilleul P, Potvin C (1995) Among-environment heteroscedasticity and genetic autocorrelation: implications for the study of phenotypic plasticity. *Genetics* 139:1815–1829
- Erceg-Hum DM, Mirosevich VM (2008) An easy way to maximise the accuracy and power of your research. *Am Psychol* 63:591–601
- Fox J, Weisberg S (2010) *An R companion to applied regression*. Sage, California
- Furno M (1996) Small sample behaviour of a robust heteroskedasticity consistent covariance matrix estimator with improved finite-sample properties. *J Stat Comput Simul* 54:115–128
- Gelman A (2005) Analysis of variance – why it is more important than ever. *Ann Stat* 33:1–53
- Gelman A, Hill J (2007) *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, New York
- Goldfeld SM, Quandt RE (1965) Some tests for homoscedasticity. *J Am Stat Assoc* 60:539–547
- Griffiths SC, Owens IPF, Burke T (1999) Environmental determination of a sexually selected trait. *Nature* 400:358–360
- Grisson RJ, Kim JJ (2005) Effect sizes for research: A broad practical approach. Erlbaum, New Jersey
- Hadfield JD (2010) MCMC methods for multi-response Generalised Linear Mixed Models: the MCMCglmm R package. *J Stat Soft* 33:1–22
- Hayes AF, Cai L (2007) Using heteroskedasticity-consistent standard error estimators in OLS regression: an introduction and software implementation. *Behav Res Methods* 39:709–722
- Herberich E, Sikorski J, Hothorn T (2010) A robust procedure for comparing multiple means under heteroskedasticity in unbalanced designs. *PLoS ONE* 5(3):e9788. doi:10.1371/journal.pone.0009788
- Hill WG (1984) On selection among groups with heterogeneous variance. *Anim Prod* 39:473–477
- Hill WG, Zhang X (2004) Effects on phenotypic variability of directional selection arising through genetic differences in residual variability. *Genet Res* 83:121–132
- Honkanen T, Jormalainen V (2005) Genotypic variation in tolerance and resistance to fouling in the brown Alga *Fucus vesiculosus*. *Ecology* 144:196–205
- Jones KS, Nakagawa S, Sheldon BC (2009) Environmental sensitivity in relation to size and sex in birds: meta-regression analysis. *Am Nat* 174:122–133
- Kauermann G, Carroll RJ (2001) A note on the efficiency of sandwich covariance matrix estimation. *J Am Stat Assoc* 96:1387–1396
- Keppel G, Wickens TD (2004) *Design and analysis: A researcher's handbook*, 4th edn. Pearson, New Jersey
- Long JS, Ervin LH (2000) Using heteroskedasticity consistent standard errors in the linear regression model. *Am Stat* 54:217–224
- Mardia KV, Kent JT, Bibby JM (1979) *Multivariate analysis*. Academic Press, New York

- Nakagawa S, Cuthill IC (2007) Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev* 82:591–605
- Nakagawa S, Gillespie DOS, Hatchwell BJ, Burke T (2007a) Predictable males and unpredictable females: sex difference in repeatability of care in a wild bird population. *J Evol Biol* 20:1674–1681
- Nakagawa S, Ockendon N, Gillespie DOS, Hatchwell BJ, Burke T (2007b) Assessing the function of house sparrows' bib size using a flexible meta-analysis method. *Behav Ecol* 18:831–840
- Pinheiro JC, Bates DM (2000) Mixed-effects models in S and S-PLUS. Springer, New York
- Pinheiro JC, Bates DM, DebRoy S, Sarkar D (2010) nlme: linear and nonlinear mixed effects models. R package version 3:1–97
- Qian L, Wang S (2001) Bias-corrected heteroscedasticity robust covariance matrix (sandwich) estimators. *J Stat Comput Simul* 70:161–174
- R Development Core Team (2010) R: A language and environment for statistical computing, version 2.12.1. R Foundation for Statistical Computing, Vienna, Austria
- Reale D, Dingemanse NJ (2010) Personality and individual social specialisation. In Szekely T, et al (eds) *Social behaviour: Genes, ecology and evolution*. Cambridge University Press
- Rönnegård L, Felleki M, Fikse F, Mulder HA, Strandberg E (2010) Genetic heterogeneity of residual variance—estimation of variance components using double hierarchical generalized. linear models. *Genet Selection Evol* 42:8
- Ruxton GD (2006) The unequal variance *t*-test is an underused alternative to Student's *t*-test and the Mann–Whitney *U* test. *Behav Ecol* 17:688–690
- Schwagmeyer PL, Mock DW (2003) How consistently are good parents good parents? Repeatability of parental care in the house sparrow, *Passer domesticus*? *Ethology* 109:303–313
- Sinn DL, Gosling SD, Moltschanivskyj NA (2008) Development of shy/bold behaviour in squid: context-specific phenotypes associated with developmental plasticity. *Anim Behav* 75:433–442
- Teder T, Tammara T, Esperk T (2008) Dependence of phenotypic variance in body size on environmental quality. *Am Nat* 172:223–232
- Venables WN (2000) Exegeses on linear models. Paper presented to S-PLUS User's Conference, Washington, DC, 8–9 October, 1998
- White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48:817–838
- Wilcox RR (1998) How many discoveries have been lost by ignoring modern statistical methods? *Am Psychol* 53:300–314
- Wilcox RR (2005) Introduction to robust estimation and hypothesis testing. Academic Press, New York
- Wilkin TA, King LE, Sheldon BC (2009) Habitat quality, nestling diet, and provisioning behaviour in great tits *Parus major*. *J Avian Biol* 40:135–145
- Wooldridge JM (2000) Introductory econometrics: a modern approach. South-Western College Publishing, Ohio
- Zeileis A (2004) Econometric computing with HC and HAC covariance matrix estimators. *J Stat Software* 11:1–17
- Zeileis A, Hothorn T (2002) Diagnostic checking in regression relationships. *R News* 2:7–10
- Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM (2009) Mixed effects models and extensions in ecology with R. Springer, New York