CHAPTER

4

Linear regression

Thomas M.H. Hope

Wellcome Centre for Human Neuroimaging, University College London, London, United Kingdom

4.1 Introduction

Linear regression analysis is probably the simplest and certainly the most common way to measure the relationships between continuous variables. Like most statistical methods, it is best understood through practical examples. Imagine that we sit on the admissions board of a selective school. Our job is to sift the applications and decide who to accept. Our aim is to find those students who will, in the fullness of time, get a PhD from a top US university. Most of these universities (and some outside the United States too) require applicants to take Graduate Record Examinations (GRE) as part of their application. The GRE include tests of verbal reasoning, quantitative reasoning, and analytical writing: they aim to be tests of intelligence, rather than of specific knowledge. So our aim is to find those students who will, years from now, score well on an intelligence test: i.e., whose intelligence quotient (IQ) will be high. One tempting approach to this problem is to pick students who already have high IQ, our hypothesis being that those with high IQ now (or recently) will have high IQ later as well. To test this hypothesis, we look at IQ data from past students—both as measured before they applied to our school and as measured years later. Our question is how well did those earlier IQ scores predict the later IQ scores?

One way to answer this question is with a scatter plot. Each student contributes one data point to the plot, which relates their earlier IQ to their later IQ, as in Fig. 4.1. If the data points all line up neatly (Fig. 4.1A), the implication is that students with higher IQ earlier also had higher IQ later, and vice versa. If there is no real relationship, the graph should look much messier (Fig. 4.1B). Regression models let us turn these visual impressions into numbers, often called effect sizes. The process fits a straight line, or linear function, to the data: the effect size is stronger when the line passes

https://doi.org/10.1016/B978-0-12-815739-8.00004-3

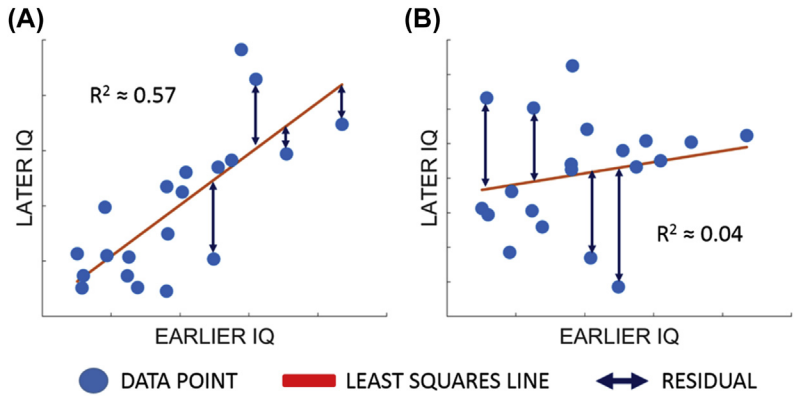

FIGURE 4.1 Scatter plots of earlier intelligence quotient (IQ) versus later IQ. (A) An example in which the relationship is relatively strong; earlier IQ explains around 57% of the variance in later IQ. (B) An example in which the relationship is weak; earlier IQ explains around 4% of the variance in later IQ.

close to the data points (Fig. 4.1A) and weaker when the data points are further away (Fig. 4.1B).

Linear regression tells us how much of the variability in the students' later IQs is explained by variability in their earlier IQs. If the number is large (i.e., close to 1), the implication is that we can use new applicants' earlier IQs to predict those same applicants' later IQs. If the number is small (i.e., close to zero), this suggests that earlier IQs are not a good predictor of later IQs. In this sense, regression models can be thought of simply as useful, practical decision-making tools.

We can also do real science with results like this because different theories of intelligence make different predictions about how strong the relationship should be. One simple way to frame this issue is as a "nature versus nurture" debate. On the nature side are those who think of 'intelligence' as a fundamental and broadly constant feature of people; like their shoe size and height, this may change as they grow during childhood, but it is still mainly fixed by their genes. On the nurture side are those who believe that a person's home and school environment drive real change in their intelligence, regardless of (or despite) any genetic influence.

This debate has real, practical implications and is still hotly contested. For example, a recent, large-scale study supported the "nature" theory of intelligence by showing that differences in exam results between selective and nonselective schools were almost entirely explained by genetic and socioeconomic differences between their students (Smith-Woolley et al., 2018). Other recent work supports the "nurture" theory by showing that the variability in later IQs that is *not* explained by earlier IQs is itself predicted by structural change in brain anatomy during adolescence (Price, Ramsden, Hope, Friston, and Seghier, 2013; Ramsden et al., 2011).

Notably, those predictions, of IQ change given brain structure change, were made with a voxel-based morphometry analysis, which is essentially a collection of linear regressions.

Regression models are often employed in much the same way—for both practical decision-making and more theoretical or scientific enquiry—when applied both to medicine in general and to brain disorders in particular. More complex methods, many of which are introduced in later chapters of this book, might produce more accurate predictions, but this is often at the expense of transparency: i.e., we cannot easily understand why those more complex models are making any particular prediction. As a consequence, the explanation of the inner workings of the method and the phenomenon of interest is often at least as prominent as prediction in research employing linear regression; while in research employing more complex methods, prediction is emphasized over explanation. For this reason, among others, linear regression models are still ubiquitous in statistical medicine, and there is every reason to believe that this will continue.

4.2 Method description

4.2.1 Simple regression

Simple regression refers to regression analyses relating a single dependent or predictor variable to a single independent or response variable. In these cases, a linear regression analysis fits straight lines of the following form:

$$y = C + \beta x + \varepsilon$$

Or in other words:

$$\begin{aligned}[\text{response } (y)] \;=\; &[\text{a constant number } (C)] \\ &+ \;([\text{predictor}] \times [\text{another number } (\beta)]) \;+\; \text{noise } (\varepsilon)\end{aligned}$$

The constant number (C) is usually called the "intercept"; it defines the value of the response variable when the predictor is set to zero. The multiplier (β) is often called a "coefficient" (and sometimes a slope coefficient or a weight); it defines the ratio of change in the response to change in the predictor. In the previous example, the predictor was "earlier IQ" and the response was "later IQ." If a linear regression analysis assigned $\beta = 2$ in that case, it means that improvement of 1 point of earlier IQ corresponds to an improvement of 2 points of later IQ. Conversely, if $\beta = 0.5$, this means that an improvement of 2 points of earlier IQ corresponds to 1 point of later IQ.

Given a set of data—like pairs of IQ test results at different times—linear regression finds C and β by minimizing the squared distances between the line and the data points. In principle, one could solve this minimization problem in many different ways, but the "ordinary least squares" (OLS) method is by far the most common: so common, in fact, that practitioners might use linear regression throughout their careers and never employ any other fitting method. This is partly because the OLS method is quick, but mainly because, under a few assumptions, OLS is guaranteed to find the best possible estimates of C and β: i.e., the estimates which really do shrink the resulting squared distances between the data and the line as far as possible.

In addition to parameter estimates, regression lets us quantify the errors or residuals of the analysis: i.e., those distances between the lines and the data in Fig. 4.1A and B. These residuals encode the variability in the response variable that is *not* explained by the predictor: i.e., linear regression is a simple way to control for covariates of no interest, by re-encoding our data as the residuals after regressing those data against the covariates. This trick has definite limitations (Miller and Chapman, 2001), but it is nevertheless ubiquitous in neuroscience and in neuroimaging data analysis in particular.

4.2.2 "General" linear regression

One potentially important limitation of simple regression is the assumption that the relationship between our predictor and response variables is itself linear: i.e., a constant unit of change in the predictor is associated with a constant unit of change in the response. If this assumption is wrong, as in Fig. 4.2A, then linear regression will

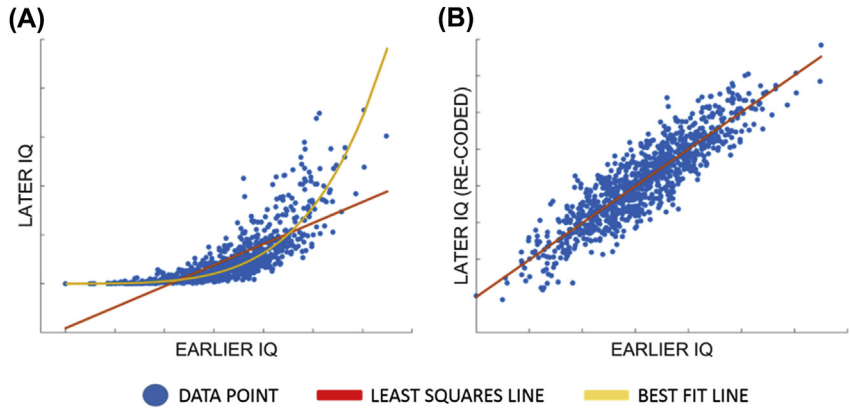

FIGURE 4.2 Linear regression with nonlinear variable transformations. (A) A nonlinear relationship, whose strength is underestimated by a linear least squares fit; (B) the same data recoded, which is much more suitable for linear regression.

underestimate the strength of the relationship. In practice, we often do not know the real relationships between our variables, and in that context, this assumption of linearity may seem too restrictive.

Notably, the linear model in Fig. 4.2A still explains a significant amount of the variance in the response variable. It might underestimate the strength of the relationship, but it still correctly identifies that there is one. This situation turns out to be quite common in practice: linear regression is usually a reasonable first step, even when we do not believe that the underlying relationships, between our variables, are really linear. However, linear regression can also accommodate known or expected nonlinearity with relative ease by working on variables which have been transformed before the regression analysis is run. In Fig. 4.2A, y is simply the fifth power of x plus noise. If we plot y against x^5, the curve in the data disappears, and we recover a clear linear relationship (Fig. 4.2B), which is now much more suitable for analysis with linear regression.

This kind of transformation is a simple example of a "kernel trick" (Ben-Hur, Ong, Sonnenburg, Scholkopf, and Ratsch, 2008): a way to transform our data so that it conforms to the assumptions made by our method (a linear regression model in this case). Many nonlinear regression and classification models—which later chapters will explore in more detail—employ essentially the same logic, albeit with more complex transformations. The point of this example is simply to make the case that, in principle at least, linear regression can be generalized to accommodate almost any type of nonlinear relationship.

4.2.3 Multiple regression

Linear regression also generalizes to multiple predictors, with relative ease. The form of the function that we aim to estimate in this case, for "n" predictors, is as follows:

$$y = C + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n + \varepsilon$$

Or in other words, the response variable is calculated as the intercept plus the sum of all of the predictor variables, each multiplied by their own coefficient plus noise. Multiple regression is incredibly powerful because it lets us explore how the combination of multiple predictors might explain a response variable and because its results are comparatively easy to interpret. And the OLS method works almost equally well in this case too.

Continuing our earlier example, imagine that we want to try to decrease the error with which we can predict "later IQ," by considering both "earlier IQ" and a second predictor variable "school ranking": e.g., the rank held by the student's school in the latest league table. This kind of

test is easy to run with multiple regression, with "earlier IQ" and "school ranking" as predictors and "later IQ" as the response. Indeed, this is effectively what was done in the recent report, mentioned previously, that selective schools add little value that is not already explained by genetics (Smith-Woolley et al., 2018). In the terms of our hypothetical example, Smith-Woolley and colleagues' result predicts that the addition of "school ranking" will improve our model of "later IQ" (i.e., increase its R^2) only slightly. Note that a multiple regression with earlier IQ and school ranking as predictors is effectively equivalent to a simple regression with school ranking as a single predictor, and the response variable calculated as the residuals of later IQ after regression against earlier IQ.

4.2.4 Limitations: the curse(s) of dimensionality

Multiple regression has an important limitation: it does not scale particularly well as the number of predictors grows. One example of this is the problem of "multicollinearity," which emerges when two predictors are strongly correlated or when one predictor is a simple combination (e.g., an arithmetic sum or product) of others. Building on the IQ prediction problem discussed in the Introduction, imagine we added "earlier IQ" scores into our regression model twice instead of once. In this case, we would have two identical predictors, and there would be at least two equally optimal parameter estimates to choose from: i.e., fitting with either one of the two coefficients alone while setting the other coefficient to zero. In cases like this, where at least some of our predictors are redundant, there is simply no unique best regression model to find.

Multicollinearity tends to be more of a problem in practice as the number of predictors grows. This might be because we have inadvertently measured two or more things which have the same underlying cause (such as resting heart rate and blood pressure). But strong associations between predictors can be observed by chance as well, if we have enough of them to consider. To illustrate the point, we can simply define 100 series of 10 numbers drawn at random from a standard normal distribution. This models a situation in which we take 100 measurements from each of 10 people. Using linear regression to test the pairwise associations for every pair of predictors, the largest effect size is 0.86. With 1000 predictors, this rises to 0.98. This is not just an abstract problem: studies in the biomedical sciences often do encounter situations like this. For example, structural magnetic resonance imaging (MRI) data are typically represented in 2 mm^3 volumetric pixels (voxels), and at this resolution, the standard brain contains 352,328 voxels. And genetic studies make the problem even worse, with millions of predictors available for analysis.

Multicollinearity is an example of a "curse of dimensionality," a phrase originally coined by Richard Bellman in 1957 (Bellman, 1957), which refers

to a collection of problems which emerge when analyzing high-dimensional data. These problems are general, rather than specific to linear regression, and as such we will not devote much more of this chapter to them. Nevertheless, one more of these problems is particularly pertinent to the study of brain disorders and so does need to be addressed here.

4.2.5 Prediction versus explanation

In the previous discussion of the relationship between 'earlier IQ" and "later IQ," we glossed over the difference between prediction and explanation, assuming that a strong relationship between the two (i.e., effect size close to 1) implies that later IQ can be predicted from earlier IQ. In simple regression, with just one predictor to consider, this is at least approximately true—but it is not true at all when we are dealing with many predictors. To illustrate the point, we can define X as a set of 100 predictor variables, each drawn at random from the normal distribution (i.e., mean = 0, standard deviation = 1). Our y, or response variable, is defined as the first predictor plus another random normal variable: i.e., the first predictor explains ~50% of the variance in the response variable, and all of the other predictors are irrelevant. We consider a sample size of 200, so each X and y variable is a vector of 200 numbers. Next, we run a series of linear regression and cross-validation analyses (with linear regression implemented in each iteration), starting by considering just the first predictor, then considering the first two predictors, then the first three, and so on until we consider all 100 predictors together. The regression analyses are called "in-sample" analyses because the whole sample is used to calculate the model coefficients. By contrast, the cross-validation analyses are "out-of-sample" analyses because we are repeatedly (1) partitioning the data into training and test samples; (2) calculating regression models based on the training sample only; and (3) using those models to predict y for the test sample (see Chapter 2). Fig. 4.3 displays the results.

When we only consider the first predictor, we can predict (out-of-sample) roughly the same amount of the variance in the response that we can explain (in-sample). However, the two types of analysis behave very differently as we add more irrelevant predictors. The in-sample effect sizes increase because of accidental correlations between each predictor and the response. However, because those correlations are accidental, they can also mislead if used to make predictions; this is why the out-of-sample effect sizes decrease. After adding enough predictors, it becomes impossible to predict the response variable at all. Indeed, the in-sample analysis is just as unreliable a way to measure predictive power because coefficients assigned to each predictor bear no relationship at all to their real relevance to the response. The failure of linear regression

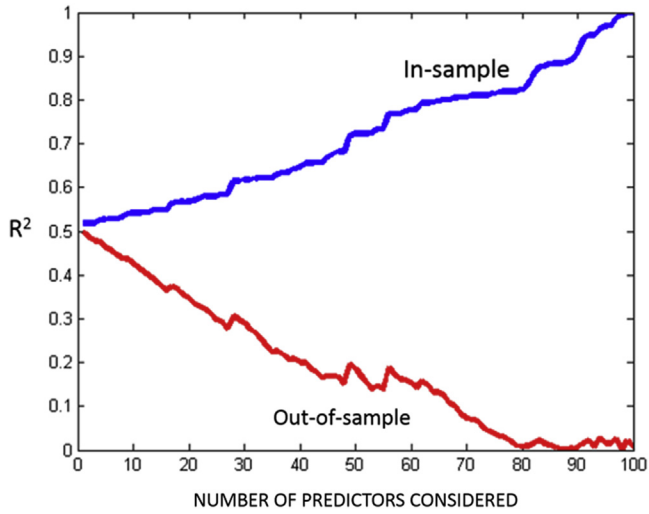

FIGURE 4.3 In-sample versus out-of-sample effect sizes. The variance explained (R^2) by predicted responses, of empirical responses, by linear regression analyses conducting with incrementally larger sets of (irrelevant) predictors. Predictions are calculated either in-sample, by regression with all observations included, or out-of-sample, by regression embedded within a leave-one-out cross-validation.

models, with too many predictors, is an important motivation for many more of the more complex methods that will be considered in later chapters of this book.

4.3 Applications to brain disorders

As mentioned previously, linear regression is ubiquitous in modern science: the study of brain disorders is no exception. In a sense, this is trivially true because regression models are part of the foundation of modern methods for analyzing neuroimaging data, such as controlling for covariates of no interest, as discussed in Section 2.1. In this section, we consider some more substantive applications of regression models such as multiple sclerosis (MS) (Section 3.1), Alzheimer's disease (AD) (Section 3.2), and aphasia (Section 3.3).

4.3.1 Predicting disease progression in multiple sclerosis

MS is a condition affecting the brain and/or spinal cord, which can cause a variety of symptoms, such as impaired vision, sensation, balance, and arm or leg movement. There are thought to be more than 100,000 people with MS in the United Kingdom alone, but the course of the

disease is very variable across patients, with some experiencing relatively minor impairments, while others suffer serious disability. That variability has encouraged the search for brain biomarkers of MS onset and progression, and many of these studies employ regression models.

In one recent study (Tedeschi et al., 2005), researchers tested brain atrophy as a potential biomarker, using an automated procedure to segment 597 MS patients' structural MRI into gray and white matter and cerebrospinal fluid. To account for differences in total head size, the authors defined WM-f and GM-f as the proportions of the total intracranial volume (ICV) that was occupied by white matter and gray matter, respectively. They then employed forward stepwise feature selection to select a multiple regression model to explain patients' levels of disability, as measured using the Expanded Disability Status Scale (EDSS). The EDSS is a standard tool in the MS literature, which rates patients' levels of disability on a 20-point scale. Features available for selection included gender, years of school attended, age at disease onset, disease duration, MS course, MS onset, and therapy, as well as the brain volume measures (WM-f and GM-f). The resulting model included GM-f, as well as MS course, disease duration, age at disease onset, number of years of schooling, and therapy. The coefficient for GM-f in that model was significant ($P < 0.001$); even after controlling for the other factors in that model, this result suggests that gray matter atrophy tracks disease progression in MS.

These results are consistent with those of another study, which also used an automated segmentation procedure (although a different one) to quantify gray and white matter volume, among other measures, in 41 MS patients (Sanfilipo, Benedict, Sharma, Weinstock-Guttman, and Bakshi, 2005). The authors added these variables to a larger, multivariable regression model, which included ICV and age as covariates of no interest: i.e., effectively regressing out the influence of those covariates, rather than dividing volumes by ICV, as in the previous study. Within this model, gray matter volume was the only MRI variable, whose addition explained significant extra variance in EDSS scores.

Another study measured the relevance of brain atrophy (including the expansion of lesions characteristic of MS onset) longitudinally, calculating their measure as a rate of change from structural MRI scans, taken at diagnosis and 1−2 years later, and relating that atrophy to the same patients' EDSS scores 10 years after diagnosis (Popescu et al., 2013). Their final model, identified via a series of forward and backward stepwise feature selection steps, included their brain atrophy measure as a significant predictor of MS severity at 10 years.

In summary, all three studies use similar methods (linear regression) and come to similar conclusions (brain atrophy tracks MS progression). However, each study calculated brain atrophy differently, and none

attempted out-of-sample analyses. Indeed, even in in-sample analysis, the longitudinal contribution of MRI variables to 10-year EDSS was small (R^2 change = 3%). For these reasons, more work is needed to show that brain atrophy is a clinically significant biomarker of disease progression in MS.

4.3.2 Predicting cognitive decline in Alzheimer's disease

AD typically manifests as a memory impairment, which progresses over time into full-blown dementia. As in MS, many studies of AD use regression to relate brain biomarkers to disease onset and progression. As memory dysfunction is often the first overt behavioral symptom of AD, research on brain biomarkers often emphasizes the hippocampus—a region located near the center of the brain, which is implicated in memory processing.

One cross-sectional study employed structural MRI to measure hippocampal volumes in 18 patients with probable AD (Deweer et al., 1995). Volumes were measured using a semiautomated procedure in which software calculated the volumes of regions based on coordinates specified by hand on the MRI images. Hippocampal volume was divided by ICV, as in Tedeschi et al. (2005). Additionally, all of the patients underwent a comprehensive battery of tests of cognitive function; the authors used simple regression to relate hippocampal volume to memory scores on those tasks (smaller volume = worse performance). This is a small cross-sectional study, but the results suggest that hippocampal atrophy may be useful to track the progression of AD.

As in MS, studies of AD have also considered whole-brain atrophy as a potential biomarker of disease progression. Supporting evidence was recently reported in a longitudinal study, pooling 47 patients with AD, 29 patients with mild cognitive impairment (MCI), and 23 controls (Sluimer et al., 2010). All participants were scanned twice with structural MRI, and whole brain volumes calculated using a semiautomated process, in which human judgment augmented the results of a standard algorithm. Differences between brain volumes at each time point were converted to a percentage change, which was divided by the time between the scans to produce a rate of change. Contemporaneously with the scans, the participants were also assessed with the Mini-Mental State Examination (MMSE), a standard assessment of cognitive function, which is sensitive to the deterioration typically observed in AD. The authors' whole-brain atrophy measure was significantly correlated with change in the MMSE scores: i.e., whole-brain atrophy tracks increasing symptom severity over time in the transition from MCI to AD.

Finally, some studies have proposed more overtly predictive models of cognitive decline in AD and other types of dementia. One recent study (Watanabe et al., 2016) tested patients with dementia at diagnosis and then again at 3, 9, and 12 months later, using a series of standard

assessment tools for dementia severity, including the MMSE. The authors used linear regression models, trained on the first two scores (at diagnosis and 3 months later) to predict the same scores at 9 and 12 months. One model was a simple regression with time as the predictor and the second employed log-transformed time: i.e., the authors used the variable re-encoding trick, described in Section 2.2. Their logarithmic model explained more of the variance in empirical MSSE scores at both 9 months ($R^2 = 0.30$ (linear) vs. 0.44 (logarithmic)) and 12 months ($R^2 = 0.34$ (linear) vs. 0.54 (logarithmic)). This small study offers preliminary evidence that dementia symptom progression is logarithmic in time postdiagnosis.

4.3.3 Predicting language impairment (aphasia) after stroke

Some of the earliest neuroscience studies were focused on aphasia, drawing inferences about the normal roles of particular brain regions in language, by studying the language impairments which followed from damage to those regions (Broca, 1861; Wernicke, 1874). Today, this work draws on large samples of stroke survivors, whose brain damage is characterized via structural neuroimaging (typically MRI) and whose language impairments are characterized via standardized batteries of language tests.

One recent example of this work employs multiple linear regression (with stepwise feature selection) to relate lesions to two response variables: the first two principle components of a large dataset of language task scores, which the authors tentatively identified with "speech production" and "speech comprehension," respectively (Fridriksson et al., 2018). The authors encoded lesions in two ways. In the first, lesions were drawn directly onto structural MRI slices by a trained neurologist. The resulting "lesion images" were then encoded as the proportion they appeared to destroy a series of anatomically defined brain regions ("regional lesion load"). In the second approach, lesions were encoded by the extent to which they interrupted the white matter tracts between those same anatomically defined regions ("tract disconnection"). The authors then used both simple and multiple regression models (the latter with forward stepwise feature selection) to identify the regional lesion load and tract disconnection variables that could best explain each response variable. The main results are illustrated in Fig. 4.4.

Fridriksson and colleagues were more concerned with analysis than prediction; in another recent study, linear regression was used to predict (Hope et al., 2015). This study focused on bilingualism, which is the norm in most parts of the world but remains poorly understood. One current debate in this field is over the extent to which second language learning changes language networks in the brain; examples of consistent processing support "neural convergence" theories (Green, 2003), while examples of

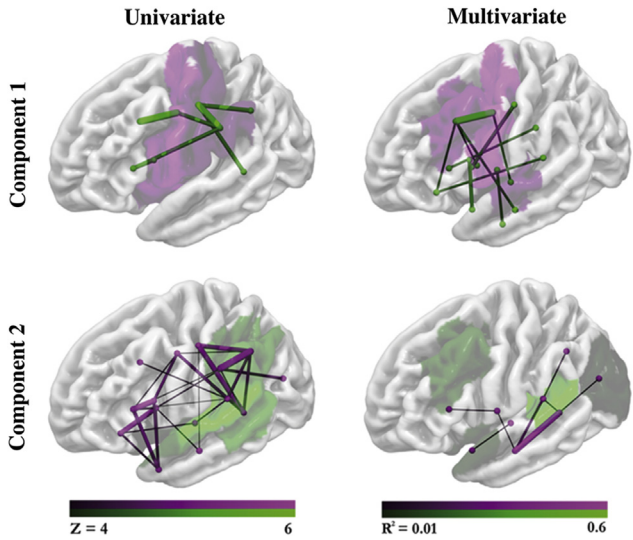

FIGURE 4.4 Lesion symptom maps for language, derived from linear regression. Simple (left columns) and multiple (right column) regression analyses of regional lesion load and connectivity disruption data, for two principle components of language task data interpreted as referring to speech production (component 1; top row) and speech comprehension (component 2; bottom row) principal component analysis (see Chapter 12) of speech comprehension and speech production tests/tasks. *Taken from Fridriksson, J., den Ouden, D.-B., Hillis, A. E., Hickok, G., Rorden, C., Basilakos, A., Yourganov, G., & Bonilha, L. (2018). Anatomy of aphasia revisited.* Brain, 141*(3), 848–862.*

differences support the "neural divergence" position (Ullman, 2001). If second language learning significantly changes language networks, models trained with monolingual patient data will not generalize to bilingual patients—and indeed models trained for Spanish–English bilinguals might not generalize to Spanish–Portuguese bilinguals. Neural divergence is bad news for prognostic research in poststroke aphasia. This was tested using multiple linear regression, with regional lesion load predictors, to learn lesion symptom associations for 22 separate language scores in 174 monolingual stroke survivors. Leave-one-out cross-validation was used to derive predicted scores for the monolingual patients, as well as used the whole monolingual sample to learn further models, predicting each language score in a separate set of 33 bilingual stroke survivors. The relationships between predicted and empirical scores were just as strong across both groups: i.e., models trained on monolingual patient data could generalize to bilingual patients. This result suggests that both patient groups share substantially similar language networks, consistent with neural convergence.

Finally, a series of recent studies have reported prognostic results that do not involve any brain data at all. These studies measure symptom

severity both acutely (within a week or so after stroke) and again around 3–6 months later and then regress baseline scores (the predictor) against recovery (outcome scores minus baseline scores). The two studies like this in language report R^2 values of 0.81 (Lazar et al., 2010) and 0.94 (Marchi, Ptak, Di Pietro, Schnider, and Guggisberg, 2017), respectively, and similar studies of recovery from other poststroke impairments report similarly strong results. None of these studies analyzed out-of-sample analyses, but they do not need to be; as in Fig. 4.3, out-of-sample and in-sample results converge when we use just one (or few) predictors. Nevertheless, it turns out that these results are fallacious because they are confounded by ceiling effects on the scales used to measure poststroke impairment (Hope et al., 2019). This is a cautionary tale: even simple regression can be misleading, and even experienced researchers can be misled.

4.4 Conclusions

Linear regression models are simple but incredibly powerful; every introduction to machine learning should start here. The key principle of this method is that the impact of each predictor variable on the response variable can be specified with just a single number, which represents the ratio of change in the predictor to change in the response. This level of simplicity is possible because of the assumption that all influences are linear: more complex functions need more numbers to describe them, and the more parameters one uses, the greater the chance that we will overfit our models to noise. Linear models are therefore usually the best choice—and sometimes the only sensible choice—when studying smaller samples. However, this does not mean that we have to assume our predictor–response relationships are themselves linear (see Section 5.2.2).

That said, linear regression models certainly have limitations; indeed, some of their more complex counterparts are expressly designed in response to those limitations. For example, the "curse of dimensionality," described in Section 5.2.5, can be addressed by forcing most of the coefficients of a multiple regression model to be zero (or close to zero); this is the basis of LASSO regression. Another way to handle this is to learn different regression models with different subsets of the available predictors for different partitions of the data; this is the motivation behind random forests.

We began this chapter by emphasizing that statistical data analyses may serve either of two functions: (1) to enhance our understanding of the data or (2) to predict those data in new cases. Linear regression models can serve both functions because they are so simple: we can use them to predict, but we can also inspect and understand them with relative ease.

Many of the more complex methods that this book will describe can make better predictions, but that extra power comes at the cost of transparency. That cost makes translational application difficult, and perhaps particularly so in psychiatry and neurology, where the rationale for a decision may be at least as important as its accuracy. For this reason, linear regression models will usually be preferred over those more complex methods, unless and until the latter demonstrate predictive advantages that are too large to ignore.

4.5 Key points

- Linear regression fits a single parameter—a slope or weight coefficient—to each predictor variable.
- Linear regression can in principle capture any (nonlinear) relationship between predictor and response variables.
- Linear regression is ubiquitous in modern science, and the science of brain disorders is no exception.
- Linear regression models provide an attractive compromise between transparency and predictive power.
- However, these models fail when we consider too many predictor variables (where "too many" really means "more than a few").
- Nevertheless, linear regression models will likely remain popular in the study of brain disorders.

References

Bellman, R. E. (1957). *Dynamic programming*. Princeton: Princeton University Press.

Ben-Hur, A., Ong, C. S., Sonnenburg, S., Scholkopf, B., & Ratsch, G. (2008). Support vector machines and kernels for computational biology. *PLoS Comput. Biol., 4*(10). https://doi.org/10.1371/journal.pcbi.1000173. e1000173.

Broca, P. (1861). Remarques sur le siége de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bulletin de la Société Anatomique, 6*, 330–357.

Deweer, B., Lehericy, S., Pillon, B., Baulac, M., Chiras, J., Marsault, C., et al. (1995). Memory disorders in probable Alzheimer's disease: The role of hippocampal atrophy as shown with MRI. *Journal of Neurology, Neurosurgery and Psychiatry, 58*(5), 590–597.

Fridriksson, J., den Ouden, D.-B., Hillis, A. E., Hickok, G., Rorden, C., Basilakos, A., et al. (2018). Anatomy of aphasia revisited. *Brain, 141*(3), 848–862.

Green, D. W. (2003). Neural basis of lexicon and grammar in L2 acquisition. In R. van Hout, A. Hulke, O. Kuiken, & R. J. Towell (Eds.), *The Lexicon-Syntax Interface in Second Language Acquisition* (pp. 197–218). John Benjamins Publishing Company.

Hope, T. M., Parker, J., Grogan, A., Crinion, J., Rae, J., Ruffle, L., et al. (2015). Comparing language outcomes in monolingual and bilingual stroke patients. *Brain, 138*(Pt 4), 1070–1083. https://doi.org/10.1093/brain/awv020.

Hope, T. M. H., Friston, K., Price, C. J., Leff, A. P., Rotshtein, P., & Bowman, H. (2019). Recovery after stroke: not so proportional after all? *Brain, 142*(1), 15–22.

Lazar, R. M., Minzer, B., Antoniello, D., Festa, J. R., Krakauer, J. W., & Marshall, R. S. (2010). Improvement in aphasia scores after stroke is well predicted by initial severity. *Stroke, 41*(7), 1485–1488. https://doi.org/10.1161/strokeaha.109.577338.

Marchi, N. A., Ptak, R., Di Pietro, M., Schnider, A., & Guggisberg, A. G. (2017). Principles of proportional recovery after stroke generalize to neglect and aphasia. *European Journal of Neurology, 24*(8), 1084–1087. https://doi.org/10.1111/ene.13296.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychol, 110*(1), 40–48.

Popescu, V., Agosta, F., Hulst, H. E., Sluimer, I. C., Knol, D. L., Sormani, M. P., et al. (2013). Brain atrophy and lesion load predict long term disability in multiple sclerosis. *Journal of Neurology, Neurosurgery and Psychiatry, 84*(10), 1082–1091. https://doi.org/10.1136/jnnp-2012-304094.

Price, C. J., Ramsden, S., Hope, T. M. H., Friston, K. J., & Seghier, M. L. (2013). Predicting IQ change from brain structure: A cross-validation study. *Developmental Cognitive Neuroscience, 5*, 172–184. https://doi.org/10.1016/j.dcn.2013.03.001.

Ramsden, S., Richardson, F. M., Josse, G., Thomas, M. S. C., Ellis, C., Shakeshaft, C., et al. (2011). Verbal and non-verbal intelligence changes in the teenage brain. *Nature, 479*(7371), 113–116. http://www.nature.com/nature/journal/v479/n7371/abs/nature10514.html#supplementary-information.

Sanfilipo, M. P., Benedict, R. H. B., Sharma, J., Weinstock-Guttman, B., & Bakshi, R. (2005). The relationship between whole brain volume and disability in multiple sclerosis: A comparison of normalized gray vs. white matter with misclassification correction. *Neuroimage, 26*(4), 1068–1077. https://doi.org/10.1016/j.neuroimage.2005.03.008.

Sluimer, J. D., Bouwman, F. H., Vrenken, H., Blankenstein, M. A., Barkhof, F., van der Flier, W. M., et al. (2010). Whole-brain atrophy rate and CSF biomarker levels in MCI and AD: A longitudinal study. *Neurobiology of Aging, 31*(5), 758–764. https://doi.org/10.1016/j.neurobiolaging.2008.06.016.

Smith-Woolley, E., Pingault, J.-B., Selzam, S., Rimfeld, K., Krapohl, E., von Stumm, S., et al. (2018). Differences in exam performance between pupils attending selective and non-selective schools mirror the genetic differences between them. *Npj Science of Learning, 3*(1), 3. https://doi.org/10.1038/s41539-018-0019-8.

Tedeschi, G., Lavorgna, L., Russo, P., Prinster, A., Dinacci, D., Savettieri, G., et al. (2005). Brain atrophy and lesion load in a large population of patients with multiple sclerosis. *Neurology, 65*(2), 280–285. https://doi.org/10.1212/01.wnl.0000168837.87351.1f.

Ullman, M. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism, 4*(02), 105.

Watanabe, A., Suzuki, M., Kotaki, H., Sasaki, H., Kawaguchi, T., Tanaka, H., et al. (2016). Predicting cognitive and behavioral functions in patients with dementia: Practical prognostic models of logarithmic and linear regression. *Edorium Journal of Disability and Rehabilitation, 2*, 144–153.

Wernicke, C. (1874). *Der aphasische symptomencomplex. Breslau: Cohn und Weigart*.