

# How to report a regression analysis

This document explains what needs to be included where when reporting a regression model(s) in a scientific article. An example of each of the sentences and paragraphs required in the introduction, method, results and discussion is given. On the right (in sans-serif text) is a commentary and explanation of each part.

The example chosen here is deliberately simple to keep things as short as possible: that the variables and measurements don't require special explanations. In a real article you should assume that your reader has completed a psychology degree but is not a specialist in your topic. Any jargon or specialist terms should be briefly explained at the point they are first used.

## INTRODUCTION

[...]

We predict that both car weight and the number of cylinders will be independent predictors of fuel economy.

The introduction contains the theoretical rationale for this study design. It usually has an inverted pyramid structure in which the opening is a very broad description of the domain or problem; each paragraph examines the existing literature, narrowing things down so that the final paragraph of the introduction ...

The final paragraph identifies something that we *currently don't know* the answer to and need to collect more data on. It makes at least one specific prediction that motivates running the study, and is tested in the sections that follow.

## METHODS

[...]

We used linear regression to predict fuel economy from weight and number of engine cylinders. We compared alternative models using a Bayes Factor. [...]

Based on the introduction it should be obvious why we want to run *this* study, and the choice of should be an 'obvious' way of answering the question posed.

Towards the end of the methods section there will be a paragraph describing the analysis to be used in *general* terms. You don't need to describe each specific model at this stage — just state the general approach. Optionally, this paragraph has a subheading like "Statistical methods" or "Data analysis".

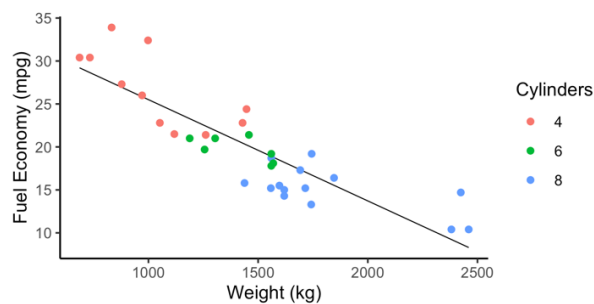
## RESULTS

Fuel economy data were obtained for 32 cars. Car weights ranged from 686 to 2460kg (mean = 1459, SD=443), and models in this sample had between 4 and 8 cylinders. Fuel economy ranged from 10 to 34mpg (mean = 20, SD=6).

Figure 2 suggests...

... that both cylinders and weight were correlated with fuel economy.

*Figure 1. Relationship between weight and fuel economy, coloured by number of cylinders.*



To test whether cylinders and weight were independent predictors of fuel economy, and to estimate the strength of these relationships...

...we ran two linear regression models. Model A included only weight as a predictor; Model B included both weight and the number of cylinders.

Results are presented in Table 1.

We begin by describing the sample. In this example we give figures in the text, but this data could also be provided as a table (e.g. if there were many more variables to describe).

Always refer to all figures somewhere within the text and...

...describe the general pattern of results that each figure shows. Imagine you are telling someone about the plot over the telephone: they don't need to know details about what it looks like, or what colours are used. But they do want to know what you learned from it.

Figures should be included roughly where they are mentioned in the text (although some journals ask for them to be at the very end, after the references). Each figure should have a descriptive title with a figure number. All axes should be labelled clearly and include the units of measurement.

It is important to give content and explain why a particular analysis is being used. This sentence links our hypothesis, methods, and the insights from Figure 1 with the analysis to come.

We need to describe i) what technique(s) were used and ii) how they were applied. We state explicitly that we are using linear regression, that we ran 2 models, and which variables are the outcomes and predictors in each model.

Always refer to each table within the text. The APA states that model coefficients should be rounded to 2 or 3 decimal places. In this case, because the predictors are on a scale that leads to very small values, I have used 3 dp. Where 2dp is sufficient (e.g. for the beta values) this is preferred. It's OK to use a slightly smaller font for tables where this helps with the format (minimum 10pt). See the workshop notes for more detail on the *content* of this table and the notes.

Table 1. Results for models A and B

Coefficient	Model A		$\beta$	Model B*		$\beta$
	Estimate	95% CI		Estimate	95% CI	
Intercept	37.287	[33.452, 41.122]		39.687	[36.179, 43.195]	
Weight	-0.012	[-0.014, -0.009]	-10.46	-0.007	[-0.010, -0.004]	-6.25
Cylinders	—	—		-1.508	[-2.356, -0.659]	-5.39
$R^2_{\text{adjusted}}$	0.75			0.82		

\* Bayes Factor in favour of Model B (vs Model A) = 22.7

Both additional weight and additional cylinders exerted a substantial negative influence on fuel economy.

APA style suggests that coefficients included in tables are *not* repeated in the text: Just summarise the size and direction of the effect. Here I write that both predictors exerted “substantial” influence because both the beta values are relatively large (-6 and -5 mpg respectively). We also know that both were independent predictors, based on the Bayes Factor.

Given these data—model B was 22 times more probable than model A, suggesting weight and cylinders are independent predictors of economy.

This sentence is based on the Bayes Factor. Here there is positive evidence in favour of model B so the BF is > 1. Had that not been the case I would have calculated the inverse Bayes Factor—that is, the evidence in favour of model A—because people seem to find Bayes factors > 1 easier to interpret.

Diagnostic plots for model B did not indicate any problems with model fit or violations of assumptions.

This sentence would often be omitted (and simply assumed) in real articles, especially so in short reports. Either way, I suggest that code for diagnostics checks should be included in the data analysis code and shared with the publication.

## DISCUSSION

As expected, we found...

We ALWAYS start the discussion with a restatement of the most important finding: in this case, that our prior prediction was confirmed.

... strong evidence that both car weight and number of cylinders were independent predictors of fuel economy.  
[...]

We do not repeat the BF value in the discussion, but the phrase “strong evidence” summarises the BF reported above.

The rest of the discussion would serve to link these new findings to the previous literature and explain implications for theory or practical applications.