

Regression Analysis in Biological Anthropology

Dr Geoff Kushnick

The Australian National University, Canberra

geoff.kushnick@anu.edu.au

Kushnick, G. (2018).
Regression analysis in
biological anthropology.
<https://geokush.com/2020/02/17/regression-analysis-in-biological-anthropology/>

What is Regression?

Regression analysis comprises several related analytical tools that are useful to researchers in a variety of disciplines, including biological anthropology. Different styles of regression are available for different types of dependent variables (see Table 1 for a non-exhaustive list) and can accommodate a wide range of contexts that are challenging for many standard statistical methods, including controlling for any number of potential confounding variables, adjusting for non-independent sampling units, and using cases that have missing data. Most types of regression analysis can be used to: (a) test hypotheses about the association between variables, (b) model the relationship between variables and provide measures of the strength of model fit and proportion of variation in a dependent variable explained by the independent variables, and (c) build equations that can be used for prediction. Regression analysis can accommodate any number of independent variables, ranging from the simplest cases where you are interested in the effect of a single independent variable on the dependent variable, to the case of multiple regression where you are interested in two or more independent variables. The simplest case is where the independent variables are continuous, but other types of independent variables can be accommodated. Table 2 introduces some resources for issues in the application of regression analysis.

Table 1. Some types of regression.

Type	Dependent Variable	Difficulty	Reference(s)
Linear	Continuous (linear)	★	Montgomery (2013); Weisberg (2014)
Logistic	Binary	★★	Hilba (2009); Hosmer et al. (2013)
Multinomial	Nominal	★★★	Agresti (2012)
Ordered logit	Ordinal	★★★	Agresti (2012)
Poisson	Count	★★	Coxe et al. (2009)
Negative binomial	Count (overdispersed)	★★	Coxe et al. (2009)
Cox	Time to event	★★	Klein (2016); Allison (2014); Hosmer (2008)

Table 2. Resources for issues in the application of regression analysis.

Issue	Description	Reference(s)
Confounding	Regression can be used to statistically control for confounding variables.	McNamee 2005; Westfall and Yarkoni (2016)
Non-Independence	There are a number of extensions to standard regression methods that account for non-independent data points.	Gelman & Hill (2007); Graubaud & Korn (1997); Harrison et al (2018)
Multicollinearity	Correlation of independent variables can lead to inflation of standard errors for individual beta values.	Graham (2003); O'Brien (2007)
Dummy Variables	Categorical independent variables can be treated as dummy variables in regression analysis.	Hardy (1993)
Interaction Effects	You can use regression to test for interaction effects.	Jaccard (2001); Jaccard et al. (2003)
Model Selection	Methods for choosing which independent variables should be used in multiple regression.	Burnham & Anderson (2004); Johnson & Omland (2004)
Prediction	Regression models can, when certain assumptions are met, be used to make predictions that go beyond the range of modeled data.	Palmer & O'Connell (2009)
Missing Data	Imputation and other methods for dealing with the biases inherent in excluding cases with missing data points.	Allison (2001); McKinnon (2010)

References

- Agresti, A. (2013). *Categorical data analysis* (3E). NY: Wiley.
- Allison, P.D. (2001). *Missing data*. Thousand Oaks, CA: Sage Publications.
- Allison, P.D. (2014). *Event history and survival analysis*. Thousand Oaks, CA: Sage Publications.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33, 261-304.
- Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment*, 91(2), 121-136.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. NY: Cambridge University Press.
- Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, 84(11), 2809-15.
- Graubard, B. I., & Korn, E. L. (1994). Regression analysis with clustered data. *Statistics in Medicine*, 13, 509-522.
- Hardy, M. A. (1993). *Regression with dummy variables*. Newbury Park, CA: Sage Publications.
- Harrison, X. A., Donaldson, L., Correa-Cano, et al. (2018). A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ*, 6, e4794.
- Hilba, J.M. (2009). *Logistic regression models*. Boca Raton, FL: CRC Press.
- Hosmer, D.W. (2008). *Applied survival analysis: Regression modeling of time-to-event data*. NJ: Wiley.
- Hosmer, D.W., Lemeshow, S., & Sturdivant, R.X. (2013). *Applied logistic regression* (3E). NY: John Wiley.
- Klein, J.P. (2016) *Handbook of survival analysis*. London: CRC Press.
- Jaccard, J. (2001). *Interaction effects in logistic regression*. Newbury Park, CA: Sage Publications.
- Jaccard, J., Turrisi, R., & Wan, C. K. (1990). *Interaction effects in multiple regression*. Newbury Park, CA: Sage Publications.
- Johnson, J., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, 19(2), 101-108.
- Mckinnon, A. (2010). The use and reporting of multiple imputation in medical research -- a review. *Journal of Internal Medicine*, 268, 586-593.
- McNamee, R. (2005). Regression modelling and other methods to control confounding. *Occupational and Environmental Medicine*, 62, 500-506.
- Montgomery, D.C. (2013). *Introduction to linear regression analysis* (5E). Somerset: John Wiley and Sons.
- O'Brien, R. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41, 673.
- Palmer, P. B., & O'Connell, D. G. (2009). Regression analysis for prediction: Understanding the process. *Cardiopulmonary Physical Therapy Journal*, 20(3), 23-26.
- Weisberg, S. (2014). *Applied linear regression*. Hoboken, NJ: Wiley.
- Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLoS ONE*, 11(3), e0152719.



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

Kushnick, G. (2018). Regression analysis in biological anthropology. <https://geokush.com/2020/02/17/regression-analysis-in-biological-anthropology/>