Writing a good “statistical methods paragraph” requires balancing length, detail, and providing sufficient information for someone else to understand your decisions and repeat your analysis. The ideal paragraph is concise and complete.

Only two people included a statement of the software. This is necessary because different programs can (and distressingly frequently) implement the “same method” differently. That can give different answers. That happens especially frequent when there are default values (e.g. for optimizer convergence criteria, binning of an empirical variogram, computational shortcuts in kriging). The one person who mentioned software only said RStudio. Much more important is which library (or libraries) were used. We’re using gstat. geoR and kriging are two other options and they may give different results.

Here are two examples (lightly edited) of very nice paragraphs from this last exam.

Universal kriging was used with a linear trend model on latitude and longitude as predictor for SST. A variogram was computed from the observed points using a Cressie-Hawkins estimator, and a Matern model (k=1) without nugget was fit to this using a starting sill of 2.5 and a starting range of 350. This produced a Matern model with a partial sill of 3.2 and a range of 149.

I used universal kriging methods with an additive latitude and longitude linear trend model to predict sea-surface temperature (°C) throughout the northern Gulf of Mexico while accounting for spatial correlation. I used a Cressie-Hawkins semivariance estimator with a linear trend model to estimate empirical semivariogram using 15 bins and at least 100 pairs of location in each bin. I fit a series of spherical, exponential, and Matern variogram models to the empirical variogram. I visually inspected the fit of each variogram model by plotting the model over the empirical variogram and I selected a model based on the lowest weighted sum of squared errors. I selected a Matern variogram model with a nugget and k=2 as the best fit. I tested the fit of the Matern model using 10-fold cross-validation. Fit of the cross-validated predicted values was evaluated based on the mean square error of prediction (MSEP) and homoscedasticity of residuals. Based on these criteria, the Matern model appeared to be an adequate fit with MSEP equal to 0.176 and a residual plot that showed equal spread. There was evidence of a single outlier, but that was left in for the analysis. I used universal kriging with a linear trend model and Matern variogram model to predict sea-surface temperature throughout the northern Gulf of Mexico.

To either of these, I would add: Computations were done using R and functions in the gstat library, version 2.0-4. Unless otherwise mentioned, default options were used.

Note: How can you find the version number? For R, look at the very top of the Console window. R prints the version when it initializes. For a package, what I do is library(help=gstat) which prints a summary of the package (replace gstat with the appropriate package name). This requires that the package is installed (i.e., install.packages() has been used), but not that it is activated, by library(). While I most frequently use this to see what functions are in the package and their names, it also prints the version number at the very top. The RStudio version number actually doesn’t matter – that’s just the windowing system to interact with R; RStudio doesn’t do any computations. The output from library(help=gstat) appears in the File pane, not in the Help pane. That’s not the same place as results from ?function.