**MCQ8 – Model Answer**

1. If instead of laying down a quadrat you had purchased 20 ammonites collected by the commercial collector which analyses could still be performed? *Bivariate analysis of the relationship between diameter and chamber number. It should be obvious that spatial and orientation data are lost immediately once an ammonite is no longer in situ. However, we can additionally not rely on the data being an unbiased sample for univariate analysis of size, and indeed based on early investigation we might logically assume that specifically the commercial collector’s data would lead us to erroneously large values. However, we can be relatively safe in running bivariate analyses of size versus chamber count as long as the number of chambers (our explanatory variable) is variable.*
2. What is the primary goal of multivariate analysis? *To reduce the dimensionality of the data. Adding as many variables as possible is generally not a good strategy as we typically want to reduce labour or expense and concentrate on variables that are tractable to collect and likely to be useful in answering whatever our question is. Similarly, we do not generally in science seek greater and greater complexity for its’ own sake. More typically we want simplicity – to explain as much as possible with as little as possible. We definitely don’t want to generate data that are difficult to visualise. Indeed, reducing the dimensionality of the data is the goal in part because it makes the data easier to visualise. Remember Lesson #1.*
3. When calculating distances in a hyperdimensional (four or more dimension) space what common mathematical theorem can be used with minimal modification? *Pythagorean theorem. Bayes’ theorem concerns probability and Descartes’ and Euclid’s have nothing to do with statistics at all. Pythagoras can be used to calculate the distances between objects in high-dimensional space and hence generate the distance matrices used in Classic Multi-Dimensional Scaling.*
4. What rule must ordination axes always adhere to? *They must be orthogonal to each other. This should hopefully have come across quite clearly in the ammonite and UK city data from practical where, for example, the c. North-South axis of the UK (PC1) is orthogonal to the c. East-West axis (PC2). If there were as many of them as there are axes in the original data then the ordination will have failed. They must also not begin at zero, but instead the data must have a mean of zero on each axis (i.e., the average point in the space must always be at the origin (0, 0, 0…). Additionally, if the ordination axes were always parallel to the original axes of variance the ordination would not have done anything useful.*
5. If using spherical statistics what relationship between the three eigenvalues suggests points are distributed around a great circle? *The first and second eigenvalues are of approximately equal size and much larger than the third eigenvalue. This should make logical sense as if data fall on a plane they have extent in two directions (e.g., width and depth), but not in another (e.g., height). Whereas, all eigenvalues are of approximately equal size this suggests the data are randomly distributed throughout the space (have width, depth, and height), and if the first eigenvalue is much larger than the other two the data only really have extent in one direction. The final option, there are big falls in size from the first to the second to the third eigenvalue is the most nuanced as eigenvalues must always be smaller going from first to second to third, but if the spacing here is regular this suggests they correspond to primarily a line, but with a clear secondary extent in a plane.*
6. What did the major axis of variance (PC1) for the UK city data represent? *The long axis of the UK (i.e., roughly the North-South axis). Hopefully this should have been very obvious with the Scottish cities at one end and the South Coast cities at the other. The second axis represented the short axis of the UK (i.e., roughly the East-West axis), and there was definitely a clear major axis (majority of variance fell on PC1). The length (in characters) of each cities name was a red herring, but in any case, this would be a weird outcome given this information was not present in the raw data.*
7. When looking at the flag distance matrix which flag is most different from the Japanese flag? *You should have seen that Japan had a distance of 2 from France, Italy, and the Netherlands, of 2.5 (1dp) from Germany and Ireland, but of 3 from India. Thus India is the most different from Japan, at least using the measurements applied here.*
8. When allowing the variance in the flag data to be spread across seven axes how many actually had non-zero variance? *Seven (i.e., all of them). We forced this onto two later, and the ammonite data clearly only really required two axes to explain it. The vast majority of variance being on the first axis was only evident in the UK city data.*
9. When forcing the variance on to only two axes what aspects of flags did the major axis (PC1) represent? *Primarily colour (green and orange at one end, red and blue at the other). Hopefully this should have been obvious, although the uniqueness of the German flag may have confused you. Stripe orientation (horizontal stripes at one end, vertical stripes at the other) was the major feature of the second axis (PC2). Whereas the presence of a central symbol did not seem important as we already saw that the Japan flag was most different from the Indian. Only the UK flag contained a saltire and although it was found at one end of the plot there was not a clear break between this and the other, non-saltire flags. Besides which, the colour aspect should be more prominent as it is reflected in every flag.*
10. When plotting the fold axis data on a ternary diagram which corner did the data fall closest to? *Point. Hopefully this should have been obvious and concurred with other determinations from the data (a visual inspection, the high value of the first eigenvalue). The final bivariate plot edged more towards a girdle interpretation and certainly the stereonet seemed to show a spread of data towards that scenario, but there was a clear concentration of points in the top left quadrant which fits best with a cluster interpretation.*