**MCQ2 – Model Answer**

1. In Lesson #1 in lecture we considered which of two models (A and B) best explained the relationship between x and y. Which was the best explanation? *Model B (more complex equation; curved line). Hopefully this should have been obvious as soon as the data were visualized as x and y describe a clear curvilinear relationship that Model B closely matches*. *We will encounter these types of models in more detail later in the course.*
2. In considering a scenario where there are only three possible outcomes (A, B, and C) and we already know the probabilities of two of them (A and B), what is the correct equation for calculating the probability of outcome C? *p(C) = 1 - (p(A) + p(B)). This is the rule that all possible outcomes must sum to 1 so if we know p(A) and p(B) we can calculate p(C) this way.*
3. What is the name given to the equation that describes the shape of the curve of a probability distribution? *Probability Density Function (PDF). The Cumulative Distribution Function describes the cumulative version of the PDF.*
4. In applying the Binomial (p(Increase in dinosaur species number) = 0.5) to our dinosaur example where did we find the observed value (N increases = 14) to fall within the distribution? *In the middle 50% (Increases and Decreases approximately equal). It was not exactly in the middle (N Increases = 13)*
5. At the end of the lecture we encountered a famous Mathematics problem known as the "Monty Hall problem", which is derived from a game show with a host of that name. What is the optimal strategy and associated probability of winning the prize (a car)? *Switch doors; p(Win car) = 2/3. If you struggle to understand why don't worry - many professional statisticians got the wrong answer. It is worth reading up on or see the YouTube playlist section on Minerva for a link to a video about it. If it still doesn't make sense I suggest actually pairing up with a friend and playing the game yourself, taking it in turns to be the host or the contestant. You should see after several rounds that the switch strategy converges on a 2/3 win probability.*
6. Which probability distribution is best suited to understanding the probabilities of coin flips? *The Binomial which works only for cases where there are two outcomes (e.g., heads or tails). We used the Discrete Uniform for dice in lecture and the Poisson for timing of eruptions in the practical.*
7. Based on the probability inference from the 1000 flips which value of p(Heads) seems most likely for the biased coin? *It should be clear that the two distributions (inferred and hypothesized) matched most closely at p = 0.2. Note this doesn’t mean that the true p(Heads) is exactly 0.2, but it seems likely that it is much closer to this value than either 0.1 or 0.3.*
8. In considering two hypotheses for the true value of p(Increase Dinosaur Species Number) which of the following statements is most accurate? *A value of 0.5 seems much more likely than a value of 0.833. We didn’t do enough work to bracket the likely true value or show that it must be larger than 0.5, but it definitely falls well outside the higher probability density of p(Increases) = 0.8333 and well inside the higher probability density of p(Increases) = 0.5. Note that this is real data and so nobody knows the true value.*
9. Which R function can we use to calculate the Probability Density Function for the Poisson? *dpois. The others would tell us the probability for a specific quartile (qpois), the quartile for a specific probability (ppois), or give us random values drawn from a Poisson distribution (rpois). Note that all probability distributions in R offer these four options and in the practical we actually used all of the four for the binomial (dbinom, pbinom, qbinom, and rbinom), the latter being hidden inside our coin flipping function.*
10. To calculate the probability that Etna erupts at least once in every 100-year span we used which probability rule? *The rule that the probability of all possible outcomes must sum to 1. I.e., we were able to side-step calculating a sum across all eruption numbers from 1 to infinity by calculating the value for 0 and subtracting it from one.*