BEDA BIOL2022/2922 STUDY NOTES

LECTURE SERIES 1

STATISTICS REVISION

- Mean = average
  - u = population
  - x = sample
- Mode = most occurring
- Variance = sum of (scores — mean)squared / n
- Standard deviation = square root of variance
  - ø = population
  - s = sample

COMMON TESTS (EXPLAINED PROPERLY LATER)

- t-test = difference between 2 means
- ANOVA = difference between >2 means
- Correlation = relationship between X and Y
- Regression = X predicts Y
- Frequencies/counts = G tests and Chi squared tests
- Assumption and requirements of these tests:
  - All data are independent
  - Variances must be homogenous (fixed using transformations)
  - Normal distribution (can be ignored with balanced design = same no. replicates in each treatment)
  - Decide which test before collecting data
  - Replication is crucial

VARIABILITY

- Intrinsic = natural variability e.g. body sizes of individuals
- Extrinsic = measurement error e.g. measure same animal a number of times and get different measurements
- Frequency distributions are the best way to display variability graphically -they show:
  - Location parameter - where is the middle of the distribution located (mean)
    - Normal distribution: Mean = mode = median all in middle
    - Skewed: Mode in the middle of the peak, median in the middle of the graph and mean towards skewed tail
    - Bimodal: 2 peaks where 2 modes are, median in the middle and mean depends on peaks
  - Dispersion parameter - how widely dispersed or spread out the population is (spread of deviations from mean)
    - Range: difference between largest and smallest
    - Variance: wide graph = lots of variance
    - Standard deviation: fat curve = high standard deviation
- Types of variables
  - Discrete = whole numbers e.g. counts of animals
  - Continuous = decimals/fractions e.g. heights/lengths etc.

SAMPLING CONSIDERATIONS

- Need to define the population so methods allow you to sample the population you want to describe
- Statistical definition of a population = all of the organisms belonging to the set (this is different to ecology definition of a population)
- We can't measure every individual in a population so we take a sample that is representative of a population and infer parameters for the whole population
- Therefore, need to make sure sample is representative of the whole population
- Sampling design
    - Correlative = sample what is already there and statistically analyse to see if there is a correlation (find if a pattern exists)
    - Experimental = change a variable and statistically analyse and compare to control areas (find why a pattern exists)
- Precision and accuracy
    - Precision = scores around the same number
        - Want frequency distribution curve to be thinner than population curve
        - Standard error of the mean (sd/√n)
    - Accuracy = scores are around the predicted value
        - Want frequency distribution curve for sample to be in same position as the population
        - Can't measure because we don't know true population mean but it can be estimated
- Random and representative
    - Samples must be representative - not the same as random
    - However random samples will usually be representative
    - To get a representative sample you must have some idea about the biology and relate sampling protocol to its underlying distribution
    - If in doubt do a preliminary sample
    - In areas where variables are not evenly distributed - use stratified sampling:
        - Do preliminary sample to find areas of high density and low density
        - Divide these areas into 2 stratums
        - Find mean and standard error of each sample
        - Find overall mean and standard error
        - This gives a more representative sample set and better estimate of mean and variance
        - The standard error is much reduced - samples more precise
        - Samples also more accurate
        - Advantage of this sampling increases as areas become more dissimilar

FALSIFACATIONIST APPROACH

- Make observation (see it in nature or previous results)
- Create a hypothesis (this observation occurs because)
- Postulate a testable prediction (If I …. then ….)
- Suggest a null (There will be no change)
- Gain statistics from replication
- Either falsify the null (statistically significant $p<0.05$) or retain it

CONTROLS

- Experimental control = group that has no treatment applied to them (not always necessary)
    - e.g. no fenced area
- Procedural controls also always have to be done (or explained why they don't have to be if someone else has shown they have no effect)
    - e.g. no fence but just fence posts