**Linear Discriminant Analysis of coho salmon smolt**

*Objective of study*: Exploratory analysis of what variables best predict age in coho salmon smolt (ages one and two primarily) using scales

* + Analysis by system, by lake versus river, pooled

**General Information for determining variables**

* Plus growth is springtime growth occurring just before smolt outmigration; end of last annulus to edge of scale.
* Size of scale varies somewhat depending on the scale row and column it came from. Circuli COUNTS might offer better utility than measurements alone.
* Biases of data (undercounted by 1-4 circuli)
* Center of scale, “focus”; size can vary so best to leave out
  + replace with distance from circulus 2 instead of focus.
* Variables should only measure to the end of a zone, not to the end of the plus growth (end of scale) since this distance is variable and occurs after annuli have formed
* Growth zones are defined as FW1; center of focus to end of first annulus. Followed by FW2, FW3. However if an age 1 has FW2 that is the plus growth. If an age 2 has FW3 that is its plus growth.
* There are very few age 3’s in the data; these will be excluded from analysis. These would have a FW3 zone and also FW4 if they have any plus growth.
* Before the analysis is begun it would be helpful to sift the data for problems;

Duplicated rows

If age 1 there must be zone 1 measurements also zone 2 is optional if plus is present. There must not be any zones but 1 or 2.

If age 2 there must be zone 1 and zone 2 measurements also zone 3 is optional if plus is present. There must not be any zones but 1 and 2 or 3.

If age 3 there must be zone 1 and zone 2and zone 3 measurements also zone 4 is optional if plus is present. There must not be any zones but 1 and 2 and 3 or 4.

* The original reader measured to the inside of the circuli (*before* the circuli). The ‘norm’ is to measure to the outside (*after*) of the circuli. This will not affect model results but if the model results are applied to new data that measures to the outside of the scale, this may cause problems. This needs to be considered and discussed in the methods section of the write-up. Possible fixes also need to be discussed.

**Tasks:**

1. ~~Determine variables to try in linear discriminant analysis~~ 
   1. ~~Talk to Lorna about ideas~~
   2. ~~Literature review~~
2. ~~Determine the calculation for the variables for SFAZ and NCFAZ~~
   1. ~~SAS code~~
3. ~~Change the SAS code to R code~~
4. ~~Chlesea will remeasure a certain number of scales and Lorna will run an ANOVA comparison between the original reader and Chelsea to determine confidence in original reader.~~ 
   1. If there are significant differences in the scale readers, the actual measurement differences will be determined (*e.g. is there a consistent pattern in the differences?*)
   2. Then we will determine if a subsample of the original data needs to be remeasured entirely and used as the new sample to run LDA. We will discuss the sample size if it comes to this.
   3. Result: once Chelsea accounted for the difference btw reading on the inside instead of outside of the scale-the results were the same with the original read on the subset of data (Sept. 2017)
5. Lorna will check ages in digitized dataset to the CWT database.
6. Sara will create figures based on the raw dataset to detect any outliers (*e.g. # circuli versus radius, Age versus fish length*)
7. ~~Kent and Lorna will review the 72 variables.~~
8. Starting in January 2018, Julie will begin an ADF&G report to summarize data digitized through 2005 (e.g. lake versus stream)
9. Sara will run LDA R code on the data
10. Lorna and Bev will help with the ADF&G report as a co-author.
11. In spring 2018, start a draft of either an NPRB or S-K grant proposal to digitize data through the current year and redigitized a subsample of the prior years data for consistency.

**Potential problems with data:**

Kent-1st year the cutoff length for cwt was 42mm, reduced to 38mm in the 2nd and 3rd year

-plus growth no recognized as plus growth

-circuli undercounted in database

-no comments about false checks (where checks occurred and rate at which they occur)

Lorna-seems like ages> than stated in dataset

-some scales don’t have a plus growth

**Table 1.** Scale measurement and count characters calculated from intercirculus distances.\* Before these variables are calculated, the plus group and the focus to C2 distance are removed.

Variable Description

Q1 Number of circuli in FW Annular Zone (NCFAZ); FW1 + FW2 (does not include plus group or focus to C2)

Q2 Width of FW Annular Zone (SFAZ, EOZ) (does not include plus group or focus to C2)

Q3 Average interval between circuli; (Q2/Q1)

Q4 Ratio of distance between C2 and C3; C3/Q2

Q5 Ratio of distance between C2 and C4; (C3+C4)/Q2

Q6 Ratio of distance between C2 and C5; (C3+…+C5)/Q2

Q7 Ratio of distance between C2 and C6; (C3+…+C6)/Q2

Q8 Ratio of distance between C2 and C7; (C3+…+C7)/Q2

Q9 Ratio of distance between C2 and C8; (C3+…+C8)/Q2

Q10 Ratio of distance between C2 and C9; (C3+…+C9)/Q2

Q11 Ratio of distance between C2 and C10; (C3+…+C10)/Q2

Q12 Ratio of distance between C2 and C11; (C3+…+C11)/Q2

Q13 Ratio of distance between C2 and C12; (C3+…+C12)/Q2

Q14 Ratio of distance between C2 and C13; (C3+…+C13)/Q2

Q15 Ratio of distance between C2 and C14; (C3+…+C14)/Q2

Q16 Ratio of distance between C2 and C15; (C3+…+C15)/Q2

Q17 Ratio of distance between C2 and C16; (C3+…+C16)/Q2

Q18 Ratio of distance between C2 and C17; (C3+…+C17)/Q2

Q19 Ratio of distance between C2 and C18; (C3+…+C18)/Q2

Q20 Ratio of distance between C2 and C19; (C3+…+C19)/Q2

Q21 Ratio of distance between C2 and C20; (C3+…+C20)/Q2

Q22 Ratio of distance between C2 and C21; (C3+…+C21)/Q2

Q23 Ratio of distance between C2 and C22; (C3+…+C22)/Q2

Q24 Ratio of distance between C2 and C23; (C3+…+C23)/Q2

Q25 Ratio of distance between C2 and C24; (C3+…+C24)/Q2

Q26 Ratio of distance between C2 and C25; (C3+…+C25)/Q2

Q27 Ratio of distance between C2 and C26; (C3+…+C26)/Q2

Q28 Ratio of distance between C2 and C27; (C3+…+C27)/Q2

Q29 Ratio of distance between C2 and C28; (C3+…+C28)/Q2

Q30 Ratio of distance between C2 and C29; (C3+…+C29)/Q2

Q31 Ratio of distance between C2 and C30; (C3+…+C30)/Q2

Q32 Ratio of distance between sixth-to-last circulus and end of zone; (distance from sixth to last circuli in NCFAZ)/Q2

Q33 Ratio of distance between seventh-to-last circulus and end of zone; (distance from seventh to last circuli in NCFAZ)/Q2

Q34 Ratio of circuli in first half to second half of Q2 (based on distance in first half of Q2); determine the distance of 0.5\*Q2 then count cirucli until distance <=0.5\*Q2, then take the count/(Q1-count)

Q35 Ratio of number of circuli in last 3/4 of Q2 to NCFAZ; determine the distance of 0.75\*Q2 then count circuli until distance <=0.75\*Q2, then take the count/(Q1-count)

Q36 Ratio of distance between circulus 4 and circulus 7 to Q2; (C7+…+C5)/Q2

Q37 Ratio of distance between circulus 5 and circulus 8 to Q2; (C8+…+C6)/Q2

Q38 Ratio of distance between circulus 6 and circulus 9 to Q2; (C9+…+C7)/Q2

Q39 Ratio of distance between circulus 7 and circulus 10 to Q2; (C10+…+C8)/Q2

Q40 Ratio of distance between circulus 8 and circulus 11 to Q2; (C11+…+C9)/Q2

Q41 Ratio of distance between circulus 9 and circulus 12 to Q2; (C12+…+C10)/Q2

Q42 Ratio of distance between circulus 10 and circulus 13 to Q2; (C13+…+C11)/Q2

Q43 Ratio of distance between circulus 11 and circulus 14 to Q2; (C14+…+C12)/Q2

Q44 Ratio of distance between circulus 12 and circulus 15 to Q2; (C15+…+C13)/Q2

Q45 Ratio of distance between circulus 13 and circulus 16 to Q2; (C16+…+C14)/Q2

Q46 Ratio of distance between circulus 14 and circulus 17 to Q2; (C17+…+C15)/Q2

Q47 Ratio of distance between circulus 15 and circulus 18 to Q2; (C18+…+C16)/Q2

Q48 Ratio of distance between circulus 16 and circulus 19 to Q2; (C19+…+C17)/Q2

Q49 Ratio of distance between circulus 17 and circulus 20 to Q2; (C20+…+C18)/Q2

Q50 Ratio of distance between circulus 18 and circulus 21 to Q2; (C21+…+C19)/Q2

Q51 Ratio of distance between circulus 19 and circulus 22 to Q2; (C22+…+C20)/Q2

Q52 Ratio of distance between circulus 20 and circulus 23 to Q2; (C23+…+C21)/Q2

Q53 Ratio of distance between circulus 21 and circulus 24 to Q2; (C24+…+C22)/Q2

Q54 Ratio of distance between circulus 22 and circulus 25 to Q2; (C25+…+C23)/Q2

Q55 Ratio of distance between circulus 23 and circulus 26 to Q2; (C26+…+C24)/Q2

Q56 Ratio of distance between circulus 24 and circulus 27 to Q2; (C27+…+C25)/Q2

Q57 Ratio of distance between circulus 25 and circulus 28 to Q2; (C28+…+C26)/Q2

Q58 Ratio of distance between circulus 26 and circulus 29 to Q2; (C29+…+C27)/Q2

Q59 Ratio of distance between circulus 27 and circulus 30 to Q2; (C30+…+C28)/Q2

Q60 Ratio of distance between circulus 28 and circulus 31 to Q2; (C31+…+C29)/Q2

Q61 Ratio of distance between circulus 29 and circulus 32 to Q2; (C32+…+C30)/Q2

Q62 Ratio of distance between circulus 30 and circulus 33 to Q2; (C33+…+C31)/Q2

Q63 Ratio of distance between circulus 31 and circulus 34 to Q2; (C34+…+C32)/Q2

Q64 Ratio of distance between circulus 32 and circulus 35 to Q2; (C35+…+C33)/Q2

Q65 Ratio of distance between circulus 33 and circulus 36 to Q2; (C36+…+C34)/Q2

Q66 Ratio of distance between circulus 34 and circulus 37 to Q2; (C37+…+C35)/Q2

Q67 Ratio of distance between circulus 35 and circulus 38 to Q2; (C38+…+C36)/Q2

Q68 Ratio of distance between circulus 36 and circulus 39 to Q2; (C39+…+C37)/Q2

Q69 Ratio of distance between circulus 37 and circulus 40 to Q2; (C40+…+C38)/Q2

Q70 Ratio of distance between circulus 38 and circulus 41 to Q2; (C41+…+C39)/Q2

Q71 Count of circuli smaller than 0.75\*Q3

Q72 Count of circuli larger than 1.25\*Q3

\*For variables Q4-Q31 and Q36-Q70, if one distance does not exist (or is a plus group), the variable is not calculated. For example, Q52 is not calculated for Sample ID AL00\_0515-0001 since Q52 is the ratio of distance between circulus 20 and circulus 23 to Q2, and C23 is 'NA'.