8.5.1 **Project proposal**

Author Contributions:
We discussed everything together and made key points on what we wanted to include as a group for each section. Individually we wrote the parts based on what we had discussed.
Hypothesis & Predictions: Clara
Data Description: Madeline and Penelope
Introductory Data exploration: Madeline and Kaylie
Analysis Plan: Kaylie

Hypothesis: Salmon quantities, and females at reproductive age of pod, will influence the number of lactating females in a pod.

We decided to change the hypothesis in order to distinguish our analysis from the paper by Jordaan et al.. Instead of focusing on the impacts of the matriarchal structure of the orca pod, we are narrowing our focus and studying specifically how another aspect of the orcas environment, the abundance of fish, affects their pod fecundity. Lactating females will be used as a measure of fecundity as they are physically mature and fit to produce offspring at this time.

Prediction(s):

- Higher numbers of salmon will provide more resources for orcas and give them the opportunity to reproduce successfully. Stredulinsky et al. estimates that lactating females need on average roughly 5 more fish then other orca whales.
- A higher number of females at reproductive age will result in a higher fecundity in an orca pod, as measured by the number of lactating females in a pod. However, as it is defined in Wasser et al article, in the absence of sufficient salmon quantities, we might still observe a low number of lactating females proportionally to the number of females at reproductive age. Low availability of salmon is a stressor that results in late pregnancy failure.

Data Description

This data is from Oceans and Fisheries Canada and it records various information about pods of northern resident killer whales over a period of 30 years. The goal of the study was to look at changes in group members and ‘female leadership’ over the years to determine the causes of fission. Pods are groups of orcas all related through their mothers, also called matrilines. In northern resident killer whales, the group studied in this data, all offspring remain with their natal group, including males. This is in contrast to many other matrilineal species where males will leave the group to find new mates. Members of pods generally stay with their mothers their entire lives, making a common maternal ancestor of all pod members important, as when she is alive pod members should all remain together. “Resident” killer whales refers to specific populations of killer whales that are exclusively fish-eating. In 1980, all members of northern resident killer whales were known. This population is also socially and genetically isolated from the other resident killer whales, and all live in the eastern coastal areas of the North Pacific ranging from the southeastern coast of Alaska to Washington. While this makes it easier to track groups and avoid bias towards particular groups, it is still possible to miss groups (for example: group 17 is seen in 1980, missing in 1981, and seen again in 1982).

Whales were identified in the data based on markings and scars on the dorsal fins and body. Data collection took place from June-October to avoid any effects from seasonal differences, temporary pod interactions for breeding, and to keep collection between years consistent. Due to the temporary pod interactions for breeding where males will breed with females from other pods, we chose not to include physically mature males in measures for pod reproduction. Two data columns in particular have important associated assumptions. Lactating females (FLg) where the birth was unobserved or a calf was missing/dead were not counted. How long lactation lasts is also uncertain, so the authors assumed that the female lactates while the calf is present and relies on her for energy <2 years. They also assumed orca kin in the pod do not ‘spontaneously’ lactate in the presence of a calf. Chinook salmon terminal run reconstruction (CK.tr) values were estimated by the authors to try and more accurately model the number of salmon orcas could hunt by considering ocean fishery impacts.

We modified the data in a few ways. First we removed any subunit (largest matriarchal lineage in the pod) variables, as we are choosing to focus on effects for the whole pod. We then only kept these columns: Year, Group ID, CK.oa, CK.tr, CM, DPerg, FLg, FRg, and Ng. Detailed descriptions of column names are below. We tallied the number of observations for each group and are also considering removing groups with very few observations as reading a significant pattern from them is unlikely. There were no NA values in the data so we did not have to remove any of those values. Finally, we are considering combining the various salmon values into one overall value (`salmon_sum`) to look at effects of total salmon populations. The data appears to be standardized in some way so we will not simply be able to ‘add’ all the salmon data together. We are considering emailing the researchers to see if the raw, unstandardized data is available.

Column Names:

- Year
- Group ID
- CK.oa (Chinook salmon ocean abundance)
- CK.tr (Chinook salmon terminal run reconstruction)
- CM (chum salmon terminal run abundance index)
- DPerg (estimated daily prey energetic requirement of pod)(kilocalories)
- FLg (Number of lactating females in the pod (as proportion to pod size)
- FRg (Number of reproductive-age females in the pod (proportion of pod size)
- Ng (Number of individuals in pod)

Analysis Plan

The statistical test we are going to use is a Linear Mixed Model to test both predictions. Our data is not independent of one another as the differing orca groups, salmon, and years are going to affect the results. For example, the number of salmon is dependent on the year, the lactating females is dependent on the Group.ID and the year. We are looking at making a random slope intercept model because we expect as the salmon abundance changes by a unit the number of lactating females will not increase by a constant slope; it will vary. It is still to be determined which variables are fixed and which ones are random but for now our starting point is as follows:

```
Random variables: Year, Group ID. *Salmon data

*We are worried about sample size issues if we include Year as a
random effect because we won't have enough power. We might not
include it after further analysis but different methods will have to
be looked at.

Fixed Variables: FLg (lactation),  Number of reproductive age
females/pod, *Salmon data

*salmon data will be treated as both a random and fixed.
```

We will also use an AICC because we have small sample sizes, to see which model fits the best. Our steps to do that will then be to create a saturated model, and then optimize the random and fixed variables using the AICC.

The first prediction: that quantity of salmon will affect the number of lactating females, again will be modeled with a LMM. The type of salmon will be tested separately and then a combined variable deemed 'salmon_sum' will be a combination of all of the salmon data to see if that shows more of an effect. We still need to transform the salmon data in a way that we can combine it if needed. The second prediction: number of reproductive females in pod will also be tested again with a LMM. A poisson distribution may be used because it is count data.

Assumptions:

- Random effects are normally distributed (can test using 'plot' (QQPlot))
- Response (lactating females) is normally distributed.
- Homogeneity/homoscedasticity of variance (testing using 'plot' (Scale-Location, Residual))

A Brief note on Next Steps

We are open to changing our analysis. After looking more in depth at the data a lot of transformations seem to have been applied but after reading the paper and their supplementary materials it is not clear how they transformed the data. We would expect that the number of lactating females to be a ‘count’ number however some of the values are negative or not whole numbers so it appears they are measuring it by density. We currently are emailing the authors to ask if they have the raw data available. We also have to email the Cetacean Research Program to get a ‘datause request’ for the original pod data.

1. If they have the raw count data:
   a. We will look at applying a Poisson distribution to our linear mixed model as the data would then be a count
2. If there is no response
   a. We will look at doing a General Mixed model method as the random effects ie) salmon data) do not follow a normal distribution. However, we are concerned with the sample size because if we include year as a random effect then our power will not be great enough (there will only be one point per year)
   b. However it does not feel right to look at the data as not a time series so other analytical methods will be considered.

Works Cited

Barrett-Lennard, L. G. (2000). *Population structure and mating patterns of Killer Whales (Orcinus orca) as revealed by DNA analysis* (T). University of British Columbia. Retrieved October 26, 2023, from https://open.library.ubc.ca/collections/ubctheses/831/items/1.0099652

Secretariat, T. B. of C., & Secretariat, T. B. of C. (n.d.). *Northern Resident Killer Whale Group Cohesion (1980-2010)—Open Government Portal*. Retrieved September 29, 2023, from https://open.canada.ca/data/en/dataset/8c773994-1031-411b-a1ad-933928daa4ac

Stredulinsky, E.H., Darimont, C.T., Barrett-Lennard, L. *et al.* Family feud: permanent group splitting in a highly philopatric mammal, the killer whale (*Orcinus orca*). *Behav Ecol Sociobiol* **75**, 56 (2021). https://doi.org/10.1007/s00265-021-02992-8

Wasser, S. K., Lundin, J. I., Ayres, K., Seely, E., Giles, D., Balcomb, K., Hempelmann, J., Parsons, K., & Booth, R. (2017). Population growth is limited by nutritional impacts on pregnancy success in endangered Southern Resident Killer Whales (Orcinus orca). *PLOS ONE*, *12*(6). https://doi.org/10.1371/journal.pone.0179824