DSC 520 Final Project

Name: Katie Briggs   
Title: Hatchery Adult Salmon Returns  
  
**Section 1 – Week 9**

**Introduction**

Hatcheries provide are used for stock enhancement or also known as restoration aquaculture. These fish were raised within the hatchery and then released to supplement to population of recreational, commercial, and ecologically important species [1]. One of these important species is the King Salmon. Commercial fishing farms require a steady, predictable source of juveniles from hatcheries to stay in production and provide a consistent product. Within the past decade there has been a decline in King Salmon, with less making the run up to Alaska to spawn within their rivers. To preserve these magnificent creatures, the state has banned catch and release fishing on their rivers. Wild and hatchery smolt are missing and biologist know that they made it out to sea.

**Research questions**

1. **Are King Salmon really in decline?**
2. **What other species, if any, are also in decline of returning to the hatcheries?**
3. **Are hatcheries truly showing restoration aquaculture with the multiple species of Salmon?**
4. **Which run has the largest effect of returned Salmon – Winter, Spring, Summer or Fall?**
5. **Are there other inputs may influence returning Salmon, more jacks than female?**

Approach

To address some of these research questions, I collected multiple data sources containing the needed information. The data sources used were within the last decade to answer the question that King Salmon are in decline. Variables that would need to be included are the number of species that have returned, sex, run time, which hatchery the salmon returned, which river it was retrieved and released from. First, I became familiar with the data and cleaned it, removing insignificant information and missing values. Some initial explorations such as histograms, normality assessment and outliers were conducted addressed, followed by scatterplots to visually observe any trends, and identify potential relationships between input variables.

I used 3 data sets from the Department of Fish and Wildlife which spans from 2010 to 2020. The data sets are Hatchery\_Adult\_Salmon\_Returns.csv, Hatchery\_Standards.csv, and Hatchery\_Programs.csv.

Data

* **Data source 1:** CSV data file Hatchery Adult Salmon Returns. This data set are records of Adult Salmon that have returned to the Hatchery. The events are separated into categories and include the number of fish handled. Event details include species, run, brood year, sex of fish, mark and tag information.
* **Data source 2:** CSV data file of Hatchery Program, which contains information describing the production goal and actual production of salmon species. Percentage survival rate of smolt to juvenile.
* **Data source 3:** CSV data file of Population Recovery Goals, which contains information describing all salmonid species and the population goal of each.
* **Data source 4:** CSV data file of Hatchery Standards as of 2018, which contains information about how relocation aquaculture is rated in percentage points.

I created a subset of Hatchery\_Adult\_Salmon\_Returns.CSV data file and selected the variables Event, Facility, Adult\_Count, Species, and Origin. The Adult\_Count mean was 39, with the most Origin of Salmon returning is from Hatcheries over the Wild.

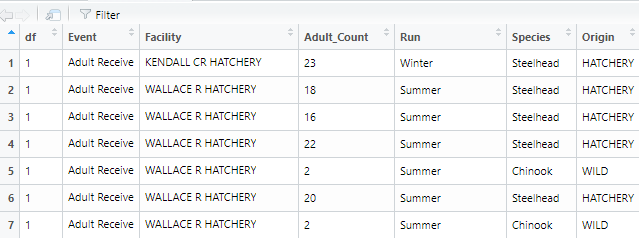
I created a subset data frame from the Hatchery\_Program.CSV data file and selecting the variables Facility\_Name, Hatchery\_Program\_Type, Species, Run, Production\_Goal, Actual\_Adult\_Hatchery\_Produced, AdultReturnsTotal, ReturningHatcheryPercent.

For the Production\_Goal the mean is 1,090,827 and Actual\_Adult\_Hatchery\_Produced mean is at 11,612 with ReturnHatcheryPercent at 47.28%.

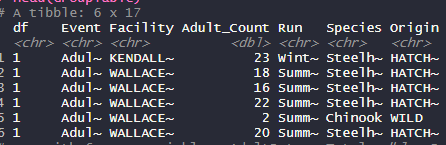
The third data set is Salmonid\_Stock\_Inventory.CVS and I created a subset, selecting the variables PopulationName, Species, Species, AbundanceQuantity, ProductionType, EscapementMethodology. I removed NAs and Unknowns from columns and sub population names.

I binded the three datasets together to create my working data and name it GroupTable.

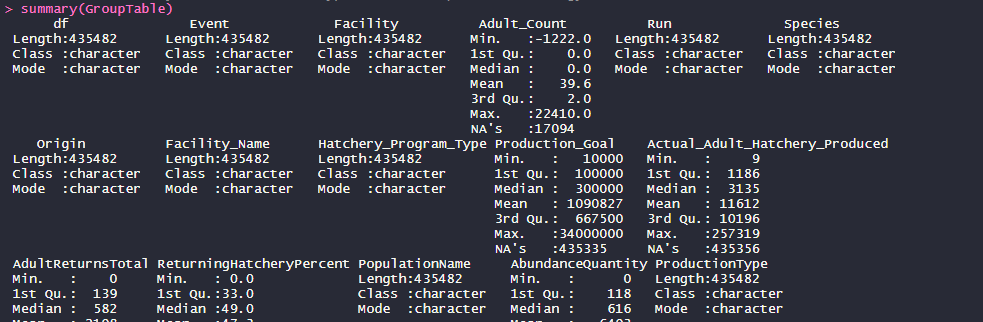
str(GroupTable)



head(GroupTable)



summary(GroupTable)



Required Packages

* Ggplot2 – for creating scatterplots and histograms
* Pastecs – descriptive statistics; quantifying shapes of distributions and outliers
* Psych – descriptive statistics; quantifying shapes of distributions and outliers
* Foreign – for binary multiple regression
* Ggm – partial correlations and its significance
* Hmisc – basic correlation coefficients
* QuantPsyc – standardized regression coefficients
* Car – regression diagnostics

Plots and Table Needs

* Histograms of each variable with normal distribution overlay to determine normality.
  + Adult\_Count

A close up of a mans face

Description automatically generated

* + Actual\_Adult\_Hatchery\_Produced

A screenshot of a cell phone

Description automatically generated

* + AbundanceQuantity

A picture containing screenshot

Description automatically generated

* + Species
    - Chinook, Coho and Chum Salmon are the most salminod returning.

A picture containing kitchen

Description automatically generated

* Q-Q plots and descriptive statistics of each variable to determine normality
  + Adult\_Count

A screenshot of a cell phone

Description automatically generated

* Abundance

A screenshot of a social media post

Description automatically generated

* Covariance, correlation coefficients, correlation matrix of all variables, coefficient of determination and scatterplots to determine relationships.
  + Actual Salmon Produced and Returns of Adults
    - We see that there is a linear relationship between these two variables.

A close up of text on a white surface

Description automatically generated

Stats

Call:

lm(formula = Actual\_Adult\_Hatchery\_Produced ~ AdultReturnsTotal,

data = Hatch\_subset)

Coefficients:

(Intercept) AdultReturnsTotal

8324.857 1.925

Shapiro-Wilk normality test

data: Hatch\_subset$Actual\_Adult\_Hatchery\_Produced

W = 0.38994, p-value < 2.2e-16

Shapiro-Wilk normality test

data: Hatch\_subset$AdultReturnsTotal

W = 0.55781, p-value < 2.2e-16

Pearson's product-moment correlation

data: Hatch\_subset$Actual\_Adult\_Hatchery\_Produced and Hatch\_subset$AdultReturnsTotal

t = 3.0254, df = 107, p-value = 0.00311

alternative hypothesis: true correlation is not equal to 0

99 percent confidence interval:

0.03825534 0.49196336

sample estimates:

cor

0.2807174

Regression analysis

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

A picture containing white, black

Description automatically generated

* Table to report results of multiple regression – change in R2, standardized betas, significance values, betas, standard errors

References

* [1] NOAA Fisheries, (n.d.) What is a Hatchery and why it is Important, Retrieved from: <https://www.fisheries.noaa.gov/node/52171>
* [2] Mauer, R. (2012), Decline in King Salmon is rooted in the Sea, State Biologist Says, Retrieved from: <https://www.adn.com/alaska-news/article/decline-king-salmon-rooted-sea-state-biologists-say/2012/06/24/>
* Open Data Network, Hatchery Adult Salmon Return, Hatchery Standards 2018, Stock Inventory Population, Retrieved from: <https://data.wa.gov/Natural-Resources-Environment/WDFW-Salmonid-Stock-Inventory-Population-Escapemen/fgyz-n3uk>
* <https://www.opendatanetwork.com/dataset/data.wa.gov/d8mu-pcf6>
* <https://www.opendatanetwork.com/dataset/data.wa.gov/xms2-7pwe>
* <https://www.opendatanetwork.com/dataset/data.wa.gov/9q4e-xhag>