```
---
title: "02_Variance_Stabilization_and_Differencing"
author: "Callum Weinberg"
date: "December 1, 2021"
output: pdf_document
---
```

````
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```
````

Libraries

````
```{r}
library(dplyr, warn.conflicts = FALSE) #Using
library(tidyr) #Using
library(knitr) #Using
library(lubridate, warn.conflicts = FALSE) #Using
library(ggplot2) #Using
library(MASS) #Uncertain
library(qpcR) #Using
library(forecast) #Using
library(cowplot) #Using
library(TSA) #Using
```
````

Part 2a: Load Data from 01

````
```{r}
load(file="Data/landings.Rdata")
```
````

Check if the Variance is Stable with a Histogram

````
```{r}
# Histogram Pounds of Red Sea Urchin
pounds_histogram = ggplot(landings, aes(x = pounds/1000)) +
  geom_histogram(binwidth = 50) +
  geom_density(aes(y = 50* ..count..),alpha = 0.05, fill = "red") +
  labs(x = "Thousands of Pounds of Red Sea Urchin", y = "Frequency") +
  #scale_x_continuous(label = comma) +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 15),
    legend.text = element_text(size = 15),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 12),
    axis.text.x = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))
pounds_histogram

# Display Variance
var(landings$pounds)/100000

## Plots for report
png(filename = "Images/histogram_raw_data.png", width = 960, height = 480)
pounds_histogram
dev.off()
```
````

Stabilize Variance

````
```{r}
## Stabalize Variance

## Apply Transformation, Homoskedastic
#find optimal lambda for Box-Cox transformation
t = 1:length(landings$pounds)
bcTransform = boxcox((landings$pounds) ~ t,plotit = TRUE)

lambda = bcTransform$x[which(bcTransform$y == max(bcTransform$y))]
print(lambda)

# Use .8 for simplicity
lambda_chosen = .8

# Box Cox Transformation, Log and Sqrt don't work well
# Confidence Interval Includes 1, using the Original Data
# and no Transformation
````

```
landings =
  landings %>%
  #mutate(pounds.bc = (1/lambda_chosen)*((pounds)^lambda_chosen-1))
  #mutate(pounds.bc = log(pounds))
  mutate(pounds.bc = pounds) # Note, if no transformation used in final just get rid of this

## Plot Box-Cox for report
png(filename = "Images/box_cox.png", width = 480, height = 480)
bcTransform = boxcox((landings$pounds) ~ t,plotit = TRUE)
dev.off()
```


Difference at Lag 12 to Remove Seasonality

```{r}
# Try 12, Since Data Retrieved Monthly

# Difference
landings_diff12 = landings[13:132,]
landings_diff12$pounds.bc_diff12 = diff(landings$pounds.bc,lag = 12)

# Plot Time Series after Differencing
full_plot_diff12 = ggplot(data = landings_diff12, mapping = aes(x = date, y = pounds.bc_diff12/1000)) +
  geom_line() +
  geom_smooth(method='lm', formula= y~x, se = FALSE) +
  labs(x = "Date", y = "Thousands of Pounds of Red Sea Urchin\nDifferenced at Lag 12") +
  #labs(title = "Red Sea Urching Landings 2008-2018\nDifferenced at Lag 12") +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 10),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 12),
    axis.text.x = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))
full_plot_diff12

# Plot Histogram after Differencing
pounds_histogram_diff12 =
  ggplot(landings_diff12, aes(x = pounds.bc_diff12/1000)) +
  geom_histogram(binwidth = 50) +
  geom_density(aes(y = 50* ..count..),alpha = 0.05, fill = "red") +
  labs(x = "Thousands of Pounds of Red Sea Urchin\nDifferenced at Lag 12", y = "Frequency") +
  #scale_x_continuous(label = comma) +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 15),
    legend.text = element_text(size = 15),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 12),
    axis.text.x = element_text(size = 12),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))
pounds_histogram_diff12

# Display Variance after This Difference
var(landings_diff12$pounds.bc_diff12)/100000


## Save the Transformed Data Out
landings_transformed_season_only = landings_diff12
i = ncol(landings_transformed_season_only)
names(landings_transformed_season_only)[i] = "pounds_transformed"
save(landings_transformed_season_only,file="Data/landings_transformed_season_only.Rdata")


## Plots for Report
png(filename = "Images/2008_2018_diff12.png", width = 960, height = 480)
plot_grid(full_plot_diff12,pounds_histogram_diff12, labels = NULL, label_size = 12, ncol = 2, nrow = 1)
dev.off()
```


Difference to Remove Trend After Seasonality Difference

```
# NOT USED IN MAIN REPORT
```{r}
# Difference at Lag 1
landings_diff12_diff1 = landings_diff12[2:120,]
```

```r
landings_diff12_diff1$pounds.bc_diff12_1 = diff(landings_diff12$pounds.bc_diff12,lag = 1)


# Plot the Data
full_plot = ggplot(data = landings_diff12_diff1, mapping = aes(x = date, y = pounds.bc_diff12_1)) +
  geom_line() +
  geom_smooth(method='lm', formula= y~x) +
  labs(x = "Date", y = "Pounds Transformed",
       title = "Red Sea Urching Landings 2008-2018\nDifferenced at Lag 12 and at Lag 1") +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 10),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 10),
    axis.text.x = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))
full_plot

# Calculate the Variance of the Differenced Data
var(landings_diff12_diff1$pounds.bc_diff12_1)/100000

# Histogram Pounds of Red Sea Urchin for Transformed Data
pounds_histogram =
  ggplot(landings_diff12_diff1, aes(x = pounds.bc_diff12_1)) +
  geom_histogram(aes(y = ..density..)) +
  geom_density(alpha = 0.1, fill = "red") +
  #labs(title = "By ZCTA") +
  labs(x = "Pounds of Red Sea Urchin", y = "Frequency") +
  #scale_x_continuous(label = comma) +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 15),
    legend.text = element_text(size = 15),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 10),
    axis.text.x = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))
pounds_histogram

## Save the Transformed Data Out
landings_transformed = landings_diff12_diff1
i = ncol(landings_transformed)
names(landings_transformed)[i] = "pounds_transformed"
save(landings_transformed,file="Data/landings_transformed.Rdata")

```


De-Trend Without The Seasonality Difference

NOT USED IN MAIN REPORT
```{r}
# Consider Other Possibilities in Code Below (i.e. difference twice no season, )
landings_diff12_diff1 = landings[2:132,]
landings_diff12_diff1$pounds.bc_diff12_1 = diff(landings$pounds.bc,lag = 1)

# Consider a second difference, but increases the variance
#landings_diff12_diff1_1 = landings[2:131,] # NEED TO SHOW IN FINAL CODE THAT THIS INCREASES THE VARIANCE
#landings_diff12_diff1_1$pounds.bc_diff12_1_1 = diff(landings_diff12_diff1$pounds.bc_diff12_1,lag = 1)


# Plot the Data
full_plot = ggplot(data = landings_diff12_diff1, mapping = aes(x = date, y = pounds.bc_diff12_1)) +
  geom_line() +
  geom_smooth(method='lm', formula= y~x) +
  labs(x = "Date", y = "Pounds Transformed",
       title = "Red Sea Urching Landings 2008-2018\nDifferenced at Lag 12 and at Lag 1") +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 10),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 10),
    axis.text.x = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))
full_plot

# Calculate the Variance of the Differenced Data
var(landings_diff12_diff1$pounds.bc_diff12_1)/100000
```

```
# Histogram Pounds of Red Sea Urchin for Transformed Data
pounds_histogram =
  ggplot(landings_diff12_diff1, aes(x = pounds.bc_diff12_1)) +
  geom_histogram(aes(y = ..density..)) +
  geom_density(alpha = 0.1, fill = "red") +
  #labs(title = "By ZCTA") +
  labs(x = "Pounds of Red Sea Urchin", y = "Frequency") +
  #scale_x_continuous(label = comma) +
  theme(text = element_text(size = 20),
    legend.title = element_text(size = 15),
    legend.text = element_text(size = 15),
    legend.key.width=unit(1,"cm"),
    axis.text.y = element_text(angle=90, hjust=1, size = 10),
    axis.text.x = element_text(size = 10),
    plot.title = element_text(hjust = 0.5, size = 12),
    axis.title=element_text(size=10,face="bold"))
pounds_histogram

## Save the Transformed Data Out
landings_transformed_no_season = landings_diff12_diff1
i = ncol(landings_transformed_no_season)
names(landings_transformed_no_season)[i] = "pounds_transformed"
save(landings_transformed_no_season,file="Data/landings_transformed_no_season.Rdata")
```