Analysis of the concentration of the plant pigment chlorophyll-a in the Coastal transect of Great Barrier Reef

Project 3

490620014

University of Sydney | DATA1001 | May 2020

1 Recommendaton

Study shows that in Coastal transects of Great Barrier Reef, mean concentration of chlorophyll-a in summer and winter season is equal. The study also compares mean concentration for all months. This could potentially be used by our clients, the marine tourism industry, and Great Barrier Reef Marine Park Authority (GBRMPA). They both play important role in ensuring sustainable use and protection of reef. Water sports have negative impact on concentration of chlorophyll-a, an important parameter in calculating Water Quality Index. Hence, they effect reefs negatively. Our clients can use this report to decide permitted marine sports in different months.

2 Evidence

The data has been collected by fully-trained community groups, tourism operators and government agencies under the chlorophyll monitoring under the Reef Plan Marine Monitoring Programme, managed by the Australian Institute of Marine Science. This study explores the samples collected in the Coastal transect of Great Barrier Reef. The data has certain limitations, and is not completely reliable. "AIMS make no representation or warranty that the data, products and services are accurate, complete, reliable or current.", as mentioned on their website. The omitting of NA's has resulted in limitation of findings. The methods used for sampling are not completely accurate, they are prone to chance and calculation errors. Some methods of sampling are harmful for marine life, this is where ethics come in. Marine biologists are working to devise better and more efficient methods of sampling.

2.1 Initial Data Analysis

Reading data.

The primary objective of this report is exploring the Coastal transect.

We are adding the month and the season in which the data was recorded column to our data.

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:lubridate':
##
##     intersect, setdiff, union
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

The dimension and structure of the variables.

2.1.1 Variable classifications

Temperature

```
##  num [1:1027] 28 28 31 31 31 31 26.1 26.1 25.4 25.4 ...
```

Salinity

```
##  num [1:1027] 35 35 37 NA 37 NA 38 38 40 40 ...
```

Concentration of the plant pigment chlorophyll-a (in μg/L)

```
##  num [1:1027] 1.02 0.92 0.54 0.54 0.62 0.62 0.56 0.54 0.45 0.42 ...
```

Transect

```
##  Factor w/ 14 levels "Burnett-Hervey",..: 4 4 4 4 4 4 4 4 4 4 ...
```

No changes in variable classification are required.

2.1.2 Data Wrangling

Finding IQR and quantile in order to remove the outliers from Temperature

Finding IQR and quantile in order to remove the outliers from Salinity

Finding IQR and quantile in order to remove the outliers from Chlorophyll

Removing the outliers

Subsetting the data;removing the outliers from specific variables

2.1.3 Final dataset

2.1.4 The dimension of the final dataset.

```
## [1] 249  17
```

2.1.5 Variable classification

Classification of variables of the final dataset are the same as before.

Variable classification of specific variables:

Temperature

```
##  num [1:249] 28 28 26.1 26.1 25.4 25.4 24.7 24.7 24.4 23.9 ...
```

Salinity

```
##  num [1:249] 35 35 38 38 40 40 36 36 38 37 ...
```

Concentration of the plant pigment chlorophyll-a (in μg/L)

```
##  num [1:249] 1.02 0.92 0.56 0.54 0.45 0.42 0.34 0.25 0.13 0.55 ...
```

Transect

```
##  Factor w/ 14 levels "Burnett-Hervey",..: 4 4 4 4 4 4 4 4 4 4 ...
```

The variable classification is the same as before.

2.2 2 sample T test (2-sided alternative hypothesis)

Is there a difference in the mean concentration of chlorophyll-a in summer and winter season?

Sample 1 is data set for which sample has been recorded in summer season. Sample 2 is data set for which sample has been recorded in winter season.

Let $\mu 1$ = mean concentration of chlorophyll-a in summer

Let $\mu 2$ = mean concentration of chlorophyll-a in winter

2.2.1 H Hypothesis

H0 : There is no difference: $\mu 1=\mu 2$, or $\mu 1-\mu 2=0$.

H1 : There is a difference: $\mu 1 \neq \mu 2$, or $\mu 1-\mu 2 \neq 0$.

2.2.2 A Assumptions

1- The 2 samples are independent

2- The 2 populations have equal spread (SD/variance)

3- The 2 populations are Normal

2.2.3 T Test Statistic

```
## [1] 0.4851811
```

2.2.4 P p-value

```
## [1] 0.6279779
```

2.2.5 C Conclusion

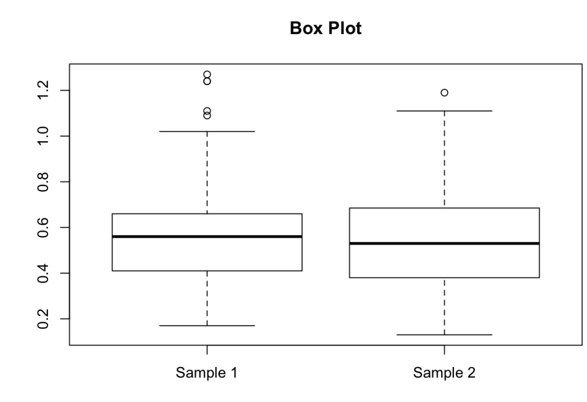
As the p-value > 0.05, we retain the null hypothesis.

2.2.6 Checking our assumptions

1- The 2 samples are independent The two samples have been recorded separately at different time periods.

2- The 2 populations have equal spread (SD/variance) 3- The 2 populations are Normal

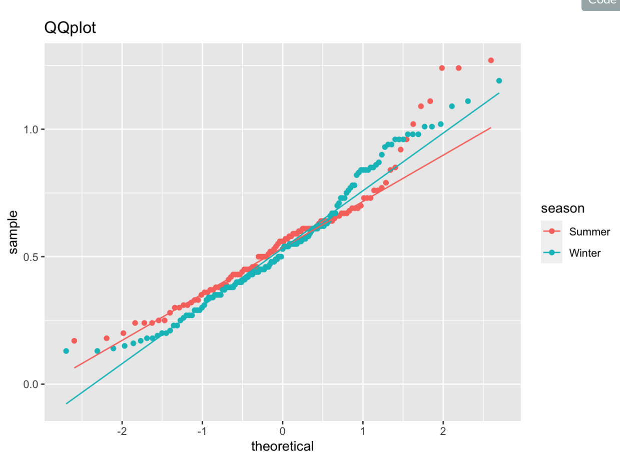
2.2.6.1 Creating Comparative Boxplots - for normality and equality of variance assumptions



Normality: the 2 samples look symmetrical and so are consistent with Normality, although the sample one looks a bit asymmetric. Equality of variance: the 2 samples have similar spread and so are consistent with the equality of variance assumption.

2.2.6.2 Q-Q Plot - for Normality

```
## Loading required package: ggplot2
```



A quantile-quantile plot (Q-Q Plot) graphs the theoretical quantiles based on the normal curve against the actual quantiles. The line formed by the points is not reasonably straight, so we can't assume that the data is normally distributed.

2.2.6.3 Shapiro-Wilk Test - for Normality

```
##
##  Shapiro-Wilk normality test
##
## data:  gbr_data$CHL_A[which(gbr_data$season == "Summer")]
## W = 0.93556, p-value = 6.448e-05
```

```
##
##  Shapiro-Wilk normality test
##
## data:  gbr_data$CHL_A[which(gbr_data$season == "Winter")]
## W = 0.96911, p-value = 0.00255
```

Both tests give small p-values (p-value<0.05), so considering the previous plots, we conclude that the population might not be Normally distributed.

2.2.6.4 Levene's Test (F-Test) - for equal spread

```
##
##  F test to compare two variances
##
## data:  gbr_data$CHL_A[which(gbr_data$season == "Summer")] and gbr_data
$CHL_A[which(gbr_data$season == "Winter")]
## F = 0.84557, num df = 105, denom df = 142, p-value = 0.3648
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.5937218 1.2160470
## sample estimates:
## ratio of variances
##           0.845573
```

The test gives large p-value (p-value>0.05),suggesting that data is consistent with equal variance.

2-The 2 populations have equal spread (SD) The boxplots suggest that the variation is similar between the 2 populations, as is confirmed by the F-Test. 3-The 2 populations are Normal The 2 samples suggest the 2 populations might not be Normal, from both the boxplots, the Q-Q Plot and the Shapiro-Wilk Test. Hence we are not certain that the conclusion that we derived from the 2 Sample T Test is true.

Since our 2 samples suggest Non-Normality, we can use Transformations or Non Parametric Tests.

2.2.7 Mann-Whitney-Wilcoxon test:

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  gbr_data$CHL_A[which(gbr_data$season == "Summer")] and gbr_data
$CHL_A[which(gbr_data$season == "Winter")]
## W = 7999.5, p-value = 0.4547
## alternative hypothesis: true location shift is not equal to 0
```

The test gives large p-value (p-value>0.05),so we retain the Null hypothesis.

2.2.8 Conclusion

Statistical conclusion:

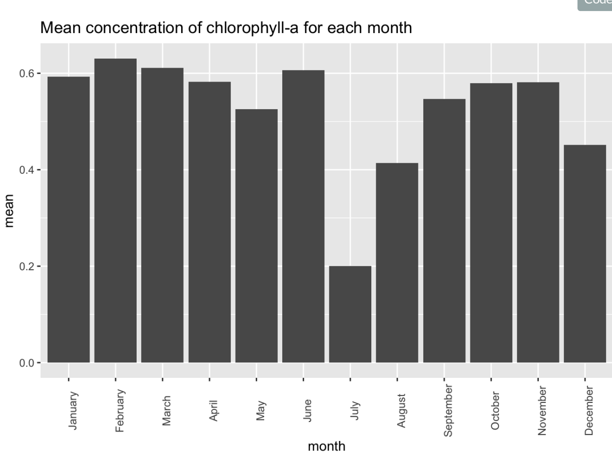
As the p-value > 0.05, we retain the null hypothesis.

Thus data is consistent with hypothesis that the mean concentration of chlorophyll-a is equal.

Scientific conclusion:

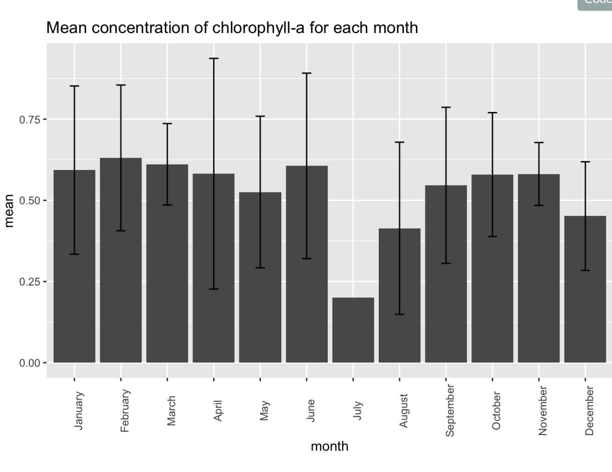
The data suggests that the change in season does not has an effect on mean concentration of chlorophyll-a.

2.2.9 Mean concentration of chlorophyll-a for every month



The above plot suggests that the mean concentration of chlorophyll-a for July is considerably low as compared to other months.

2.2.9.1 Adding error bars to the above plot



There is no error bar for month of July. This is because there is only one sample collected in July in dataset. Thus we cannot say that mean concentration of chlorophyll-a for July is considerably low as compared to other months. The mean for all months except July are close, but the error bars are large and overlapping hence we are not confident that mean concentration of chlorophyll-a for all months except July is close.

3 Summary

We have low confidence that mean concentration of chlorophyll-a for every month are close to each other, but mean chlorophyll concentration for summer and winter season is equal in Coastal transects. The World Heritage listed reef attracts tourists throughout the year. People participate in numerous water sports. Motorised water sports lead to depletion of nutrients and concentration of chlorophyll-a. Water sports like scuba diving and snorkelling can damage corals. Marine Tourism Industries, and GBRMPA can use this data to control number of motor sport activities allowed every month in order to protect Great Barrier Reef. This way we can protect it without prohibiting all recreational marine activities.

4 Other Evidence

This study reviews the impact of the major components of marine tourism on the GBR and tourism management, and concludes that reef tourism produces generally localized impacts and is intensively managed on the GBR relative to other reef uses (Harriott, 2001). This study agrees with our report, but it states that effect of tourism has small impact as compared to other factors.

It showcases how commitment and long term-cooperation between tourism sector, scientists, and World Heritage management authorities can benefit the overall management performance of a World Heritage site. (UNESCO, 2006). This study is consistent with our recommendation.

4.0.1 References

eAtlas,2004,The Great Barrier Reef Long-Term Chlorophyll Monitoring System,Eatlas.org.au,https://eatlas.org.au/gbr/ltmp-data

ggplot visualisations, http://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html

Harriott, 2001, Marine tourism impacts on the Great Barrier Reef, researchGate , https://www.researchgate.net/publication/233568910_Marine_tourism_impacts_on_the_Great_Barrier_Re

Harriott,2003, Rrrc.org.au, https://www.rrrc.org.au/wp-content/uploads/2014/04/Technical-Report-46.pdf

Reef Plan,2016,Great Barrier Reef Report Card 2016| Reef Water Quality Protection Plan,Reefplan.qld.gov.au,https://www.reefplan.qld.gov.au/__data/assets/pdf_file/0021/46164/report-card-2016-marine-methods.pdf

Research Data Australia,The Great Barrier Reef Long-term Chlorophyll Monitoring,researchdata.ands.org,https://researchdata.ands.org.au/great-barrier-reef-1992-2009/677311

Schaffelke et al.,2009,Reef Rescue Marine Monitoring Program,rrrc.org.au,http://rrrc.org.au/wp-content/uploads/2014/05/371b-378_AIMS_2008-09_Final-report.pdf

UNESCO,2006,Guide 4 – Case study: Great Barrier Reef (Australia) | UNESCO Sustainable Tourism Toolkit, Whc.unesco.org, http://whc.unesco.org/sustainabletourismtoolkit/guide-4-%E2%80%93-case-study-great-barrier-reef-australia