T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
FEN FAKÜLTESİ
İSTATİSTİK BÖLÜMÜ

PROJE BAŞLIĞI

Bitirme Projesi Raporu

Pelin PEKER
Merve AK
Edanur Binnaz DURSUN
Ahmet ÇALI

Şubat 2024

Rapor Değerlendirme

"PROJE BAŞLIĞI" başlıklı bitirme projesi raporu tarafımdan okunmuş, kapsamı ve niteliği açısından bir Bitirme Projesi raporu olarak kabul edilmiştir.

Dr. Engin YILDIZTEPE

Teşekkür

Tüm çalışma süresince yönlendiriciliği, katkıları ve yardımları ile yanımızda olan danışmanımız Dr. Engin YILDIZTEPE 'ye ve böyle bir çalışmayı yapmamız için bize fırsat tanıyan Dokuz Eylül Üniversitesi Fen Fakültesi İstatistik Bölümüne teşekkür ederiz.

Pelin PEKER
Merve AK

Özet

Özet, çalışmanın önemini ve faydasını anlatan bir bölüm degildir. Çalısmayı ana hatlarıyla anlatacak ve 300 kelimeyi aşmayacak şekilde hazırlanmalıdır. En az üç en çok beş anahtar kelime ilgili yere yazılmalıdır.

ikinci paragraf buradan başlar

Anahtar Kelimeler: anahtar kelime 1, anahtar kelime 2, anahtar kelime 3

Abstract

The preface pretty much says it all.

Second paragraph of abstract starts here.

Keywords: keyword1, keyword2, keyword3

Contents

List of Tables

List of Figures

Introduction

Welcome to the *R Markdown* thesis template. This template is based on (and in
many places copied directly from) the Reed College LaTeX template, but hopefully
it will provide a nicer interface for those that have never used TeX or LaTeX before.
Using *R Markdown* will also allow you to easily keep track of your analyses in **R**
chunks of code, with the resulting plots and output included as well. The hope is this
R Markdown template gets you in the habit of doing reproducible research, which
benefits you long-term as a researcher, but also will greatly help anyone that is trying
to reproduce or build onto your results down the road.

Hopefully, you won't have much of a learning period to go through and you will
reap the benefits of a nicely formatted thesis. The use of LaTeX in combination with
Markdown is more consistent than the output of a word processor, much less prone to
corruption or crashing, and the resulting file is smaller than a Word file. While you
may have never had problems using Word in the past, your thesis is likely going to be
about twice as large and complex as anything you've written before, taxing Word's
capabilities. After working with *Markdown* and **R** together for a few weeks, we are
confident this will be your reporting style of choice going forward.

Why use it?

R Markdown creates a simple and straightforward way to interface with the beauty
of LaTeX. Packages have been written in **R** to work directly with LaTeX to produce
nicely formatting tables and paragraphs. In addition to creating a user friendly
interface to LaTeX, *R Markdown* also allows you to read in your data, to analyze
it and to visualize it using **R** functions, and also to provide the documentation and
commentary on the results of your project. Further, it allows for **R** results to be
passed inline to the commentary of your results. You'll see more on this later.

Who should use it?

Anyone who needs to use data analysis, math, tables, a lot of figures, complex
cross-references, or who just cares about the final appearance of their document
should use *R Markdown*. Of particular use should be anyone in the sciences, but the
user-friendly nature of *Markdown* and its ability to keep track of and easily include
figures, automatically generate a table of contents, index, references, table of figures,
etc. should make it of great benefit to nearly anyone writing a thesis project.

Chapter 1

R Markdown Basics

Here is a brief introduction into using *R Markdown*. *Markdown* is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. *R Markdown* provides the flexibility of *Markdown* with the implementation of **R** input and output. For more details on using *R Markdown* see `http://rmarkdown.rstudio.com`.

Be careful with your spacing in *Markdown* documents. While whitespace largely is ignored, it does at times give *Markdown* signals as to how to proceed. As a habit, try to keep everything left aligned whenever possible, especially as you type a new paragraph. In other words, there is no need to indent basic text in the Rmd document (in fact, it might cause your text to do funny things if you do).

1.1   Lists

It's easy to create a list. It can be unordered like

- Item 1
- Item 2

or it can be ordered like

1. Item 1
2. Item 2

Notice that I intentionally mislabeled Item 2 as number 4. *Markdown* automatically figures this out! You can put any numbers in the list and it will create the list. Check it out below.

To create a sublist, just indent the values a bit (at least four spaces or a tab). (Here's one case where indentation is key!)

1. Item 1
2. Item 2
3. Item 3
    - Item 3a
    - Item 3b

1.2   Line breaks

Make sure to add white space between lines if you'd like to start a new paragraph. Look at what happens below in the outputted document if you don't:

Here is the first sentence. Here is another sentence. Here is the last sentence to end the paragraph. This should be a new paragraph.

Now for the correct way:

Here is the first sentence. Here is another sentence. Here is the last sentence to end the paragraph.

This should be a new paragraph.

1.3   R chunks

When you click the **Knit** button above a document will be generated that includes both content as well as the output of any embedded **R** code chunks within the document. You can embed an **R** code chunk like this (`cars` is a built-in **R** dataset):

```r
summary(cars)
```

```
    speed            dist
Min.   : 4.0   Min.   :  2.00
1st Qu.:12.0   1st Qu.: 26.00
Median :15.0   Median : 36.00
Mean   :15.4   Mean   : 42.98
3rd Qu.:19.0   3rd Qu.: 56.00
Max.   :25.0   Max.   :120.00
```

1.4   Inline code

If you'd like to put the results of your analysis directly into your discussion, add inline code like this:

> The `cos` of 2π is 1.

Another example would be the direct calculation of the standard deviation:

> The standard deviation of `speed` in `cars` is 5.2876444.

One last neat feature is the use of the `ifelse` conditional statement which can be used to output text depending on the result of an **R** calculation:

> The standard deviation is less than 6.

Note the use of `>` here, which signifies a quotation environment that will be indented.

As you see with `$2 \pi$` above, mathematics can be added by surrounding the mathematical text with dollar signs. More examples of this are in Mathematics and Science if you uncomment the code in Math.

1.5   Including plots

You can also embed plots. For example, here is a way to use the base **R** graphics package to produce a plot using the built-in `pressure` dataset:
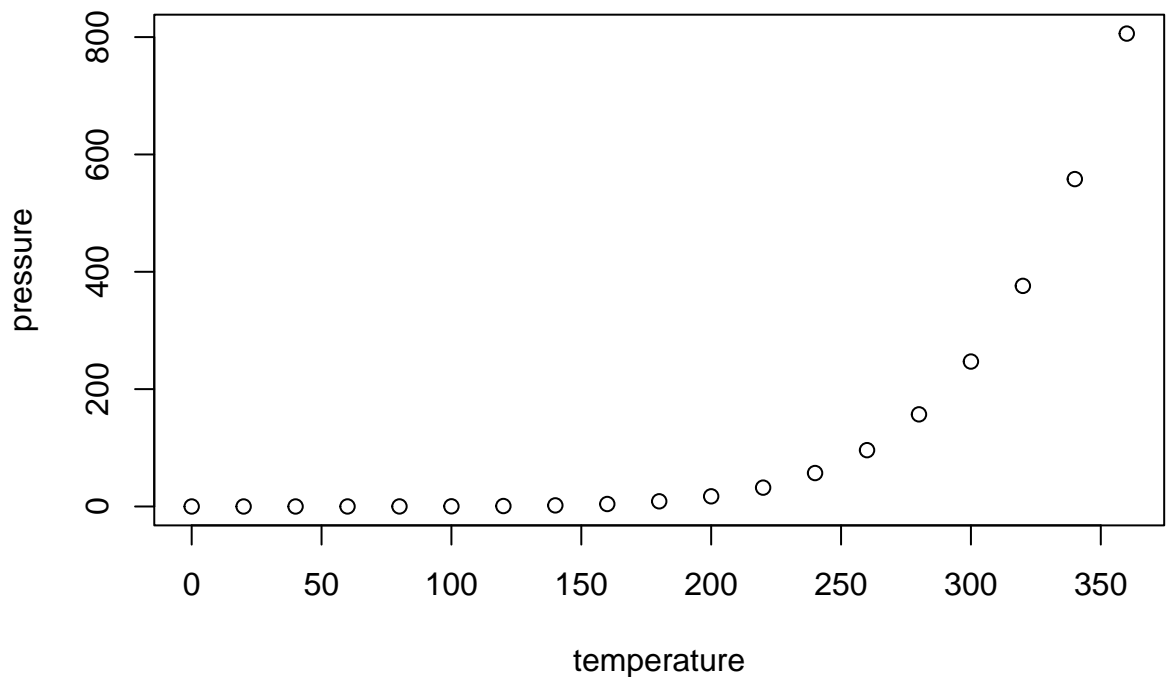
Note that the `echo=FALSE` parameter was added to the code chunk to prevent printing of the **R** code that generated the plot. There are plenty of other ways to add chunk options. More information is available at `http://yihui.name/knitr/options/`.

Another useful chunk option is the setting of `cache=TRUE` as you see here. If document rendering becomes time consuming due to long computations or plots that are expensive to generate you can use knitr caching to improve performance. Later in this file, you'll see a way to reference plots created in **R** or external figures.

1.6   Loading and exploring data

Included in this template is a file called `flights.csv`. This file includes a subset of the larger dataset of information about all flights that departed from Seattle and Portland in 2014. More information about this dataset and its **R** package is available at `http://github.com/ismayc/pnwflights14`. This subset includes only Portland flights and only rows that were complete with no missing values. Merges were also done with the `airports` and `airlines` data sets in the `pnwflights14` package to get more descriptive airport and airline names.

We can load in this data set using the following command:

```r
flights <- read.csv("data/flights.csv")
```

The data is now stored in the data frame called `flights` in **R**. To get a better feel for the variables included in this dataset we can use a variety of functions. Here we can see the dimensions (rows by columns) and also the names of the columns.

```r
dim(flights)
```

```
[1] 52808    16
```

```r
names(flights)
```

```
 [1] "month"        "day"         "dep_time"    "dep_delay"   "arr_time"
 [6] "arr_delay"    "carrier"     "tailnum"     "flight"      "dest"
[11] "air_time"     "distance"    "hour"        "minute"      "carrier_name"
[16] "dest_name"
```

Another good idea is to take a look at the dataset in table form. With this dataset having more than 50,000 rows, we won't explicitly show the results of the command here. I recommend you enter the command into the Console *after* you have run the **R** chunks above to load the data into **R**.

```r
View(flights)
```

While not required, it is highly recommended you use the `dplyr` package to manipulate and summarize your data set as needed. It uses a syntax that is easy to understand using chaining operations. Below I've created a few examples of using `dplyr` to get information about the Portland flights in 2014. You will also see the use of the `ggplot2` package, which produces beautiful, high-quality academic visuals.

We begin by checking to ensure that needed packages are installed and then we load them into our current working environment:

```r
# List of packages required for this analysis
#pkg <- c("dplyr", "ggplot2", "knitr", "bookdown", "remotes")
pkg <- c("tidyverse", "knitr", "bookdown", "remotes")
# Check if packages are not installed and assign the
# names of the packages not installed to the variable new.pkg
new.pkg <- pkg[!(pkg %in% installed.packages())]
# If there are any packages in the list that aren't installed,
# install them
if (length(new.pkg))
  install.packages(new.pkg)
```

```r
# Load packages (thesisdown will load all of the packages as well)
library(thesisdown)
#library(dplyr)
#library(ggplot2)
library(tidyverse)
library(bookdown)
library(knitr)
```

The example we show here does the following:

- Selects only the `carrier_name` and `arr_delay` from the `flights` dataset and then assigns this subset to a new variable called `flights2`.

- Using `flights2`, we determine the largest arrival delay for each of the carriers.

```
flights2 <- flights %>%
  select(carrier_name, arr_delay)
max_delays <- flights2 %>%
  group_by(carrier_name) %>%
  summarize(max_arr_delay = max(arr_delay, na.rm = TRUE))
```

A useful function in the `knitr` package for making nice tables in *R Markdown* is called `kable`. It is much easier to use than manually entering values into a table by copying and pasting values into Excel or LaTeX. This again goes to show how nice reproducible documents can be! (Note the use of `results="asis"`, which will produce the table instead of the code to create the table.) The `caption.short` argument is used to include a shorter title to appear in the List of Tables.

```
kable(max_delays,
      col.names = c("Airline", "Max Arrival Delay"),
      caption = "Maximum Delays by Airline",
      caption.short = "Max Delays by Airline",
      longtable = TRUE,
      booktabs = TRUE)
```

Table 1.1: Maximum Delays by Airline

Airline	Max Arrival Delay
Alaska Airlines Inc.	338
American Airlines Inc.	1539
Delta Air Lines Inc.	651
Frontier Airlines Inc.	575
Hawaiian Airlines Inc.	407
JetBlue Airways	273
SkyWest Airlines Inc.	421
Southwest Airlines Co.	694
US Airways Inc.	347
United Air Lines Inc.	472
Virgin America	366

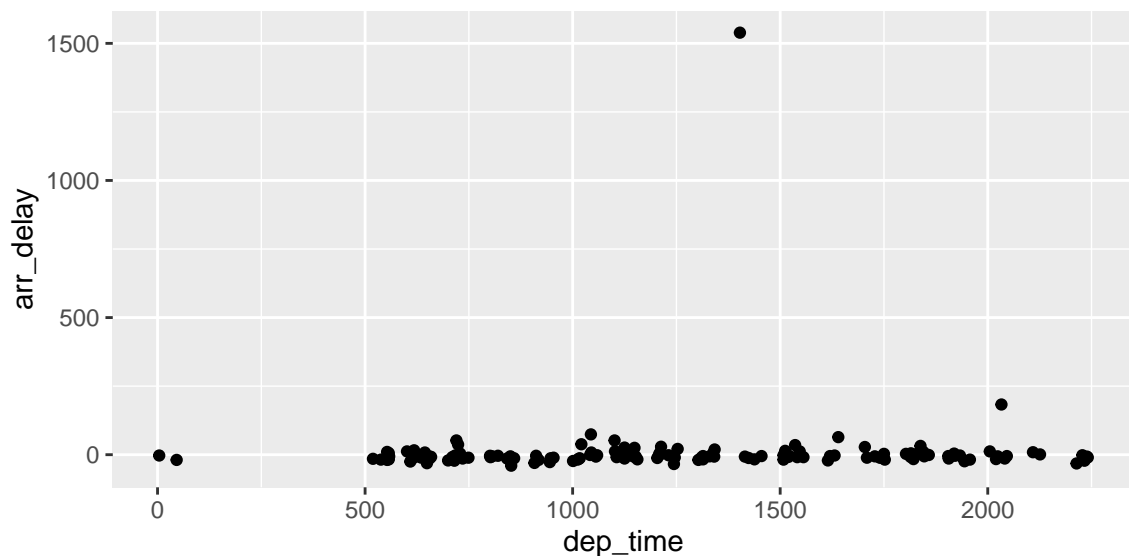The last two options make the table a little easier-to-read.

We can further look into the properties of the largest value here for American Airlines Inc. To do so, we can isolate the row corresponding to the arrival delay of 1539 minutes for American in our original `flights` dataset.

```
flights %>% filter(arr_delay == 1539,
                   carrier_name == "American Airlines Inc.") %>%
  select(-c(month, day, carrier, dest_name, hour,
            minute, carrier_name, arr_delay))
```

```
  dep_time dep_delay arr_time tailnum flight dest air_time distance
1     1403      1553     1934  N595AA   1568  DFW      182     1616
```

We see that the flight occurred on March 3rd and departed a little after 2 PM on its way to Dallas/Fort Worth. Lastly, we show how we can visualize the arrival delay of all departing flights from Portland on March 3rd against time of departure.

```
flights %>% filter(month == 3, day == 3) %>%
  ggplot(aes(x = dep_time, y = arr_delay)) + geom_point()
```


1.7   Additional resources

- *Markdown* Cheatsheet - `https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet`

- *R Markdown* Reference Guide - `https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf`

- Introduction to `dplyr` - `https://cran.rstudio.com/web/packages/dplyr/vignettes/introduction.html`

- `ggplot2` Documentation - `http://docs.ggplot2.org/current/`

Chapter 2

Mathematics and Science

2.1 Math

TeX is the best way to typeset mathematics. Donald Knuth designed TeX when he got frustrated at how long it was taking the typesetters to finish his book, which contained a lot of mathematics. One nice feature of *R Markdown* is its ability to read LaTeX code directly.

If you are doing a thesis that will involve lots of math, you will want to read the following section which has been commented out. If you're not going to use math, skip over or delete this next commented section.

2.2 ÇOKLU DEĞİŞİM NOKTASI

Chemical formulas will look best if they are not italicized. Get around math mode's automatic italicizing in LaTeX by using the argument `$\mathrm{formula here}$`, with your formula inside the curly brackets. (Notice the use of the backticks here which enclose text that acts as code.)

So, $\mathrm{Fe_2^{2+}Cr_2O_4}$ is written `$\mathrm{Fe_2^{2+}Cr_2O_4}$`.

Exponent or Superscript: O^-

Subscript: CH_4

To stack numbers or letters as in Fe_2^{2+}, the subscript is defined first, and then the superscript is defined.

Bullet: $CuCl \bullet 7H_2O$

Delta: Δ

Reaction Arrows: $\longrightarrow$ or $\xrightarrow{solution}$

Resonance Arrows: $\leftrightarrow$

Reversible Reaction Arrows: $\rightleftharpoons$

2.2.1 İkili Segmentasyon Algoritması

İkili segmentasyon algoritması (BinSeg), değişim noktası literatüründe kullanılan en köklü arama yöntemidir. İkili segmentasyon arama algoritmasının erken uygulamaları

arasında Scott ve Knott (1974) ile Sen ve Srivastava (1975) bulunmaktadır.

İkili segmentasyon, herhangi bir tek değişim noktası yöntemini ardışık olarak farklı veri setlerinde tekrarlayarak çoklu değişim noktalarına genişletmek için kullanılabilir. İlk olarak, tek bir değişim noktası test istatistiğini tüm veri setine uygular. $y_1, y_2, ..., y_n$ şeklindeki veri seti üzerinde bir başlangıç noktası belirlenir. Bu başlangıç noktası, veri setinin ortalaması, medyanı veya başka bir özelliği olabilir. Belirlenen başlangıç noktasında bir değişim noktası testi yapılır. Bu test, veri setini iki alt küme olarak böldüğünde, oluşan alt kümelerin toplam maliyetinin belirli bir kritere göre düşük olup olmadığını kontrol eder, yani bir τ'nin aşağıdaki koşulu sağlayıp sağlamadığını test eder:

$$C(y_{1:\tau}) + C(y_{(\tau+1):n}) + \beta < C(y_{1:n})$$

Burada:

- **C**:Bir segment için maliyet fonksiyonu

- $y_{1:\tau}$:Başlangıçtan değişim noktasına kadar olan veri seti

- $y_{(\tau+1):n}$:Değişim noktasından sona kadar olan veri seti

- β:Aşırı uyum karşısında koruma sağlayan ceza terimi

Eğer bu koşul sağlanmıyorsa, o zaman herhangi bir değişim noktası tespit edilememiştir ve algoritma durur. Aksi takdirde veri, belirlenen değişim noktasından önce ve sonra olmak üzere iki segmente bölünür. Tek değişim noktası tespit yöntemi, değişiklikten önce ve sonra iki yeni segmente de tekrarlanır. Her iki segmentte de değişim noktaları belirlenirse, bunları yeni belirlenen değişim noktasında daha fazla segmentlere böler ve her yeni segmente değişim noktası tespit yöntemini uygular. Bu süreç, verinin herhangi bir bölümünde değişim noktası bulunamayana kadar devam eder.

Çoklu değişim noktalarını belirlemek için yaygın olarak kullanılan bir yaklaşım, aşağıdaki ifadeyi minimize etmektir:

$$\sum_{i=1}^{m+1} C(y(\tau_{i-1} + 1) : \tau_i) + \beta f(m)$$

Burada, C bir segment için bir maliyet fonksiyonu ve $\beta f(m)$ aşırı uyum karşısında koruma sağlayan bir ceza terimidir.

İkili segmentasyon, herhangi bir değişim noktasının konumu önceden belirlenmiş değişim noktalarına bağlı olduğu için ($f(m) = m$ olarak) yukarıdaki denklemin yaklaşık bir minimize edilmesidir. Algoritmanın her adımı, bu denklemi azaltıyorsa ek bir değişim noktası eklemeye çalışır. İkili segmentasyon algoritmasının avantajı, n'nin veri uzunluğu olduğu durumda $O(n)$ hesaplama maliyeti ile uygulanabilen hızlı bir algoritma olmasıdır. Ancak, C'yi uygun bir şekilde seçmek zor olabilir ve farklı C seçimleri, değişim noktalarının sayısının tahmininde önemli farklara neden olabilir.

2.2.2 PELT

2.2.3 Parçalı Regresyon

Parçalı veya kesik çizgili modeller, yanıt ile bir veya daha fazla açıklayıcı değişken arasındaki ilişkilerin parçalı doğrusal olduğu, yani iki veya daha fazla değişkenle temsil edildiği regresyon modelleridir.Bu ilişkiler, genellikle bilinmeyen değerlerde birleştirilen iki veya daha fazla düz çizgi tarafından temsil edilir, bu değerlere genellikle kırılma noktaları, değişim noktaları veya birleşim noktaları denir. Bu yöntemde bağımsız değişken, aralıklara bölünür ve her bir aralığa ayrı bir çizgi segmenti uyarlanır.Parçalı regresyon analizi ayrıca çeşitli bağımsız değişkenlerle yapılan çok değişkenli veriler üzerinde de gerçekleştirilebilir. Parçalı regresyon analizi, bağımsız değişkenlerin belirli gruplara ayrıldığı durumlarda, bu gruplardaki değişkenler arasındaki ilişkilerin farklı olduğuna inanıldığında kullanışlıdır. Bu parçalar arasındaki sınırlar, değişim noktaları olarak adlandırılır.

Matematiksel olarak, model şu şekilde ifade edilebilir:

$y = \beta_{0i} + \beta_{1i}x + \epsilon$

Bu denklemde, β_{0i} ve β_{1i} sırasıyla i-inci segmentin kesişim noktası ve eğimini temsil eder.

Parçalı regresyon, ekonomi, biyoloji, çevre bilimleri, epidemiyoloji gibi çeşitli alanlarda kullanılır.Kalite iyileştirme müdahaleleriyle ilgili çalışmalarda parçalı regresyon analizlerinin kullanımına dair birçok örnek yayınlanmıştır.

Temel bir parçalı regresyon analizinde zaman periyodu müdahale öncesi ve sonrası parçalara bölünür ve her parçada ayrı ayrı kesişme noktaları ve eğimler tahmin edilir. Müdahaleden önce ve sonra kesişmelerde ve eğimlerdeki değişiklikleri test etmek için istatistiksel testler gerçekleştirilir. Veriler ve model spesifikasyonunda bazı basit değişikliklerle parçalı regresyon analizi, genellikle standart istatistiksel yazılım paketlerinde kolayca uygulanabilir. Genellikle, zaman içinde alınan gözlemler ilişkilidir, bu nedenle otokorelasyonu hesaba katmak için genellikle ek bir düzeltme yapılması gerekir.

Parçalı doğrusal regresyon, parçalı regresyonun doğrusal regresyon kullanılarak elde edilen bir alt türüdür.İki parçalı doğrusal regresyon, bir değişim noktası ile ayrılmış iki parçayla, değişken bir etkenin (x) yanıt fonksiyonunun (Y_r) ani bir değişikliğini nicelendirmek için kullanışlı olabilir. Değişim noktası, yanıt fonksiyonunun kritik, güvenli veya eşik değeri olarak yorumlanabilir; bu değerin ötesinde veya altında (istenmeyen) etkiler meydana gelebilir. Değişim noktası, karar verme süreçlerinde önemlidir.

Her bir parça için ayrı ayrı uygulanan en küçük kareler yöntemi, iki regresyon çizgisini veriye mümkün olduğunca iyi uyacak şekilde oluştururken, gözlemlenen (y) ve hesaplanan (Yr) bağımlı değişken değerleri arasındaki farkın karesini en aza indirir. Bu yöntemle şu iki denklem elde edilir:

$Yr = A1.x + K1 < BP$ (değişim noktası için)

$Yr = A2.x + K2 > BP$ (değişim noktası için)

Burada:

- Y_r, x'in belirli bir değeri için beklenen (tahmin edilen) y değeridir;

- A1 ve A2 regresyon katsayılarıdır (çizgi segmentinin eğimini gösterir);

- K1 ve K2 regresyon sabitleridir (y-ekseninde kesişimi gösterir).

Bu yöntem aynı zamanda iki korelasyon katsayısı (R) da üretir:

$R_1^2 = 1 - \frac{\sum(y-Y_r)^2}{\sum(y-Y_{a1})^2}$ x<BP (değişim noktası için)

ve

$R_2^2 = 1 - \frac{\sum(y-Y_r)^2}{\sum(y-Y_{a2})^2}$ x>BP (değişim noktası için)

Burada:

- $\sum(y - Y_r)^2$ her bir parça için minimize edilmiş SSD'yi temsil eder

- Y_{a1} ve Y_{a2} ilgili parçalarda y'nin ortalamasıdır.

En uygun eğilimi belirlemede, bu eğilimin güvenilir (anlamlı) olduğundan emin olmak için istatistiksel testler gerçekleştirilmelidir.

Eğer anlamlı bir değişim noktası tespit edilemezse, değişim noktası olmadan bir regresyona geçilmelidir. Aşağıdaki istatistiksel testler, eğilim türünü belirlemek için kullanılır:

1. **A1 ve A2'nin Anlamlılığı:** A1 ve A2'nin anlamlılığı, A1 ve A2'nin standart hata SE'si ve Student'ün t-distribution'ı kullanılarak belirlenir.

2. **A1 ve A2'nin Farkının Anlamlılığı:** A1 ve A2'nin farkının anlamlılığı, farklarının standart hatası SE ve Student'ün t-distribution'ı kullanılarak belirlenir.

3. **Y1 ve Y2'nin Farkının Anlamlılığı:** Y1 ve Y2'nin farkının anlamlılığı, farklarının standart hatası SE ve Student'ün t-distribution'ı kullanılarak belirlenir.

4. **Değişim Noktasının Varlığını Test Etme:** Sözde skor testi, parçalı çizginin tahminini gerektirmez ve kırılma noktasının varlığını test etmek için daha formal bir istatistiksel yaklaşımdır.

Ayrıca, tüm verilerin korelasyon katsayısı (R_a), belirleme katsayısı veya açıklama katsayısı, regresyon fonksiyonlarının güven aralıkları ve ANOVA analizi kullanılmaktadır.

C_d katsayısı, tüm veri seti için belirlenen şartlar altında maksimize edilmesi gereken bir belirleme katsayısıdır ve şu formülle hesaplanır:

$C_d = 1 - \frac{\sum(y-Y_r)^2}{\sum(y-Y_a)^2}$

Burada Y_r, önceki regresyon denklemlerine göre beklenen (tahmin edilen) y değeridir ve Y_a, tüm y değerlerinin ortalamasıdır.

C_d katsayısı, 0 ile 1 arasında değer alır. Saf, parçalanmamış doğrusal regresyonda, C_d ve R_a^2 değerleri eşittir. Parçalı regresyonda, C_d'nin R_a^2'den anlamlı derecede büyük olması, parçalanmanın haklı olduğunu gösterir.

Değişim noktasının en uygun değeri, C_d katsayısının maksimum olduğu noktada bulunabilir.

You may wish to put your reaction in an equation environment, which means that LaTeX will place the reaction where it fits and will number the equations for you.

$$C_6H_{12}O_6 + 6O_2 \longrightarrow 6CO_2 + 6H_2O \tag{2.1}$$
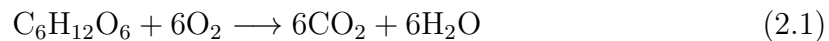
We can reference this combustion of glucose reaction via Equation (2.1).

2.2.4   Other examples of reactions

$NH_4Cl_{(s)} \rightleftharpoons NH_{3(g)} + HCl_{(g)}$
$MeCH_2Br + Mg \xrightarrow[below]{above} MeCH_2 \bullet Mg \bullet Br$
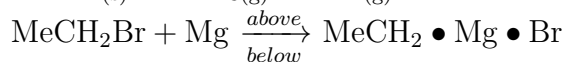
2.3   Physics

Many of the symbols you will need can be found on the math page `http://web.reed.edu/cis/help/latex/math.html` and the Comprehensive LaTeX Symbol Guide (`http://mirror.utexas.edu/ctan/info/symbols/comprehensive/symbols-letter.pdf`).

2.4   Biology

You will probably find the resources at `http://www.lecb.ncifcrf.gov/~toms/latex.html` helpful, particularly the links to bsts for various journals. You may also be interested in TeXShade for nucleotide typesetting (`http://homepages.uni-tuebingen.de/beitz/txe.html`). Be sure to read the proceeding chapter on graphics and tables.

Chapter 3

Tables, Graphics, References, and Labels

3.1 Tables

In addition to the tables that can be automatically generated from a data frame in **R** that you saw in R Markdown Basics using the `kable` function, you can also create tables using *pandoc*. (More information is available at `http://pandoc.org/README.html#tables`.) This might be useful if you don't have values specifically stored in **R**, but you'd like to display them in table form. Below is an example. Pay careful attention to the alignment in the table and hyphens to create the rows and columns.

We can also create a link to the table by doing the following: Table **??**. If you go back to Loading and exploring data and look at the `kable` table, we can create a reference to this max delays table too: Table 1.1. The addition of the (`\#tab:inher`) option to the end of the table caption allows us to then make a reference to Table `\@ref(tab:label)`. Note that this reference could appear anywhere throughout the document after the table has appeared.

3.2   Figures

If your thesis has a lot of figures, *R Markdown* might behave better for you than that other word processor. One perk is that it will automatically number the figures accordingly in each chapter. You'll also be able to create a label for each figure, add a caption, and then reference the figure in a way similar to what we saw with tables earlier. If you label your figures, you can move the figures around and *R Markdown* will automatically adjust the numbering for you. No need for you to remember! So that you don't have to get too far into LaTeX to do this, a couple **R** functions have been created for you to assist. You'll see their use below.

In the **R** chunk below, we will load in a picture stored as `ist.png` in our "figure" directory. We then give it the caption of "IST logo", the label of "istlogo", and specify that this is a figure. Make note of the different **R** chunk options that are given in the R Markdown file (not shown in the knitted document).

```
include_graphics(path = "figure/ist.png")
```


Figure 3.1: IST logo

Here is a reference to the IST logo: Figure 3.1. Note the use of the `fig:` code here. By naming the **R** chunk that contains the figure, we can then reference that

figure later as done in the first sentence here. We can also specify the caption for the figure via the R chunk option `fig.cap`.

Below we will investigate how to save the output of an **R** plot and label it in a way similar to that done above. Recall the `flights` dataset from Chapter 1. (Note that we've shown a different way to reference a section or chapter here.) We will next explore a bar graph with the mean flight departure delays by airline from Portland for 2014. Note also the use of the `scale` parameter which is discussed on the next page.

```
flights %>% group_by(carrier) %>%
  summarize(mean_dep_delay = mean(dep_delay)) %>%
  ggplot(aes(x = carrier, y = mean_dep_delay)) +
  geom_bar(position = "identity", stat = "identity", fill = "red")
```
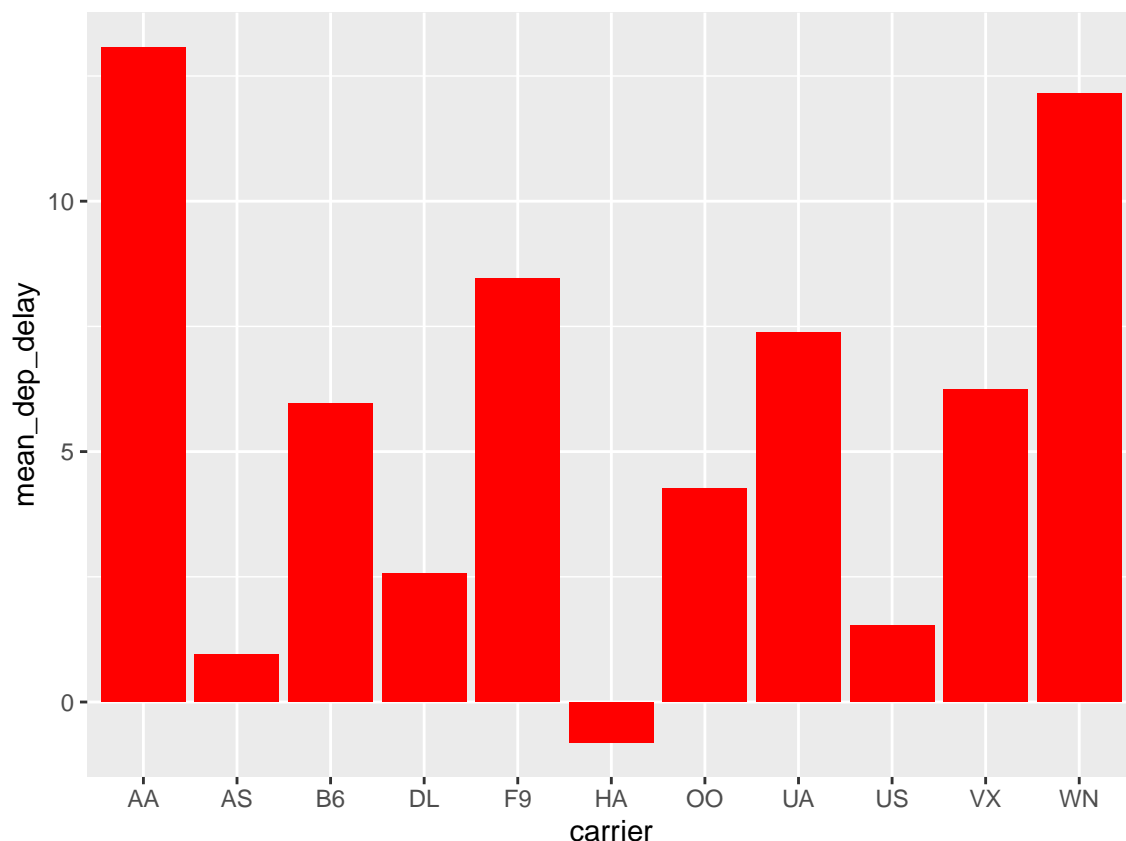

Figure 3.2: Mean Delays by Airline

Here is a reference to this image: Figure 3.2.

A table linking these carrier codes to airline names is available at `https://github.com/ismayc/pnwflights14/blob/master/data/airlines.csv`.

Next, we will explore the use of the `out.extra` chunk option, which can be used to shrink or expand an image loaded from a file by specifying `"scale= "`. Here we use the mathematical graph stored in the "subdivision.pdf" file.
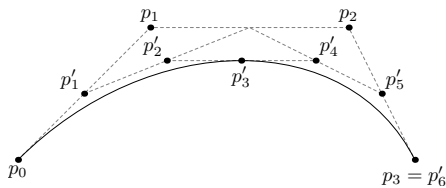
Figure 3.3: Subdiv. graph

Here is a reference to this image: Figure 3.3. Note that `echo=FALSE` is specified so that the **R** code is hidden in the document.

More Figure Stuff

Lastly, we will explore how to rotate and enlarge figures using the `out.extra` chunk option. (Currently this only works in the PDF version of the book.)
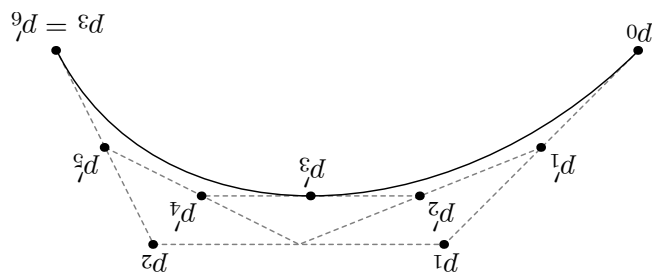
Figure 3.4: A Larger Figure, Flipped Upside Down

As another example, here is a reference: Figure 3.4.

3.3 Footnotes and Endnotes

You might want to footnote something. [1] The footnote will be in a smaller font and placed appropriately. Endnotes work in much the same way. More information can be found about both on the CUS site.

3.4 Bibliographies

Of course you will need to cite things, and you will probably accumulate an armful of sources. There are a variety of tools available for creating a bibliography database (stored with the .bib extension). In addition to BibTeX suggested below, you may want to consider using the free and easy-to-use tool called Zotero. The librarians have created

---

[1]footnote text

Zotero documentation at `http://libguides.reed.edu/citation/zotero`. In addition, a tutorial is available from Middlebury College at `http://sites.middlebury.edu/zoteromiddlebury/`.

R Markdown uses *pandoc* (`http://pandoc.org/`) to build its bibliographies. One nice caveat of this is that you won't have to do a second compile to load in references as standard LaTeX requires. To cite references in your thesis (after creating your bibliography database), place the reference name inside square brackets and precede it by the "at" symbol. For example, here's a reference to a book about worrying: (Molina ve Borkovec, 1994). This `Molina1994` entry appears in a file called `thesis.bib` in the `bib` folder. This bibliography database file was created by a program called BibTeX. You can call this file something else if you like (look at the YAML header in the main .Rmd file) and, by default, is to placed in the `bib` folder.

For more information about BibTeX and bibliographies, see the CUS site `http://web.reed.edu/cis/help/latex/index.html`. There are three pages on this topic: *bibtex* (which talks about using BibTeX, at `http://web.reed.edu/cis/help/latex/bibtex.html`), *bibtexstyles* (about how to find and use the bibliography style that best suits your needs, at `http://web.reed.edu/cis/help/latex/bibtexstyles.html`) and *bibman* (which covers how to make and maintain a bibliography by hand, without BibTeX, at `http://web.reed.edu/cis/help/latex/bibman.html`). The last page will not be useful unless you have only a few sources.

If you look at the YAML header at the top of the main .Rmd file you can see that we can specify the style of the bibliography by referencing the appropriate csl file. You can download a variety of different style files at `https://www.zotero.org/styles`. Make sure to download the file into the csl folder.

Tips for Bibliographies

- Like with thesis formatting, the sooner you start compiling your bibliography for something as large as thesis, the better. Typing in source after source is mind-numbing enough; do you really want to do it for hours on end in late April? Think of it as procrastination.
- The cite key (a citation's label) needs to be unique from the other entries.
- When you have more than one author or editor, you need to separate each author's name by the word "and" e.g. `Author = {Noble, Sam and Youngberg, Jessica},`.
- Bibliographies made using BibTeX (whether manually or using a manager) accept LaTeX markup, so you can italicize and add symbols as necessary.
- To force capitalization in an article title or where all lowercase is generally used, bracket the capital letter in curly braces.
- You can add a Thesis citation[2] option. The best way to do this is to use the phdthesis type of citation, and use the optional "type" field to enter "Undergraduate thesis."

---

[2]Noble (2002)

3.5 Anything else?

If you'd like to see examples of other things in this template, please contact the .... @ .... team (email `email@email.email`) with your suggestions. We love to see people using *R Markdown* for their theses, and are happy to help.

Chapter 4

Bölüm 4 Başlık

4.1 Bu bir alt başlık

Bu bölümde şu konular yer almaktadır. . .

4.1.1 Bu ikinci seviye bir alt başlık

Sonuç

If we don't want Conclusion to have a chapter number next to it, we can add the `{-}` attribute.

More info

And here's some other random info: the first paragraph after a chapter title or section head *shouldn't be* indented, because indents are to tell the reader that you're starting a new paragraph. Since that's obvious after a chapter or section title, proper typesetting doesn't add an indent there.

Kaynaklar

10 Angel, E. (2000). *Interactive Computer Graphics : A Top-Down Approach with OpenGL.* Boston, MA: Addison Wesley Longman.

Angel, E. (2001a). *Batch-file Computer Graphics : A Bottom-Up Approach with QuickTime.* Boston, MA: Wesley Addison Longman.

Angel, E. (2001b). *test second book by angel.* Boston, MA: Wesley Addison Longman.

Deussen, O. ve Strothotte, T. (2000). Computer-Generated Pen-and-Ink Illustration of Trees. *"Proceedings of" SIGGRAPH 2000*, Computer Graphics Proceedings, Annual Conference Series, 13-18.

Fisher, R., Perkins, S., Walker, A. ve Wolfart, E. (1997). *Hypermedia Image Processing Reference.* New York, NY: John Wiley & Sons.

Gooch, B. ve Gooch, A. (2001a). *Non-Photorealistic Rendering.* Natick, Massachusetts: A K Peters.

Gooch, B. ve Gooch, A. (2001b). *Test second book by gooches.* Natick, Massachusetts: A K Peters.

Hertzmann, A. ve Zorin, D. (2000). Illustrating Smooth Surfaces. *Proceedings of SIGGRAPH 2000*, Computer Graphics Proceedings, Annual Conference Series, $5(17)$, 517-526.

Jain, A. K. (1989). *Fundamentals of Digital Image Processing.* Englewood Cliffs, New Jersey: Prentice-Hall.

Molina, S. T. ve Borkovec, T. D. (1994). The Penn State Worry Questionnaire: Psychometric properties and associated characteristics. G. C. L. Davey ve F. Tallis (Ed.), *Worrying: Perspectives on theory, assessment and treatment* içinde (ss. 265-283). New York: Wiley.

Noble, S. G. (2002). *Turning images into simple line-art.* (Yayımlanmamış undergraduate thesis). Reed College.

Reed College. (2007). LaTeX Your Document. `http://web.reed.edu/cis/help/LaTeX/index.html` adresinden erişildi.

Russ, J. C. (1995). *The Image Processing Handbook, Second Edition.* Boca Raton, Florida: CRC Press.

Salisbury, M. P., Wong, M. T., Hughes, J. F. ve Salesin, D. H. (1997). Orientable Textures for Image-Based Pen-and-Ink Illustration. *"Proceedings of" SIGGRAPH 97*, Computer Graphics Proceedings, Annual Conference Series, 401-406.

Savitch, W. (2001). *JAVA: An Introduction to Computer Science & Programming.* Upper Saddle River, New Jersey: Prentice Hall.

Tukey, J. W. (1984). *The collected works of John W. Tukey* (C. 1). Taylor & Francis.

Wong, E. (1999). *Artistic Rendering of Portrait Photographs.* (Yayımlanmamış mathesis). Cornell University.

Appendix A

İlk Ek Başlığı

This first appendix includes all of the R chunks of code that were hidden throughout the document (using the `include = FALSE` chunk tag) to help with readability and/or setup.

In the main Rmd file

```r
# This chunk ensures that the thesisdown package is
# installed and loaded. This thesisdown package includes
# the template files for the thesis.
 if(!require(remotes)) install.packages("remotes", repos = "http://cran.rstudio.co
 if(!require(thesisdown))remotes::install_github("ismayc/thesisdown")
 library(thesisdown)
```

In Chapter 3:

```r
# This chunk ensures that the thesisdown package is
# installed and loaded. This thesisdown package includes
# the template files for the thesis and also two functions
# used for labeling and referencing
if(!require(remotes))
  install.packages("remotes", repos = "http://cran.rstudio.com")
if(!require(dplyr))
    install.packages("dplyr", repos = "http://cran.rstudio.com")
if(!require(ggplot2))
    install.packages("ggplot2", repos = "http://cran.rstudio.com")
if(!require(ggplot2))
    install.packages("bookdown", repos = "http://cran.rstudio.com")
if(!require(thesisdown)){
  library(remotes)
  remotes::install_github("ismayc/thesisdown")
```

```r
  }
library(thesisdown)
flights <- read.csv("data/flights.csv")
```

Appendix B

İkinci Ek Başlığı

İkinci Ek