Predict Number of Procedures To Start Up

Dileesh Venugopal, Shashank Hosahally & Shirley Verghese

Contents

1 Executive Summary

A startup company is an entrepreneurial venture or a new business in the form of a company a partnership or temporary organization designed to search for a repeatable and scalable business model.

In every country, we found the number of procedure officially required for a formal establishment of startup considerably varies. These procedures include obtaining all necessary licenses and permits and completing any required notifications, verifications or inscriptions for the company and employees with relevant authorities

A procedure is defined as any interaction of the company founders with external parties (for example, government agencies, lawyers, auditors or notaries). Interactions between company founders or company officers and employees are not counted as procedures. Procedures that must be completed in the same building but in different offices or at different counters are counted separately.

This is an effort to analyze the below factors that influence the number of procedures that would be required in the respective country.

2 Summary Statistics & Exploration (Stage 1)

2.1 Introduction

The Data was retrieved from World Bank Data. The dataset has more than 300 thousand rows of data and we needed to clean the data to arrive at this dataset to predict the No of Procedures.

The link to the dataset is: http://data.worldbank.org/data-catalog/world-development-indicators?cid=GPD_WDI

2.2 Summary on All Data

Structure and summary statistics on the entire data.

```
## 'data.frame':    442 obs. of  15 variables:
## $ Country.Code                    : chr  "AFG" "AFG" "AGO" "AGO" ...
## $ Year                            : int  2013 2014 2014 2013 2013 2014 2013 2014 2013 2014 ...
## $ EaseOfDoingBusinessIndex        : num  182 183 181 180 108 ...
## $ NoOfProceduresToStartUp         : num  3 3 8 8 5 ...
## $ NumberOfProceduresToEnforceContract: num  46 46 46 46 39 ...
## $ ProfitTax                       : num  0 0 25.3 25.3 9.5 ...
## $ StrengthLegalRights             : num  9 9 1 1 8 ...
## $ TimeRequiredToGetElectricityDays: num  114 114 145 145 177 ...
## $ TimeRequiredToStartBusinessInDays: num  5 7 66 66 4.5 ...
## $ TimeToPrepareAndPayTaxesHours   : num  275 275 282 282 357 ...
## $ TotalTaxRate                    : num  35.8 35.8 52 52 30.9 ...
## $ Short.Name                      : Factor w/ 248 levels "Afghanistan",..: 1 1 6 6 2 2 8 8 234 23
## $ Region                          : Factor w/ 8 levels "","East Asia & Pacific",..: 7 7 8 8 3 3 
## $ Income.Group                    : Factor w/ 6 levels "","High income: nonOECD",..: 4 4 6 6 6 6 
## $ Government.Accounting.concept    : Factor w/ 3 levels "","Budgetary central government",..: 3 3 

##   Country.Code            Year        EaseOfDoingBusinessIndex
##   Length:442         Min.   :2013   Min.   :  1.00
##   Class :character   1st Qu.:2013   1st Qu.: 49.36
##   Mode  :character   Median :2014   Median : 98.73
##                      Mean   :2014   Mean   : 95.38
##                      3rd Qu.:2014   3rd Qu.:140.00
##                      Max.   :2014   Max.   :189.00
##
##  NoOfProceduresToStartUp NumberOfProceduresToEnforceContract   ProfitTax
##  Min.   : 1.000          Min.   :21.00                       Min.   : 0.00
##  1st Qu.: 5.000          1st Qu.:34.00                       1st Qu.:10.12
##  Median : 7.000          Median :38.00                       Median :16.99
##  Mean   : 7.047          Mean   :38.13                       Mean   :16.27
##  3rd Qu.: 8.243          3rd Qu.:42.00                       3rd Qu.:21.68
##  Max.   :18.000          Max.   :55.00                       Max.   :65.80
##
##  StrengthLegalRights TimeRequiredToGetElectricityDays
##  Min.   : 0.000      Min.   : 18.00
##  1st Qu.: 3.000      1st Qu.: 64.15
##  Median : 5.000      Median : 90.00
##  Mean   : 4.903      Mean   :105.65
##  3rd Qu.: 6.000      3rd Qu.:125.00
```
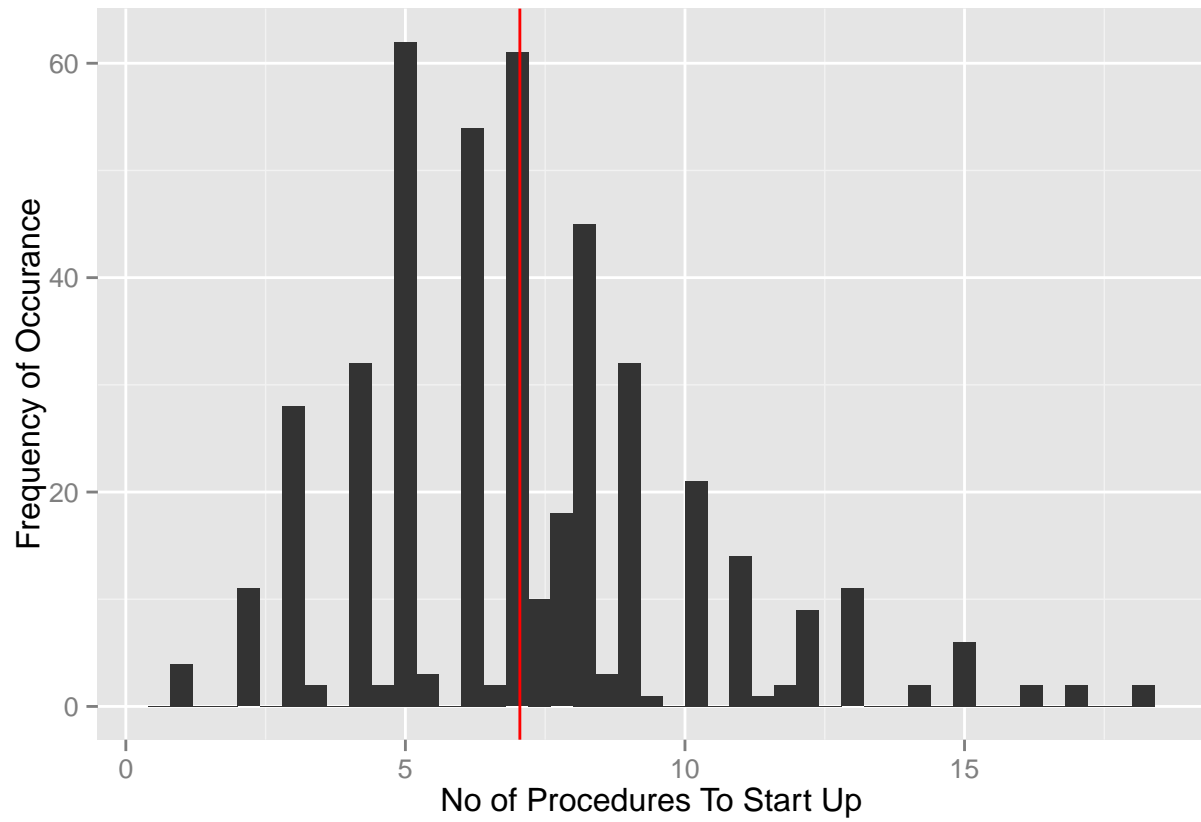
```
##   Max.    :12.000     Max.     :470.00
##
##   TimeRequiredToStartBusinessInDays TimeToPrepareAndPayTaxesHours
##   Min.    :  0.50                    Min.    :   12.0
##   1st Qu.:  9.00                     1st Qu.: 159.2
##   Median : 16.00                     Median : 218.0
##   Mean    : 23.51                    Mean    : 264.1
##   3rd Qu.: 29.24                     3rd Qu.: 312.0
##   Max.    :204.00                    Max.     :2600.0
##
##    TotalTaxRate                    Short.Name                        Region
##   Min.    :  7.40   Afghanistan         : 2   Europe & Central Asia      :98
##   1st Qu.: 31.50   Albania             : 2   Sub-Saharan Africa         :94
##   Median : 39.30   Algeria             : 2                              :66
##   Mean    : 41.68   Angola              : 2   Latin America & Caribbean :66
##   3rd Qu.: 48.23   Antigua and Barbuda: 2   East Asia & Pacific        :56
##   Max.    :275.40   Arab World          : 2   Middle East & North Africa:42
##                     (Other)             :430   (Other)                    :20
##               Income.Group                    Government.Accounting.concept
##                    : 66                                     :122
##   High income: nonOECD: 54   Budgetary central government    :132
##   High income: OECD   : 64   Consolidated central government:188
##   Low income          : 58
##   Lower middle income :102
##   Upper middle income : 98
##
```
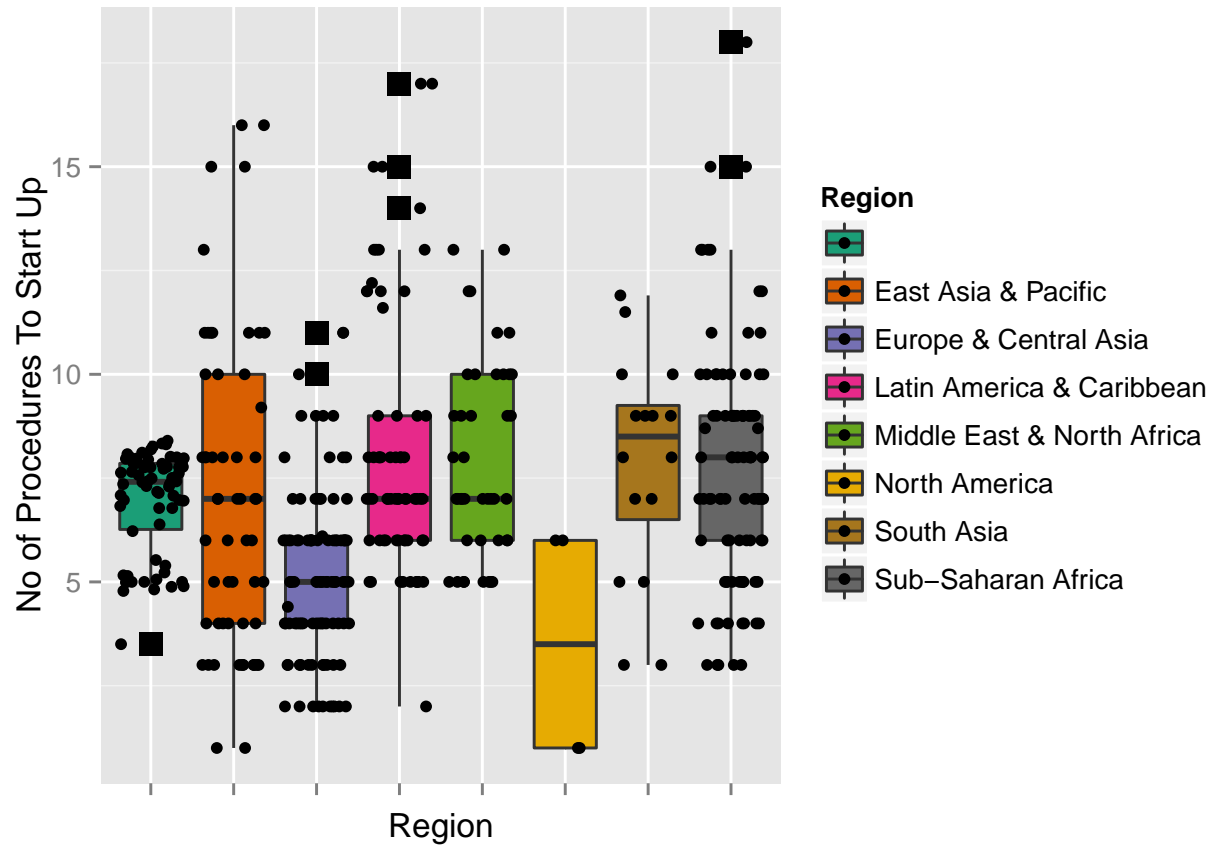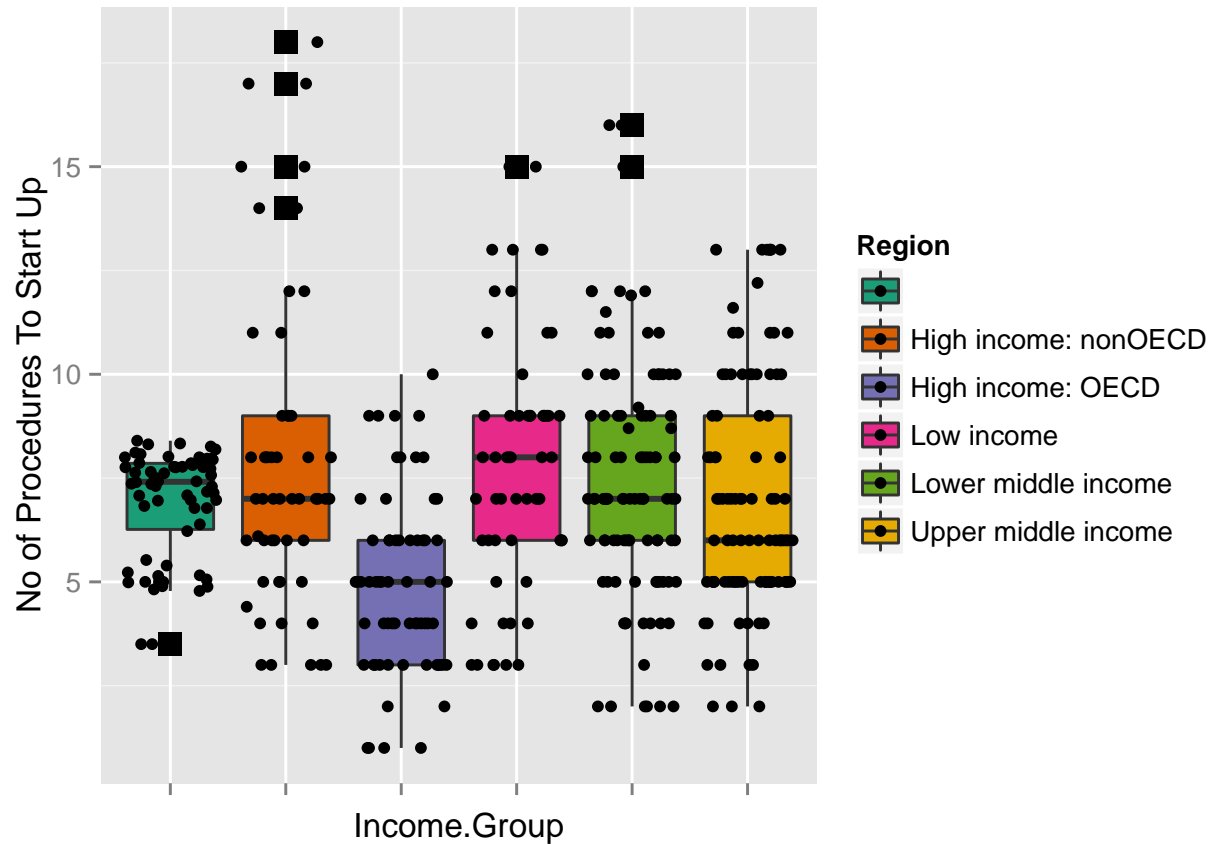
4

2.3 Distribution of Number of Procedures to StartUp

2.4 Boxplot of Number of Procedures based on Country Region

2.5 Boxplot of Number of Procedures based on Country's Income Group

Salient Features:

- Graphically, it appears High Income OECD Countries have lesser No of Procedures

3 Formal Statistical Inferences (Stage 2)

3.1 On Average, Do High Income OECD Countries Have Lesser Procedures to StartUp Business

3.1.1 The Hypotheses

- Null Hypothesis (H_0) : On average, there is no difference in between the Number of Procedures to start up a business in any region
- Alternate Hypothesis (H_A): On average, High Income OECD Countries have lesser Number of Procedures (one-sided test)

Mathematically, the hypothese are expressed below:

- H_0: $_{\text{diff}} = 0$
- H_A: $_{\text{diff}} > 0$

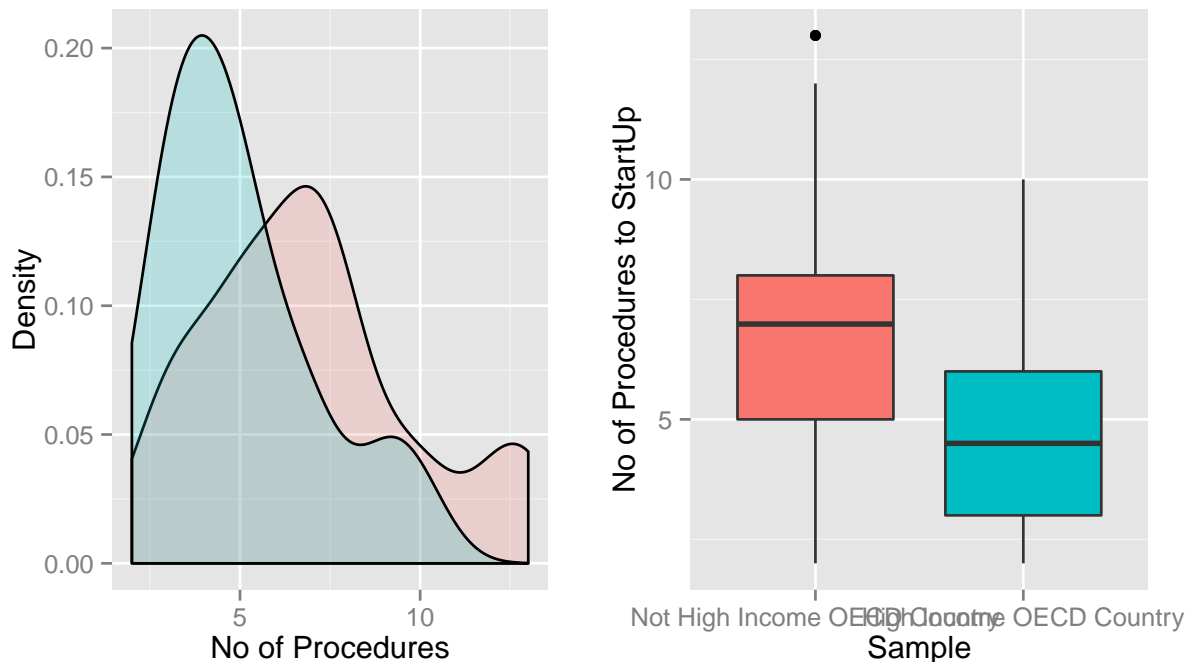### 3.1.2 The Data

We need to separate the datasets

```
# Creating the datasets for Holiday and NotHoliday
Hyp_NotOECD <- subset( StartUpCountryData , Income.Group != 'High income: OECD' )
Hyp_OECD <- subset( StartUpCountryData , Income.Group == 'High income: OECD' )
```

3.1.3 Central Limit Theorm: Checking the Conditions for Hypothesis Testing for Paired Data

The conditions for hypothesis testing:

- Independence: Sampled observations must be independent. Random sample must be collected and if it is without replacement then the sample size must be less than 10% of the Population
- Sample Size / Skew: The no of elements must be more than 30.

We select a size of 40 which is more than than 10% of the data, therefore, we will sample with replacement


```
## Warning in rm(xVar, colorVar, sampleDensityDf, plottingDensity,
## boxPlotDensity, : object 'Hyp_Holiday' not found
```

```
## Warning in rm(xVar, colorVar, sampleDensityDf, plottingDensity,
## boxPlotDensity, : object 'Hyp_NotHoliday' not found
```

3.1.4 Calculating the Test Statistic

```
# Calculating the Difference
Diff = OECD_Sample - NotOECD_Sample
# Calculating the Test Statistic
xBar <- mean(Diff)
# Calculating the Test Statistic
zScore <- xBar / standardError(Diff)
# Calculating p-value
# pnorm() because we are doing a one-sided test - less than than
pValue <- pnorm( zScore )
```

```
xBar
```

```
## [1] -1.918717
```

```
zScore
```

```
## [1] -2.875818
```

```
pValue
```

```
## [1] 0.002014909
```

3.1.5 Decision: Null Hypothesis (H_0) is Rejected

The **Null Hypothesis (H_0)** is rejected because the pValue is less than the significance value of 0.05.

This imples that the Alternate Hypothesis (H_A) is NOT rejected and the no of procedures to start a business in High Income OECD Countries is lesser than countries from other regions

3.1.6 Real World Application

This confirms with reality where it is easy to start business quickly in countries that belong to the High income OECD Countries.

4 Multiple Regression - Predicting The Number of Procedures

```
modelNoOfProcedures <- lm(
  NoOfProceduresToStartUp ~ . -Country.Code -Short.Name ,
  data = StartUpCountryData )
```

The model built has an R^2=0.5560471.

The following factors are correlated the number of procedures to start a business:

- Time Required To Start Business In Days
- Total Tax Rate
- Strength of Legal Rights
- Income Group of the Country
- Strength of Legal Rights

We can predict the No of days using the predictor model.