Unit 1 Project

Data in R, Fall 2020

Dave Brown

Sep 25, 2020

# Dataset

The **forestfires.csv** dataset contains daily information on meteorology, forest conditions and wildfires for a park in Portugal from 2000 to 2003.

Variables include:

1. X - x-axis spatial coordinate within the Montesinho park map

2. Y - y-axis spatial coordinate within the Montesinho park map

3. month - month of the year: "jan" to "dec"

4. day - day of the week: "mon" to "sun"

5. FFMC - FFMC index from the FWI system

6. DMC - DMC index from the FWI system

7. DC - DC index from the FWI system

8. ISI - ISI index from the FWI system

9. temp - temperature in Celsius degrees

10. RH - relative humidity in %

11. wind - wind speed in km/hr

12. rain – precipitation in mm/m2

13. area - the burned area of the forest (in ha)

The following links contain background information on this dataset.

<https://www.nwcg.gov/publications/pms437/cffdrs/fire-weather-index-system>

<http://www3.dsi.uminho.pt/pcortez/fires.pdf>

# Question 1

Read, examine and format the data set.

## (i) Read in the data

Read the dataset into the dataframe “***forestfire.dat***” your R working environment. Examine the data using commands like View, str and head. Comment briefly on the dimensions and types of data.

## (ii) Data type

Take the steps necessary to ensure that all variables in your dataframe are of the correct data type. Comment.

## (iii) Range of values

Generate data on the ranges of values for each variable. Are there any missing values? Comment briefly, making sure to note the distribution of values for **area** and **month**.

# Question 2

Restructure data to address possible research questions.

## (i) weekend

Researchers theorize that as fires are often initiated by human activities, and people camp more often on weekends, there might be more fires on weekends relative to weekdays.

Construct a new logical **weekend** variable to replace **day** in your data.frame (should be in the same column position) such that **weekend** equals TRUE if the day is Sat or Sun and **FALSE** otherwise.

Examine the resulting data.frame and comment on the distribution of values for weekend and month particularly.

## (ii) fire

Compute a new logical variable termed **fire.event** which is true if area > 0 and false otherwise. Rename **area** as **fire.area**.

## (iii) re-sort

Using the order function, sort the data first by **month** (alphabetical) and next by **weekend** and last by inverse **area**. Examine the top most values with the **head** function.

# Question 3

Compute dewpoint (***td***) using the formula below.



Where ***A1*** = 7.625, ***Bl*** = 243.04°C, **RH** = Observed Relative Humidity, and ***t*** = temperature.

## (i) Write an R function

Write an R function ***dewpoint*** to compute the variable dewpoint (**DP**) as a function of **temp** and **RH.**

Note that the R **log** function can be used to compute the natural log.

To test your function, run it for current temp and relative humidity and compare your result to an online calculator. There are different ways to estimate dew point so there might be some difference, but the results should be close.

## (ii) Compute dewpoint variable

Compute a dew point variable using your part (i) function via three different methods.

### vector operation

Using vector operations, compute dew point for all observations in **forestfires.dat**. Summarize the result and comment briefly.

### for loop

Using a **for** loop, compute dew point for all observations in **forestfires.dat**. Summarize the result and comment briefly.

### apply

Using the **apply** function, compute dew point for all observations in in **forestfires.dat**. Summarize the result and comment briefly.

Do all three methods give the same result?

# Question 4

For this question, we will conduct exploratory data analysis on **forestfires.dat**.

## (i) weekend effect

Compute the mean values of all numerical variables and logical variables as grouped by **weekend**.

What does the mean function actually compute when applied to the logical **fire.event** variable?

Looking at **fire.area** and **fire.event** in particular comment on a possible weekend effect.

## (ii) North-south gradient

To examine a possible north-south gradient, first transform **Y** into an ordered factor.

Next compute mean fire.event and fire.area as a function of grouping variable **Y**.

Interpret your results and comment.