R1 - Data analyis

Richard T. Watson

2020-12-26 14:27:56

```
library(tidyverse)
library(lubridate)
```

a.

Access the data Atlanta weather (a timestamp, air temperature, humidity, and precipitation), and electricity prices (a timestamp and cost in cents per kWh). Merge the files for electricity, price, and weather.

```
url <-  "http://www.richardtwatson.com/data/ATLweather.csv"
w <- read_delim(url,delim=',')
url <-  "http://www.richardtwatson.com/data/electricityprices.csv"
e <- read_delim(url,delim=',')
m <- inner_join(w,e,by=c("Timestamp" = 'timestamp'))
```

b.

Compute the correlation between temperature and electricity price. What do you conclude?

```
cor.test(m$Temperature,m$cost)
```

```
##
##  Pearson's product-moment correlation
##
## data:  m$Temperature and m$cost
## t = 42.434, df = 52506, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.1738064 0.1903459
## sample estimates:
##      cor
## 0.182089
```

Conclusion

- As the p-value is less than .05, conclude that there is a relationship between temperature and electricity cost. As the temperature increases, electricity prices increase.
- The correlation is *small* (0.18) as it is between .1 and .3, as shown in the following table:

Correlation coefficient	Effect size
.10 - .30	Small
.30 - .50	Moderate
> .50	Large

- Given the small correlation, there are likely other factors that influence electricity cost, such as generation capacity

Note: The p-value tells you whether there is a relationship and the correlation coefficient indicates the size of that relationship.

c.

c. Compute the monthly average temperature for Atlanta. Name columns appropriately. Show all months.

```
means <-  m%>% group_by(month(Timestamp)) %>%
  summarize(MeanTemp = mean(Temperature, na.rm=TRUE))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
colnames(means) <- c('Month','Mean temperature')
print(means, n = 12)
```

```
## # A tibble: 12 x 2
##    Month `Mean temperature`
##    <dbl>              <dbl>
##  1     1               43.6
##  2     2               45.5
##  3     3               54.7
##  4     4               64.5
##  5     5               71.6
##  6     6               78.6
##  7     7               79.8
##  8     8               79.2
##  9     9               74.1
## 10    10               63.7
## 11    11               53.0
## 12    12               48.9
```

Extract the data for August and redo the correlation What do you conclude?

```
Aug <- m %>% filter((month(Timestamp) == 8))
cor.test(Aug$Temperature,Aug$cost)
```

```
##
##  Pearson's product-moment correlation
##
```

```
## data:  Aug$Temperature and Aug$cost
## t = 44.754, df = 4454, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5363643 0.5768841
## sample estimates:
##       cor
## 0.5569555
```

Conclusion

- As the p-value is less than .05, conclude that there is a relationship between temperature and electricity cost during August
- The correlation is *large* (0.56). There appears to be generation capacity problem in August, usually the hottest months of the year, so electricity prices are raised to dampened demand on hot days.

d.

Extract the data for January and redo the correlation What do you conclude?

```
Jan <- m %>% filter((month(Timestamp) == 1))
cor.test(Jan$Temperature,Jan$cost)
```

```
##
##  Pearson's product-moment correlation
##
## data:  Jan$Temperature and Jan$cost
## t = -31.192, df = 5204, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.4195245 -0.3737465
## sample estimates:
##        cor
## -0.3968823
```

Conclusion

- As the p-value is less than .05, conclude that there is a relationship between temperature and electricity cost during winter
- The correlation is negative and *moderate* (-.4). There appears to be generation capacity problem in January, usually the coldest month of the year, so electricity prices are raised to dampened demand on cold days.

e.

Download the data for solar radiation (a timestamp and solar radiation in watts/m2) for Athens. Compute the average, min, and max for solar radiation for the year (one value for each of the three measures).

Average, min, and max

```
url <- "http://www.richardtwatson.com/data/SolarRadiationAthens.csv"
s <- read_delim(url,delim=',')
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   TimeStamp = col_datetime(format = ""),
##   SolarWatt = col_double()
## )
```

```
mean(s$SolarWatt)
```

```
## [1] 193.2395
```

```
min(s$SolarWatt)
```

```
## [1] 0
```

```
max(s$SolarWatt)
```

```
## [1] 1457
```

f.

Assuming the total area for capturing solar energy by PV cells for a house is 25m2 (269 square feet)) and solar cells are 20% efficient (i.e., 20% of the radiation received is converted into electricity). How much electricity in kWh will be generated on average each day in August and January?

Explanation

- A solar panel of 20% efficiency receiving 200 Watts/mˆ2 of solar radiation will generate 40 Watts/mˆ2 of electricity
- A solar panel system generating 1000 Watts consistently for one hour generates 1 kWh. In a day it would generate 24kWh (1000*24/1000). Of course, the sun does not shine consistently for 24 hours.

```
Jul <- s %>% filter((month(TimeStamp) == 7))
round(mean(Jul$SolarWatt)*24/1000*25*.2,2)
```

```
## [1] 30.81
```

```
Jan <- s %>% filter((month(TimeStamp) == 1))
round(mean(Jan$SolarWatt)*24/1000*25*.2,2)
```

```
## [1] 14.43
```

Findings

About 30.8 kWh will be generated each day in July. About 14.2 kWh will be generated each day in January.