Portfolio of Research Questions

Seunghun Lee

Sep 22, 2020

Contents

Research Question

Researchers were interested in whether the correlation between temperature of each day and the number of registered shared-bike users were positivley related. They obtained a random sample of 731 days and their values for temperature (Celcius), which was derived via (t-t_min)/(t_max-t_min), t_min = -8, t_max = +39 (only in hourly scale), and for registered users and investigated their research question using an alpha of 0.05. (Fanaee-T, Hadi and Gama, Joao, 2013)

```
# https://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset
day <- read.csv('./data/day.csv')
day_temp_reg <- subset(day, select = c('temp', 'registered')) # subsetting the necessary columns
head(day_temp_reg)
```

```
##       temp registered
## 1 0.344167        654
## 2 0.363478        670
## 3 0.196364       1229
## 4 0.200000       1454
## 5 0.226957       1518
## 6 0.204348       1518
```

```
summary(day_temp_reg)
```

```
##       temp          registered
##  Min.   :0.05913   Min.   :  20
##  1st Qu.:0.33708   1st Qu.:2497
##  Median :0.49833   Median :3662
##  Mean   :0.49538   Mean   :3656
##  3rd Qu.:0.65542   3rd Qu.:4776
##  Max.   :0.86167   Max.   :6946
```

Hypotheses

H1: There is a positive relationship betwen temperature and the number of registered bike user. $\rho > 0$
H0: There is not a positive relationship between temperature and the number of registered bike user. $\rho \leq 0$

Critical test statistic

df = 731 - 2 = 729, α-level = 0.05, one-tailed,

```
abs(qt(0.05, 729))
```

```
## [1] 1.646947
```

Critical t(0.05, df = 729) = 1.647

Test statistic

```
# Defining necessary variables
temp <- day$temp
registered <- day$registered
```

1. Computation by hand

```
tr_byhand <- function(data1, data2){
  r <- cor(data1, data2)
  Sr <- sqrt((1-r^2)/(length(data1)-2))
  Tr <- r/Sr
  print(c(r = r, t = Tr))
}
tr_byhand(temp, registered)
```

```
##         r         t
##  0.540012 17.323348
```

```
# r = 0.54, t = 17.323
```

2. Test statistic via `cor.test` fuction

```
cor.test(temp, registered, alternative = 'greater')
```

```
##
##   Pearson's product-moment correlation
##
## data:  temp and registered
## t = 17.323, df = 729, p-value < 2.2e-16
## alternative hypothesis: true correlation is greater than 0
## 95 percent confidence interval:
##  0.4954141 1.0000000
## sample estimates:
##      cor
## 0.540012
```

Conclusion

- Reject the null hypothesis (H0)
- There is a significant positive relationship between temperature and the number of registered users
- [r= 0.54, $t(729) = 17.323$, $p < 0.05$]