CR_Portfolio_3

Seunghun Lee

2020-10-28 12:20:43

Contents

Research Question

A researcher was interested in investigating whether temperature in Celsius degree was a predictor of the burned area of the forest (in ha) in Portugal. The researcher hypothesized that the prediction relationship would be positive. Using the sample of data for the 517 instances found in forestfires.csv, test the researcher's hypothesis using α of 0.05

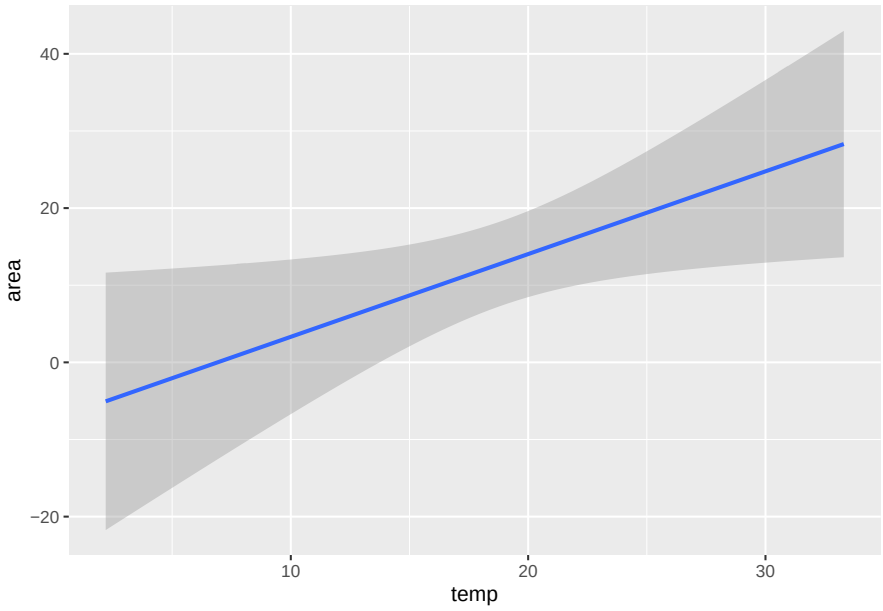
```
# http://archive.ics.uci.edu/ml/datasets/Forest+Fires
fire <- read.csv("./data/forestfires.csv") %>%
   select(temp, area)

# Descriptive statistics
fire %>%
  describe()
```

```
##      vars   n  mean    sd median trimmed  mad min     max   range  skew
## temp    1 517 18.89  5.81  19.30   19.09 5.34 2.2   33.30   31.10 -0.33
## area    2 517 12.85 63.66   0.52    3.18 0.77 0.0 1090.84 1090.84 12.77
##      kurtosis   se
## temp     0.11 0.26
## area   191.50 2.80
```

```
fire %>%
  ggplot(aes(y=area, x=temp)) +
  geom_smooth(method = 'lm')
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Hypotheses

- H_1: Temperature in Celsius degree is a positive predictor of the burned area of the forest (ha) in Portugal
- H_0: Temperature in Celsius degree is not a positive predictor of the burned area of the forest (ha) in Portugal

Critical test statistic

```
qt(p = 0.05, df = 515, lower.tail = FALSE)
```

```
## [1] 1.647818
```

- $\alpha = 0.05$, one-tailed, df = 517 -2 = 515, t(df= 515) = +1.648

Sample test statistic results

```
# Unstandardized coefficient estimates
mod <- lm(area ~ temp, data = fire)
summary(mod)
```

```
##
## Call:
## lm(formula = area ~ temp, data = fire)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -27.34  -14.68  -10.39   -3.42 1071.33
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -7.4138     9.4996  -0.780   0.4355
## temp          1.0726     0.4808   2.231   0.0261 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.41 on 515 degrees of freedom
## Multiple R-squared:  0.009573,   Adjusted R-squared:  0.00765
## F-statistic: 4.978 on 1 and 515 DF,  p-value: 0.0261
```

- **The simple test statistic t = 2.231**

```
# Standardized coefficient estimates
scale_mod <- lm(scale(area)~scale(temp), data = fire)
summary(scale_mod)
```

```
##
## Call:
## lm(formula = scale(area) ~ scale(temp), data = fire)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.4295 -0.2307 -0.1633 -0.0537 16.8301
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.870e-17  4.381e-02   0.000   1.0000
## scale(temp)  9.784e-02  4.385e-02   2.231   0.0261 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9962 on 515 degrees of freedom
## Multiple R-squared:  0.009573,   Adjusted R-squared:  0.00765
## F-statistic: 4.978 on 1 and 515 DF,  p-value: 0.0261
```

- **Standardized slope = 9.74e-02**

Conclusion

- Reject the null hypothesis
- Temperature in Celsius degree is a significant positive predictor of the burned area of the forest (ha) in Portugal
- [B = 1.07, $\beta = 0.0978$, t(515) = 2.231, $p < 0.05$]
- The results indicate that for every one Celsius degree, the burned area of the forest in Portugal increased by 1.07 ha