

regressionmodel

Javier

19/10/2020

```
library(nycflights13)

## Warning: package 'nycflights13' was built under R version 4.0.3

data<-flights
clean_data<-na.omit(data)

head(clean_data)

## # A tibble: 6 x 19
##   year month   day dep_time sched_dep_time dep_delay arr_time
##   <int> <int> <int>   <int>         <int>      <dbl>   <int>
## 1  2013     1     1     517           515         2     830
## 2  2013     1     1     533           529         4     850
## 3  2013     1     1     542           540         2     923
## 4  2013     1     1     544           545        -1    1004
## 5  2013     1     1     554           600        -6     812
## 6  2013     1     1     554           558        -4     740
## # ... with 11 more variables: arr_delay <dbl>, carrier <chr>, flight
## #   tailnum <chr>, origin <chr>, dest <chr>, air_time <dbl>, distance
## #   hour <dbl>, minute <dbl>, time_hour <dtm>

flights_model_0 <-
lm(clean_data$arr_delay~clean_data$distance+clean_data$air_time)
summary(flights_model_0)

##
## Call:
## lm(formula = clean_data$arr_delay ~ clean_data$distance +
## clean_data$air_time)
##
## Residuals:
```

```

##      Min      1Q  Median      3Q      Max
## -69.27 -22.91 -12.28   5.44 1284.47
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.4559364   0.1728867   -8.421  <2e-16 ***
## clean_data$distance -0.0876549   0.0007613 -115.145  <2e-16 ***
## clean_data$air_time  0.6652632   0.0059795  111.256  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.73 on 327343 degrees of freedom
## Multiple R-squared:  0.04012, Adjusted R-squared:  0.04012
## F-statistic: 6842 on 2 and 327343 DF, p-value: < 2.2e-16

linearMod <- lm(clean_data$arr_delay ~ clean_data$dep_delay)
summary(linearMod)

##
## Call:
## lm(formula = clean_data$arr_delay ~ clean_data$dep_delay)
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -107.587 -11.005  -1.883   8.938  201.938
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -5.8994935   0.0330195  -178.7  <2e-16 ***
## clean_data$dep_delay  1.0190929   0.0007864  1295.8  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.03 on 327344 degrees of freedom
## Multiple R-squared:  0.8369, Adjusted R-squared:  0.8369
## F-statistic: 1.679e+06 on 1 and 327344 DF, p-value: < 2.2e-16

f1 <- summary(flights_model_0)$r.squared
f1

## [1] 0.04012395

f2 <- summary(linearMod)$r.squared
f2

## [1] 0.8368641

```