assignment_06_KanaparthiVenkata

Venkata Kanaparthi

5/13/2021

```
library(ggm)
```

```
## Warning: package 'ggm' was built under R version 4.0.5
```

```
library(ggplot2)
```

Set the working directory to the root of your DSC 520 directory

```
setwd('E:/MSDS-SEM2/DSC520/CodingAssignments/DSC520KANAPARTHI')
## Load the `data/r4ds/heights.csv` to
heights_df <- read.csv("data/r4ds/heights.csv")
```

Fit a linear model using the `age` variable as the predictor and `earn` as the outcome

```
age_lm <- lm(earn ~ age, data = heights_df , na.action = na.exclude)
age_lm
```

```
##
## Call:
## lm(formula = earn ~ age, data = heights_df, na.action = na.exclude)
##
## Coefficients:
## (Intercept)          age
##    19041.53        99.41
```

View the summary of your model using `summary()`

```
summary(age_lm)
```

```
##
## Call:
## lm(formula = earn ~ age, data = heights_df, na.action = na.exclude)
##
## Residuals:
##    Min      1Q Median     3Q    Max
## -25098 -12622  -3667   6883 177579
##
## Coefficients:
```

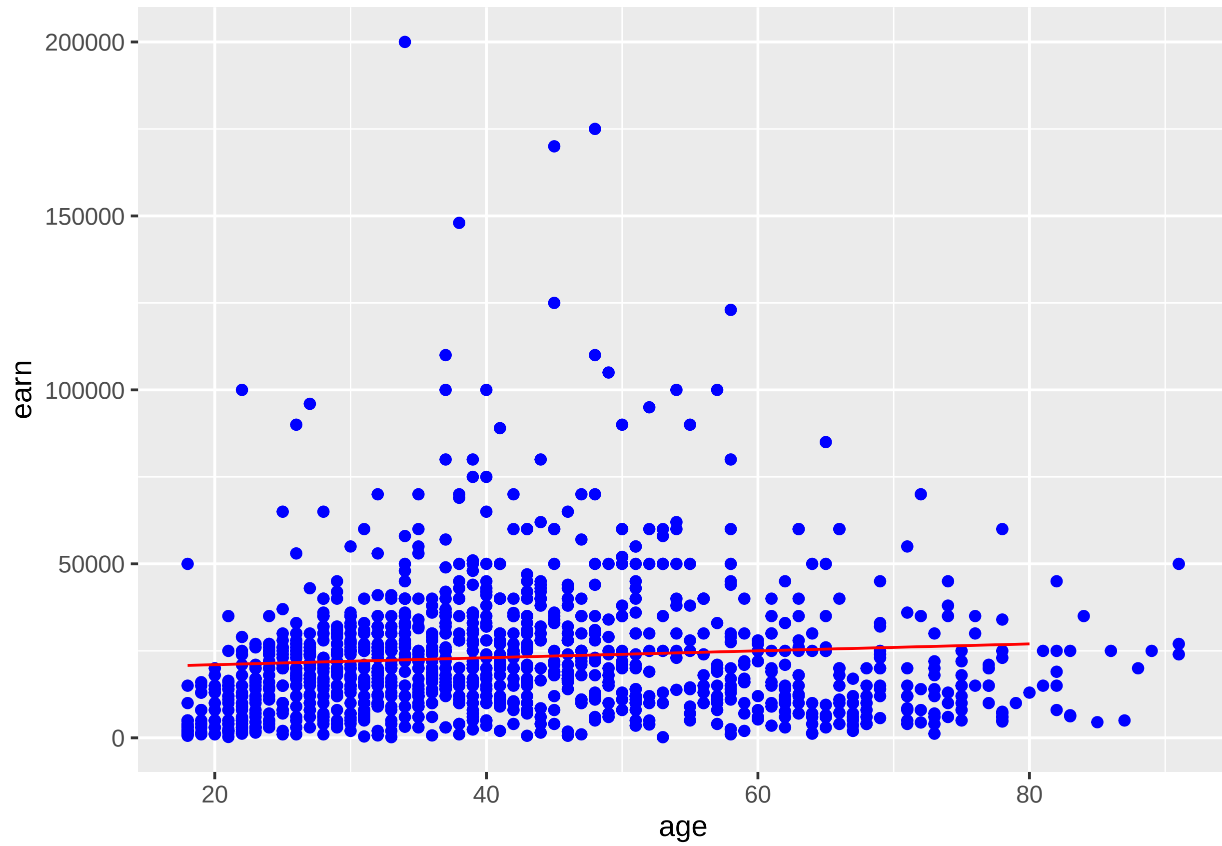
```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19041.53    1571.26  12.119  < 2e-16 ***
## age            99.41      35.46   2.804  0.00514 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19420 on 1190 degrees of freedom
## Multiple R-squared:  0.006561,   Adjusted R-squared:  0.005727
## F-statistic:  7.86 on 1 and 1190 DF,  p-value: 0.005137
```

Creating predictions using `predict()`

```
age_testing_df <- data.frame(age = c(18,36,54,72,80))
age_predict_df <- data.frame(earn = predict(age_lm, newdata = age_testing_df), age=age_testing_df)
```

Plot the predictions against the original data

```
ggplot(data = heights_df, aes(y = earn, x = age)) + geom_point(color='blue') +  geom_line(color='red',da
```



Mean on earn

```
mean_earn <- mean(heights_df$earn)
mean_earn
```

```
## [1] 23154.77
```

Corrected Sum of Squares Total

```r
sst <- sum((mean_earn - heights_df$earn)^2)
sst
```

```
## [1] 451591883937
```

Corrected Sum of Squares for Model

```r
ssm <- sum((mean_earn - age_predict_df$earn)^2)
ssm
```

```
## [1] 31266044
```

Residuals

```r
residuals <- heights_df$earn - age_predict_df$earn
```

```
## Warning in heights_df$earn - age_predict_df$earn: longer object length is not a
## multiple of shorter object length
```

```r
#residuals
```

Sum of Squares for Error

```r
sse <- sum(residuals^2)
sse
```

```
## [1] 457150421385
```

R Squared

```r
r_squared <- ssm/sst
r_squared
```

```
## [1] 6.923518e-05
```

Number of observations

```r
n <- 5
n
```

```
## [1] 5
```

Number of regression parameters

```r
p <- 2
p
```

```
## [1] 2
```

Corrected Degrees of Freedom for Model (p-1)

```r
dfm <- p-1
dfm
```

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

Degrees of Freedom for Error (n-p)

```r
dfe <- n-p
dfe
```

```
## [1] 3
```

Corrected Degrees of Freedom Total: DFT = n - 1

```r
dft <- n-1
dft
```

```
## [1] 4
```

Mean of Squares for Model: MSM = SSM / DFM

```r
msm <- ssm/dfm
msm
```

```
## [1] 31266044
```

Mean of Squares for Error: MSE = SSE / DFE

```
mse <- sse/dfe
mse
```

```
## [1] 152383473795
```

Mean of Squares Total: MST = SST / DFT

```
mst <- sst/dft
mst
```

```
## [1] 1.12898e+11
```

F Statistic F = MSM/MSE

```
f_score <- msm/mse
f_score
```

```
## [1] 0.00020518
```

Adjusted R Squared R2 = 1 - (1 - R2)(n - 1) / (n - p)

```
adjusted_r_squared <- 1-((1 - r_squared)*(n - 1) / (n - p))
adjusted_r_squared
```

```
## [1] -0.333241
```

Calculate the p-value from the F distribution

```
p_value <- pf(f_score, dfm, dft, lower.tail=F)
p_value
```

```
## [1] 0.9892574
```