inference_regression_lab

Brooke Wheeler

2/2/2022

```
library(tidyverse)
```

```
## ── Attaching packages ─────────────────────────────────────── tidyverse
1.3.1 ──
```

```
## ✓ ggplot2 3.3.5     ✓ purrr   0.3.4
## ✓ tibble  3.1.6     ✓ dplyr   1.0.7
## ✓ tidyr   1.1.4     ✓ stringr 1.4.0
## ✓ readr   2.1.1     ✓ forcats 0.5.1
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
women <- read.csv("Mass_Calorie_Data-1.csv", header = TRUE, sep = ",")
attach(women)
```

we want to fit a regression model that predicts Calorie Rate (RATE) in calories per day based on Lean Body Mass (LBM) in kg.

1. Start by fitting the regression model. Find the values of the least-squares estimators (regression coefficients) and write the equation of the LSR line.

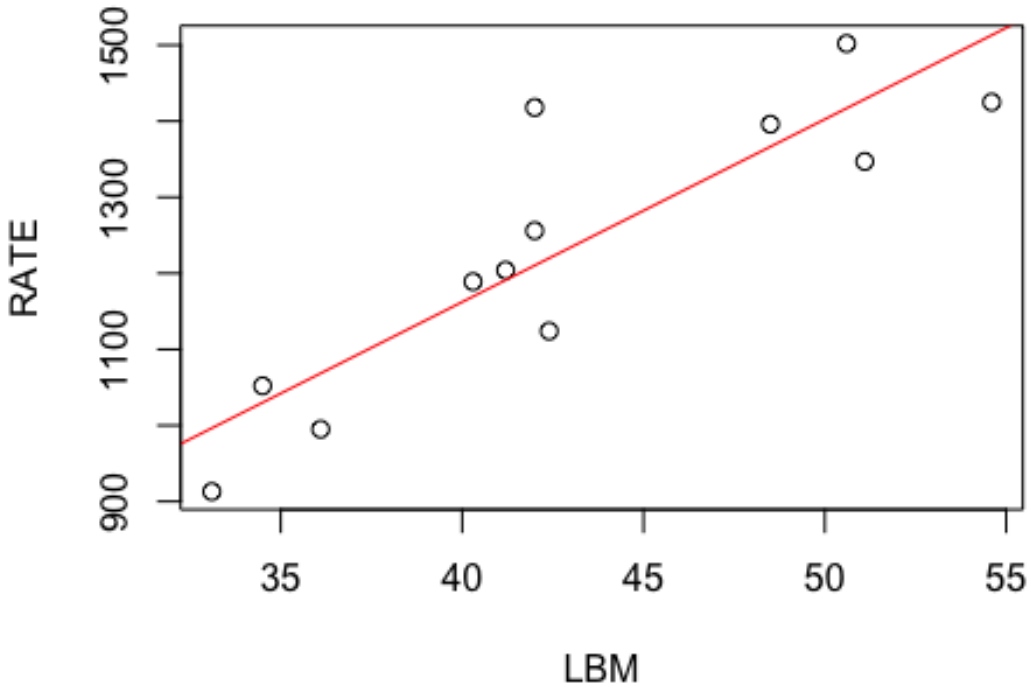
```
# RATE = y
# LBM = X

# fitting a regression model
reg <- lm(RATE ~ LBM)

reg
```

```
##
## Call:
## lm(formula = RATE ~ LBM)
##
## Coefficients:
```

```
## (Intercept)          LBM
##      201.16        24.03

# least square estimators: -81.433

# Least squares regression line: y= -81.433 + .159x
```

2. Plot the regression line (in a color other than black) with the data. Include the plot here.

```
reg

##
## Call:
## lm(formula = RATE ~ LBM)
##
## Coefficients:
## (Intercept)          LBM
##      201.16        24.03

plot(LBM, RATE)
abline(reg, col = "red")
```



3. Find a 98% confidence interval for the slope, β_1. Paste your results below. Verify that your lower and upper limits match those from the notes.

```
confint(reg, level = .98)
```

```
##                    1 %     99 %
## (Intercept) -301.01781 703.3410
## LBM           12.49043  35.5617
```

```
# 98% confidence interval for slop of B_1 is 12.49043 to 35.5617
```

4. What does the 98% confidence interval for the intercept, β_0, show? State the lower and upper limits and discuss what they mean in context.

```
confint(reg, level = .98)
```

```
##                    1 %     99 %
## (Intercept) -301.01781 703.3410
## LBM           12.49043  35.5617
```

```
# the 98% confidence interval for the intercept means that with 98%
confidence we estimate that the average (y) value of the intercept is between
```

```
-301.01781 and 703.3410 meaning that when x is zero (if a women was to weigh
0 kg) then the rate of calories burned per day is between -301.01781 and
703.3410
```

5. Is there a positive association between LBM and Calorie Rate? Use α=0.02. State the hypotheses. Provide the value of s(b_1), the t statistic and p-value. Do they match those found in the notes? Explain.

```
# null hyp:  slope = 0
# alt. hyp:  slope >0

summary(reg)

##
## Call:
## lm(formula = RATE ~ LBM)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -95.87 -82.28  16.28  33.62 207.74
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  201.162    181.701   1.107 0.294169
## LBM           24.026      4.174   5.756 0.000184 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 95.08 on 10 degrees of freedom
## Multiple R-squared:  0.7682, Adjusted R-squared:  0.745
## F-statistic: 33.13 on 1 and 10 DF,  p-value: 0.0001836

# s(b1)= 4.174

# looking at the confidence interval for number 3 we found our 98% confidence
interval for slop of B_1 is 12.49043 to 35.5617. Since zero is not in this
interval we know we can reject the null hypothesis at level alpha=.02 and
support that there is a positive association between LBM and calorie rate.
```

6. Note the value of "Residual standard error" shown in the summary of the regression results. What does this correspond to? (Hint: Look in the notes.) Based on this value, what is the value of the MSE?

```
summary(reg)
```

```
##
## Call:
## lm(formula = RATE ~ LBM)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -95.87 -82.28  16.28  33.62 207.74
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  201.162    181.701   1.107 0.294169
## LBM           24.026      4.174   5.756 0.000184 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 95.08 on 10 degrees of freedom
## Multiple R-squared:  0.7682, Adjusted R-squared:  0.745
## F-statistic: 33.13 on 1 and 10 DF,  p-value: 0.0001836
```

```
# residual standard error is 95.08
# the residual standard error measures the average distance of residuals from
the regression line. The residual standard eror corresponds to RMSE. This
means that on average the residual is 95.08 away from the regression line.

# MSE is 9040 calories per day. RMSE= 95.08
```