Module 7: Linear regression

Yuan Tian

07/20/2023

Linear regression in R

```
library(tidyverse) #ggplot2, dplyr, etc.
library(reshape2) #need this for melt()
library(knitr) #need this for kable
library(MASS) #contains dataset
```

Load the birthwt data. This data contains 189 observations, 9 predictors, and an outcome, birthweight, available both as a continuous measure and a binary indicator for low birth weight.

```
data(birthwt)
head(birthwt)
```

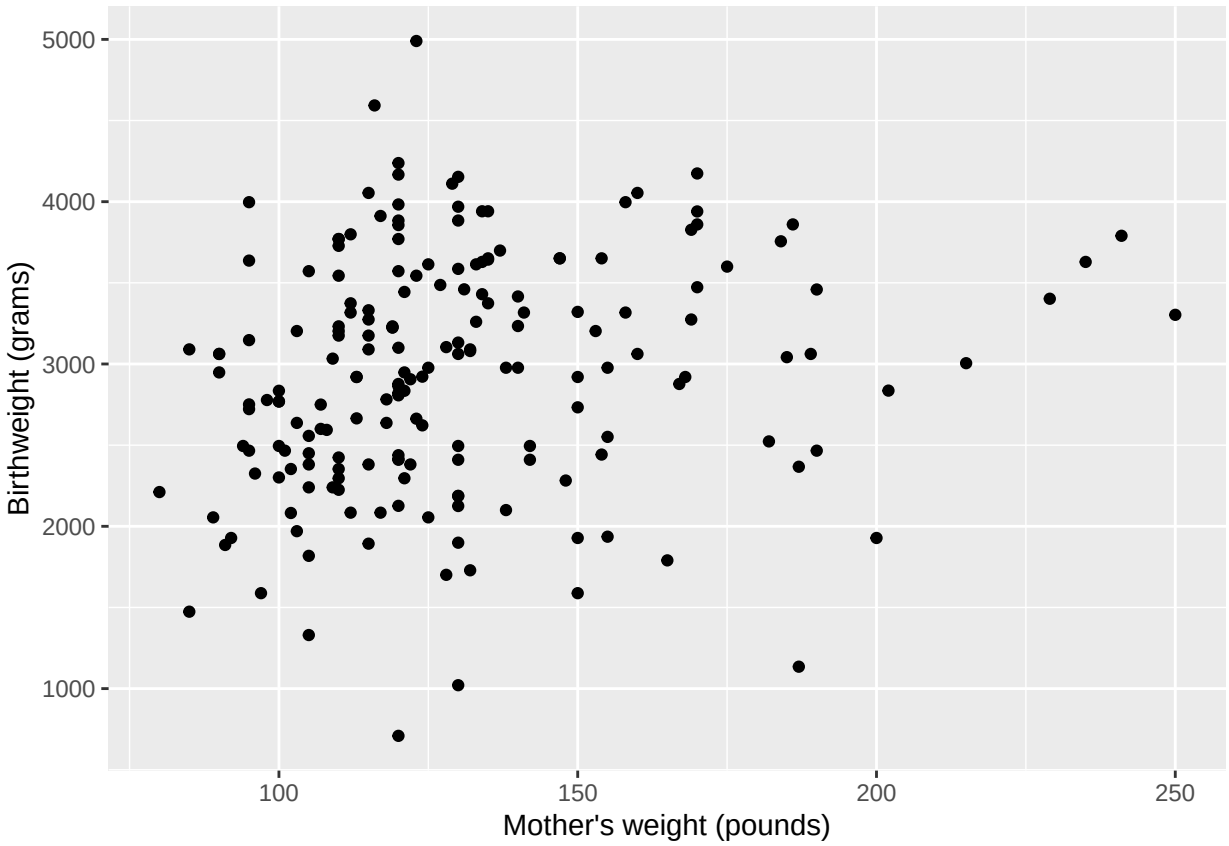
```
##    low age lwt race smoke ptl ht ui ftv  bwt
## 85   0  19 182    2     0   0  0  1   0 2523
## 86   0  33 155    3     0   0  0  0   3 2551
## 87   0  20 105    1     1   0  0  0   1 2557
## 88   0  21 108    1     1   0  0  1   2 2594
## 89   0  18 107    1     1   0  0  1   0 2600
## 91   0  21 124    3     0   0  0  0   0 2622
```

1. Plot a scatterplot of birthweight (bwt) and mother's weight (lwt).
2. Use OLS to fit the regression of birthweight on mother's weight.
3. Extract the following: estimated coefficients, standard errors, variance-covariance matrix, and confidence intervals.
4. Plot the regression line and interpret the intercept and slope
5. Does the interpretation of the intercept make sense? How might we change this?
6. Now, we want to fit a model that includes race, mother's age, and smoking status in the model. Race takes on value 1 for white, 2 for black, and 3 for other. Mother's age is continuous. Smoking status is binary. Write out the regression function we may be interested in.
7. Use OLS to calculate the coefficient estimates in this model.
8. Interpret all the coefficient estimates.
9. Print the results in Rmarkdown using kable().

Solution

1.

```
ggplot(birthwt, aes(x = lwt, y = bwt)) +
  geom_point() +
  labs(x = "Mother's weight (pounds)", y = "Birthweight (grams)")
```



2.

```
fit <- lm(bwt ~ lwt, data = birthwt)
summary(fit)
```

```
##
## Call:
## lm(formula = bwt ~ lwt, data = birthwt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2192.12  -497.97    -3.84   508.32  2075.60
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2369.624    228.493  10.371   <2e-16 ***
## lwt            4.429      1.713   2.585   0.0105 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 718.4 on 187 degrees of freedom
## Multiple R-squared:  0.0345, Adjusted R-squared:  0.02933
## F-statistic: 6.681 on 1 and 187 DF,  p-value: 0.0105
```

3.

```
# Estimated coefficients.
coefficients(fit)
```

```
## (Intercept)         lwt
## 2369.623518    4.429108
```

```
# Standard errors.
summary(fit)$coeff[, 2]
```

```
## (Intercept)         lwt
##  228.493206    1.713494
```

```
# Variance-covariance matrix.
vcov(fit)
```

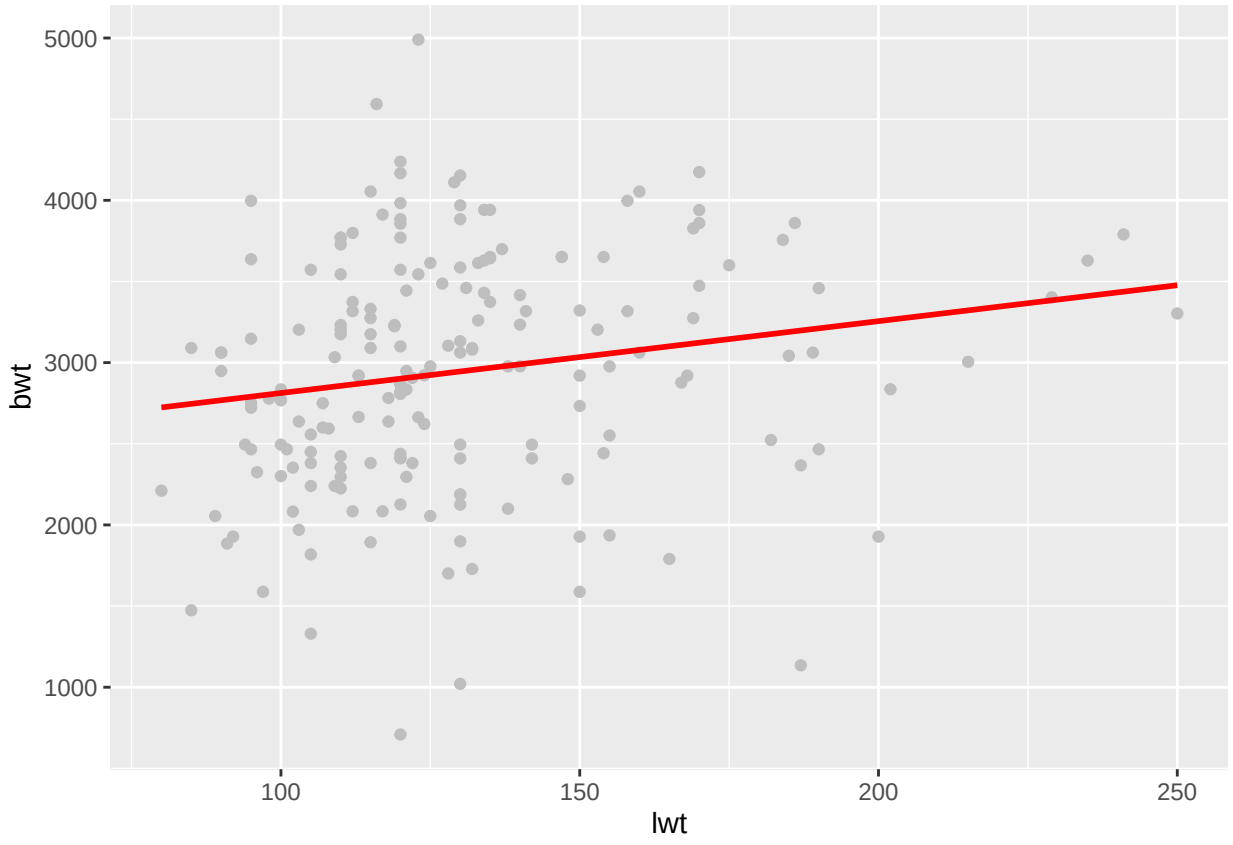
```
##             (Intercept)         lwt
## (Intercept)  52209.1453 -381.144214
## lwt           -381.1442    2.936061
```

```
# Confidence intervals.
confint(fit)
```

```
##                   2.5 %     97.5 %
## (Intercept) 1918.867879 2820.37916
## lwt            1.048845    7.80937
```

4.

```
ggplot(birthwt, aes(x = lwt, y = bwt)) +
  geom_point(color = "grey") +
  stat_smooth(method = "lm", col = "red", se = FALSE)
```



5.

```
birthwt <- birthwt %>% mutate(lwt_star = lwt - mean(lwt))

fit.new <- lm(bwt ~ lwt_star, data = birthwt)
summary(fit.new)
```

```
##
## Call:
## lm(formula = bwt ~ lwt_star, data = birthwt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2192.12  -497.97    -3.84   508.32  2075.60
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2944.587     52.259  56.346   <2e-16 ***
## lwt_star       4.429      1.713   2.585   0.0105 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 718.4 on 187 degrees of freedom
## Multiple R-squared:  0.0345, Adjusted R-squared:  0.02933
## F-statistic: 6.681 on 1 and 187 DF,  p-value: 0.0105
```

7.

```
fit2 <- lm(bwt ~ as.factor(race) + age + smoke, data = birthwt)
summary(fit2)
```

```
##
## Call:
## lm(formula = bwt ~ as.factor(race) + age + smoke, data = birthwt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2322.6  -447.3    28.4   502.2  1612.3
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       3281.673    260.664  12.590  < 2e-16 ***
## as.factor(race)2 -444.069    156.194  -2.843 0.004973 **
## as.factor(race)3 -447.858    119.017  -3.763 0.000226 ***
## age                 2.134      9.771   0.218 0.827326
## smoke            -426.093    109.988  -3.874 0.000149 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 690 on 184 degrees of freedom
## Multiple R-squared:  0.1236, Adjusted R-squared:  0.1046
## F-statistic:  6.49 on 4 and 184 DF,  p-value: 6.592e-05
```

8.

- The estimated mean birthweight among infants born to mothers who are non-smokers are zero years old (weird!), and white is 3281.7 grams.
- The estimated mean birthweight among infants with black mothers is 444.07 grams lower than the

mean birthweight among infants with white mothers, holding all other variables constant

- The estimated mean birthweight among infants with mothers in the "other" race category is 444.86 grams lower than the mean birthweight among infants with white mothers, holding all other variables constant
- The estimated change in mean birthweight corresponding to a one year change in mother's age is 2.134 grams, holding all other variables constant
- The estimated mean birthweight among infants with mothers that smoke is 426.09 grams lower than the mean birthweight among infants with mothers that do not smoke, holding all other variables constant.

9.

```
table <- data.frame(summary(fit2)$coef)
row.names(table) <- c("Intercept","White","Black","Mother's age", "Smoker")

knitr::kable(table,digits=3,align=rep('c', 2),
  col.names = c("estimate","standard error","test statistic","p-value"))
```

	estimate	standard error	test statistic	p-value
Intercept	3281.673	260.664	12.590	0.000
White	-444.069	156.194	-2.843	0.005
Black	-447.858	119.017	-3.763	0.000
Mother's age	2.134	9.771	0.218	0.827
Smoker	-426.093	109.988	-3.874	0.000