BabyWeightSmoking

Ray Chen

2025-05-28

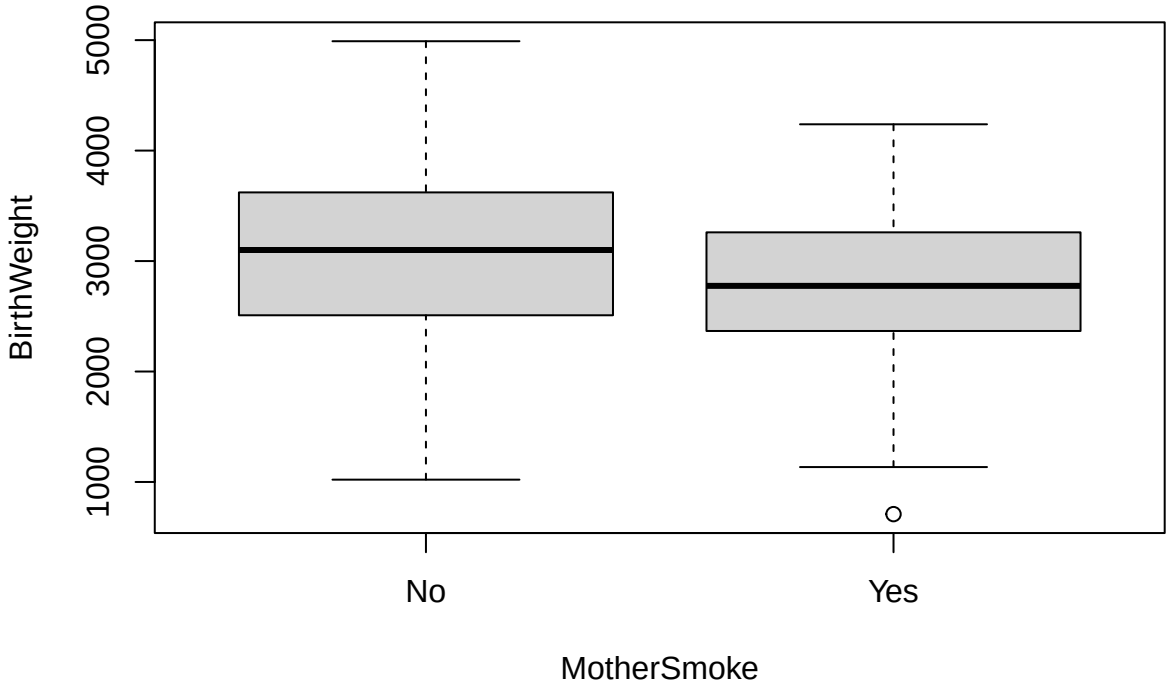
Load Data

```
data <- read.csv("BabyWeightSmoking.csv")
head(data)
```

```
##   LowBirthWeight MothersAge  Race MotherSmoke FTV BirthWeight
## 1            Yes         28 Other         Yes   0         709
## 2            Yes         29 White          No   2        1021
## 3            Yes         34 Black         Yes   0        1135
## 4            Yes         25 Other          No   0        1330
## 5            Yes         25 Other          No   0        1474
## 6            Yes         27 Other          No   0        1588
```

Boxplot of MotherSmoke and BirthWeight

```
boxplot_smoke <- boxplot(BirthWeight ~ MotherSmoke, data = data)
```



Two Sample t-test

null hypothesis: There is no difference in population mean age of mothers who smoke and those who do not smoke

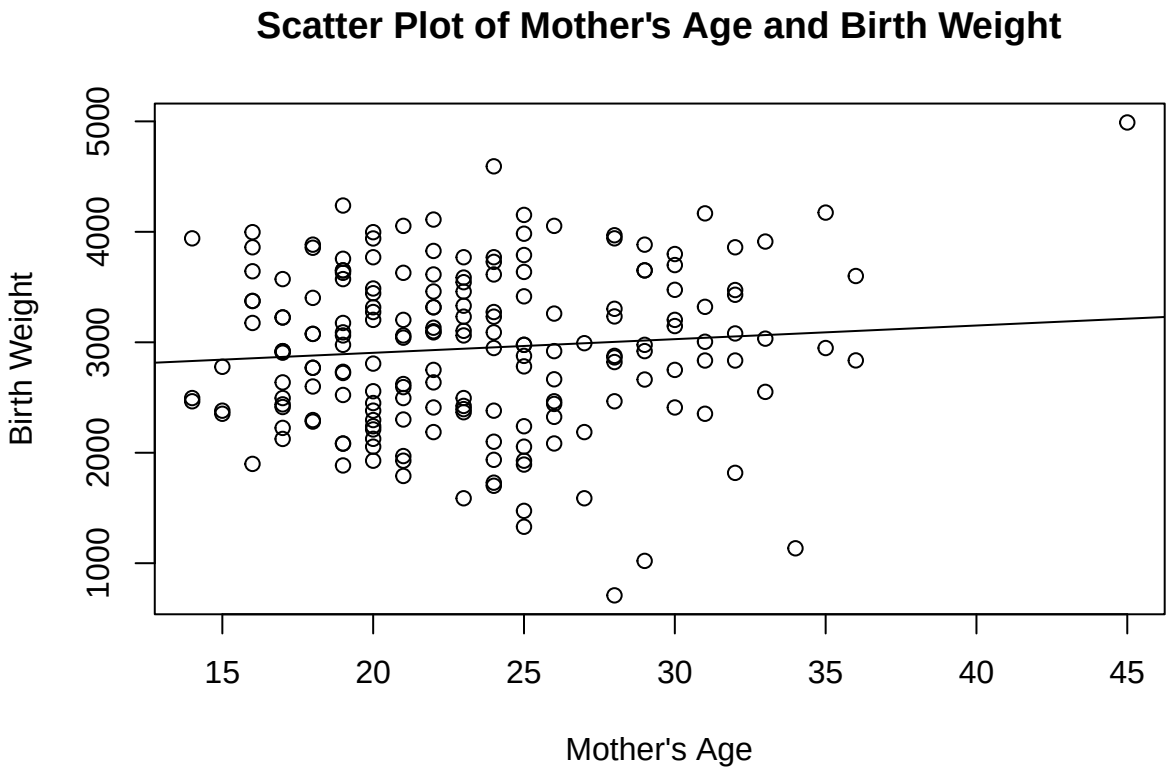
```
t_test_smoking <- t.test(MothersAge ~ MotherSmoke, data = data)
t_test_smoking
```

```
##
##  Welch Two Sample t-test
##
## data:  MothersAge by MotherSmoke
## t = 0.61768, df = 164.72, p-value = 0.5376
## alternative hypothesis: true difference in means between group No and group Yes is not equal to 0
## 95 percent confidence interval:
##  -1.054672  2.014954
## sample estimates:
##  mean in group No mean in group Yes
##          23.42609          22.94595
```

Since p value is significantly large, we fail to reject the null hypothesis

Scatter graph of MothersAge versus BirthWeight and linear regression model for prediction

```
plot(data$MothersAge, data$BirthWeight,
     xlab = "Mother's Age",
     ylab = "Birth Weight",
     main = "Scatter Plot of Mother's Age and Birth Weight",)
abline(lm(BirthWeight ~ MothersAge, data = data))
```



Prediction of BirthWeight given MothersAge 40 years old using linear regression

```
model <- lm(BirthWeight ~ MothersAge, data = data)
summary(model)
```

```
##
## Call:
## lm(formula = BirthWeight ~ MothersAge, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2294.53  -517.71    10.56   530.65  1776.27
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2657.33     238.80  11.128   <2e-16 ***
## MothersAge     12.36      10.02   1.234    0.219
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 728 on 187 degrees of freedom
## Multiple R-squared:  0.008076,   Adjusted R-squared:  0.002772
## F-statistic: 1.523 on 1 and 187 DF,  p-value: 0.2188
```

```
predict(model, data.frame(MothersAge = 40))
```

```
##        1
## 3151.906
```