Exercise_2

Mohammad Imtiaz Nur

2/6/2021

ID: 1878074

1. Comparison of infants birth weight between smoking and non-smoking mothers

```r
library(tidyverse)
```

```
## -- Attaching packages -------------------------------------------------------------- t
```

```
## v ggplot2 3.3.2     v purrr   0.3.3
## v tibble  3.0.3     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts -------------------------------------------------------------------- tidyvers
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
# Load data from MASS into a tibble
birthwt <- as_tibble(MASS::birthwt)
#?birthwt

# Rename variables
colnames(birthwt) <- c("birthwt.below.2500",
                       "mother.age",
                       "mother.weight",
                       "race",
                       "mother.smokes",
                       "previous.prem.labor",
                       "hypertension",
                       "uterine.irr",
                       "physician.visits",
                       "birthwt.grams")

# Change factor level names
birthwt <- birthwt %>%
    mutate(race = recode_factor(race, `1` = "white", `2` = "black", `3` = "other")) %>%
    mutate_at(c("mother.smokes", "hypertension", "uterine.irr", "birthwt.below.2500"),
              ~ recode_factor(.x, `0` = "no", `1` = "yes"))
```
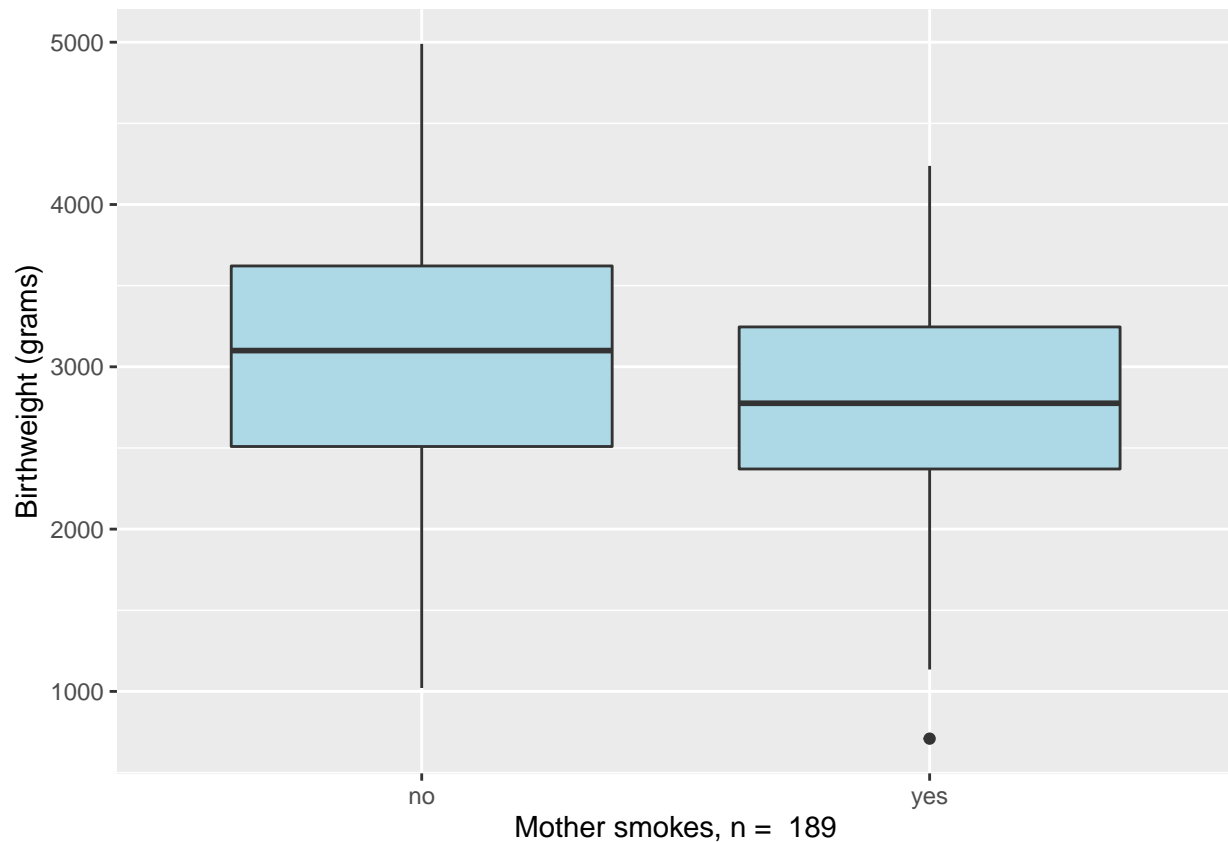
```
# Create boxplot showing how birthwt.grams varies between
# history of smoking
qplot(x = mother.smokes, y = birthwt.grams,
      geom = "boxplot", data = birthwt,
      xlab = paste("Mother smokes, n = ", nrow(birthwt)),
      ylab = "Birthweight (grams)",
      fill = I("lightblue"))
```


Above boxplot shows that, though in the smoking mother group the birthweight values including the mean are lower than the non-smoking mother group; we can see some significant overlapping between both groups. Hence, this boxplot distribution can't signifies that lower birthweight of infants occured due to smoking of their mother. So, we can define our null hypothesis as the mothers having smoking habit don't need to be concerned about baby's birthweight.

Testing of first hypothesis

We tried two-tailed hypothesis test for our case.

```
birthwt.t.test <- t.test(birthwt.grams ~ mother.smokes, data = birthwt)
birthwt.t.test
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  birthwt.grams by mother.smokes
## t = 2.7299, df = 170.1, p-value = 0.007003
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##    78.57486 488.97860
## sample estimates:
##  mean in group no mean in group yes
##          3055.696          2771.919
```

```
birthwt.t.test$p.value
```

```
## [1] 0.007002548
```

Our study finds that birth weights are on average greater in the non-smoking group compared to the smoking group.

Here, t-statistic = 2.73, p-value = 0.007 = 0.7%, confidence interval(95%) = 78.575, 488.979.

As the p-value is less than 5%, the null hypothesis can be rejected. So, the overlapping in the boxplot is not significant which means that, mothers having smoking habit are prone to give birth to infant with lower birthweight than the non-smoking mothers.

2. Comparison of infants birthweight between a group of mothers with a presence of uterine irritability and a group of mothers without it.

```
# Create boxplot showing how birthwt.grams varies between
# history of hypertension
qplot(x = uterine.irr, y = birthwt.grams,
      geom = "boxplot", data = birthwt,
      xlab = paste("Mother with a presense of uterine irritability, n = ", nrow(birthwt)),
      ylab = "Birthweight (grams)",
      fill = I("lightblue"))
```
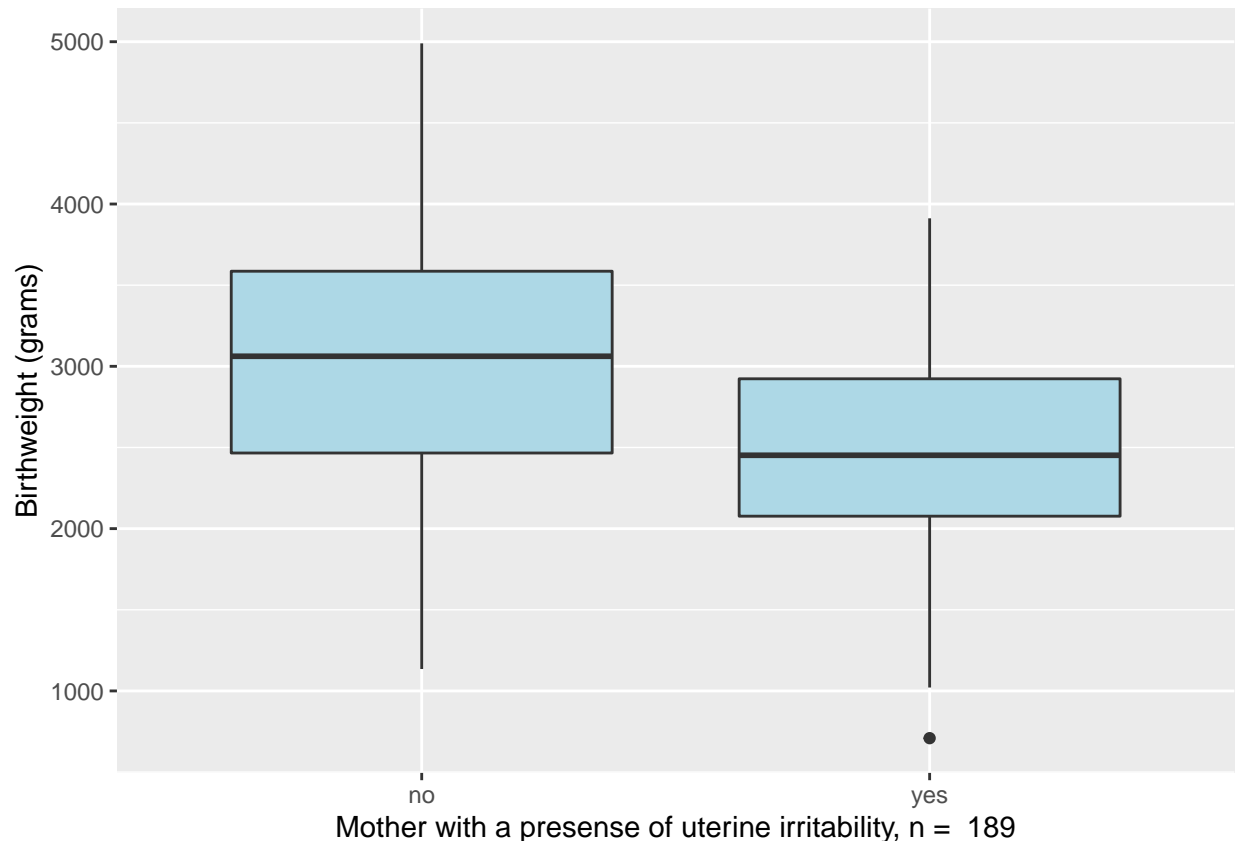
Above boxplot shows that, though in the uterine irritability mother group the birthweight values including the mean are lower than the non-uterine irritability mother group; we can see some significant overlapping between both groups. Hence, this boxplot distribution can't signifies that lower birthweight of infants occured due to uterine irritability of their mother. So, we can define our null hypothesis as the mothers having uterine irritability don't need to be concerned about baby's birthweight.

Testing of second hypothesis

We tried two-tailed hypothesis test for our case.

```
birthwt.t.test <- t.test(birthwt.grams ~ uterine.irr, data = birthwt)
birthwt.t.test
```

```
##
##  Welch Two Sample t-test
##
## data:  birthwt.grams by uterine.irr
## t = 3.8615, df = 35.696, p-value = 0.000455
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  275.8913 886.6553
## sample estimates:
##  mean in group no mean in group yes
##          3030.702          2449.429
```

```
birthwt.t.test$p.value
```

[1] 0.0004550226

Our study finds that birth weights are on average greater in the non-uterine irritability group compared to the uterine irritability group.

Here, t-statistic = 3.861, p-value = 5×10^{-4} = 0.046%, confidence interval(95%) = 275.891, 886.655.

As the p-value is less than 5%, the null hypothesis can be rejected. So, the overlapping in the boxplot is not significant which means that, mothers having uterine irritability are prone to give birth to infant with lower birthweight than the non-uterine irritability mothers.