Assignment 8: Birth Times

Alizeh Murtaza

2023-03-31

Exercise 1

```
live_births <- nsfg6 %>%
  filter(outcome == 1) %>%
  mutate(
    birth_order = if_else(
      birthord == 1,
      "first",
      "other"
    )
  )
```

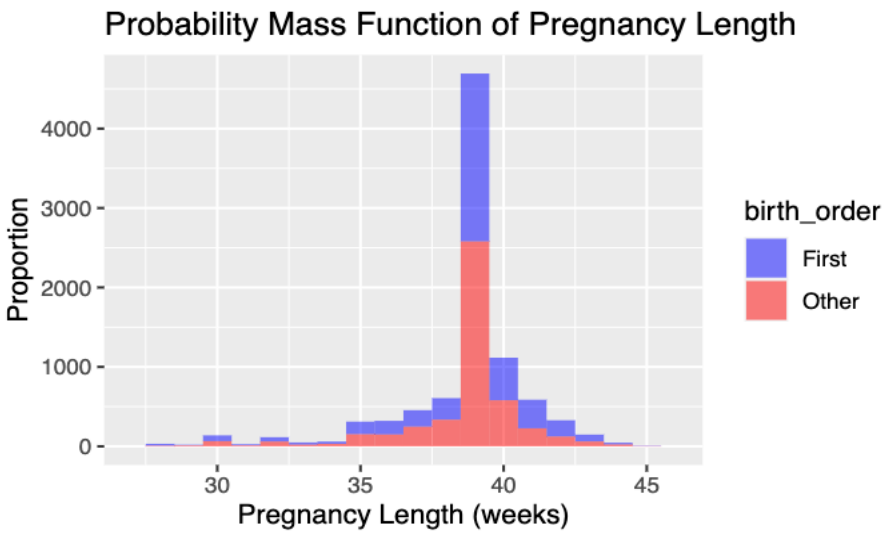
```
pregnancy_length <- live_births %>%
  select(prglngth, birth_order)
```

Exercise 2

```
ggplot(pregnancy_length, aes(x = prglngth, fill = birth_order)) +
  geom_histogram(binwidth = 1, alpha = 0.5) +
  xlim(27, 46) +
  labs(title = "Probability Mass Function of Pregnancy Length",
       x = "Pregnancy Length (weeks)",
       y = "Proportion") +
  scale_fill_manual(values = c("blue", "red"),
                    labels = c("First", "Other"))
```

```
## Warning: Removed 80 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 4 rows containing missing values (geom_bar).
```



i. The mode is at ariund 37 weeks.

ii. To confirm the statement that "first born children either arrive early or arrive late when compared with non-first-borns," we should most definitely use a hypothesis test because correlation does not equal causation. This dataset may have other variables that can contribute to its results.

Exercise 3

```
pregnancy_length %>%
  group_by(birth_order) %>%
  summarize(
    mean = mean(prglngth),
    median = median(prglngth),
    sd = sd(prglngth),
    IQR = IQR(prglngth),
    min = min(prglngth),
    max = max(prglngth)
  )
```

birth_order	mean	median	sd	IQR	min	max
first	38.60095	39	2.791901	1	0	48
other	38.52291	39	2.615852	0	4	50

i. The different summary statistics between the two distributions are not significantly different from each other. Everything execept the minimum are close in number to each other.

Exercise 4

i. The test statistic to this experiment is the difference in means of pregnancy lengths between first borns and non- first borns.

ii. Null Hypothesis: There is no difference between pregnancy lengths between first borns and non- first borns.

iii. Alternative Hypothesis: There is a difference between pregnancy lengths between first borns and non- first borns.

First borns have a longer pregnancy length that non-first borns. First borns have a shorter pregnancy length that non-first borns.

iv. This is a two sided hypothesis test.

Exercise 5

```
##null_distribution <- pregnancy_length %>%

##specify(prglngth ~ birth_order)  %>%

##hypothesize(null = "independence")  %>%

##generate(reps = 10000, type = "")  %>%

##calculate(stat = "diff in means", order = c("first", "other"))

## get_p_value(x = pregnancy_length, obs_stat = observed_statistic, direction = "two-sided")

##shade_p_value(obs_stat = observed_statistic, direction = ##"two-sided") +
##labs(title = "Distribution of Difference in Pregnancy Length ##by Birth Order",
##       x = "Difference in Pregnancy Length (weeks)",
##       y = "Probability")
```

Exercise 6

```
birth_bootstraps <- pregnancy_length %>%
  specify(prglngth ~ birth_order) %>%
  generate(10000, type = "bootstrap") %>%
  calculate(stat = "diff in means", order = c("first", "other"))
```

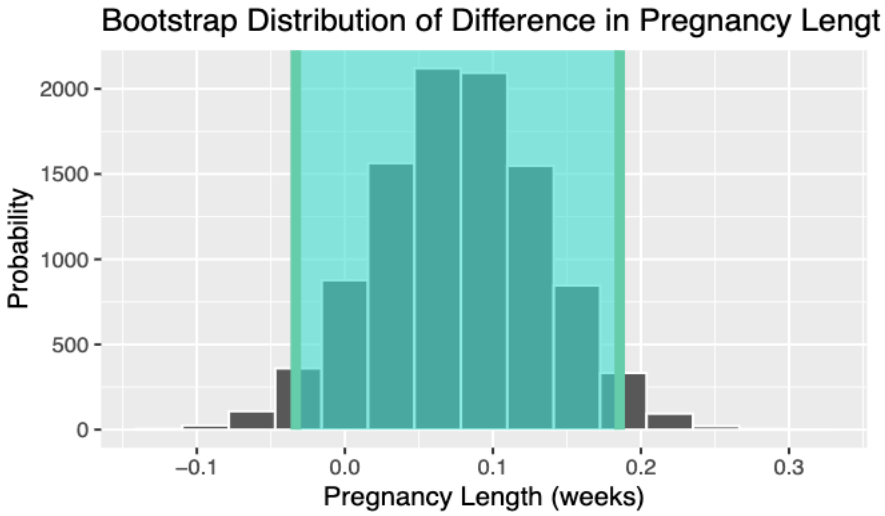
```
bootstrap_ci <- birth_bootstraps %>%
  get_confidence_interval()
```

```
## Using `level = 0.95` to compute confidence interval.
```

```
bootstrap_ci
```

lower_ci	upper_ci
-0.0331656	0.1856359

```r
birth_bootstraps %>%
  visualize() +
  shade_confidence_interval(bootstrap_ci) +
  labs(title = "Bootstrap Distribution of Difference in Pregnancy Length",
       x = "Pregnancy Length (weeks)",
       y = "Probability")
```



Exercise 7

```r
bootstrap_results <- cohens_d_bootstrap(data = pregnancy_length, model = prglngth ~ birth_orde
```

```r
bootstrap_report(bootstrap_results)
```

```
## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 5000 bootstrap replicates
##
## CALL :
## boot::boot.ci(boot.out = cohens_d_bootstrap_sim, type = c("perc"))
##
## Intervals :
## Level     Percentile
## 95%   (-0.0112,  0.0713 )
## Calculations and Intervals on Original Scale
##
## Response variable
## prglngth
##
## Explanatory variable
## birth_order
##
## Explanatory category with larger mean
## first
```

```
##
## Explanatory category with smaller mean
## other
##
## Cohen's d observed value
## 0.0288791
```

```
plot_ci(bootstrap_results)
```