STAT 22000: Final Part 3

Zixi Li

Q1

With the following R codes, we know that

- the number of cases that the mother is married: 613
- the number of cases that the mother is unmarried: 386
- the averae age of the married mothers in the data: 29.14
- the averae age of the unmarried mothers in the data: 23.56
- the standard deviation (SD) of the age of the married mothers in the data: 5.52
- the standard deviation (SD) of the age of the unmarried mothers in the data: 5.66

```
favstats(mage ~ marital, data=nc)
```

```
##       marital min Q1 median Q3 max    mean      sd   n missing
## 1     married  18 25     29 33  50 29.1419 5.52396 613       0
## 2 not married  13 19     22 27  41 23.5622 5.65800 386       0
```

Q2

With $df = min(613 - 1, 386 - 1) = 385$, the critical value for 95% CI is $t^* \approx 1.966$.

```
qt(0.05/2, df=385, lower.tail = F)
```

```
## [1] 1.96614
```

The 95% CI for the mean difference $\mu_m - \mu_n$ is

$$\bar{x}_m - \bar{x}_n \pm t^* \sqrt{\frac{s_m^2}{n_m} + \frac{s_n^2}{n_n}} \approx 29.14 - 23.56 \pm 1.966 \sqrt{\frac{(5.52)^2}{613} + \frac{(5.66)^2}{386}} \approx (4.864, 6.296)$$

Here the subscript "m" represents married, "n" represents not married.

Q3

The age of married mothers is independent with the age of unmarried mothers.

The sample size $n_m = 613$ and $n_n = 386$ are both sufficiently large so that the t-approximation is good even when the data are clearly skewed.