HW 03: ANOVA + Multiple linear regression

Jwalin Patel

15/03/2021

```
library(tidyverse)
library(broom)
library(knitr)
```

```
babies <- read.csv
```

Question 1

$H0 : u_1 = u_2 = u_3 = u_4 = u_5$ *vs* $Ha :$ *atleast one* u_i *is not equal to another*

Question 2

To evaluate these hypothesis, the p-value is calculated using an F distribution with K-1 and n-K degrees of freedom. Therefore: $(5-1) = 4$, and $(45-5) = 40$ $degreesof freedom.$

Question 3

Since the p-value of 0.0168 in consideration is less than our significance level of $a = 0.05$, we can say that the data provides sufficient evidence that at least one of the teaching methods has/have a mean test score that is different from the other teaching methods. Therefore, we can reject the null hypothesis.

Question 4

$$D.O.F : A = K - 1 = 5 - 1 = 4$$

$$Residuals\ D.O.F : D = n - K = 1172 - 5 = 1167$$

$$Total\ D.O.F : G = n - 1 = 1171$$

$$Sum\ Sq : B = 501.54 * A = 501.54 * 4 = 2006.16$$

$$Residuals\ Mean\ Sq : E = 267382/D = 267382/1167 = 229.1191$$

$$F - statistic : C = 501.54/E = 501.54/229.1191 = 2.188993$$

$$Total\ Sum\ of\ Squares : H = B + 267382 = 2006.16 + 267382 = 269388.2$$

$$F = no\ value$$

Question 5

1) Normality The mean and median value appear to be approximately 40 for every education attainment as shown by the box-plot, thus we do not expect to observe high level of skewness in any particular direction. Moreover, while the box plot seems to suggest that there are few outliers within each educational attainment, none of the outliers seem to be extreme. Due to the sample size, it is also unlikely that these minor outliers cause our estimates to be up- or down-biased. Therefore we can suggest that normality is satisfied.

2) Constant Variance Using the summary statistics, we can check to see if constant variance is satisfied. If(f) the ratio between the largest and smallest standard deviations among the education attainment groups is less than 2, we can state that the condition is satisfied. The smallest standard deviation is observed in the Bachelor's group (13.62), whereas the largest is obbserved in the Jr. College group (18.1). Dividing the two yields 1.328928. Thus, the ratio is less than 2. We can say that the constant variance condition is satisfied.

3) Independence So far, based on what we know about the study, we do not seem to have reason to believe that one educational attainment groups is not independent of another. Thus, we can suggest that the independence condition is satisfied.

Question 6

Given that the p-value of 0.0682 is more than our significance level of $a = 0.05$, we can conclude that the data do not provide sufficient evidence that at least one educational attainment has (mean hours worked) that is different from another educational attainment group. Thus, we fail to reject the null hypothesis.

...