Anova

Washim Ahmed

12/25/2016

```
library(readxl)
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Conflicts with tidy packages ----------------------------------------------
```

```
## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
age_of_workers <- read_excel("~/manipal_practice/anova.xls")
summary(age_of_workers)
```

```
##  Himachal Pradesh  Uttarakhand      Tamil Nadu
##  Min.   :21.00    Min.   :18.00   Min.   :28.00
##  1st Qu.:26.00    1st Qu.:23.00   1st Qu.:35.00
##  Median :31.00    Median :29.00   Median :41.50
##  Mean   :31.53    Mean   :28.77   Mean   :41.01
##  3rd Qu.:37.00    3rd Qu.:35.00   3rd Qu.:47.00
##  Max.   :42.00    Max.   :39.00   Max.   :53.00
```

```
combine_all_ages <- gather(age_of_workers, "Factory", "Age", 1:3)
anova <- aov(formula = Age ~ Factory, data = combine_all_ages)
summary(anova)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Factory       2   9726    4863   102.8 <2e-16 ***
## Residuals   351  16601      47
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

I can see p value is **very very tiny** which is less than significance value of 0.01 (0.01 given in assignment) therefore I can **reject null hypothesis** which means there is a **significant age difference across factories**.

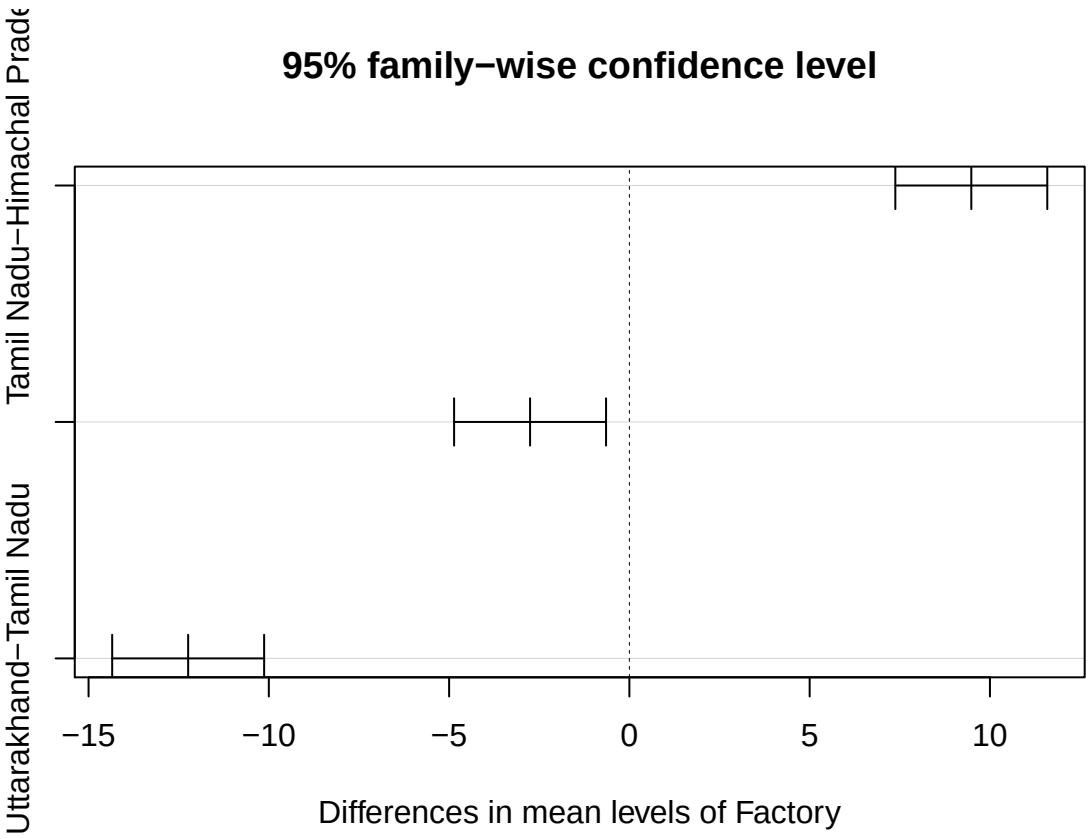
I will do **TukeyHSD** test to identify which factory workers does have significance age difference.

```
TukeyHSD <- TukeyHSD(anova)
TukeyHSD
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Age ~ Factory, data = combine_all_ages)
##
## $Factory
##                                   diff        lwr         upr     p adj
```

```
## Tamil Nadu-Himachal Pradesh    9.483051   7.375677  11.5904249 0.0000000
## Uttarakhand-Himachal Pradesh  -2.754237  -4.861611  -0.6468632 0.0063842
## Uttarakhand-Tamil Nadu       -12.237288 -14.344662 -10.1299141 0.0000000
```

```
plot(TukeyHSD)
```



1. For Tamil Nadu and Himachal Pradesh where P value $0 < 0.01$ which means Tamil Naru's workers are older than Himachal Pradesh
2. For Uttarakhand and Tamil Nadu where P value $0 < 0.01$ which means Tamil Naru's workers are older than Uttarakhand

The workers at Tamil Naru factory are older than the ones at the other factories.