A4Q3

Yuqi Gu

7/22/2024

(a)

```
water.dat <- read.table("~/Desktop/STAT 332/Assignments/A4/water.txt")
names(water.dat) <- c("burner", "salt", "time")
water.dat$burner <- as.factor(water.dat$burner)
water.dat$salt <- as.factor(water.dat$salt)
water.aov <- aov(time ~ burner + salt + burner:salt, data = water.dat)
summary(water.aov)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## burner        3 110.50   36.83  73.667 1.86e-14 ***
## salt          3   6.83    2.28   4.556  0.00909 **
## burner:salt   9   3.33    0.37   0.741  0.66931
## Residuals    32  16.00    0.50
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(b)

```
qf(0.95, 9, 32) #Test for interaction
```

```
## [1] 2.188766
```

```
qf(0.95, 3, 32) #Test for main effects of burner
```

```
## [1] 2.90112
```

```
qf(0.95, 3, 32) #Test for main effects of salt
```

```
## [1] 2.90112
```

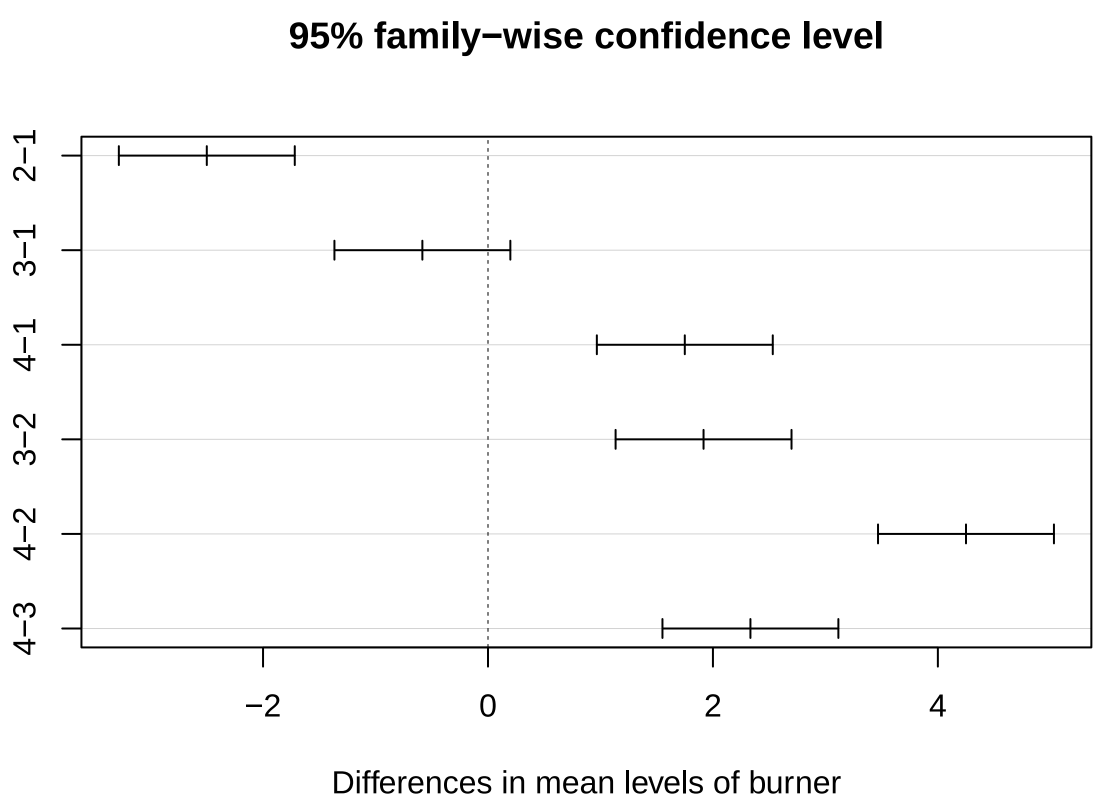
Since $0.741 < 2.188766$ and the corresponding p-value is 0.66931 , we do not reject the null hypothesis. Consequently, the interactions between burner and salt are not significant for the mean time to boiling. Thus, the mean time to boiling stay the same if the interactions between salt and burner are different. Since $73.667 > 2.90112$ and the corresponding p-value is 1.86e-14, which is much smaller than 0.05, we reject H0. Thus, we conclude that the effect of burner on mean time to boiling is significantly important. Thus, at the 5% level, the mean time to boiling differs if the burners are different. Since $4.556 > 2.90112$ and the corresponding p-value is 0.00909, which is smaller than 0.05, we reject H0. Thus, we conclude that the effect of salt on mean time to boiling is significantly important. Thus, at the 5% level, the mean time to boiling differs if the numbers of teaspoons of salt are different.

(c)

```r
TukeyHSD(water.aov, "burner", conf.level=0.95)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = time ~ burner + salt + burner:salt, data = water.dat)
##
## $burner
##           diff        lwr       upr     p adj
## 2-1 -2.5000000 -3.2821253 -1.717875 0.0000000
## 3-1 -0.5833333 -1.3654587  0.198792 0.2016864
## 4-1  1.7500000  0.9678747  2.532125 0.0000052
## 3-2  1.9166667  1.1345413  2.698792 0.0000010
## 4-2  4.2500000  3.4678747  5.032125 0.0000000
## 4-3  2.3333333  1.5512080  3.115459 0.0000000
```

```r
plot(TukeyHSD(water.aov, "burner"))
```



The right-front burner, which is 2 burner, is the fastest at boiling water. This is because the difference of 2-1 is -2.5000000, that of 3-2 is 1.9166667, and that of 4-2 is 4.2500000. In other words, the mean time to boiling of burner 2 is 2.5 minutes shorter than that of burner 1. The mean time to boiling of burner 2 is 1.92 minutes shorter than that of burner 3. The mean time to boiling of burner 2 is 4.25 minutes short than that of burner 4. Also, the mean time to boiling of burner 3 is 0.58 minutes shorter than that of burner 1 since the difference of 3-1 is -0.5833333. The mean time to boiling of burner 4 is 1.75 longer than that of burner 1 and the mean time to boiling of burner 4 is 2.33 longer than that of burner 3 since the difference of 4-1 is 1.7500000 and that of 4-3 is 2.3333333. Moreover, we could observe the confidence interval of 2-1 does not contain 0 and the upper bound is smaller than 0. The confidence interval of 3-2 does not contain 0 and the lower bound is larger than 0. The confidence interval of 4-2 does not contain 0 and the lower bound is larger than 0. These all indicate that burner 2 is faster than all the other three burners. Thus, the right-front burner is the fastest at boiling water. Also, observing the plots, we find that the 95% confidence interval of 2-1 does not contain 0 and it is negative and the 95% confidence intervals of 3-2 and 4-2 are positive and do not contain 0. Thus, the time to boiling of burner 2 is shorter than that of 1, 3, and 4. Thus, the right-front burner is the fastest at boiling water.

(d)

```
tmp <- model.tables(water.aov,type = "means")
tmpburner <- tmp[[1]][[2]]

burner.tab <- data.frame(names (tmp$tables[[2]]), c(tmp$tables[[2]]) )
colnames(burner.tab) <- c(names (tmp$tables[2]), "means")
rownames(burner.tab)<-NULL

MSE <- 0.5
J <- 4
n_0 <- 3
L_hat <- (burner.tab$means[1]+burner.tab$means[2])/2 - (burner.tab$means[3]+burner.tab$means[4])/2
var <- MSE*((1/2)^2/(J*n_0)+(1/2)^2/(J*n_0)+(-1/2)^2/(J*n_0)+(-1/2)^2/(J*n_0))

t <- qt(0.975, 32)
lwr <- L_hat - t*sqrt(var)
upr <- L_hat + t*sqrt(var)
lwr
```

```
## [1] -2.249121
```

```
upr
```

```
## [1] -1.417546
```

The corresponding 95% confidence interval is [-2.249121, -1.417546]. Since it does not contain 0 and the upper bound of this confidence interval is lower than 0, we conclude that the two right burners does not take the same time to boil water than the left ones. Specifically, the two right burners take shorter time to boil water than the left ones.

(e)

```
tmpsalt <- tmp[[1]][[3]]

salt.tab <- data.frame(names(tmp$tables[[3]]), c(tmp$tables[[3]]) )
colnames(salt.tab) <- c(names (tmp$tables[3]), "means")
```

```
rownames(salt.tab)<-NULL

I <- 4
weightsalt <- c(0, 2, 4, 6)
cisalt <- weightsalt-mean(weightsalt)

L_hatsalt <- salt.tab$means[1]*cisalt[1]+salt.tab$means[2]*cisalt[2]+
  salt.tab$means[3]*cisalt[3]+salt.tab$means[4]*cisalt[4]
varsalt <- MSE*((cisalt[1])^2/(I*n_0)+(cisalt[2])^2/(I*n_0)+
              (cisalt[3])^2/(I*n_0)+(cisalt[4])^2/(I*n_0))

lwrsalt <- L_hatsalt - t*sqrt(varsalt)
uprsalt <- L_hatsalt + t*sqrt(varsalt)
lwrsalt
```

```
## [1] -3.359457
```

```
uprsalt
```

```
## [1] 0.3594572
```

The corresponding 95% confidence interval is [-3.359457, 0.3594572]. Since it contains 0, we do not reject null hypothesis and so we conclude that there is no linear trend in the time to boil water due to the level of salt.

```
means <- aggregate(time ~ salt, data = water.dat, mean)

contrast_coeff <- cisalt
contrast_value <- sum(contrast_coeff * means$time)

n <- n_0*I
SS_contrast <- (contrast_value^2) / sum(contrast_coeff^2 / n)

anova_model <- aov(time ~ burner * salt, data = water.dat)
anova_summary <- summary(anova_model)

F_contrast <- SS_contrast / MSE

p_value <- pf(F_contrast, df1 = 1, df2 = anova_model$df.residual, lower.tail = FALSE)

# Output the results
contrast_value
```

```
## [1] -1.5
```

```
SS_contrast
```

```
## [1] 1.35
```

```r
F_contrast
```

```
## [1] 2.7
```

```r
p_value
```

```
## [1] 0.1101398
```

```r
b1 <- contrast_value/(n*sum(contrast_coeff^2))
b1
```

```
## [1] -0.00625
```

Since the p-value calculated is 0.1101398, which is larger than 0.05, we do not reject the null hypothesis and conclude that there is no trend in the time to boil water due to the level of salt.