STP 531

Applied Analysis of Variance

Homework 2

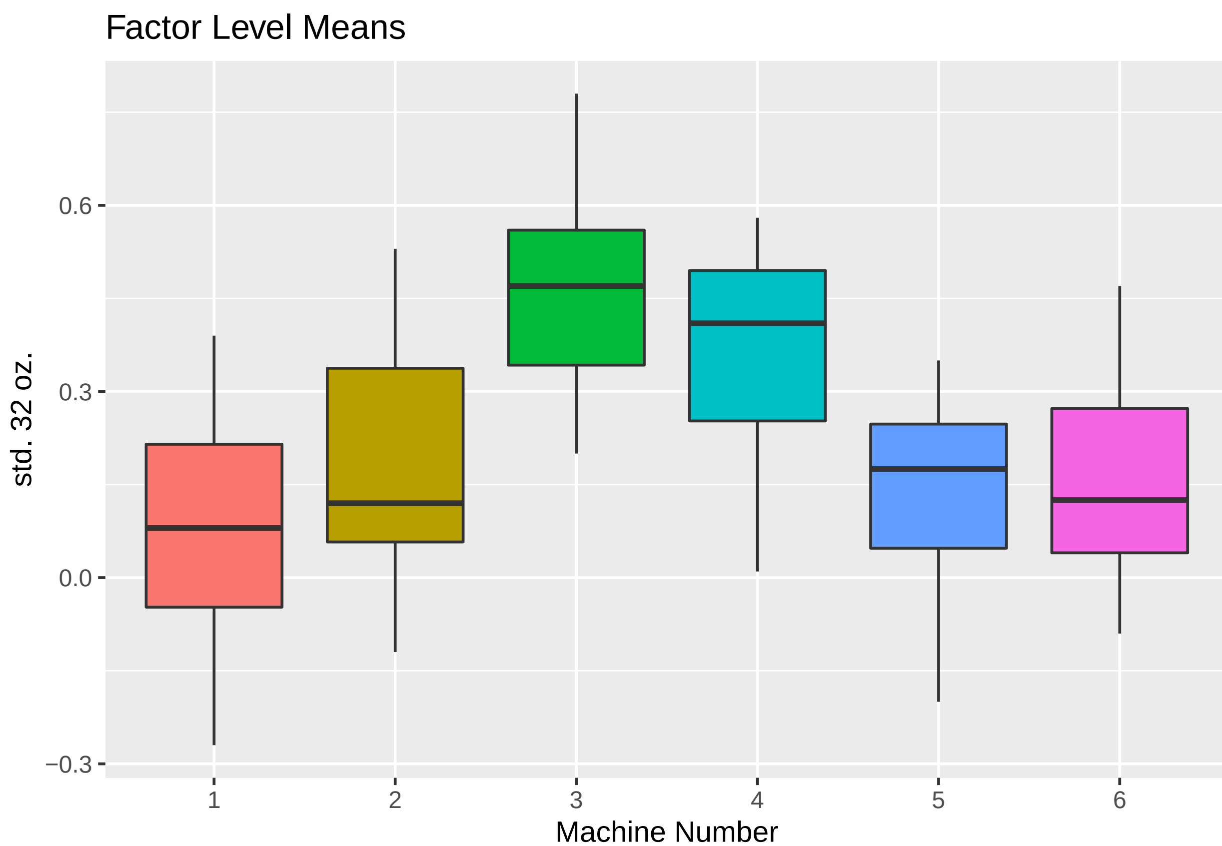
Nathan A. Nguyen

10 February 2021

Question 16.11

$$Y_{ij} = \mu_{ij} + \epsilon_{ij}$$
$$n_T = 120$$
$$n_i = 20$$
$$i = 1, 2, \ldots, 6 \text{ and } j = 1, 2, \ldots, 20$$

Part A



Part B

Table 1: Fitted Values

group	mean
1	0.0735
2	0.1905
3	0.4600
4	0.3655
5	0.1250
6	0.1515

Part C

Table 2: Deviations

Group_1	Group_2	Group_3	Group_4	Group_5	Group_6
-0.2135	0.2695	-0.25	0.1245	-0.315	-0.1015
0.1265	-0.0805	0.32	0.2145	0.145	-0.2015
-0.0035	-0.0705	-0.14	0.1545	-0.065	0.1285
0.1065	0.2795	-0.01	-0.0755	-0.015	0.3185
0.3065	0.0495	-0.24	-0.0955	0.105	-0.0315
0.0265	-0.1305	-0.11	0.1845	0.025	0.1185
-0.1135	-0.3105	0.08	0.0345	-0.115	-0.0715
-0.3435	0.1395	-0.22	-0.2255	0.095	0.0185
0.1965	-0.1305	0.01	0.1145	0.165	0.2785
-0.2835	-0.2205	0.16	-0.0255	0.015	-0.2215
0.3165	-0.1405	0.01	-0.3555	0.075	0.0485
-0.1435	0.3395	0.09	-0.0355	0.175	-0.1415
-0.0935	0.2295	0.13	-0.1855	-0.235	-0.0515
0.2065	0.0995	0.25	-0.2355	0.145	0.0085
0.0165	0.1695	-0.01	0.1145	-0.325	-0.2115
0.0565	-0.1505	0.02	0.1745	0.115	-0.0215
0.1865	-0.0205	-0.02	0.1445	0.075	0.2785
-0.0035	-0.1705	0.04	0.0545	0.015	0.1985
-0.0835	-0.0805	-0.26	0.0845	0.225	-0.2415
-0.2635	-0.0705	0.15	-0.1655	-0.305	-0.1015

The residuals sum to zero $-2.3592239 \times 10^{-16}$, which is essentially 0.

Part D

Table 3: ANOVA table

Source	SS	df	MS
Between Machines	2.29	5	0.46
Error	3.53	114	0.03
Total	5.82	119	

Part E

Let $\alpha = 0.05$

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$$
$$H_a : \text{not all } \mu_i \text{ are equal}$$
$$F^* = \frac{MSTR}{MSE}$$

- if $F^* \leq F(1-\alpha; r-1; n_T - r)$, conclude H_0.
- if $F^* > F(1-\alpha; r-1; n_T - r)$, conclude H_a.

$$F^* = \frac{MSTR}{MSE} = \frac{0.46}{0.03} \approx 15.33$$
$$F(1-\alpha; r-1; n_T - r) = F(0.95, 5, 114) \approx 2.29$$

Since $F^* > F_{critical}$ there is sufficient evidence to suggest that the factor level means are all not equal. In other words, there exists a difference between the treatments.

Part F

The p-value is $3.6363746 \times 10^{-11}$, which is is consistent with part E – that there is sufficient evidence to reject the null hypothesis and proceed with proceed with the notion that not all of the factor level means are equal.

Part G

The variation variation between groups is indeed large compared the variation within groups. Based on the plots alone, we can see that there is nothing really standing out with regards to the width of each of the groups' box. In other words, the interval of each box are approximately the same.

As for the variation between groups. That is obviously seen. if we isolate group 2 and 3, their 1st-3rd quartiles don't even fall within the other 4 groups' distributions. Also, $E\{MSE\} < E\{MSTR\}$ where $E\{MSTR\}$ is larger than an order of magnitude – more evidence to suggest that the variation between groups is more important than variation within groups.

Appendix

Question 16.11

Part A

```
getdata <- function(...){
  e = new.env()
  name = data(..., envir = e)[1]
  e[[name]]
}

data <- getdata("FillingMachines")

data2 <- data
# drop redundant row column
data2 <- subset(data2, select = -row)
#force group to factor variables (indicators)
data2$group <- as.factor(data2$group)

p1 <- data2 %>% ggplot(aes(x = group, y = y)) +
  geom_boxplot(aes(fill = factor(group))) +
  theme(legend.position = "none") +
  labs(title = "Factor Level Means",
       x = "Machine Number",
       y = "std. 32 oz.")
p1
```

Part B

```
factor_level_means <- data2 %>%
  group_by(group) %>%
  summarise(mean = mean(y))

kable(factor_level_means[], caption = "Fitted Values",
      format = "markdown") %>%
  kable_styling(position = "center")
```

Part C

```
deviation <- data2 %>%
  group_by(group) %>%
  summarise(resid = y-mean(y))

corgi = as.data.frame(cbind(Group_1 = deviation$resid[deviation$group == 1],
                          Group_2 = deviation$resid[deviation$group == 2],
                      Group_3 = deviation$resid[deviation$group == 3],
                      Group_4 = deviation$resid[deviation$group == 4],
                      Group_5 = deviation$resid[deviation$group == 5],
                      Group_6 = deviation$resid[deviation$group == 6]))
zero = sum(corgi$Group_1) + sum(corgi$Group_2) + sum(corgi$Group_3)
+ sum(corgi$Group_4) + sum(corgi$Group_5) + sum(corgi$Group_6)
machine1_dev <- as.data.frame(subset(deviation, group == 1))
```

```r
kable(corgi[], caption = "Deviations", format = "markdown") %>%
  kable_styling(position = "center")
```

Part D

```r
n_T <- 120
n_i <- 20
Y.. <- sum(data2$y)
YBar.. <- Y.. / n_T
SSTR <- n_i*((0.0735-YBar..)^2+(0.1905-YBar..)^2+(0.46-YBar..)^2+
               (0.3655-YBar..)^2+(0.125-YBar..)^2+(0.1515-YBar..)^2)
SSE <- sum((deviation$resid)^2)
MSTR <- SSTR/5
MSE <- SSE/(n_T-6)
```

Part F

```r
p.value <- pf(F.star, df1 = 5, df2 = 114, lower.tail = FALSE)
```