# STAT565_Lab

Shen Qu

'Mar 11, 2019

(a). Plot the data and report the plot here (A plot with data and means of treatment combinations). Do not report code here. Describe the observed relationship between two factors.

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    90 obs. of  6 variables:
## $ method : chr  "a" "a" "a" "a" ...
## $ variety: num  1 1 1 1 1 1 2 2 2 2 ...
## $ y      : num  221 241 191 221 251 181 271 151 206 286 ...
## $ m      : Factor w/ 3 levels "a","b","c": 1 1 1 1 1 1 1 1 1 1 ...
## $ v      : Factor w/ 5 levels "v1","v2","v3",..: 1 1 1 1 1 1 2 2 2 2 ...
## $ mv     : Factor w/ 15 levels "a.v1","b.v1",..: 1 1 1 1 1 1 4 4 4 4 ...
```
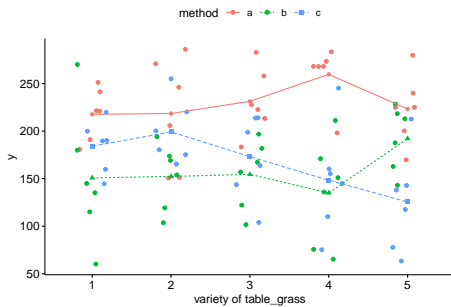

(b). Obtain the numerical summary for each treatment combination and factor levels separately. Report them here in a tabular form

method	min	Q1	median	Q3	max	mean	sd	n	missing
a	151	201.5	226.5	268	286	230.1	39.86	30	0
b	60	125.2	160	186.5	270	157	48.93	30	0
c	63	143.2	164.5	200	255	166.1	49.28	30	0

variety	min	Q1	median	Q3	max	mean	sd	n	missing
a.1	181	198.5	221	236	251	217.7	27.33	6	0
b.1	60	120	140	171.2	270	150.8	70.53	6	0
c.1	145	167.5	190	197.5	220	184.2	27.28	6	0
a.2	151	164.8	226	264.8	286	218.5	58.89	6	0
b.2	104	127.8	161.5	172.8	194	152.3	34.45	6	0
c.2	165	176.2	190	215	255	199.2	33.68	6	0
a.3	183	215.5	225.5	250.5	283	231.3	35.02	6	0
b.3	102	130.8	162	178.2	197	154.5	36.16	6	0
c.3	104	149	181.5	210.2	214	173.2	44.09	6	0
a.4	198	268	268	271.8	283	259.7	30.77	6	0
b.4	65	91	143.5	166	211	135	56.05	6	0
c.4	75	118.8	150	158.8	245	148.3	57.24	6	0
a.5	170	206.2	225	236.2	280	223.3	37.1	6	0
b.5	143	169.2	200.5	216.8	228	192.2	33.68	6	0
c.5	63	88	128	141.8	213	125.5	53.55	6	0
1	60	148.8	190	220.8	270	184.2	51.86	18	0
2	104	156.8	177.5	216.5	286	190	50.19	18	0
3	102	158.8	190	214	283	186.3	49.5	18	0
4	65	138.2	165.5	262.2	283	181	73.99	18	0
5	63	143	194	223.2	280	180.3	57.84	18	0

(c). Fit the two-factor factorial model and report the complete ANOVA table here. Do not report code here. The complete ANOVA table should have a row for each of the following: main effects of each treatment, two-factor interaction effects, error and total

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
m	2	95316	47658	24.25	7.525e-09
v	4	1138	284.5	0.1448	0.9648
m:v	8	37449	4681	2.382	0.02409
Residuals	75	147377	1965	NA	NA
Total	89	281279.2	3160.44	NA	NA

(d). Based on the ANOVA table write your conclusion appropriately. Perform all the necessary tests and report the conclusion along with the p-value.

The line plot shows that not all lines are parallel. Difference in y between methods is not same for different varieties. There could be an interaction effect.

According to ANOVA table, there is a significant interaction effect from methods and varieties on the y at 5% significance level (P-value=0.02409). That means, effect of method and effect of methods and varieties on y is not independent. Therefore, examie the simple effects.

The table shows the multiple comparisons of levels in varieties and methods.

The difference in y between varieties 2 and 5 is significant when method c is applies (P-value=0.0295).

The difference in y between method a and b are significant when varieties 1 - 4 is applies (P-value=0.0290, 0.0310, 0.0101, <.0001, respectively).

The difference in y between method a and c are significant when varieties 4, 5 is applies (P-value=0.0001, 0.0008, respectively).

The difference in y between method b and c is significant when varieties 5 is applies (P-value=0.0295).

The Tukey test indicates the difference in y between method a and b, a and c are significant when varieties 4 is applies (P-value=0.0002, 0.0015, respectively), the difference in y between method a and c is significant when varieties 5 is applies (P-value=0.0096).

The difference in y between varieties 1-5 applied method a and varieties 5 applied method c are significant (P-value=0.0383, 0.0349, 0.0074, 0.0001, 0.0199, respectively).

The difference in y between varieties 1-4 applied method b and varieties 4 applied method a are significant (P-value=0.0050, 0.0061, 0.0080, 0.0006, respectively).

The difference in y between varieties 4 applied method b and varieties 3 applied method a are significant (P-value=0.0238).

The Scheffe methods indicates the difference in y between varieties 4 applied method a and varieties 5 applied method c are significant (P-value=0.0326).

contrast	estimate	SE	df	t.ratio	p.value	P.adj
c: v2-v5	73.667	25.6	75	2.878	0.0404	
v1: a-b	66.8	25.6	75	2.611	0.0290	
v2: a-b	66.2	25.6	75	2.585	0.0310	
v3: a-b	76.8	25.6	75	3.002	0.0101	Tukey
v4: a-b	124.7	25.6	75	4.871	<.0001	0.0002
v4: a-c	111.3	25.6	75	4.350	0.0001	0.0016
v5: a-c	97.8	25.6	75	3.823	0.0008	0.0096
v5: b-c	66.7	25.6	75	2.605	0.0295	
v1,a-v5,c	92.167	25.6	75	3.601	0.0383	
v2,a-v5,c	93.000	25.6	75	3.634	0.0349	
v3,a-v4,b	96.333	25.6	75	3.764	0.0238	
v3,a-v5,c	105.833	25.6	75	4.135	0.0074	
v4,a-v1,b	108.833	25.6	75	4.252	0.0050	
v4,a-v2,b	107.333	25.6	75	4.194	0.0061	
v4,a-v3,b	105.167	25.6	75	4.109	0.0080	
v4,a-v4,b	124.667	25.6	75	4.871	0.0006	
v4,a-v4,c	111.333	25.6	75	4.350	0.0036	Scheffe
v4,a-v5,c	134.167	25.6	75	5.242	0.0001	0.0326
v5,a-v5,c	97.833	25.6	75	3.823	0.0199	

(e). Provide the plots of residuals here. Do not report code here.
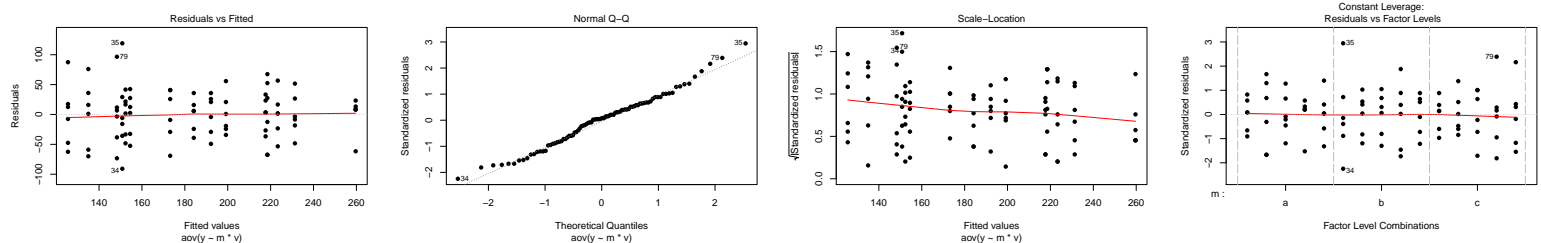
{(f). Based on the residual plots, clearly explain whether assumptions in the model are satisfied or violated.

The plot of studentized residual versus predicted (fitted) value shows that except few outliers, the residuals are evenly distributed about zero at each prededict value (zero mean) and vertical deviations of residuals from zero are about same for each predicted value (constant variance).

The plots of studentized residual versus factor levels didn't show obvious violation of zero mean and constant variance.

The QQ plot shows that some data points are not on the line and flattening at the extremes, which is a litte violation of normality.

(g). The researchers are interested in testing the following contrasts. Test each of the using software, and report conclusion along with p value.

c1. Mean yield for variety 1 versus variety 3 at method b

The $H0$ for contrast is $\mu_{b1.} - \mu_{b3.} = 0 \implies \beta_1 - \beta_3 + (\tau\beta)_{b1} - (\tau\beta)_{b3} = 0$. The $H1$ is they don't equal zero.

The P-value of this contrast test is 0.8865, which is large enough. We cannot reject the $H0$ and conclude that the mean yield for variety 1 versus variety 3 at method b may be same at 5% significance level. The contrast tests with Boferroni or tukey's adjustment have the same results (P-calue=1.0000, 0.9998, respectively).

c2. Mean yield for method a versus average of methods b and c

The $H0$ for contrast is $\mu_{a..} - \frac{1}{2}(\mu_{b..} + \mu_{c..}) = 0 \implies 2\tau_a - \tau_b - \tau_c = 0$. The $H1$ is they don't equal zero.

The P-value of this contrast test is 0.0000, which is small enough. We can reject the $H0$ and conclude that the mean yield for method a versus average of methods b and c are significantly different at 5% significance level. The contrast tests with Boferroni or tukey's adjustment have the same results (P-calue=0.0000, 0.0000, respectively).

c3. Mean yield for method a versus average of methods b and c for variety 1

The $H0$ for contrast is $\mu_{a1.} - \frac{1}{2}(\mu_{b1.} + \mu_{c1.}) = 0$. The $H1$ is they don't equal zero.

The P-value of this contrast test is 0.0265, which is small enough. We can reject the $H0$ and conclude that the mean yield for method a versus average of methods b and c for variety 1 are significantly different at 5% significance level. The contrast tests with Boferroni or tukey's adjustment give the opposite results (P-calue=0.1060, 0.1019, respectively).

c4. Mean yield for method a versus average of methods b and c between variety 1 and variety 2

The $H0$ for contrast is $[\mu_{a1.} - \frac{1}{2}(\mu_{b1.} + \mu_{c1.})] - [\mu_{a2.} - \frac{1}{2}(\mu_{b2.} + \mu_{c2.})] = 0$. The $H1$ is they don't equal zero.

The P-value of this contrast test is 0.5819, which is large enough. We cannot reject the $H0$ and conclude that the mean yield for method a versus average of methods b and c for variety 1 may be same at 5% significance level. The contrast tests with Boferroni or tukey's adjustment give the same results (P-calue=1.0000, 0.9694, respectively).

(h). Report the code here without output.

```
table_grass <- read_excel("Grass.xlsx")
table_grass$m <- as.factor(table_grass$method)
table_grass$v <- factor(table_grass$variety, levels = c(1, 2, 3, 4, 5), labels = c("v1",
    "v2", "v3", "v4", "v5"))
table_grass$mv <- interaction(table_grass$m, table_grass$v)
model_grass <- aov(y ~ m * v, data = table_grass)
model_grass_inter <- aov(y ~ mv, data = table_grass)
str(table_grass)
ggline(data = table_grass, x = "variety", y = "y", add = c("mean", "jitter"),
    shape = "method", color = "method", linetype = "method", ylab = "y", xlab = "variety of table_grass")
pander(favstats(y ~ method, data = table_grass))
pander(favstats(y ~ method | variety, data = table_grass))
summary(model_grass)
plot(model_grass, pch = 16)


# Test simple effects # Method-1# 1.1 Multiple comparison of levels in
```

```r
# variety factor for a given level of method factor #
v_m <- pairs(lsmeans(object = model_grass, specs = ~v | m))
# 1.1 Multiple comparison of levels in method factor for a given level of
# variety factor #
m_v <- pairs(lsmeans(object = model_grass, specs = ~m | v))
# 1.1 same with
lsmeans(model_grass, list(pairwise ~ v | m, pairwise ~ m | v))
# 1.2 All the above Multiple comparisons with Tukey's adjustment # (strict)
test(rbind(v_m, m_v), adjust = "tukey")

# Test all the interaction effect# 1.
pairs(lsmeans(object = model_grass, specs = ~v + m))
pairs(lsmeans(object = model_grass, specs = ~m * v))
pairs(lsmeans(model_grass_inter, "mv"))
lsmeans(model_grass, pairwise ~ v + m)
TukeyHSD(model_grass, conf.level = 0.95)
ScheffeTest(model_grass, conf.level = 0.95)
# 2. another type of interaction effect
contrast(lsmeans(model_grass, ~v * m), interaction = "pairwise")

# Test specific contrasts# 1. Create a inteaction variable by multiplying
# the two factors #
table_grass$mv <- interaction(table_grass$m, table_grass$v)
# 2. Fit the model again using only that new variable #
model_grass_inter <- aov(y ~ mv, data = table_grass)
summary(model_grass_inter)  # Check the ANOVA#
summary.lm(model_grass_inter)  # Check the estimated model coefficients. Same with previous#

# 2. Obtain least square estimates for treatment combinations #
lsmf_grass <- lsmeans(model_grass_inter, "mv")

# 2. Check the order of terms# 3. Write the vectors of contrasts#
contrast_list_grass <- list(b1_b3 = c(0, 1, rep(0, 5), -1, rep(0, 7)), a_b.c = c(rep(c(1,
    -0.5, -0.5), 5)/5), a1_b1.c1 = c(1, -0.5, -0.5, rep(0, 12)), a1_b1.c1__a1_b1.c1 = c(1,
    -0.5, -0.5, 0, 0, 0, -1, 0.5, 0.5, rep(0, 6)))
# Variety 1 versus Variety 3 at Method b # Method a versus average of
# Methods b and c averaged across Varieties. # Method a versus average of
# Methods b and c for Variety 1 # Method a versus average of Methods b and c
# between Variety 1 and Variety 2 #

contrast(lsmf_grass, contrast_list_grass)  # without adjustment #
summary(contrast(lsmf_grass, contrast_list_grass), adjust = "bonferroni")  # Bonferroni#
summary(contrast(lsmf_grass, contrast_list_grass), adjust = "tukey")  #Tukey's#
```