

STAT 3022 Homework 5

Problem 6.22.

Source	DF	SS	MS	F	p-value
Face(Yes/No)	1	12915	12915	129.15	0
Gender(M/F)	1	2500	2500	25.0	2.580991e-16
Interaction	1	400	400	4.0	0.04832569
Residuals	96	9600	100		
Total	99	25415			

```
> 1-pf(129.15, 1, 96)
[1] 0
> 1-pf(25.0, 1, 96)
[1] 2.580991e-06
> 1-pf(4.0, 1, 96)
[1] 0.04832569
```

Problem 6.25.

```
> library(Stat2Data)
> data(Alfalpa)
> aggregate(Ht4~Row, data = Alfalfa,FUN= "mean" )
```

```
Row Ht4
1 a 1.16
2 b 1.57
3 c 1.25
4 d 2.26
5 e 2.46
```

6.25 (a): means of each row of cups

```
> aggregate(Ht4~Acid, data = Alfalfa,FUN= "mean" )
```

```
Acid Ht4
1 1.5HCl 1.466
2 3.0HCl 1.084
3 water 2.670
```

6.25 (a): means of each row of treatment

```
> mean(Alfalpa$Ht4)
```

```
[1] 1.74
```

```
> sd(Alfalpa$Ht4)
```

```
[1] 1.105396
```

6.25 (a): mean and standard deviation for the growth in all cups

```
> lm1 = aov(Ht4~Acid+Row, data=Alfalpa)
```

```
> summary(lm1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Acid	2	6.852	3.426	4.513	0.0487 *
Row	4	4.183	1.046	1.378	0.3235
Residuals	8	6.072	0.759		

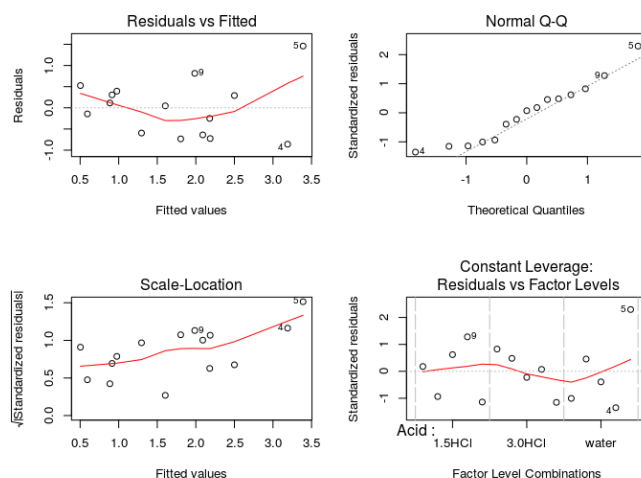
```
---
```

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

6.25 (b): two-way ANOVA table for differences in average growth due to acid using rows as a blocking variable.

```
> par(mfrow = c(2, 2))
```

```
> plot(lm1)
```



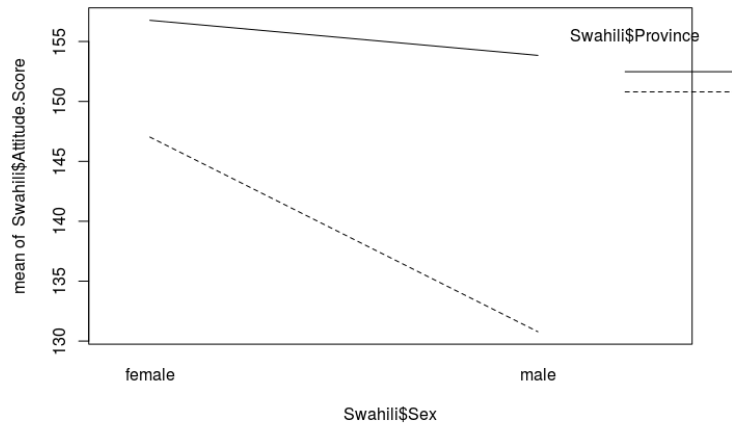
6.25 (c): From the normal q-q plot, the normality is not very good for it shows a general linear pattern while there exists a curve. The residuals vs fitted value plot shows that the zero mean assumption (equal variances) holds. So generally, this model is acceptable.

6.25 (d): Yes. Because the p-value of acid is $0.0487 < 0.05$ which indicates that Acid effect is significant, and the observed variability is because of the treatment.

6.25 (e): No. Because the p-value of Row is $0.3235 > 0.05$ which indicates that distance effect is not significant, and the observed variability is not because of the distance.

Problem 6.28

```
> data(Swahili)
> par(mfrow = c(1, 1))
> interaction.plot(Swahili$Sex, Swahili$Province, Swahili$Attitude.Score)
```



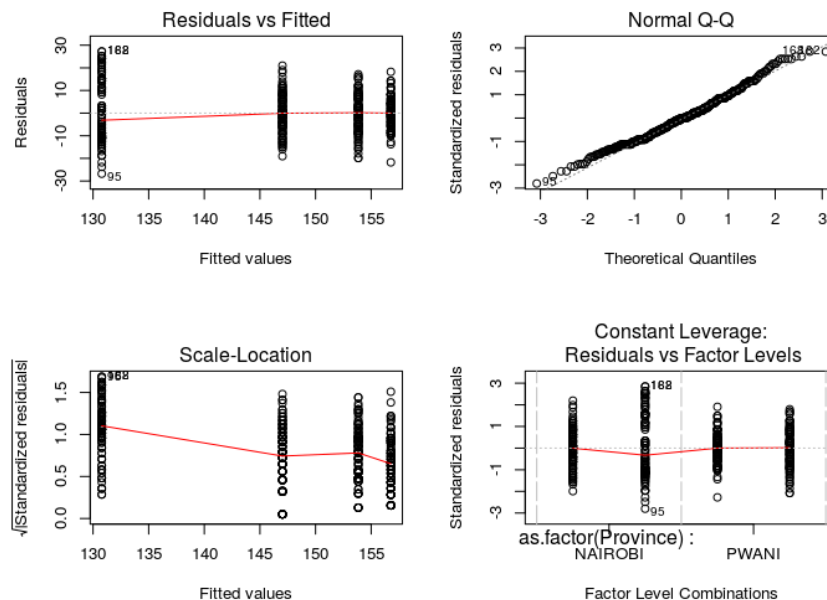
```
> lm2=aov(Attitude.Score~as.factor(Province) * as.factor(Sex), data = Swahili)
> summary(lm2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
as.factor(Province)	1	32275	32275	349.82	< 2e-16 ***
as.factor(Sex)	1	11021	11021	119.45	< 2e-16 ***
as.factor(Province):as.factor(Sex)	1	5320	5320	57.66	1.66e-13 ***
Residuals	476	43917	92		

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

6.28 (a): The plot shows that the average of Nairobi is lower than the average of Pwani. Male has lower average than female. The difference between male and female is larger in Nairobi than in Pwani. The table shows that both p-value of province factor and sex factor are very smaller than 0.05, which indicates that both these two factor are very important to this model.

```
> par(mfrow = c(2, 2))
> plot(lm2)
```

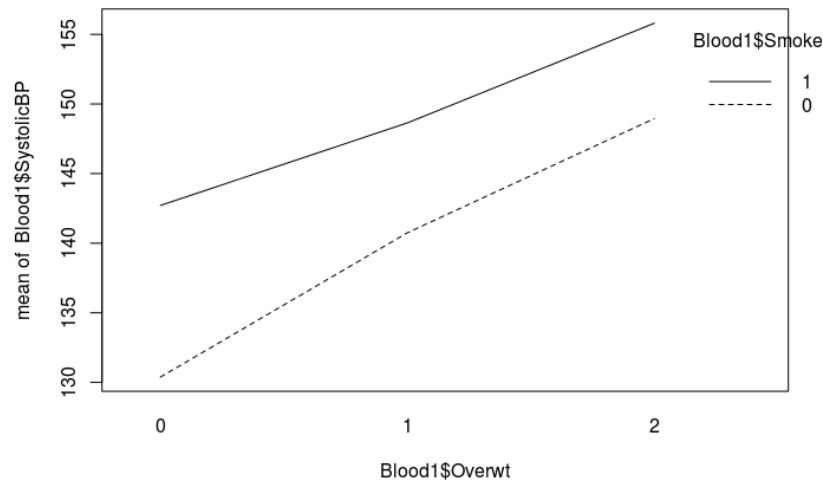


6.28 (b): The normal q-q plot shows a general linear pattern which indicates that the normality is met. The residuals vs fitted value plots shows that the zero mean assumption (equal variance) holds. So both normality and equal variance conditions are satisfied.

6.28 (c): The balanced design needs equal number of observations from two factors' each combinations of levels. In this problem, there is only one school from each province, which means if we have students from the school of one province as our observations, we will have no other observations (students) from both this school and the school in another province at same time. So this can't be a balanced complete factorial design.

Problem 7.12:

```
> data(Blood1)
> Blood1$Overwt = as.factor(Blood1$Overwt)
> Blood1$Smoke = as.factor(Blood1$Smoke)
> interaction.plot(Blood1$Overwt, Blood1$Smoke, Blood1$SystolicBP)
```



```
> lm3 = lm(SystolicBP ~ Smoke * Overwt, data = Blood1)
> anova(lm3)
```

Analysis of Variance Table

Response: SystolicBP

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Smoke	1	14635	14634.9	20.5255	7.387e-06 ***
Overwt	2	23443	11721.6	16.4396	1.224e-07 ***
Smoke:Overwt	2	769	384.4	0.5391	0.5836
Residuals	494	352228	713.0		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> source("LSDtest.R")
```

```
> lout = LSDtest(Blood1$SystolicBP, Blood1$Overwt, alpha = 0.05)
```

```
> lout
```

	pair	est	lwr	upr
1	0~1	-8.051464	-14.45261	-1.650318
2	0~2	-16.865865	-22.24359	-11.488141
3	1~2	-8.814400	-15.11657	-2.512235

```
> subset(lout, sign(lwr)==sign(upr))
```

	pair	est	lwr	upr
1	0~1	-8.051464	-14.45261	-1.650318
2	0~2	-16.865865	-22.24359	-11.488141
3	1~2	-8.814400	-15.11657	-2.512235

```
> ncomp = choose(nlevels(Blood1$Overwt), 2)
```

```
> bout = LSDtest(Blood1$SystolicBP, Blood1$Overwt, alpha = 0.05/ncomp)
```

```
> bout
```

	pair	est	lwr	upr
1	0~1	-8.051464	-15.87754	-0.2253901
2	0~2	-16.865865	-23.44070	-10.2910325
3	1~2	-8.814400	-16.51946	-1.1093404

```
> subset(bout, sign(lwr)==sign(upr))
```

	pair	est	lwr	upr
1	0~1	-8.051464	-15.87754	-0.2253901
2	0~2	-16.865865	-23.44070	-10.2910325
3	1~2	-8.814400	-16.51946	-1.1093404

7.12 (a): The Bonferroni intervals shows that there is not interval cover 0, so all

intervals are significant and that all weight groups are different from each other.

```
> lm4 = aov(SystolicBP ~ Overwt, data = Blood1)
> summary(lm4)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Overwt	2	27801	13900	19.02	1.1e-08 ***
Residuals	497	363274	731		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> TukeyHSD(lm4, "Overwt", data = Blood1)
  Tukey multiple comparisons of means
    95% family-wise confidence level
```

Fit: aov(formula = SystolicBP ~ Overwt, data = Blood1)

```
$Overwt
```

	diff	lwr	upr	p adj
1-0	8.051464	0.3927115	15.71022	0.0366867
2-0	16.865865	10.4316024	23.30013	0.0000000
2-1	8.814400	1.2740746	16.35473	0.0170703

7.12 (b): The TukeyHSD shows that all weight groups are different from each other. And they are all significant.

7.12 (c): From the two table we knows that there is no significant difference between these two test.