Assignment 5 CS 4070

Nirmal Roy, 4724429

January 11, 20178

Question 1

1a

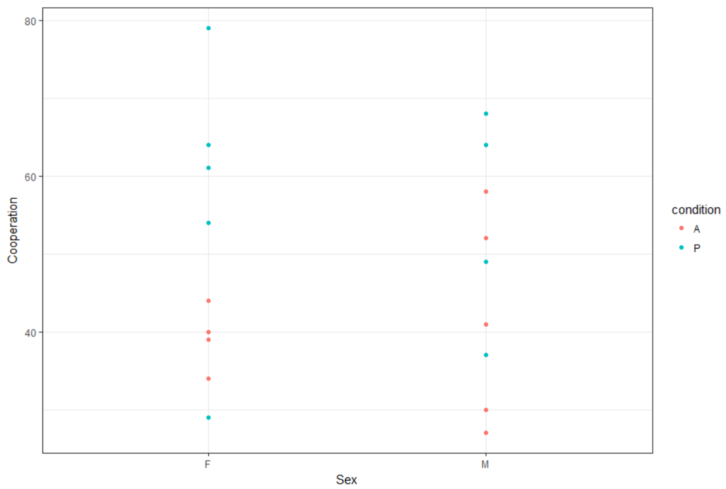

Figure 1: Scatterplot of cooperation versus sex, the type of condition is shown in different colors as mentioned in the legend

1b

Table 1 shows the mean and standard deviation of cooperation for each interaction of the factors.

Table 1

		Sex	
Condition		**Male**	**Female**
Anonymous	Mean	41.6	40.2
	SD	13.46	4.15
Public Choice	Mean	54	57.4
	SD	12.39	18.31

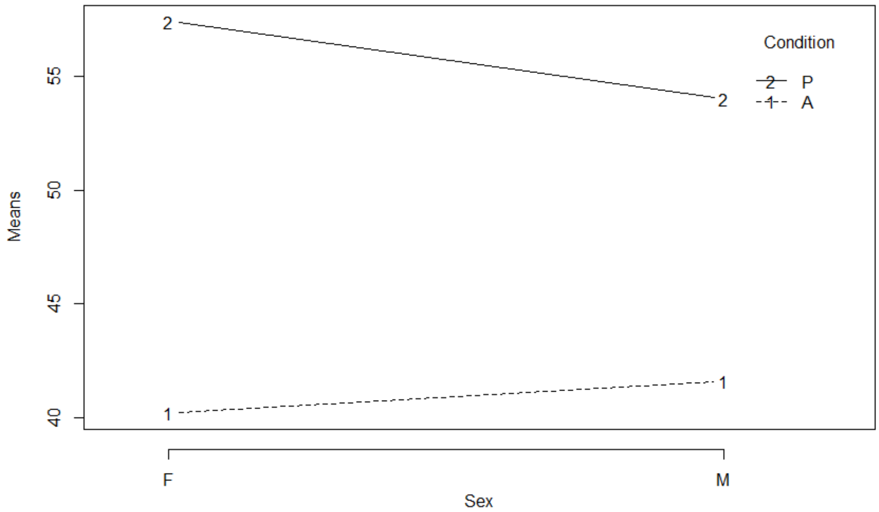

Figure 2: Plot showing interaction in the two way classifier, the profiles are not parallel but the order of means don't change either

Fig 2 shows that when condition is Public choice women tend to have higher number of co operative choices than men. The experiment as expected has a higher cooperation for Public choice condition. Interestingly, for women when it's anonymous, the number of cooperative choices is less than that of men. Also the difference in standard deviation of cooperation based on the two conditions was way higher for women than men.

1c

Let us consider the two way ANOVA model

$$Y_{ijk} = \mu + \alpha_j + \beta_k + \gamma_{jk} + \epsilon_{ijk} \tag{1}$$

where α_j and β_k are main-effect parameters for Conditions(R_1) and Sex(C_1) respectively and γ_{jk} are interaction parameters for the factors. Now using the usual sigma constraints, specifying that each set of parameters sums to 0 over each of its coordinates we can form the deviation regressors as shown in Table 2

Table 2

Combinations	(α_1) R_1	(β_1) C_1	(γ_{11}) R_1C_1
Anonymous-Male	$+1$	$+1$	$+1$
Anonymous-Female	$+1$	-1	-1
PublicChoice-Male	-1	$+1$	-1
PublicChoice-Female	-1	-1	$+1$

The correlation among the regressors are zero since for every point they are assigned 1 or -1 and there are equal number of points(for each factor and interaction), consequently there are equal number of 1s and -1s. It can be proved using the following code snippet in fig 3

```
> Guyer$dregCon <-ifelse(Guyer$condition == "P", 1, -1)
> Guyer$dregSex <-ifelse(Guyer$sex == "M", 1, -1)
> Guyer$dregInt <-Guyer$dregSex*Guyer$dregCon
> cor(Guyer$dregCon, Guyer$dregSex)
[1] 0
> cor(Guyer$dregCon, Guyer$dregInt)
[1] 0
> cor(Guyer$dregSex, Guyer$dregInt)
[1] 0
```

Figure 3: Correlation among these regressors are all zero

1d

i

The default anova function in R does provides sequential sum of squares (type I) sum of square. Which means in fig 4, from the first function we get SS(α) = 1095.2, SS($\beta|\alpha$) = 5.0 and SS($\gamma|\beta, \alpha$) = 28.8. Similarly from the second function we get, SS(β) = 5, SS($\alpha|\beta$) = 1095.2 and SS($\gamma|\alpha, \beta$) = 28.8. Since the values are same we can say that the dataset is balanced.

Hence, SS($\alpha|\beta$) = 1095.2, SS($\beta|\alpha$) = 5.0 and SS($\gamma|\alpha, \beta$) = 28.8 is obtained.

```
> anova(lm(cooperation ~condition*sex, Guyer))
Analysis of Variance Table

Response: cooperation
              Df Sum Sq Mean Sq F value  Pr(>F)
condition      1 1095.2 1095.20  6.3739 0.02253 *
sex            1    5.0    5.00  0.0291 0.86669
condition:sex  1   28.8   28.80  0.1676 0.68767
Residuals     16 2749.2  171.83
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(lm(cooperation ~sex*condition, Guyer))
Analysis of Variance Table

Response: cooperation
              Df Sum Sq Mean Sq F value  Pr(>F)
sex            1    5.0    5.00  0.0291 0.86669
condition      1 1095.2 1095.20  6.3739 0.02253 *
sex:condition  1   28.8   28.80  0.1676 0.68767
Residuals     16 2749.2  171.83
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

Figure 4

ii

The Anova function in the R library car, gives us the opportunity to do a type III test from which we can get the values SS($\alpha|\beta, \gamma$) = 1095 , SS($\beta|\alpha, \gamma$) = 5 and SS($\gamma|\alpha, \beta$) = 29 as can be seen from fig 5

```
> Anova(lm(cooperation ~ dregCon + dregSex + dregInt, data=Guyer, contrasts=list(topic=contr.sum, sys=contr.sum)), type=3)
Anova Table (Type III tests)

Response: cooperation
            Sum Sq Df  F value    Pr(>F)
(Intercept)  46658  1 271.5426 1.851e-11 ***
dregCon       1095  1   6.3739   0.02253 *
dregSex          5  1   0.0291   0.86669
dregInt         29  1   0.1676   0.68767
Residuals     2749 16
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 5

iii

For this, separate linear models for condition, sex and condition*sex was fit to obtain SS(α) = 1095.2, SS(β) = 5 and SS(γ) = 28.8

iv

For the Guyer Data , Sums of squares for various models fit to the data are as follows:
SS(α, β, γ) = 1129
SS(α, β) = 1100.2
SS(α, γ) = 1129
SS(β, γ) = 1129
SS(α) = 1095.2
SS(β) = 5

Using the special formulas given in Section 8.2.5 in Fox. we get
SS($\gamma|\alpha, \beta$) = SS(α, β, γ) - SS(α, β) =(1095.2 +=28.8
SS($\alpha|\beta, \gamma$) = SS(α, β, γ) - SS(β, γ) =
SS($\beta|\alpha, \gamma$) = SS(α, β, γ) - SS(α, γ) =
SS($\alpha|\beta$) = SS(α, β) - SS(β) = 1095.2
SS($\beta|\alpha$) = SS(α, β) - SS(α) = 5

1e

As can be seen from table 3 the null hypothesis that condition has no effect on Cooperation doesn't hold owing to very low p value and hence it is statistically significant. Whereas, both sex and the interaction of sex and condition has very high value and their respective null hypothesis that they don't have any effect on condition holds. In this instance, where the interactions are negligible, just SS($\alpha|\beta$) and SS($\beta|\alpha$) without taking interaction into consideration, will test hypothesis about the main effects. This is the type 2 approach and statistically more powerful for this event. With this approach, we find that p-value for condition is 0.019 and that for sex is 0.863.Hence, we can conclude that only ‘condition’ is significant for this data.

Table 3: ANOVA table

Source	*SS*	*df*	*MS*	*F*	*p*
Condition	1095.2	1	1095.2	6.37	0.02253
Sex	5	1	5	0.0291	0.87
Condition * Sex	28.8	1	28.8	0.1676	0.687
Residuals	3878.2 - 1129 = 2749.2	16	171.83		
Total	3878.2	19			

Question 2

2a

Table 4 shows the mean and standard deviation of scores and also the number of observations in cell for each interaction of the factors of the Adler data.

Table 4

		expectation	
instruction		**HIGH**	**LOW**
GOOD	Mean	4.067	- 18.059
	SD	16.65	10.57
	Frequency	15	17
SCIENTIFIC	Mean	-6.94	1.92
	SD	8.45	11.71
	Frequency	18	13
NONE	Mean	-10	-3.5
	SD	15.64	11.63
	Frequency	16	18

From fig 6 we can see that photographs shown by assistants who were told to expect good ratings and asked to collect good data got highest ratings on their photographs followed by assistants who collected scientific data but were told to expect low ratings. On an average, researchers expecting low ratings received higher ratings maybe because they put in that extra effort to find data. Whereas researchers expecting good data probably got complacent when they were not told any method of collecting data.

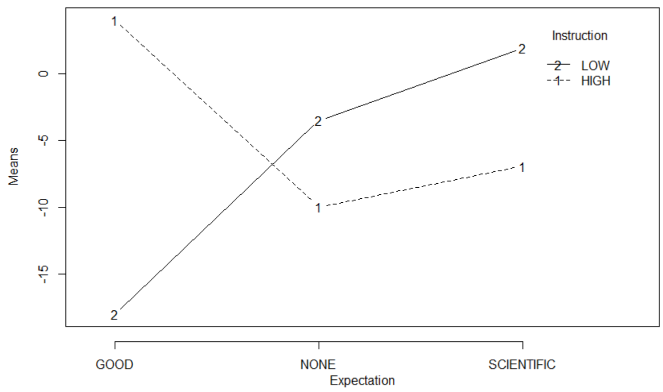

Figure 6: Plot showing interaction in the two way classifier, the profiles are not parallel and the order of means change too, hence it's "disordinal"

2b

The null hypothesis of the main-effects holds due to the high p-value, while the interaction is statistically significant and the null hypothesis that the interaction is not important fails because of the very low p-value(table 5). The factors together explains the rating while individually the fail to do so. In this instance, where the interactions probably are not negligible, SS($\alpha|\beta$) and SS($\beta|\alpha$) do not test hypothesis about the main effects. The type 3 approach, that is using SS($\alpha|\beta,\gamma$) and SS($\beta|\alpha,\gamma$), is interesting and can be used to test the hypothesis about main-effects. None of the main effects are statistically significant and only the interaction describes the data properly.

Table 5: ANOVA table

Source	*SS*	*df*	*MS*	*F*	*p*
Instruction		2			
$\alpha|\beta,\gamma$	338.6		169.3	1.0627	0.3498
Expectation		1			
$\beta|\alpha,\gamma$	123		123	0.7721	0.3819
Instruction * Expectation	4729.4	2	2364.7	14.844	2.63*e-06
Residuals	19687.26-5191=14496.2	91	159.3		
Total	19687.26.2	96			

Appendix

```r
ggplot(Guyer, aes (x = sex, y = cooperation, color= condition))+geom_point()+
  labs(x = "Sex", y = "Cooperation") + theme_bw()


 m11 <- mean(subset(Guyer, condition == "A"& sex== "M")$cooperation)
 m12 <- mean(subset(Guyer, condition == "A" & sex== "F")$cooperation)
 m21 <- mean(subset(Guyer, condition == "P"& sex== "M")$cooperation)
 m22 <- mean(subset(Guyer, condition == "P"& sex== "F")$cooperation)

 std11 <- sd(subset(Guyer, condition == "A"& sex== "M")$cooperation)
 std12 <- sd(subset(Guyer, condition == "A"& sex== "F")$cooperation)
 std21 <- sd(subset(Guyer, condition == "P"& sex== "M")$cooperation)
 std22 <- sd(subset(Guyer, condition == "P"& sex== "F")$cooperation)

 interaction.plot(Guyer$sex, Guyer$condition, Guyer$cooperation, fun = mean,
                  trace.label= "Condition", xlab= "Sex",
                  ylab= "Means",  type = c("b"), xtick = TRUE)


 Guyer$dregCon <-ifelse(Guyer$condition == "P", 1, -1)
 Guyer$dregSex <-ifelse(Guyer$sex == "M", 1, -1)
 Guyer$dregInt <-Guyer$dregSex*Guyer$dregCon



 m11 <- mean(subset(Adler, expectation == "HIGH"& instruction== "GOOD")$rating)
 m21 <- mean(subset(Adler, expectation == "HIGH"& instruction== "SCIENTIFIC")$rating)
 m31 <- mean(subset(Adler, expectation == "HIGH"& instruction== "NONE")$rating)
 m12 <- mean(subset(Adler, expectation == "LOW"& instruction== "GOOD")$rating)
 m22 <- mean(subset(Adler, expectation == "LOW"& instruction== "SCIENTIFIC")$rating)
 m32 <- mean(subset(Adler, expectation == "LOW"& instruction== "NONE")$rating)


 s11 <- sd(subset(Adler, expectation == "HIGH"& instruction== "GOOD")$rating)
 s21 <- sd(subset(Adler, expectation == "HIGH"& instruction== "SCIENTIFIC")$rating)
 s31 <- sd(subset(Adler, expectation == "HIGH"& instruction== "NONE")$rating)
 s12 <- sd(subset(Adler, expectation == "LOW"& instruction== "GOOD")$rating)
 s22 <- sd(subset(Adler, expectation == "LOW"& instruction== "SCIENTIFIC")$rating)
 s32 <- sd(subset(Adler, expectation == "LOW"& instruction== "NONE")$rating)

 n11 <- nrow(subset(Adler, expectation == "HIGH"& instruction== "GOOD"))
 n21 <- nrow(subset(Adler, expectation == "HIGH"& instruction== "SCIENTIFIC"))
 n31 <- nrow(subset(Adler, expectation == "HIGH"& instruction== "NONE"))
 n12 <- nrow(subset(Adler, expectation == "LOW"& instruction== "GOOD"))
 n22 <- nrow(subset(Adler, expectation == "LOW"& instruction== "SCIENTIFIC"))
 n32 <- nrow(subset(Adler, expectation == "LOW"& instruction== "NONE"))
```

Figure 7: Relevant Codes