STAT 456 Homework 5

Hailey Lee

03/24/2024

Instructions:

1. For Q1 below, you need to carry out a one-way ANOVA to test the hypothesis that the treatments will have different effects. [You have to do it by hand, however, you can use R to run the analysis to help you check your ANOVA table results.]
2. You are liable for missing points if you don't include output;
3. Whenever possible, please run saved variables, so our TA knows if your code goes the right way and assigns partial credits even if your final answer is wrong.
4. For the parts manually calculated by hand, you can either type it in R markdown or write your solutions on paper and scan it to pdf and submit it. Please submit your solutions using a zip file, and name your files as Lastname-456-Hw5.zip and Lastname-456-Hw5.zip.

1. In one research study, 20 young pigs are assigned at random among 4 experimental groups. Each group is fed a different diet. (This design is a completely randomized design.) The data are saved in `diet.csv`, which contains the pigs' weights in kg after being raised on these diets for 10 months. We wish to ask whether mean pig weights are the same for all 4 diets.

Data from study of pigs weights

Feed 1	Feed 2	Feed 3	Feed 4
60.8	68.3	102.6	87.9
57.1	67.7	102.2	84.7
65.0	74.0	100.5	83.2
58.7	66.3	97.5	85.8
61.8	69.9	98.9	90.3

Carry out a one-way ANOVA **by hand** to test the hypothesis that the treatments will have different effects. [You have to do it **by hand**, however, you can use R to run the analysis to help you check your ANOVA table results.]

```r
pigs <- read.csv("diet.csv")
pigs$Feed <- as.factor(pigs$Feed)

feed1 <- dplyr::filter(pigs, Feed == 1)
feed2 <- dplyr::filter(pigs, Feed == 2)
feed3 <- dplyr::filter(pigs, Feed == 3)
feed4 <- dplyr::filter(pigs, Feed == 4)

y1 <- mean(feed1$Weight)
y2 <- mean(feed2$Weight)
y3 <- mean(feed3$Weight)
y4 <- mean(feed4$Weight)
y <- mean(c(y1, y2, y3, y4))

print(paste("Feed 1 mean -> ", y1))
```

```
## [1] "Feed 1 mean ->  60.68"
```

```r
print(paste("Feed 2 mean -> ", y2))
```

```
## [1] "Feed 2 mean ->  69.24"
```

```r
print(paste("Feed 3 mean -> ", y3))
```

```
## [1] "Feed 3 mean ->  100.34"
```

```r
print(paste("Feed 4 mean -> ", y4))
```

```
## [1] "Feed 4 mean ->  86.38"
```

```r
print(paste("Overall mean -> ", y))
```

```
## [1] "Overall mean ->  79.16"
```

```r
SST <- 5*(y1-y)^2 + 5*(y2-y)^2 + 5*(y3-y)^2 + 5*(y4-y)^2
dfT <- 4 - 1
MST <- SST / dfT

print(paste("SST -> ", SST))
```

```
## [1] "SST ->  4703.188"
```

```r
print(paste("dfT -> ", dfT))
```

```
## [1] "dfT ->  3"
```

```r
print(paste("MST -> ", MST))
```

```
## [1] "MST ->  1567.72933333333"
```

```r
SSE <- sum((feed1$Weight-y1)^2) + sum((feed2$Weight-y2)^2) + sum((feed3$Weight-y3)^2) +
dfE <- 4 * (5 - 1)
MSE <- SSE / dfE

print(paste("SSE -> ", SSE))
```

```
## [1] "SSE ->  121.34"
```

```r
print(paste("dfE -> ", dfE))
```

```
## [1] "dfE ->  16"
```

```r
print(paste("MSE -> ", MSE))
```

```
## [1] "MSE ->  7.58375"
```

```r
f <- MST / MSE

print(paste("f -> ", f))
```

```
## [1] "f ->  206.722180099995"
```

```r
#pigs.aov <- aov(Weight ~ Feed, data = pigs)
#summary(pigs.aov)
```

2.**Weight gain and junk food.** An experiment randomly assigned lab mice to one of three groups. The data are saved in `deermice.csv`. Group 1 received a standard diet, group 2 received a diet of junk food, and group 3 received a diet of health food. Weight gains (grams) after 5 weeks were:

```
Group 1 (standard diet):   11.8   12.0   10.7    9.1   12.1
```

```
Group 2 (junk food):        13.6   14.4   12.8   13.0   13.4
Group 3 (health food):       9.2    9.6    8.6    8.5    9.8
```

(a) Conduct an ANOVA and report the ANOVA table. Is there any evidence that the population means differ for at least two of the three groups?

```r
mice <- read.csv("deermice.csv")
one.way <- aov(WTGAIN ~ GROUP, data = mice)
summary(one.way)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)    
## GROUP        2  46.30   23.15   29.69 2.26e-05 ***
## Residuals   12   9.36    0.78                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since p-value (2.26e-05) is significantly small (much closer to zero), we can reject the null hypothesis. There is enough evidence that the population means differ for at least two of the three groups.

(b) Calculate the means and standard deviations for each group. Conduct a pairwise comparison for the group means without any adjustment, and report the p-value for each test.

```r
standard <- dplyr::filter(mice, GROUP == "standard")
mean_standard <- mean(standard$WTGAIN)
sd_standard <- sd(standard$WTGAIN)

print(paste("Mean for Group 1 -> ", mean_standard))
```

```
## [1] "Mean for Group 1 ->  11.14"
```

```r
print(paste("Standard Deviation for Group 1 -> ", sd_standard))
```

```
## [1] "Standard Deviation for Group 1 ->  1.27003936946852"
```

```r
junk <- dplyr::filter(mice, GROUP == "junk")
mean_junk <- mean(junk$WTGAIN)
sd_junk <- sd(junk$WTGAIN)
print(paste("Mean for Group 2 -> ", mean_junk))
```

```
## [1] "Mean for Group 2 ->  13.44"
```

```r
print(paste("Standard Deviation for Group 2 -> ", sd_junk))
```

```
## [1] "Standard Deviation for Group 2 ->  0.622896460095897"
```

```r
health <- dplyr::filter(mice, GROUP == "health")
mean_health <- mean(health$WTGAIN)
sd_health <- sd(health$WTGAIN)
```

```r
print(paste("Mean for Group 3 -> ", mean_health))
```

```
## [1] "Mean for Group 3 ->  9.14"
```

```r
print(paste("Standard Deviation for Group 3 -> ", sd_health))
```

```
## [1] "Standard Deviation for Group 3 ->  0.581377674149946"
```

```r
pvals <- pairwise.t.test(mice$WTGAIN,
                         mice$GROUP,
                         p.adjust.method = "none"
)$p.value
pvals
```

```
##                health        junk
## junk     5.549123e-06          NA
## standard 3.772975e-03 0.001424363
```

The p-value for junk and health is 5.549123e-06. The p-value for standard and health is 3.772975e-03. The p-value for standard and junk is 0.001424363.

(c) Incorporate Bonferroni's adjustment into each of the post hoc comparisons. Report the p-value based on the Bonferroni's adjustment.

```r
pvals <- pairwise.t.test(mice$WTGAIN,
                         mice$GROUP,
                         p.adjust.method = "bonferroni"
)$p.value
pvals
```

```
##                health        junk
## junk     1.664737e-05          NA
## standard 1.131893e-02 0.004273088
```

The p-value for junk and health is 1.664737e-05. The p-value for standard and health is 1.131893e-02. The p-value for standard and junk is 0.004273088.

(d) Consider the TukeyHSD (Tukey Honest Significant Differences) post hoc test. Report the p-value based on the TukeyHSD post-hoc test.

```r
one.way2 <- aov(WTGAIN ~ GROUP, data = mice)
TukeyHSD(one.way2)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = WTGAIN ~ GROUP, data = mice)
##
## $GROUP
```

```
##                    diff        lwr        upr     p adj
## junk-health      4.3  2.8101309  5.7898691 0.0000154
## standard-health  2.0  0.5101309  3.4898691 0.0097641
## standard-junk   -2.3 -3.7898691 -0.8101309 0.0037612
```

The p-value for junk and health is 0.0000154. The p-value for standard and health is 0.0097641. The p-value for standard and junk is 0.0037612.

3. Is there an interaction? Each of the following tables gives means for a two-way ANOVA. Make a plot of the means with the levels of Factor A on the x axis. State whether there is an interaction, and if there is, describe it.

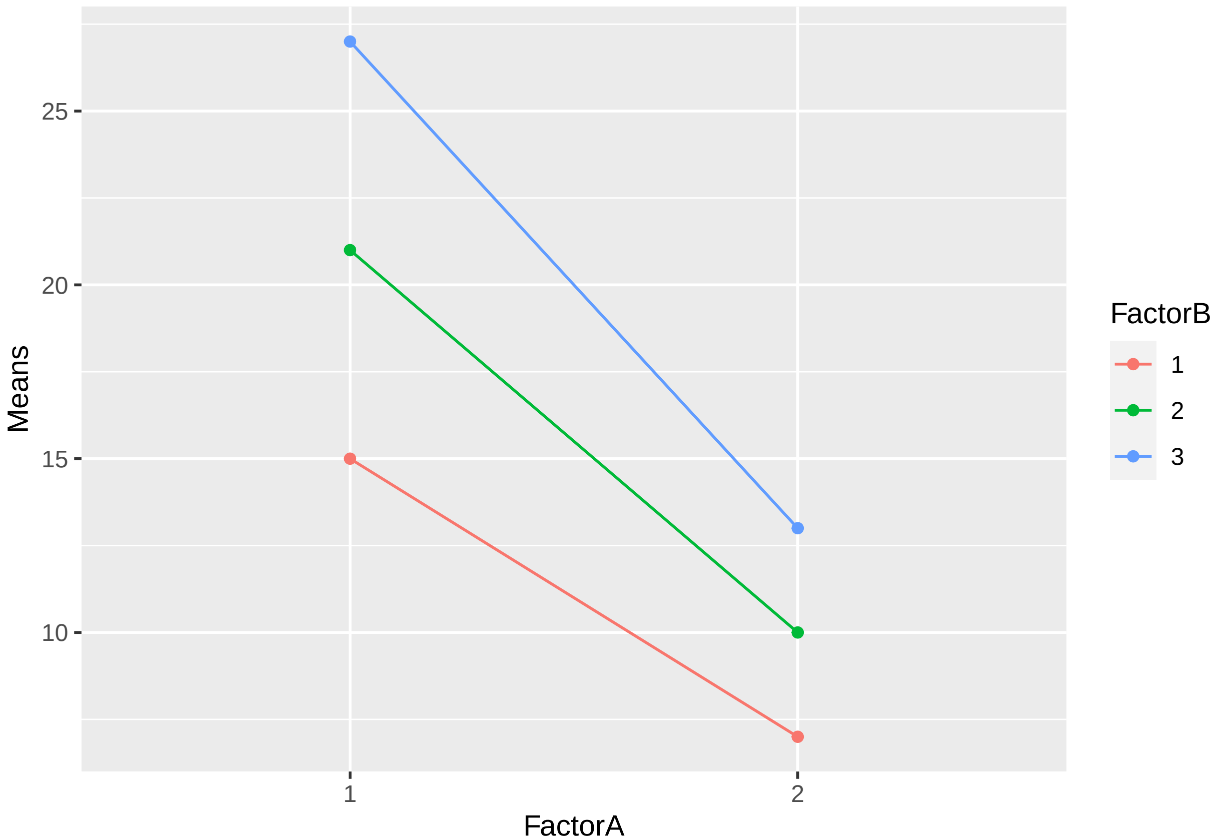
(a)

```
           Factor B
Factor A    1   2   3
       1   15  21  27
       2    7  10  13
```

```r
library(ggplot2)

q3a <- data.frame(
  FactorA = c("1", "2", "1", "2", "1", "2"),
  FactorB = c("1", "1", "2", "2", "3", "3"),
  Means = c(15, 7, 21, 10, 27, 13)
)

ggplot(q3a, aes(x = FactorA, y = Means, color = FactorB, group=FactorB)) +
  geom_point() +
  geom_line()
```

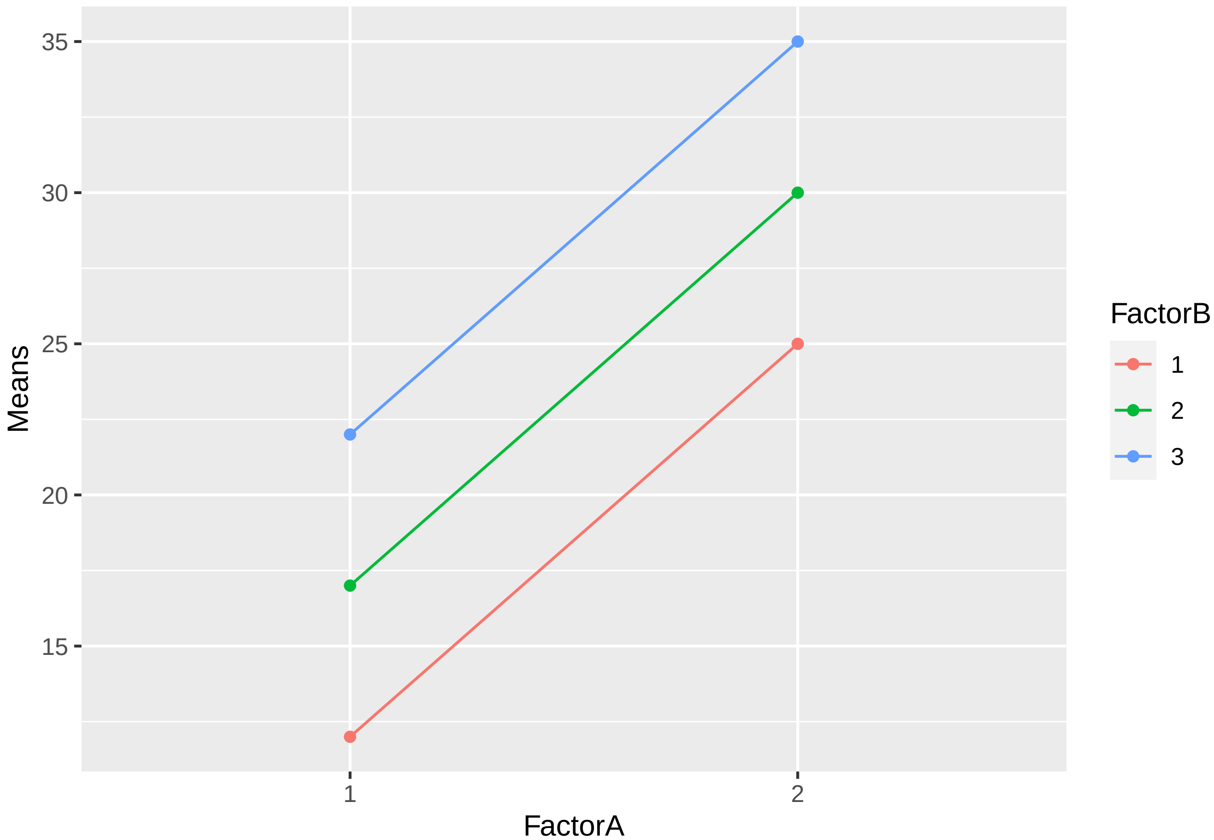


There is an interaction between Factor A and B: the effect of Factor B appears to be greater for the first level of Factor A.

(b)
```
        Factor B
Factor A    1    2    3
        1  12   17   22
        2  25   30   35
```

```
q3b <- data.frame(
  FactorA = c("1", "2", "1", "2", "1", "2"),
  FactorB = c("1", "1", "2", "2", "3", "3"),
  Means = c(12, 25, 17, 30, 22, 35)
)

ggplot(q3b, aes(x = FactorA, y = Means, color = FactorB, group=FactorB)) +
  geom_point() +
  geom_line()
```



There is no interaction between Factor A and B.

4. Based on the following partial ANOVA table results, complete the ANOVA table and draw conclusions at the 5% significance level.

Effect	df	SS	MS	F
A	2	90		
B	4	165		
AB	8	204		
Error	60	900		

```r
MSA <- 90 / 2
MSB <- 165 / 4
MSAB <- 204 / 8
MSE <- 900 / 60

print(paste("MSA -> ", MSA))
```

```
## [1] "MSA ->  45"
```

```r
print(paste("MSB -> ", MSB))
```

```
## [1] "MSB ->  41.25"
```

```r
print(paste("MSAB -> ", MSAB))
```

```
## [1] "MSAB ->  25.5"
```

```r
print(paste("MSE -> ", MSE))
```

```
## [1] "MSE ->  15"
```

```r
fA <- MSA / MSE
fB <- MSB / MSE
fAB <- MSAB / MSE

print(paste("fA -> ", fA))
```

```
## [1] "fA ->  3"
```

```r
print(paste("fB -> ", fB))
```

```
## [1] "fB ->  2.75"
```

```r
print(paste("fAB -> ", fAB))
```

```
## [1] "fAB ->  1.7"
```

5. **Discounts and Expected Prices.** Does the frequency with which a supermarket product is offered at a discount affect the price that customers expect to pay for the product? Does the percent reduction also affect this expectation? These questions were examined by researchers in a study conducted on students enrolled in an introductory management course at a large midwestern university. For 10 weeks, 160 subjects received information about the products. The treatment conditions corresponded to the number of promotions (one, three) during this 10-week period, and the percent that the product was discounted (20% or 40%). Ten students were randomly assigned to each of the $2 \times 2 = 4$ treatments. Here are the data (save in `supermarket.csv`):

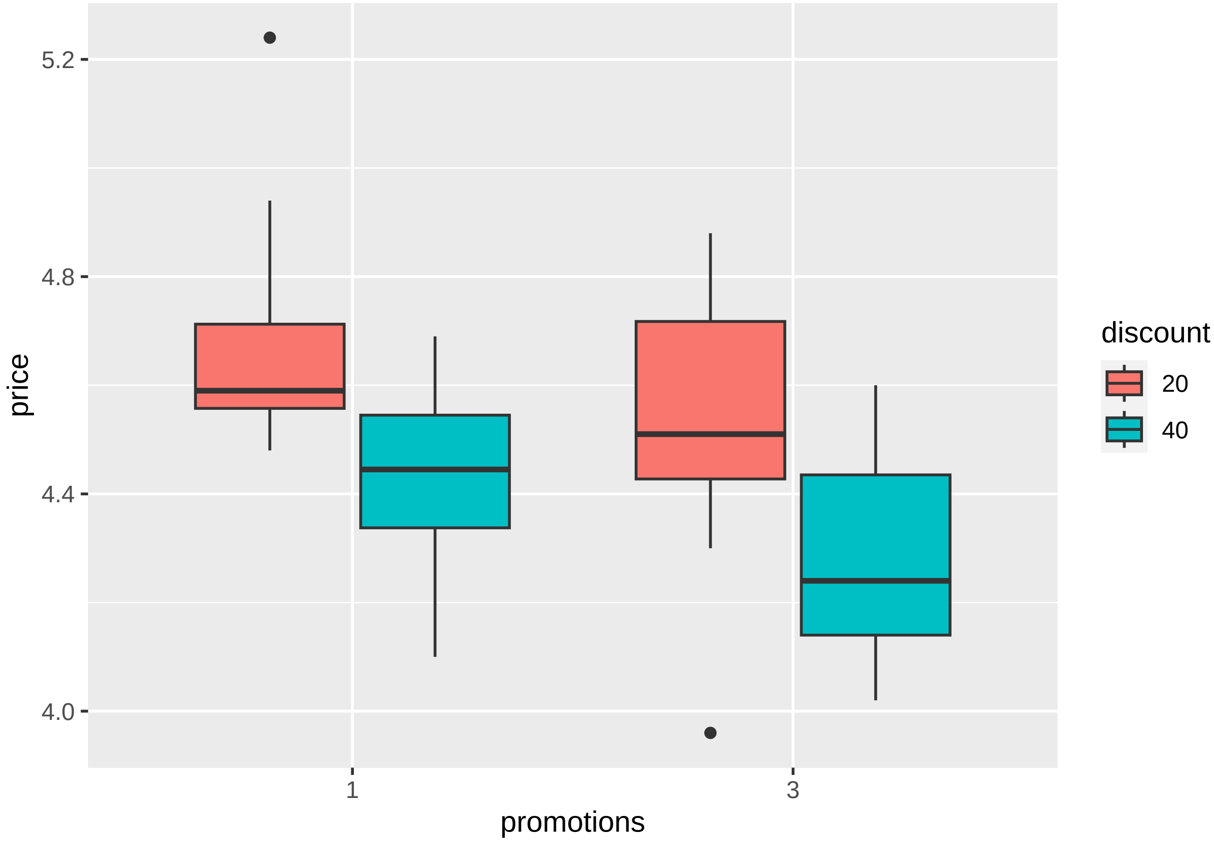
Number of promotions	discount	Expected price									
1	40	4.10	4.50	4.47	4.42	4.56	4.69	4.42	4.17	4.31	4.59
1	20	4.94	4.59	4.58	4.48	4.55	4.53	4.59	4.66	4.73	5.24
3	40	4.07	4.13	4.25	4.23	4.57	4.33	4.17	4.47	4.60	4.02
3	20	4.88	4.80	4.46	4.73	3.96	4.42	4.30	4.68	4.45	4.56

(a) Create boxplots for each combination of the number of promotions and discount.

```r
supermarket <- read.csv("supermarket.csv")
supermarket$promotions <- as.factor(supermarket$promotions)
supermarket$discount <- as.factor(supermarket$discount)

ggplot(supermarket, aes(x = promotions, y = price, fill = discount)) +
```

```r
  geom_boxplot()
```



(b) Obtain a two-way ANOVA table with main factors "promotions" and "discount" and draw conclusions at the 5% significance level.

```r
two.way <- aov(price ~ promotions + discount, data = supermarket)
summary(two.way)
```

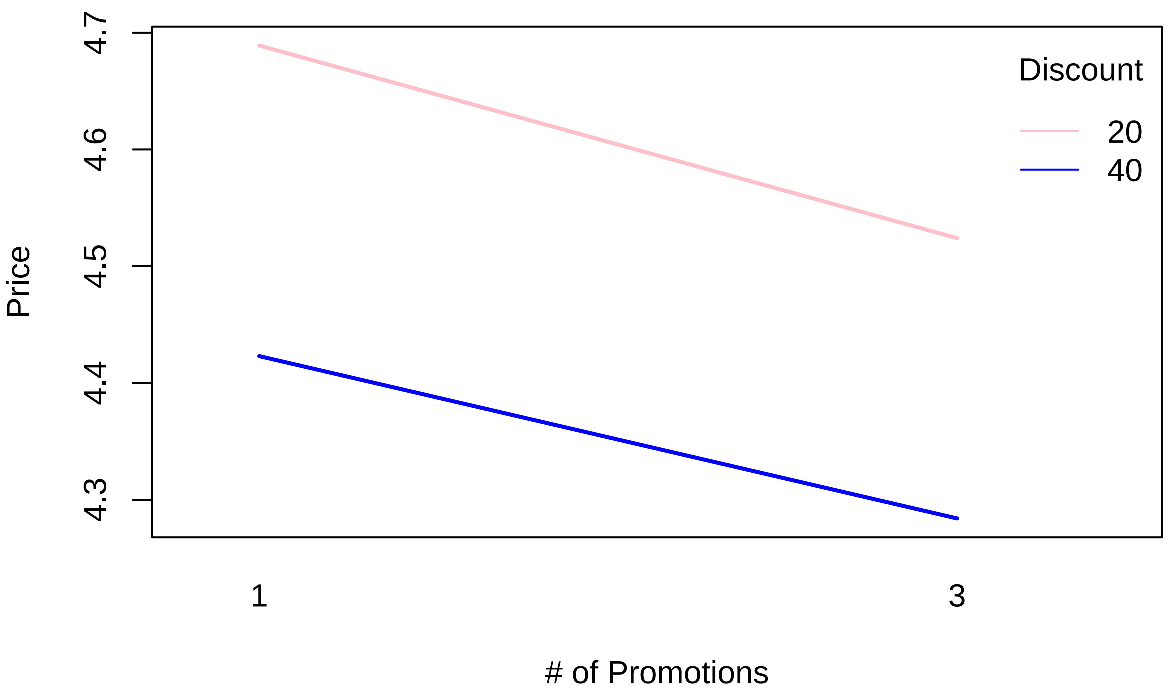
```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## promotions   1 0.2310  0.2310   4.666 0.03731 *
## discount     1 0.6401  0.6401  12.927 0.00094 ***
## Residuals   37 1.8321  0.0495
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-values are less than the significance level, we reject the null hypothesis. We can conclude that both promotions and discount are statistically significant.

(c) Construct an appropriate interaction plot, and describe the patterns that you see.

```r
interaction.plot(x.factor = supermarket$promotions, #x-axis variable
                 trace.factor = supermarket$discount, #variable for lines
                 response = supermarket$price, #y-axis variable
                 fun = mean, #metric to plot
```

```r
                 ylab = "Price",
                 xlab = "# of Promotions",
                 col = c("pink", "blue"),
                 lty = 1, #line type
                 lwd = 2, #line width
                 trace.label = "Discount")
```



As the number of promotions increases, for both discount level, the price that customer expected to pay decreases at similar rate.

(d) Does the plot suggest that there is an interaction between the number of promotions and discount?

The effect of Discount on the number of promotions appear to be consistent.

(e) Obtain the ANOVA table with the number of promotions, discount and their interaction effect.

```r
two.way2 <- aov(price ~ promotions * discount, data = supermarket)
summary(two.way2)
```

```
##                     Df Sum Sq Mean Sq F value Pr(>F)
## promotions           1 0.2310  0.2310   4.544 0.0399 *
## discount             1 0.6401  0.6401  12.589 0.0011 **
## promotions:discount  1 0.0017  0.0017   0.033 0.8564
## Residuals           36 1.8304  0.0508
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(f) Are there any significant main effects or an interaction effect from (e) part?

Since the p-values of promotions and discount are greater than the significance level, we fail to reject the null hypothesis on main effects. There are not enough evidence to conclude that the two main effects are significant. The p-value of the interaction effect is also significantly high (close to 1). Therefore, there are not enough evidence for significant interaction effect.