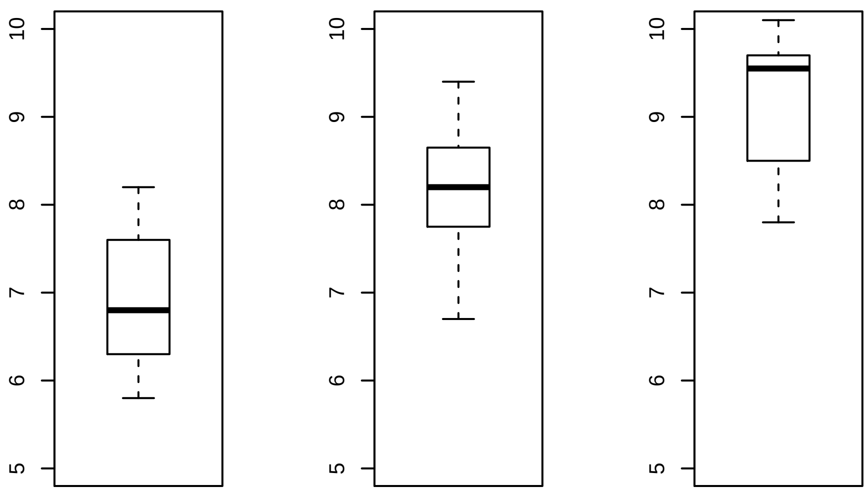
Michael Leibert
Math 651
Homework 9

16.7. Productivity improvement. An economist compiled data on productivity improvements last year for a sample of firms producing electronic computing equipment. The firms were classified according to the level of their average expenditures for research and development in the past three years (low, moderate, high). The results of the study follow (productivity improvement is measured on a scale from 0 to 100). Assume that ANOVA model (16.2) is appropriate.

a. Prepare aligned dot plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

```
par(mar=c(2,3,2,3),mfrow=c(1,3))
boxplot(prod[which(prod[2] == 1),1],ylim=c(5,10))
boxplot(prod[which(prod[2] == 2),1],ylim=c(5,10))
boxplot(prod[which(prod[2] == 3),1],ylim=c(5,10))
```



They appear to have similar levels of variability, but the the factor level means do appear to differ.

b. Obtain the fitted values.

```
 prod.r<-max(as.numeric(prod[,2]))
prod.n<-nrow(prod)

Y<-as.matrix(prod[,1])
X<-matrix(0,nrow(prod),3)
B<-matrix(NA,3,1)
E<-matrix(NA,nrow(prod),1)
Ybar<-rep(NA,nrow(prod))

for (i in 1:3){
        Ybar[which(prod[,2] == i)]<-mean(prod[which(prod[,2] == i),1 ])
        X[which(prod[,2] == i),i]<-     prod[which(prod[,2] == i),2]/
                prod[which(prod[,2] == i),2]
        B[i,1]<-mean(prod[which(prod[,2] == i),1])
```

```
        E[which(prod[,2] == i),1]<-prod[which(prod[,2] == i),1]-B[i,1]}
Ybar

##  [1] 6.877778 6.877778 6.877778 6.877778 6.877778 6.877778 6.877778
##  [8] 6.877778 6.877778 8.133333 8.133333 8.133333 8.133333 8.133333
## [15] 8.133333 8.133333 8.133333 8.133333 8.133333 8.133333 8.133333
## [22] 9.200000 9.200000 9.200000 9.200000 9.200000 9.200000
```

c. Obtain the residuals. Do they sum to zero in accord with (16.21)?

```
  as.vector(E)

##  [1]  0.72222222  1.32222222 -0.07777778 -1.07777778  0.02222222
##  [6] -0.27777778 -0.57777778  0.82222222 -0.87777778 -1.43333333
## [11] -0.03333333  1.26666667  0.46666667 -0.33333333 -0.43333333
## [16]  0.76666667 -0.23333333  0.16666667  0.56666667 -1.03333333
## [21]  0.26666667 -0.70000000  0.50000000  0.90000000 -1.40000000
## [26]  0.40000000  0.30000000

 options(scipen=999)
 sum(E)

## [1] 0.0000000000001154632
```

They do sum to zero.

d. Obtain the analysis of variance table.

```
prod$X1<-as.factor(prod$X1)
anova(lm(prod$Y~prod$X1))

## Analysis of Variance Table
##
## Response: prod$Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## prod$X1    2 20.125 10.0626   15.72 4.331e-05 ***
## Residuals 24 15.362  0.6401
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

e. Test whether or not the mean productivity improvement differs according to the level of research and development expenditures. Control the a risk at .05. State the alternatives, decision rule, and conclusion.

H_0: all μ_i are equal $(i = 1, 2, 3)$, H_a: not all μ_i are equal.

```
 qf(1-.05,prod.r-1,prod.n-prod.r)

## [1] 3.402826

summary(aov(prod$Y~prod$X1))[[1]][1,4]

## [1] 15.72053
```

$F^* = 15.72053$; $F(.95; 2, 24) = 3.402826$. Fail to reject H_0 if $F^* \leq 3.402826$, otherwise reject H_0. Reject H_0.

f. What is the P-value of the test in part (e)? How does it support the conclusion reached in part (e)?

```
1-pf(summary(aov(prod$Y~prod$X1))[[1]][1,4],prod.r-1,prod.n-prod.r)
## [1] 0.00004330692
```

It is much lower than $\alpha = .05$ level in part (e).

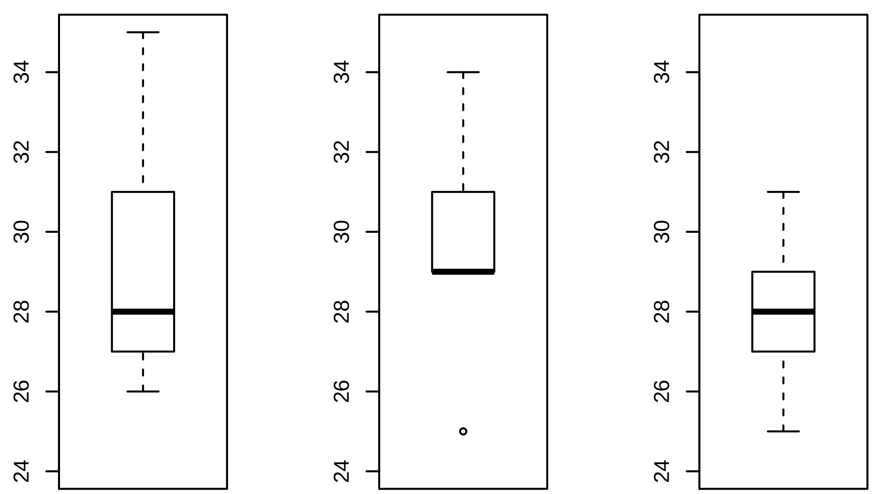
g. What appears to be the nature of the relationship between research and development expenditures and productivity improvement?

It looks like increases in research and development expenditures increase productivity.

16.8. Questionnaire color. In an experiment to investigate the effect of color of paper (blue, green, orange) on response rates for questionnaires distributed by the "windshield method" in supermarket parking lots, 15 representative supermarket parking lots were chosen in a metropolitan area and each color was assigned at random to five of the lots. The response rates (in percent) follow. Assume that ANOVA model (16.2) is appropriate.

a. Prepare aligned dot plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

```
par(mar=c(2,3,2,3),mfrow=c(1,3))
boxplot(QC[which(QC[2] == 1),1],ylim=c(24,35))
boxplot(QC[which(QC[2] == 2),1],ylim=c(24,35))
boxplot(QC[which(QC[2] == 3),1],ylim=c(24,35))
```



The factor means appear to be similar, however the variability among the levels differs considerably.

b. Obtain the fitted values.

```
Y<-as.matrix(QC[,1])
X<-matrix(0,nrow(QC),3)
B<-matrix(NA,3,1)
E<-matrix(NA,nrow(QC),1)
Ybar<-rep(NA,nrow(QC))
```

```
for (i in 1:3){
        Ybar[which(QC[,2] == i)]<-mean(QC[which(QC[,2] == i),1 ])
        X[which(QC[,2] == i),i]<-       QC[which(QC[,2] == i),2]/
                QC[which(QC[,2] == i),2]
        B[i,1]<-mean(QC[which(QC[,2] == i),1])
        E[which(QC[,2] == i),1]<-QC[which(QC[,2] == i),1]-B[i,1]
}
Ybar
##  [1] 29.4 29.4 29.4 29.4 29.4 29.6 29.6 29.6 29.6 29.6 28.0 28.0 28.0 28.0
## [15] 28.0
```

c. Obtain the residuals.

```
as.vector(E)
##  [1] -1.4 -3.4  1.6 -2.4  5.6  4.4 -0.6 -4.6  1.4 -0.6  3.0 -3.0 -1.0  1.0
## [15]  0.0
options(scipen=999)
sum(E)
## [1] 0
```

d. Obtain the analysis of variance table.

```
QC$X1<-as.factor(QC$X1)
anova(lm(QC$Y~QC$X1))
## Analysis of Variance Table
##
## Response: QC$Y
##           Df Sum Sq Mean Sq F value Pr(>F)
## QC$X1      2    7.6     3.8  0.3918 0.6842
## Residuals 12  116.4     9.7
```

e. Conduct a test to determine whether or nor the mean response rates for the three colors differ. Use level of significance $\alpha = .10$. State the alternatives. decision rule, and conclusion. What is the P-value of the test?

H_0: all μ_i are equal $(i = 1, 2, 3)$, H_a: not all μ_i are equal.

```
qf(1-.1,QC.r-1,QC.n-QC.r)
## [1] 2.806796
summary(aov(QC$Y~QC$X1))[[1]][1,4]
## [1] 0.3917526
1-pf(summary(aov(QC$Y~QC$X1))[[1]][1,4],QC.r-1,QC.n-QC.r)
## [1] 0.6842074
```

$F^* = 0.3917526$; $F(.90; 2, 24) = 3.402826$. Fail to reject H_0 if $F^* \leq 3.402826$, otherwise reject H_0. Fail to reject H_0. P-value = 0.6842074.

f. When informed of the findings, an executive said: "See? I was right all along. We might as well print the questionnaires on plain white paper, which is cheaper." Does this conclusion follow from the findings of the study? Discuss.

Considering white paper is not included in the study, I am not sure she can draw this conclusion. Further, our analysis only tested whether or not the mean response rates of the three colors differ.

16.18. Refer to Productivity improvement Problem 16.7. Regression model (16.75) is to be employed for testing the equality of the factor level means.

a. Set up the **Y**, **X**, and $\boldsymbol{\beta}$ matrices.

$$\mathbf{Y} = \begin{pmatrix} 7.60 \\ 8.20 \\ 6.80 \\ 5.80 \\ 6.90 \\ 6.60 \\ 6.30 \\ 7.70 \\ 6.00 \\ 6.70 \\ 8.10 \\ 9.40 \\ 8.60 \\ 7.80 \\ 7.70 \\ 8.90 \\ 7.90 \\ 8.30 \\ 8.70 \\ 7.10 \\ 8.40 \\ 8.50 \\ 9.70 \\ 10.10 \\ 7.80 \\ 9.60 \\ 9.50 \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix} \qquad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix}$$

b. Obtain $\mathbf{X}\boldsymbol{\beta}$. Develop equivalent expressions of the elements of this vector in terms of the cell means μ_i.

$$\boldsymbol{X\beta} = \begin{pmatrix} \mu+\tau_1 \\ \mu+\tau_1 \\ \mu+\tau_1 \\ \mu+\tau_1 \\ \mu+\tau_1 \\ \mu+\tau_1 \\ \mu+\tau_1 \\ \mu+\tau_1 \\ \mu+\tau_1 \\ \mu+\tau_2 \\ \mu+\tau_2 \\ \mu+\tau_2 \\ \mu+\tau_2 \\ \mu+\tau_2 \\ \mu+\tau_2 \\ \mu+\tau_2 \\ \mu+\tau_2 \\ \mu+\tau_2 \\ \mu+\tau_2 \\ \mu+\tau_2 \\ \mu+\tau_2 \\ \mu-\tau_1-\tau_2 \\ \mu-\tau_1-\tau_2 \\ \mu-\tau_1-\tau_2 \\ \mu-\tau_1-\tau_2 \\ \mu-\tau_1-\tau_2 \\ \mu-\tau_1-\tau_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_3 \\ \mu_3 \\ \mu_3 \\ \mu_3 \end{pmatrix}$$

c. Obtain the fitted regression function. What is estimated by the intercept term?

```
X[which(prod$X1 == max(as.numeric(prod$X1 )) ), ]<- (-1)
 lm(Y ~ X[,1]+ X[,2]    )
```

```
##
## Call:
## lm(formula = Y ~ X[, 1] + X[, 2])
##
## Coefficients:
## (Intercept)       X[, 1]       X[, 2]
##     8.07037     -1.19259      0.06296
```

$\hat{Y} = 8.07037 - 1.19259X_1 + 0.06296X_2$. $\mu_{\cdot} = \frac{\sum_{i=1}^{r} \mu_i}{r}$

d. Obtain the regression analysis of variance table.

```
prod$X1<-as.factor(prod$X1)
anova( lm(Y ~ prod$X1 ))#£
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## prod$X1    2 20.125 10.0626   15.72 4.331e-05 ***
## Residuals 24 15.362  0.6401
## ---
```

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

e. Conduct the test for equality of factor level means; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

H_0: $\tau_1 = \tau_2 = 0$, H_a: not both τ_1 and τ_2 equal zero.

```
(sum(anova(lm(Y ~ X[,1]+ X[,2]    ))[1:(prod.r-1),3])/(prod.r-1))/
sum(anova(lm(Y ~ X[,1]+ X[,2]    ))[ (prod.r ),3])
## [1] 15.72053
qf(1-.05,prod.r-1,prod.n-prod.r)
## [1] 3.402826
```

$F^* = 15.72053$, $F(.95; 2, 24) = 3.402826$. If $F^* \leq 3.402826$ fail to reject H_0, otherwise reject H_0. Reject H_0.

16.19. Refer to Questionnaire color Problem 16.8. Regression model (16.75) is to be employed for testing the equality of the factor level means.

a. Set up the **Y**, **X**, and $\boldsymbol{\beta}$ matrices.

$$\mathbf{Y} = \begin{pmatrix} 28 \\ 26 \\ 31 \\ 27 \\ 35 \\ 34 \\ 29 \\ 25 \\ 31 \\ 29 \\ 31 \\ 25 \\ 27 \\ 29 \\ 28 \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix} \qquad \boldsymbol{\beta} = \begin{pmatrix} \mu \\ \tau_1 \\ \tau_2 \end{pmatrix}$$

b. Obtain $\mathbf{X}\boldsymbol{\beta}$. Develop equivalent expressions of the elements of this vector in terms of the cell means μ_i.

$$\boldsymbol{X\beta} = \begin{pmatrix} \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_1 \\ \mu + \tau_2 \\ \mu + \tau_2 \\ \mu + \tau_2 \\ \mu + \tau_2 \\ \mu + \tau_2 \\ \mu - \tau_1 - \tau_2 \\ \mu - \tau_1 - \tau_2 \\ \mu - \tau_1 - \tau_2 \\ \mu - \tau_1 - \tau_2 \\ \mu - \tau_1 - \tau_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \\ \mu_3 \\ \mu_3 \\ \mu_3 \end{pmatrix}$$

c. Obtain the fitted regression function. What is estimated by the intercept term?

```
 lm(Y ~ X[,1]+ X[,2]   )
##
## Call:
## lm(formula = Y ~ X[, 1] + X[, 2])
##
## Coefficients:
## (Intercept)       X[, 1]       X[, 2]
##        29.0          0.4          0.6
```

$\hat{Y} = 29 + 0.4X_1 + 0.6X_2$. $\mu. = \frac{\sum_{i=1}^{r} \mu_i}{r}$

d. Obtain the regression analysis of variance table.

```
QC$X1<-as.factor(QC$X1)
anova(lm( Y ~QC$X1)) #£
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value Pr(>F)
## QC$X1      2    7.6     3.8  0.3918 0.6842
## Residuals 12  116.4     9.7
```

e. Conduct the test for equality of factor level means; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

H_0: $\tau_1 = \tau_2 = 0$, H_a: not both τ_1 and τ_2 equal zero.

```
(sum(anova(lm(Y ~ X[,1]+ X[,2]   ))[1:(QC.r-1),3])/(QC.r-1))/
sum(anova(lm(Y ~ X[,1]+ X[,2]   ))[ (QC.r ),3])
## [1] 0.3917526
qf(1-.1,QC.r-1,QC.n-QC.r)
## [1] 2.806796
```

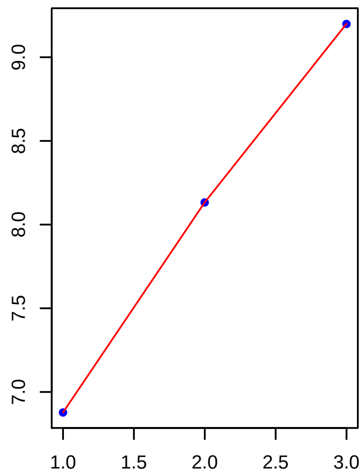
$F^* = 0.3917526$, $F(.95; 2, 12) = 2.806796$. If $F^* \leq 2.806796$ fail to reject H_0, otherwise reject H_0. Fail to reject H_0.

17.8. Refer to Productivity improvement Problem 16.7.

a. Prepare a line plot of the estimated factor level means $\hat{Y}_i$. What does this plot suggest regarding the effect of the level of research and development expenditures on mean productivity improvement.

```
Ybar;1:prod.r
## [1] 6.877778 6.877778 6.877778 6.877778 6.877778 6.877778 6.877778
## [8] 6.877778 6.877778 8.133333 8.133333 8.133333 8.133333 8.133333
## [15] 8.133333 8.133333 8.133333 8.133333 8.133333 8.133333 8.133333
## [22] 9.200000 9.200000 9.200000 9.200000 9.200000 9.200000
## [1] 1 2 3
```

```
par(mar=c(2,2,2,2),mfrow=c(1,3) )
plot(1:prod.r,unique(Ybar),col="blue",pch=16)
lines(1:prod.r, unique(Ybar)  , col='red')
```



This plot clearly suggests that with more expenditures for research and development there is an increase in productivity.

d. Obtain confidence intervals for all pairwise comparisons of the treatment means; use the Tukey procedure and a 90 percent family confidence coefficient. State your findings and prepare a graphic summary by underlining nonsignificant comparisons in your line plot in part (a).

```
TukeyHSD(      aov(prod$Y ~ as.factor(prod$X1) ),      conf.level = 0.9)
##   Tukey multiple comparisons of means
##     90% family-wise confidence level
##
## Fit: aov(formula = prod$Y ~ as.factor(prod$X1))
##
## $`as.factor(prod$X1)`
##         diff       lwr      upr     p adj
## 2-1 1.255556 0.4954165 2.015695 0.0043755
## 3-1 2.322222 1.4136823 3.230762 0.0000335
## 3-2 1.066667 0.2047500 1.928583 0.0347870
```

All of our p-values are less than 0.1 which means there are no nonsignificant comparisons in part (a).