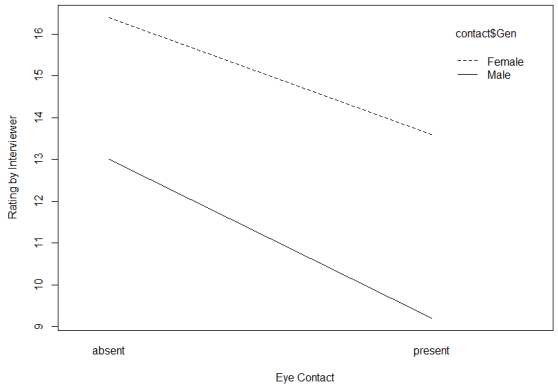
STA 101 HW 6 Solutions

Dr. Erin K. Melcon

1. (a) The plot follows:



There does not appear to be significant interaction effect, since the lines on the plot are almost parallel.

(b) H_0: The model with no interactions is a statistically better fit than the one with interactions (I.e., all $(\gamma\delta)_{ij} = 0$), vs. H_A: The model with no interactions is not a statistically better fit than the one with interactions (I.e., at least one $(\gamma\delta)_{ij} \neq 0$)
The test-statistic is : 0.2057613, with corresponding p-value 0.656202.
We fail to reject the null, and conclude that the model without interactions between gender and eye contact does not improve the statistical fit of the model (in other words, an interaction term is not needed).

(c) H_0: The model with no factor A (eye contact) effect is a statistically better fit than the one with a factor A (eye contact) effect (I.e., all $\gamma_i = 0$), vs. H_A: The model with no factor A (eye contact) is not a statistically better fit than the one with a factor A (eye contact) effect (I.e., at least one $\gamma_i \neq 0$)
The test-statistic is : 9.4022346, with corresponding p-value 0.0069911.
We reject the null, and conclude that the model with a factor A (eye contact) effect does improve the statistical fit of the model (in other words, a factor A (eye contact) effect should be included).

(d) H_0: The model with no factor B (gender) effect is a statistically better fit than the one with a factor B (gender) effect (I.e., all $\delta_j = 0$), vs. H_A: The model with no factor B (gender) effect is not a statistically better fit than the one with a factor B (gender) effect (I.e., at least one $\delta_j \neq 0$)
The test-statistic is : 13.1320467, with corresponding p-value 0.0020976.
We reject the null, and conclude that the model with a factor B (gender) effect does improve the statistical fit of the model (in other words, a factor B (gender) effect should be included).

---

2. (a) The highest rating can be found by looking at the interaction plot, or by calculating all combinations of factor means. If we do that, we see that females who had no eye contact have the highest rating.

(b) The estimates of γ_i are:

	absent	present
$\hat{\gamma}_i$	1.65	-1.65

(c) The estimates of δ_j are:

	Female	Male
$\hat{\gamma}_j$	1.95	-1.95

	Female	Male
absent	0.00	0.00
present	0.00	0.00

(d) The estimates for $(\gamma\delta)_{ij}$ are:

Note, the above is appropriate since we concluded there were no interaction effects.

---

3. (a) The 90% confidence interval using the Bonferroni multiplier for $\mu_{absent} - \mu_{present}$ is: (1.4273887, 5.1726113).

(b) Since the confidence interval bounds do not contain zero, it does suggest a significant factor A effect, and more specifically, that interviews with absent eye contact had higher ratings than interviews with eye contact (on average).

(c) The 90% confidence interval using the Bonferroni multiplier for $\mu_{female} - \mu_{male}$ is: (2.0273887, 5.7726113).

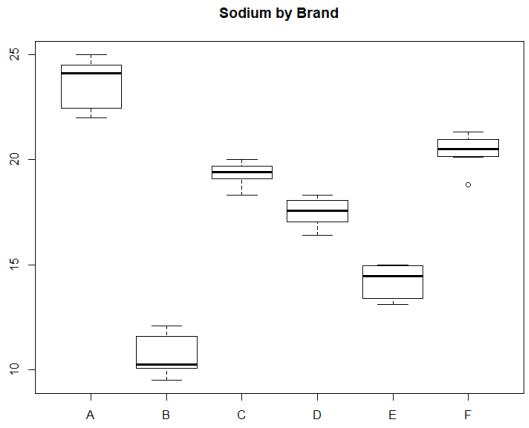
(d) Since the confidence interval bounds do not contain zero. it does suggest a significant factor B effect, and more specifically, that females tended to score higher on average than males.

(e) The 99% confidence interval using the Bonferroni multiplier for $\mu_{present,female} - \mu_{present,male}$ is : (-0.0107432, 8.8107432).

(f) We are 99% confident that there is no significant difference in average interview ratings for females who made significant eye contact and males who made significant eye contact.

---

4. (a) The plot follows:



Since the variance is quite unequal by group, we may use random effects. Or, since it is very likely we did not sample all the brands (we have a random subset of brands), whatever effect we see on sodium is random.

(b) The estimate for σ_A is :21.2737352. The estimate for σ_ϵ is :0.7159524.

(c) The proportion of reduction in variance is : 0.9988687. This suggests that 99.886867% of the variance in sodium is explained by brand.

(d) $H_0 : \sigma_A^2 = 0$ vs. $H_A : \sigma_A^2 \neq 0$.
The test-statistic is: 126.2727745, with corresponding p-value of approximately zero.

(e) We reject the null and conclude that a random effect for brand should be included in the model (or that brand has a significant effect on the sodium content).

(a) The students are randomly selected from a larger group of students, and not all students were measured.
There are many possible vitamin supplements that could have been used, but this model only considers three.

(b) $\hat{\sigma}_A^2 = 3799.4896$, $\hat{\sigma}_\epsilon^2 = 2.073024 \times 10^4$, $\hat{\sigma}_Y^2 = 2.452973 \times 10^4$

(c) $\hat{\sigma}_B^2 = 3695.4241$, $\hat{\sigma}_\epsilon^2 = 2.1100468 \times 10^4$, $\hat{\sigma}_Y^2 = 2.4795892 \times 10^4$

(d) $\hat{\sigma}_A^2 = 4424.9104$, $\hat{\sigma}_B^2 = 4373.1769$, $\hat{\sigma}_\epsilon^2 = 2.193361 \times 10^4$, $\hat{\sigma}_Y^2 = 3.0731697 \times 10^4$

(e) $H_0 : \sigma_A^2 = 0$ vs $H_A : \sigma_A^2 > 0$.
The test-statistic is: 8.2296499, with corresponding p-value 2
(using $\alpha = 0.05$ or $\alpha = 0.10$) We reject the null and conclude there is a signifcant effect of student (or that a random effect for student is needed)
(using $\alpha = 0.01$) We fail reject the null and conclude there is not a signifcant effect of student (or that a random effect for student is not needed)

(f) $H_0 : \sigma_B^2 = 0$ vs $H_A : \sigma_B^2 > 0$.
The test-statistic is: 6.6346387, with corresponding p-value 1
(using $\alpha = 0.05$ or $\alpha = 0.10$) We reject the null and conclude there is a signifcant effect of supplement (or that a random effect for supplement is needed)
(using $\alpha = 0.01$) We fail reject the null and conclude there is not a signifcant effect of supplement (or that a random effect for supplement is not needed)

```r
find.means = function(the.data,fun.name = mean){
  a = length(unique(the.data[,2]))
  b = length(unique(the.data[,3]))
  means.A = by(the.data[,1], the.data[,2], fun.name)
  means.B = by(the.data[,1],the.data[,3],fun.name)
  means.AB = by(the.data[,1],list(the.data[,2],the.data[,3]),fun.name)
  MAB = matrix(means.AB,nrow = b, ncol = a, byrow = TRUE)
  colnames(MAB) = names(means.A)
  rownames(MAB) = names(means.B)
  MA = as.numeric(means.A)
  names(MA) = names(means.A)
  MB = as.numeric(means.B)
  names(MB) = names(means.B)
  MAB = t(MAB)
  results = list(A = MA, B = MB, AB = MAB)
  return(results)
}

get.gamma.delta = function(the.model,the.data){
  nt = nrow(the.data)
  a = length(unique(the.data[,2]))
  b = length(unique(the.data[,3]))
  the.data$hat = the.model$fitted.values
  the.ns = find.means(the.data,length)
  a.vals = sort(unique(the.data[,2]))
  b.vals= sort(unique(the.data[,3]))
  muij = matrix(nrow = a, ncol = b)
  rownames(muij) = a.vals
  colnames(muij) = b.vals
  for(i in 1:a){
    for(j in 1:b){
      muij[i,j] = the.data$hat[which(the.data[,2] == a.vals[i] & the.data[,3] == b.vals[j])[1]]
    }
  }
  mi. = rowMeans(muij)
  m.j = colMeans(muij)
  mu.. = sum(muij)/(a*b)
  gammai = mi. - mu..
  deltaj = m.j - mu..
  gmat = matrix(rep(gammai,b),nrow = a, ncol = b, byrow= FALSE)
  dmat = matrix(rep(deltaj,a),nrow = a, ncol = b,byrow=TRUE)
  gamma.deltaij =round(muij -(mu.. + gmat + dmat),8)
  results = list(Mu.. = mu.., Gam = gammai, Del = deltaj, GamDel = gamma.deltaij)
  return(results)
}

find.mult = function(alpha,a,b,dfSSE,g,group){
  if(group == "A"){
    Tuk = round(qtukey(1-alpha,a,dfSSE)/sqrt(2),3)
    Bon = round(qt(1-alpha/(2*g), dfSSE ) ,3)
    Sch = round(sqrt((a-1)*qf(1-alpha, a-1, dfSSE)),3)
  }else if(group == "B"){
    Tuk = round(qtukey(1-alpha,b,dfSSE)/sqrt(2),3)
    Bon = round(qt(1-alpha/(2*g), dfSSE ) ,3)
    Sch = round(sqrt((b-1)*qf(1-alpha, b-1, dfSSE)),3)
  }else if(group == "AB"){
```

```
    Tuk = round(qtukey(1-alpha,a*b,dfSSE)/sqrt(2),3)
    Bon = round(qt(1-alpha/(2*g), dfSSE ) ,3)
    Sch = round(sqrt((a*b-1)*qf(1-alpha, a*b-1, dfSSE)),3)
  }
  results = c(Bon, Tuk,Sch)
  names(results) = c("Bonferroni","Tukey","Scheffe")
  return(results)
}

scary.CI = function(the.data,MSE,multiplier,group,cs){
   if(sum(cs) != 0 & sum(cs !=0 ) != 1){
    return("Error - you did not input a valid contrast")
  }else{
    the.means = find.means(the.data)
    the.ns =find.means(the.data,length)
    nt = nrow(the.data)
    a = length(unique(the.data[,2]))
    b = length(unique(the.data[,3]))
    if(group =="A"){
        a.means = rowMeans(the.means$AB)
        est = sum(a.means*cs)
        mul = rowSums(1/the.ns$AB)
        SE = sqrt(MSE/b^2 * (sum(cs^2*mul)))
        N = names(a.means)[cs!=0]
        CS = paste("(",cs[cs!=0],")",sep = "")
        fancy = paste(paste(CS,N,sep =""),collapse = "+")
        names(est) = fancy
    }else if(group == "B"){
        b.means = colMeans(the.means$AB)
        est = sum(b.means*cs)
        mul = colSums(1/the.ns$AB)
        SE = sqrt(MSE/a^2 * (sum(cs^2*mul)))
        N = names(b.means)[cs!=0]
        CS = paste("(",cs[cs!=0],")",sep = "")
        fancy = paste(paste(CS,N,sep =""),collapse = "+")
        names(est) = fancy
    } else if(group == "AB"){
      est = sum(cs*the.means$AB)
      SE = sqrt(MSE*sum(cs^2/the.ns$AB))
      names(est) = "someAB"
    }
    the.CI = est + c(-1,1)*multiplier*SE
    results = c(est,the.CI)
    names(results) = c(names(est),"lower bound","upper bound")
    return(results)
  }
}

#Problem 1
the.data = read.csv("C:/Github/Teaching-Materials/STA-101/STA-101-2017-Spring/Datasets/HW-7/contact.csv")
#interaction.plot(contact$Eye,contact$Gen,contact$Rating, xlab = "Eye Contact", ylab = "Rating by Interviewer"
names(the.data) = c("Y","A","B")
A.B = lm(Y ~ A + B,the.data)
AB = lm(Y ~ A*B,the.data)
A = lm(Y ~ A,the.data)
B = lm(Y ~ B,the.data)
N = lm(Y ~ 1, the.data)
TAB = anova(A.B,AB)
TAB1 = anova(B, A.B)
TAB2 = anova(A, A.B)
```

```
#Problem  2
None = get.gamma.delta(A.B, the.data)
the.means = find.means(the.data,mean)
#Problem  3
nt = nrow(the.data)
a = length(unique(the.data[,2]))
b = length(unique(the.data[,3]))
c.a = c(1,-1)
B1= find.mult(0.10, a, b, nt -(a+b-1), 1, "A")[1]
c.b = c(1,-1)
B2= find.mult(0.10, a, b, nt -(a+b-1), 1, "B")[1]
c.ab = matrix(0,nrow = 2, ncol =2)
c.ab[2,1] = 1
c.ab[2,2] = -1
B3= find.mult(0.01, a, b, nt -(a+b-1), 1, "AB")[1]
SSE = sum(A.B$residuals^2)
MSE = SSE/(nt -(a + b - 1))

CI1 = scary.CI(the.data,MSE,B1,"A",c.a)
CI2 = scary.CI(the.data,MSE,B2,"B",c.b)
CI3 = scary.CI(the.data,MSE,B3,"AB",c.ab)


#Problem 4
library(nlme)
library(lme4)
salt = read.csv("C:/Github/Teaching-Materials/STA-101/STA-101-2017-Spring/Datasets/HW-7/salt.csv")
Random.model = lmer( Sodium ~ 1+ (1|Brand),data = salt)

sigmaE =  as.data.frame(VarCorr(Random.model))[2,4]
sigmaA =  as.data.frame(VarCorr(Random.model))[1,4]
prop.explained = (sigmaA^2)/(sigmaA^2 +sigmaE^2)
null.model = gls( Sodium ~ 1, data= salt)
LL0 = logLik(null.model)
LLA = logLik(Random.model)
G2 = 2*(LLA - LL0)
p.val = 0.5*pchisq(G2, df=  1, lower.tail = FALSE)
#Problem 5
the.data = read.csv("C:/Github/Teaching-Materials/STA-101-2018-Spring/Datasets/HW-6/sodium.csv")
names(the.data) = c("Y","A","B")
ran.A.model = lmer(Y ~ 1 + B + (1|A),data = the.data,REML = FALSE)
ran.B.model = lmer(Y ~ 1 + A + (1|B),data = the.data,REML = FALSE)
ran.AB.model.noI = lmer(Y ~ 1 + (1|A) + (1|B),data = the.data,REML = FALSE) #No interactions
sa = round(61.64^2,4)
se = round(143.98^2,4)

sb = round(60.79^2,4)
se2 = round(145.26^2,4)

sa3 = round(66.52^2,4)
sb3 = round(66.13^2,4)
se3 = round(148.10^2,4)

TAB1 = anova(ran.A.model,ran.AB.model.noI)
TAB2 = anova(ran.B.model,ran.AB.model.noI)
```

\end{enumerate}