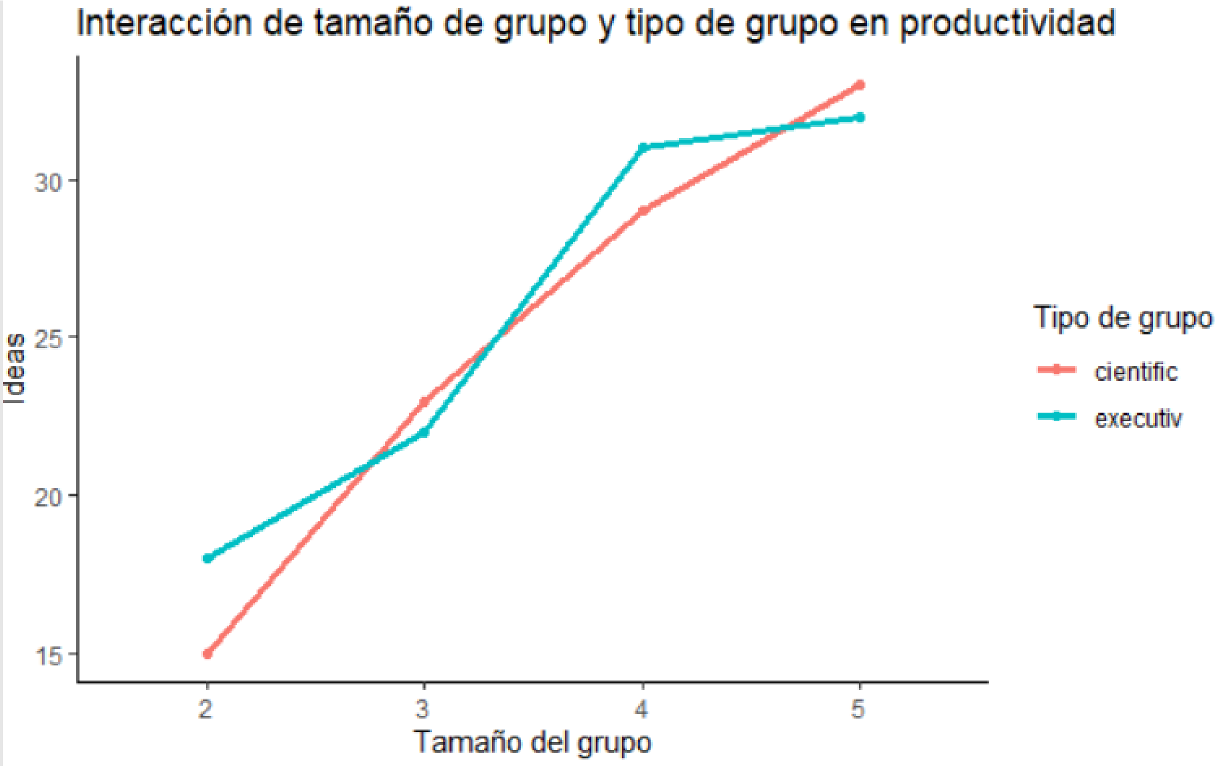
Ejercicio 1: Caso factorial una réplica
La Psicología de las Organizaciones es una disciplina científica que estudia la conducta del ser humano y sus experiencias en el contexto del trabajo a nivel individual, grupal y organizacional. Un tópico de interés de esta rama es evaluar el efecto del tamaño de un equipo de trabajo sobre productividad. Un investigador quiere determinar si las *lluvias de ideas* son más efectivas en grupos de trabajo grandes o en grupos de trabajo pequeños. Para ello formó cuatro grupos de empresarios de tamaños 2, 3, 4 y 5 personas; de manera análoga realizó el experimento con grupos de científicos. A cada grupo se le realizó la siguiente pregunta: *¿Cómo encontrar el punto de equilibro entre desarrollo económico y desarollo científico?*. Cada grupo tuvo 30 minutos para generar ideas. La variable de interés consiste en el número de ideas diferentes propuestas por grupo. Los resultados se encuentran en la siguiente tabla:

	Tamaño del grupo			
Tipo de grupo	2	3	4	5
Ejecutivo	18	22	31	32
Científicos	15	23	29	33

a) Grafique los datos para visualizar interacción y efectos principales. Comente.
Respuesta

```
(tipegroups<-factor(rep(c("executiv","cientific"),each=4)))
[1] executiv  executiv  executiv  executiv  cientific cientific cientific cientific
Levels: cientific executiv

(sizegroups<-factor(rep(c(2,3,4,5),times=2)))
[1] 2 3 4 5 2 3 4 5
Levels: 2 3 4 5

(nideas<-c(c(18,22,31,32),c(15,23,29,33)))
[1] 18 22 31 32 15 23 29 33

(data<-data.frame(tipegroups,sizegroups,nideas))
  tipegroups sizegroups nideas
1   executiv          2     18
2   executiv          3     22
3   executiv          4     31
4   executiv          5     32
5  cientific          2     15
6  cientific          3     23
7  cientific          4     29
8  cientific          5     33


#Grafico de interaccion

library(ggplot2)
ggplot(data, aes(x=sizegroups, y=nideas, color=tipegroups, group=tipegroups)) + geom_point()
+ geom_line(size=1.2) +
labs(title = "Interaccion de tamano de grupo y tipo de grupo en productividad",
x= "Tamano del grupo", y="Ideas", color="Tipo de grupo") + theme_classic()
```



- **Efecto principal del tamaño**: Se observa que, en general, a medida que aumenta el tamaño del grupo, aumenta la cantidad de ideas diferentes. Se observan diferencias notorias en la cantidad de ideas por tamaño de grupo.
- **Efecto principal del tipo de grupo**: Si se observa el comportamiento del grupo *cientific* y del grupo *executiv* se observan bastante cercanos, precisa mayor análisis.
- **Interacción tamaño y tipo de grupo**: Se observan cruces entre las tendencias, lo que sugiere cierta interacción existente entre ambas variables. Si las variables no tuvieran ninguna incidencia conjunta, no se esperarían observar intersecciones.

b) Ajuste el modelo con interacción. Al obtener la tabla ANOVA del modelo se entrega un Warning, comente por qué se arroja dicho warning. Justifique.
Respuesta

```
# Modelo con interaccion
contrasts(tipegroups)<-contr.sum
contrasts(sizegroups)<-contr.sum

(modelo<-aov(nideas~tipegroups*sizegroups))
Call:
   aov(formula = nideas ~ tipegroups * sizegroups)

Terms:
                tipegroups sizegroups tipegroups:sizegroups
Sum of Squares       1.125    318.375                 6.375
Deg. of Freedom          1          3                     3

Estimated effects may be unbalanced

anova(modelo)
Analysis of Variance Table

Response: nideas
                      Df Sum Sq Mean Sq F value Pr(>F)
tipegroups             1   1.12   1.125
sizegroups             3 318.38 106.125
```

```
tipegroups:sizegroups  3   6.37   2.125
Residuals              0   0.00
Warning message:
In anova.lm(modelo) :
  ANOVA F-tests on an essentially perfect fit are unreliable
```

El Warning indica que los tests F usuales no resultan confiables. Esto se debe a que, dado que se posee una observación por celda y en un modelo con interacción los grados de libertad asociados a SCE son $ab(n-1)$ y en este caso $n = 1$ por esto se obtiene *Residuals df=0*, se indetermina MCE y por lo tanto no se podrían calcular los tests F. Además $SCE = \sum e_{ij}^2 = \sum (y_{ij} - \hat{y}_{ij})^2 = 0$. Intuitivamente lo que sucede es que, para estimar cada parámetro se utiliza una observación, y por ende no quedarían grados de libertad disponibles.

c) ¿Cuál sería el valor ajustado por este modelo para el grupo de científicos de tamaño 2?
Respuesta
Dado lo anterior, el valor ajustado sería precisamente el valor observado del número de ideas del grupo científico de tamaño 2: $\hat{y}_{\text{cientifico tamaño 2}} = 15$.

```
fitted(modelo)
 1  2  3  4  5  6  7  8
18 22 31 32 15 23 29 33

data$nideas
[1] 18 22 31 32 15 23 29 33
```

d) Asumiendo que $\hat{D} = -0,2$ realice el test de Tukey de aditividad. ¿Qué concluye respecto a la componente interacción?
Respuesta

```
#Test de tukey

#Paso 1) Extraer vectores de los coeficientes estimados para cada factor
contrasts(tipegroups)<-contr.sum
contrasts(sizegroups)<-contr.sum

aditive<-aov(nideas~tipegroups+sizegroups)
coef(aditive)
(Intercept) tipegroups1 sizegroups1 sizegroups2 sizegroups3
     25.375      -0.375      -8.875      -2.875       4.625


(avalues<-ifelse(tipegroups=="cientific",unname(coef(aditive)[2]),-unname(coef(aditive)[2])))
[1]  0.375  0.375  0.375  0.375 -0.375 -0.375 -0.375 -0.375

(bvalues<-ifelse(sizegroups=="2",unname(coef(aditive)[3]),ifelse(sizegroups=="3",unname(coef(aditive)
[1] -8.875 -2.875  4.625  7.125 -8.875 -2.875  4.625  7.125


#Paso 2) Define Dhat

Dhat<--0.2


#Paso 3) Definir SCAB*

SCABp<-Dhat^{2}*sum(avalues^{2}*bvalues^{2})

#Paso 4) Extraer SCA, SCB y SCT del modelo aditivo

SCA<-anova(aditive)[1,2]
SCB<-anova(aditive)[2,2]
SCT<-sum(anova(aditive)[,2])
```

```
#Paso 5) Definir SCE*

SCEp<-SCT-SCA-SCB-SCABp

#Paso 6) Definir el estadistico Fp

(a<-length(levels(tipegroups)))   #Niveles del factor tipo
[1] 2

(b<-length(levels(sizegroups)))   #Niveles del factor size
[1] 4

Fp=(SCABp/1)/(SCEp/(a*b-a-b))

#Paso 7) Regla de decision

Fp>qf(0.95,1,a*b-a-b)    #No se rechaza la hipotesis nula

#Se tiene la hipotesis de aditividad
```

e) Realice análisis del modelo aditivo, ¿podría decir que el tamaño influye de manera significativa en las *lluvias de ideas*?. ¿Qué pasos a seguir le sugeriría al investigador? Comente.
Respuesta

```
summary(aditive)
            Df Sum Sq Mean Sq F value  Pr(>F)
tipegroups   1    1.1    1.12   0.529 0.51950
sizegroups   3  318.4  106.13  49.941 0.00464 **
Residuals    3    6.4    2.12
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1      1
```

Es posible observar que la suma cuadrática asociada a la variable tipo de grupos siendo 1.1 es drásticamente menor a la suma cuadrática asociada a la variable tamaño de grupos siendo ésta de 318.4, asimismo, realizando los tests F usuales, se puede observar que la variable tamaño de grupo resulta significativa al 95% mientras que, dado lo anterior, no ocurre lo mismo con el tipo de grupos. Por ende, el investigador intuía bien que la productividad en término de ideas se ve asociada de manera importante al tamaño de grupos, y en este caso, el tipo de grupos no presenta un mayor efecto.

```
coef(aditive)
(Intercept) tipegroups1 sizegroups1 sizegroups2 sizegroups3
     25.375      -0.375      -8.875      -2.875       4.625

levels(tipegroups)
[1] "cientific" "executiv"

unname(-coef(aditive)[2]) #efecto de tipegroup2
[1] 0.375

levels(sizegroups)
[1] "2" "3" "4" "5"

sum(-coef(aditive)[3:5])  #efecto de sizegroup4
[1] 7.125
```

Es posible observar que respecto al tamaño de grupo, se observa que al tener un tamaño de 5 personas en el equipo de observa un incremento de 7.125 en el número de ideas promedio. Ahora bien, se observa en general una tendencia creciente del número de ideas respecto al tamaño del equipo, lo cual pudiera ser esperable dado que más personas producirían naturalmente más ideas. Pudiera ser un buen paso a seguir, realizar un estudio similar pero con tamaños de grupo más grande y ver a partir de qué tamaño de grupo se deja de observar un incremento significativo en el número de las ideas promedio.

Ejercicio 2: Variable bloque Fuente del estudio

El Perfil de Sensibilidad No Verbal (MiniPONS) evalúa las diferencias individuales en la habilidad para reconocer emociones, actitudes interpersonales y comunicación de intenciones por canales no verbales. El MiniPONS consiste en un conjunto de videos cortos que muestran a una mujer con un tono emocional negativo y positivo manipulado de expresiones faciales, lenguaje corporal y voz. El participante debe indicar qué emoción cree que representa la mujer del video, la variable respuesta corresponde a la cantidad de aciertos del participante. Además, se conoce el estado mental del participante al momento del experimento, siendo clasificado como **UD** si presenta depresión unipolar, **BD I** si presenta desorden bipolar tipo 1, **BD II** si presenta desorden bipolar tipo 2 y **Control** si presenta un estado psíquico normal. La base de datos `mental` contiene la información.

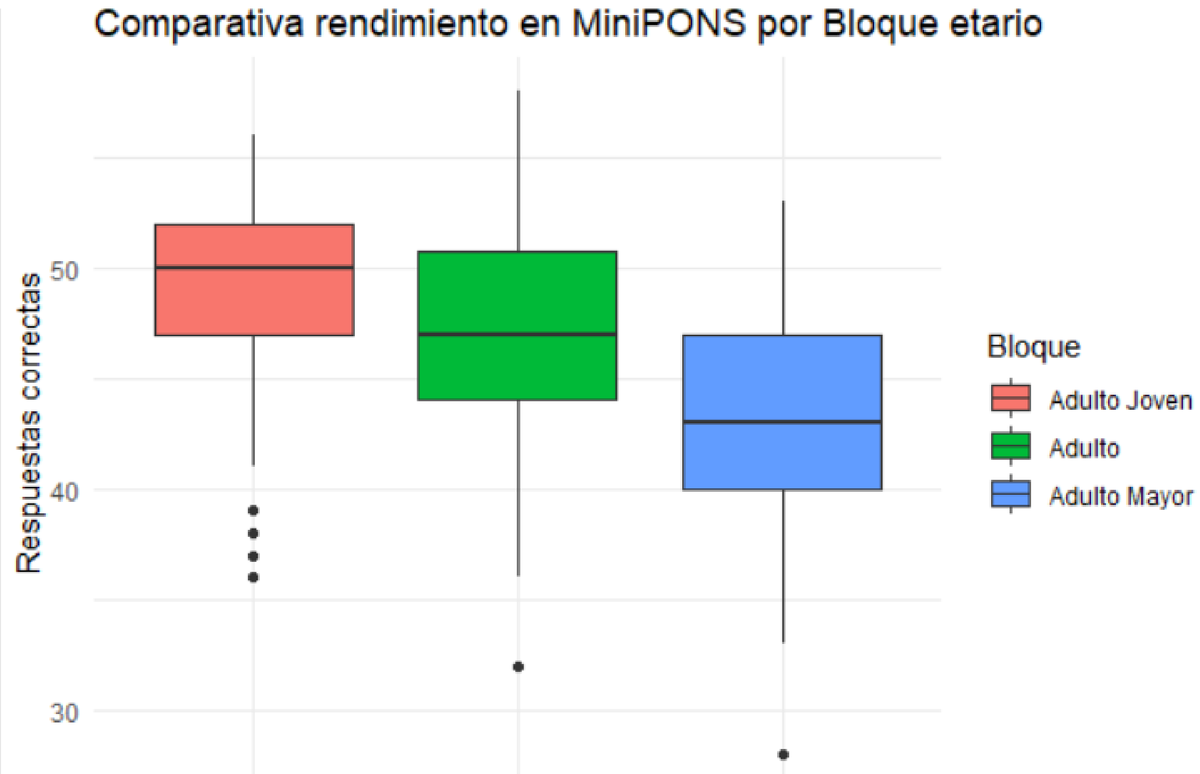
a) En la base de datos además se posee la edad de los participantes. La investigadora a cargo cree que hay diferencias naturales que pudieran explicarse por la edad pues se ha demostrado que el rendimiento en distintos tests se ve influenciado por la madurez cerebral. Cree una variable bloque de acuerdo a la edad de los participantes como sigue:

   - **Adulto Joven** Si la edad del participante es menor o igual a 45 años
   - **Adulto** Si la edad del participante es mayor a 45 pero menor que 60 años
   - **Adulto Mayor** Si la edad del participante es superior o igual a 60 años

   Evalúe la variable bloque creada. ¿Le parece que explica diferencias en términos de respuestas correctas?

   Respuesta

```
 print(head(mental))
# A tibble: 6 x 3
  Type    Age Right_answers
  <chr> <dbl>         <dbl>
1 BD I     47            40
2 BD I     49            49
3 BD I     45            43
4 BD I     53            44
5 BD II    50            50
6 BD I     31            54

summary(mental$Age)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
21.00   39.00   50.00   48.72   58.00   78.00

Bloque<-ifelse(mental$Age<=40, "Adulto Joven",ifelse(mental$Age<60,"Adulto","Adulto Mayor"))

mental$Bloque<-Bloque

mental$Bloque<-factor(Bloque,levels=c("Adulto Joven","Adulto","Adulto Mayor"))

ggplot(aes(y = Right_answers, x = Bloque, fill=Bloque), data = mental) +
  geom_boxplot()+theme_minimal()+
  theme(axis.text.x = element_blank())+ylab("Respuestas correctas")+xlab("")+
  ggtitle("Comparativa rendimiento en MiniPONS por Bloque etario")
```




```
library(dplyr)
mental %>% group_by(Bloque) %>% summarise(Media=mean(Right_answers),Minimo=min(Right_answers),
Maximo=max(Right_answers),n=n())
# A tibble: 3 x 5
  Bloque       Media Minimo Maximo     n
  <fct>        <dbl>  <dbl>  <dbl> <int>
1 Adulto Joven  49.0     36     56   107
2 Adulto        47.2     32     58   114
3 Adulto Mayor  43.1     28     53    56

#Test de comparaciones m ltiples 2 a 2

pairwise.t.test(x = mental$Right_answers, g = mental$Bloque, p.adjust.method = "bonferroni", pool.sd
alternative = "two.sided")

    Pairwise comparisons using t tests with pooled SD

data:  mental$Right_answers and mental$Bloque

             Adulto Joven Adulto
Adulto       0.022        -
Adulto Mayor 4.7e-12      8.7e-07

P value adjustment method: bonferroni
```

Analizando el test de comparaciones múltiples usando el método de bonferroni, se obtiene que las diferencias de medias entre los distintos bloques son significativas. Por lo que, la variable bloque etario explicaría diferencias relevantes respecto a la cantidad de respuestas correctas de los participantes. *Respuesta*

```
#b) Caso no balanceado
addmargins(table(mental$Bloque,mental$Type))

               BD I BD II Control  UD Sum
  Adulto Joven   37    15      53   2 107
  Adulto         24    24      53  13 114
  Adulto Mayor    9    10      13  24  56
```

```
Sum                70      49       119   39  277
```

Se puede observar que en este caso y dada la variable Bloque creada, nos encontramos en el caso no balanceado, pues dentro de cada bloque se observa un número de participantes diferente. El grupo Adulto mayor cuenta con información de 56 partipantes mientras que el grupo adulto joven cuenta con información de 107 participantes. Es importante analizar la concentración de las observaciones puesto que, podría ocurrir que una combinación Bloque factor no presentara observaciones y por lo tanto, esto podría significar irregularidades al momento de estimar la interacción.

c) Realice un análisis completo de los datos respecto del problema. Comente el modelo obtenido.
Respuesta

```
#c) Modelo con interaccion
mental$Type<-factor(mental$Type)

contrasts(mental$Bloque)<-contr.sum
contrasts(mental$Type)<-contr.sum

modelfull<-aov(Right_answers~Bloque*Type,data=mental)

anova(modelfull)
Analysis of Variance Table

Response: Right_answers
             Df Sum Sq Mean Sq F value    Pr(>F)
Bloque        2 1265.6  632.78 36.4663 1.024e-14 ***
type          3 1600.2  533.40 30.7391 < 2.2e-16 ***
Bloque:type   6   90.1   15.01  0.8652     0.521
Residuals   265 4598.4   17.35
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1      1
```

Es posible observar mirando los tests F ya conocidos, que el Bloque, tal como se intuyó presenta una alta significancia (considerando una confianza del 95%), por lo tanto, pudiera decirse que estratificar por el bloque etario creado resulta efectivo, pues logra explicar diferencias naturales en el rendimiento de los participantes en el test MiniPONS. Por otra parte, la variable estado psíquico también resulta presentar una alta significancia. Sin embargo, la interacción no resulta significativa. Se ajusta el modelo aditivo.

d) Obtenga los coeficientes asociados a cada nivel de la variable estado psíquico y bloque etario. ¿Bajo qué estado mental se presentan mejores resultados en el test?
Respuesta

```
contrasts(mental$Bloque)<-contr.sum
contrasts(mental$Type)<-contr.sum

aditivo<-aov(Right_answers~Type+Bloque,data=mental)

 coef(aditivo)
(Intercept)       Type1       Type2       Type3     Bloque1     Bloque2
 45.7461472  -1.4921574  -0.4080783   3.6825975   2.0175729   0.3819115
levels(mental$Type)
[1] "BD I"    "BD II"   "Control" "UD"

sum(-coef(aditivo)[2:4])  #efecto de estado psiquico UD
[1] -1.782362

levels(mental$Bloque)
[1] "Adulto Joven" "Adulto"       "Adulto Mayor"
```

```
sum(-coef(aditivo)[5:6])  #efecto asociado a rango etario Adulto mayor
[1] -2.399484
```

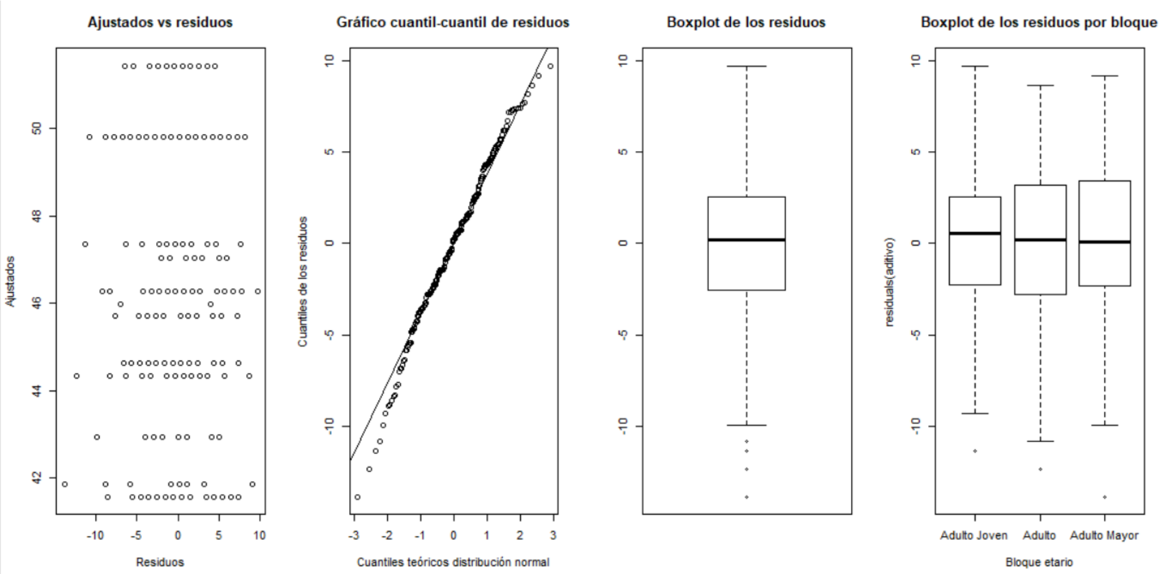
Se obtienen resultados bastante interesantes al realizar el modelo. El desorden bipolar tipo 1 presenta un efecto negativo de -1.49 y menor al efecto del desorden bipolar tipo 2 de -0.4, esto es interesante pues el desorden bipolar tipo 1 se asocia a un estado psíquico alterado de hiperactividad mental, mientras que, un desorden bipolar de tipo 2 se asocia a una exaltación psíquica más leve. Por lo tanto, se esperarían mejores resultados en el test de MiniPONS en un participante con desorden bipolar tipo 2 que con desorden bipolar tipo 1. Además, es posible observar que el mayor efecto se tiene en los participantes cuyo estado anímico se clasifica como *Control*, es decir, con estado psíquico no alterado. De lo anterior pudiera desprenderse que, un estado psíquico alterado sugiere efectos (por lo visto negativos) en el rendimiento en el test de reconocimiento afectivo MiniPONS.
Respecto al bloque etario, se obtienen los resultados que ya se esperaban a partir del análisis previo, a mayor edad se observa una disminución en el desempeño en el test de reconocimiento afectivo, pudiera deberse a la madurez cerebral, pero eso requiere otro análisis.
Se puede complementar lo anterior con una prueba post hoc de comparaciones múltiples de Tukey:

```
round(TukeyHSD(aditivo)$Type,5)
                    diff       lwr       upr   p adj
BD II-BD I       0.51020  -1.49254   2.51295 0.91247
Control-BD I     5.09244   3.47284   6.71204 0.00000
UD-BD I         -2.42491  -4.57339  -0.27643 0.01987
Control-BD II    4.58223   2.75715   6.40732 0.00000
UD-BD II        -2.93511  -5.24245  -0.62778 0.00624
UD-Control      -7.51735  -9.50126  -5.53343 0.00000
```

Es bastante interesante observar que si bien, un desorden bipolar tipo 2 sugiere mejor rendimiento que el desorden bipolar tipo 1 (comparando los coeficientes), el valor-p asociado a la diferencia de medias del número de aciertos es 0.91 y de hecho es la única diferencia de medias que no rechaza la igualdad.

e) Realice el análisis residual del modelo considerado en (c). Comente en términos de los supuestos.
Respuesta

```
#e) Analisis de supuestos

par(mfrow=c(1,4))

plot(residuals(aditivo), fitted(aditivo), main="Ajustados vs residuos",xlab="Residuos",
ylab="Ajustados")

qqnorm(residuals(aditivo),main="Grafico cuantil-cuantil de residuos",
xlab="Cuantiles teoricos distribuci n normal",ylab="Cuantiles de los residuos")
qqline(residuals(aditivo))

boxplot(residuals(aditivo),main="Boxplot de los residuos")
boxplot(residuals(aditivo)~mental$Bloque,xlab="Bloque etario",
main="Boxplot de los residuos por bloque")
```



Respecto al gráfico cuantil-cuantil de los residuos del modelo aditivo, es posible observar cierta distancia en las colas respecto al comportamiento de los cuantiles de una distribución normal teórica, aunque, en el centro, se observan bastante apegados. Se precisa realizar un test de normalidad.
Respecto al boxplot general de los residuos, es bastante importante notar que el valor máximo es de 10 y el valor mínimo es de -10, lo que indican una distancia bastante grande respecto al valor real del puntaje de reconocimiento afectivo en la prueba. Además es posible observar 3 outliers infieriores en el gráfico, dichos outliers necesitan ser detectados y ver cuál es la influencia que tienen en el modelo y la estimación de los efectos principales.
Respecto a los boxplot por bloque, se observa que las medianas de los tres bloques estan bastante cercanos al valor nulo, comportamiento esperable. Respecto a los rango intercuartílicos pudieran verse similares, por lo que, podría esperarse variabilidades similares intrabloques, pero se requiere mayor análisis. También es posible observar 3 outliers, un outlier por bloque, los que, seguramente, pudieran ser los outliers antes detectados en el boxplot general.

Ejercicio 3: Modelo de medias aleatorias de celdas
Considere un modelo de medias aleatorias de celda:

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

Donde $\mu_i \sim^{iid} N(\mu, \sigma^2_\mu)$, $\epsilon_{ij} \sim^{iid} N(0, \sigma^2)$, $\forall i = 1, ..., r$, $j = 1, ..., n$. μ_i y ϵ_{ij} son variables aleatorias independientes.

Demuestre que:

$$E(MCTrat) = \sigma^2 + n\sigma^2_\mu$$

Respuesta
Las medias por nivel i se obtienen como sigue:

$$Y_{ij} = \mu_i + \epsilon_{ij} \Longleftrightarrow \frac{\sum_{j=1}^n y_{ij}}{n} = \frac{\sum_{j=1}^n \mu_i}{n} + \frac{\sum_{j=1}^n \epsilon_{ij}}{n} \Longleftrightarrow \bar{Y}_{i\cdot} = \mu_i + \bar{\epsilon}_{i\cdot}$$

De manera análoga se puede obtener la media global:

$$\bar{Y}_{i\cdot} = \mu_i + \bar{\epsilon}_{i\cdot} \Longleftrightarrow \frac{\sum_{i=1}^{r} \bar{Y}_{i\cdot}}{r} = \frac{\sum_{i=1}^{r} \mu_i}{r} + \frac{\sum_{i=1}^{r} \bar{\epsilon}_{i\cdot}}{r} \Longleftrightarrow \bar{Y} = \mu + \bar{\epsilon}$$

Luego,

$$\bar{Y}_{i\cdot} - \bar{Y} = (\mu - \mu_i) + (\bar{\epsilon}_{i\cdot} - \bar{\epsilon}) \rightarrow \sum_i (\bar{Y}_{i\cdot} - \bar{Y})^2 = \sum_i (\mu - \mu_i)^2 + \sum_i (\bar{\epsilon}_{i\cdot} - \bar{\epsilon})^2 + 2\sum_i (\bar{\epsilon}_{i\cdot} - \bar{\epsilon}_{\cdot\cdot})(\mu - \mu_i)$$

Recuerde que por independencia asumida anteriormente,

$$2\sum_i (\bar{\epsilon}_{i\cdot} - \bar{\epsilon})(\mu - \mu_i) = 0$$

Sabiendo que $\epsilon_{ij} \sim^{iid} N(0, \sigma^2)$, es fácil ver que $\bar{\epsilon}_i \sim^{iid} N(0, \sigma^2/n)$ y un estimador insesgado para σ^2/n es $\frac{\sum_i (\bar{\epsilon}_{i\cdot} - \bar{\epsilon})^2}{r-1}$. Luego, utilizando propiedades lineales de la esperanza:

$$E(\sum_i (\bar{\epsilon}_{i\cdot} - \bar{\epsilon})^2) = \frac{(r-1)\sigma^2}{n}$$

De manera análoga:

$$E(\sum_i (\mu - \mu_i)^2) = (r-1)\sigma_\mu^2$$

Utilizando la igualdad obtenida anteriormente se expresa el MCTrat, en función de las dos esperanzas anteriores:

$$E(\frac{n}{(r-1}\sum_i (\bar{Y}_{i\cdot} - \bar{Y})^2) = \frac{n}{(r-1)}[(r-1)\sigma_\mu^2 + \frac{(r-1)}{n}\sigma^2] = n\sigma_\mu^2 + \sigma^2$$

Se concluye que:

$$E(MCTrat) = n\sigma_\mu^2 + \sigma^2$$