Mixed Model MPDC

Data

```
Datos %>%
  ungroup() %>%
  dplyr::select(Subject,Level,MPDC) %>%
  group_by(Subject,Level) %>%
  mutate(mid = 1:n()) %>%
  pivot_wider(names_from=mid,values_from=MPDC) %>%
  arrange(Subject,Level) %>%
  kable("latex", booktabs = T) %>%
  kable_styling(latex_options = c("striped", "scale_down"))
```

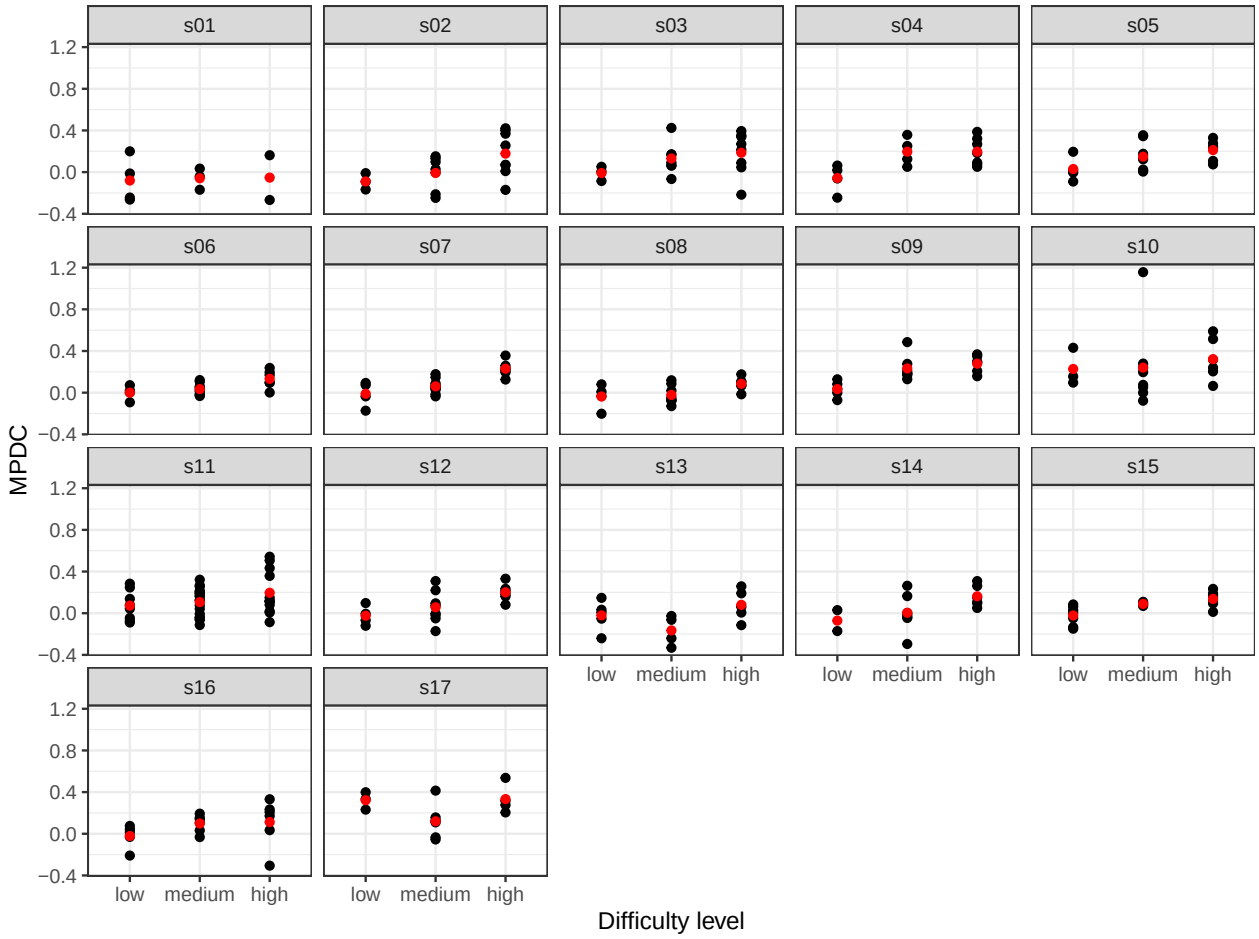
Subject	Level	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
s01	low	-0.0132	-0.2638	-0.2440	0.1995															
s01	medium	-0.0412	-0.1693	0.0336																
s01	high	0.1621	-0.2674																	
s02	low	-0.1672	-0.0101	-0.0914																
s02	medium	0.1312	-0.0040	-0.2476	0.0280	0.0981	0.1520	-0.2127												
s02	high	0.4206	0.3688	0.0700	0.0682	0.4001	0.2557	0.0096	-0.1703											
s03	low	0.0512	0.0074	-0.0854																
s03	medium	-0.0663	0.0635	0.0910	0.0626	0.4244	0.1680	0.1711												
s03	high	-0.2167	0.3398	0.3942	0.2138	0.2700	0.0889	0.0458	0.3520											
s04	low	0.0629	-0.0612	-0.2455	0.0179															
s04	medium	0.3578	0.2510	0.0508	0.1254															
s04	high	0.0507	0.3864	0.3215	0.1827	0.0916	0.0764	0.2677												
s05	low	-0.0044	0.1953	-0.0907	0.0186															
s05	medium	0.0239	0.0050	0.1758	0.0135	0.1230	0.3539	0.3450												
s05	high	0.1079	0.2358	0.2773	0.2465	0.3286	0.0740													
s06	low	0.0719	0.0204	-0.0927	0.0012															
s06	medium	0.0599	0.0137	0.0289	0.1060	-0.0139	0.0495	0.1210	-0.0233	-0.0321										
s06	high	0.1698	0.2374	0.0015	0.0940	0.0984	0.1967													
s07	low	-0.1728	-0.0356	0.0934	0.0740															
s07	medium	0.0425	-0.0169	0.0529	0.1777	0.0502	0.0697	-0.0356	0.1475	0.0285	0.0896									
s07	high	0.3562	0.2128	0.2318	0.1953	0.2580	0.1260													
s08	low	-0.2024	0.0797	0.0080	-0.0324															
s08	medium	0.1189	-0.1291	0.0861	-0.0729	-0.0758	0.0192	-0.0719	-0.0476											
s08	high	0.1759	0.0698	0.1036	0.0680	-0.0156	0.1072													
s09	low	-0.0712	-0.0002	0.0809	0.1281	0.0344														
s09	medium	0.4854	0.2765	0.1837	0.1292	0.1698	0.2055	0.1821												
s09	high	0.3507	0.2080	0.1578	0.2897	0.2974	0.3679													
s10	low	0.0965	0.4308	0.1554																
s10	medium	0.0522	0.0749	0.2791	-0.0772	0.1969	0.0002	1.1568	0.2404											
s10	high	0.3167	0.5150	0.2395	0.0646	0.5890	0.2042													
s11	low	0.0435	0.2838	0.1378	-0.0707	-0.0449	-0.0893	0.0792	0.2471											
s11	medium	0.3217	0.1240	0.2144	-0.0072	0.2574	0.0753	0.1674	0.1883	0.1617	0.1261	-0.0597	-0.0394	-0.1134	0.2668	0.2072	0.0424	-0.0415	0.1921	-0.0597
s11	high	0.5068	0.1162	0.3582	0.1145	0.0807	0.0065	0.1492	-0.0856	0.5424	0.1147	0.4334	0.0188							
s12	low	0.0975	-0.0673	-0.1206	-0.0054															
s12	medium	0.2205	0.0727	0.0946	0.3088	-0.0169	-0.0029	-0.1717	-0.0492											
s12	high	0.3309	0.2123	0.2342	0.1830	0.1614	0.0812													
s13	low	-0.2404	-0.0522	0.0080	0.1475	0.0331														
s13	medium	-0.0631	-0.2405	-0.0263	-0.3314															
s13	high	0.0063	0.2582	0.1913	0.0605	-0.1146														
s14	low	-0.1713	0.0298																	
s14	medium	-0.0380	-0.0178	0.2633	0.0019	-0.0456	-0.2953	0.1646												
s14	high	0.1521	0.0974	0.3085	0.2622	0.0504	0.1048													
s15	low	-0.0481	0.0841	-0.1488	0.0296	-0.1340	0.0620	0.0051												
s15	medium	0.1093	0.0968	0.0700	0.0826															
s15	high	0.2335	0.1278	0.0940	0.1680	0.1887	0.0130													
s16	low	-0.0301	0.0753	-0.2093	0.0431	0.0147														
s16	medium	0.1933	-0.0319	0.1536	0.0320	0.1390	0.1059													
s16	high	0.2337	0.0349	0.1714	0.3313	0.2088	-0.3060													
s17	low	0.3987	0.2318	0.3346																
s17	medium	0.4141	-0.0548	0.1569	0.1274	0.1093	-0.0342													
s17	high	0.3151	0.5366	0.2758	0.2048															

Level	n	MD	SD
low	72	0.0125	0.1421
medium	124	0.0833	0.1728
high	106	0.1840	0.1617

Summary by group

```
Datos %>%
  group_by(Level) %>%
  summarise(n=n(),MD=mean(MPDC),SD=sd(MPDC)) %>%
  kable() %>%
  kable_styling(latex_options = c("striped"))
```

```
(q <-Datos %>% ggplot(aes(x=Level,y=MPDC)) +
    geom_point() + facet_wrap(~ Subject)+
    labs(x="Difficulty level")+theme_bw()+
    stat_summary(fun="mean", geom="point",color="red"))
```



Random Intercept and Slope Model

The following model is used to investigate whether there are significant differences between the study variables:

$$y_{ij} = \mu + l_k + s_j + (sl)_{jk} + \epsilon_{ij}, \tag{1}$$

where y_{ij} is the response variable (MPDC) for the i-th observation from the j-th subject, μ is the intercept, l_k is the k-th difficulty level, s_j is the jth subject effect, $(sl)_{jk}$ is the subject-level effect, i.e., the k-th level effect at the j-th subject, ϵ_{ij} is the error term (residual) for the ith observation from the jth subject.

We called *level* l a fixed effect, and ϵ is our *error term* that represent deviations from our predictions due to *random* factors that we cannot control experimentally. However, several measurements were taken for each subject at each difficulty level and that violates the assumption of independence of a linear model. On the other hand, each individual has a different cognitive load capacity, and this will be a characteristic factor that will affect all the responses of the same subject, which will make these responses interdependent instead of independent, see figure **??**. The way we approaches this situation is adding a random effect to the subject and to the subject-level interaction. This allows us to solve this lack of independence by assuming a different intercept and slope for each subject. And finally, we assume that the residual, subject and subject-level effects are all relations of separate distributions, all with zero means:

$$\begin{aligned} \epsilon_{ij} &\sim N(0, \sigma^2), \\ s_j &\sim N(0, \sigma_s^2), \\ (sl)_{jk} &\sim N(0, \sigma_{sl}^2). \end{aligned}$$

Hence, s_j and $(sl)_{jk}$ are now random effects, and μ and l_k are fixed effects.

Using the **R** notation the model is

$$MPDC = (b_0 + u_{Subject}) + b_{Level} Level + \epsilon$$

In order to evaluate if there is an effect due to the difficulty level we will use the likelihood ratio test of the model with the *Level* effect against the model without the *Level* effect.

```
MPDC_mixed_reducido <- lme4::lmer(MPDC ~ 1 + (1+Level|Subject),data=Datos,REML=F)

MPDC_mixed_lme4 <- lme4::lmer(MPDC ~ Level + (1+Level|Subject),data=Datos,REML=F)

anova(MPDC_mixed_reducido,MPDC_mixed_lme4)
```

```
## Data: Datos
## Models:
## MPDC_mixed_reducido: MPDC ~ 1 + (1 + Level | Subject)
## MPDC_mixed_lme4: MPDC ~ Level + (1 + Level | Subject)
##                     npar  AIC  BIC logLik deviance Chisq Df  Pr(>Chisq)
## MPDC_mixed_reducido    8 -222 -192    119     -238
## MPDC_mixed_lme4       10 -251 -214    135     -271  32.6  2 0.000000082 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value of the ratio test is significant at a level of 0.001.

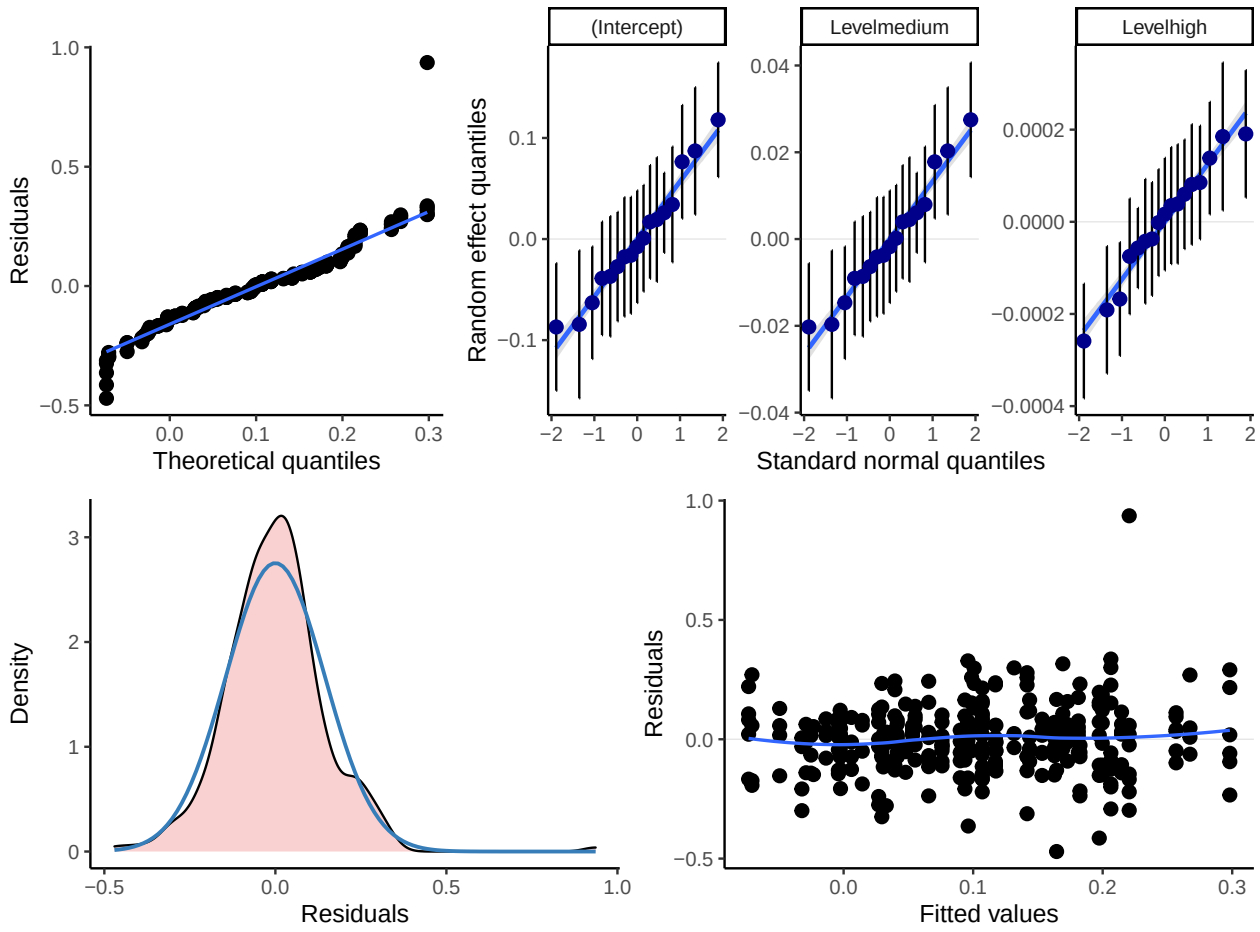
```
MPDC_mixed_lme4 <- lme4::lmer(MPDC ~ Level + (1+Level|Subject),data=Datos)
summary(MPDC_mixed_lme4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: MPDC ~ Level + (1 + Level | Subject)
##    Data: Datos
```

```
##
## REML criterion at convergence: -252.7
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -3.167 -0.588  0.037  0.502  6.304
##
## Random effects:
##  Groups   Name        Variance     Std.Dev. Corr
##  Subject  (Intercept) 0.0041705315 0.064580
##           Levelmedium 0.0002263320 0.015044  1.00
##           Levelhigh   0.0000000201 0.000142 -1.00 -1.00
##  Residual             0.0220592648 0.148524
## Number of obs: 302, groups:  Subject, 17
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   0.0136     0.0236    0.58
## Levelmedium   0.0614     0.0226    2.71
## Levelhigh     0.1668     0.0228    7.30
##
## Correlation of Fixed Effects:
##             (Intr) Lvlmdm
## Levelmedium -0.473
## Levelhigh   -0.576  0.607
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
p<-plot_model(MPDC_mixed_lme4, type = "diag")

({p[[1]]+theme(plot.title=element_blank(),plot.subtitle=element_blank())+scale_x_continuous(name="Theor
```




```
contr <- glht(MPDC_mixed_lme4,linfct=mcp(Level="Tukey"))
summary(contr, test = adjusted("holm"))
```

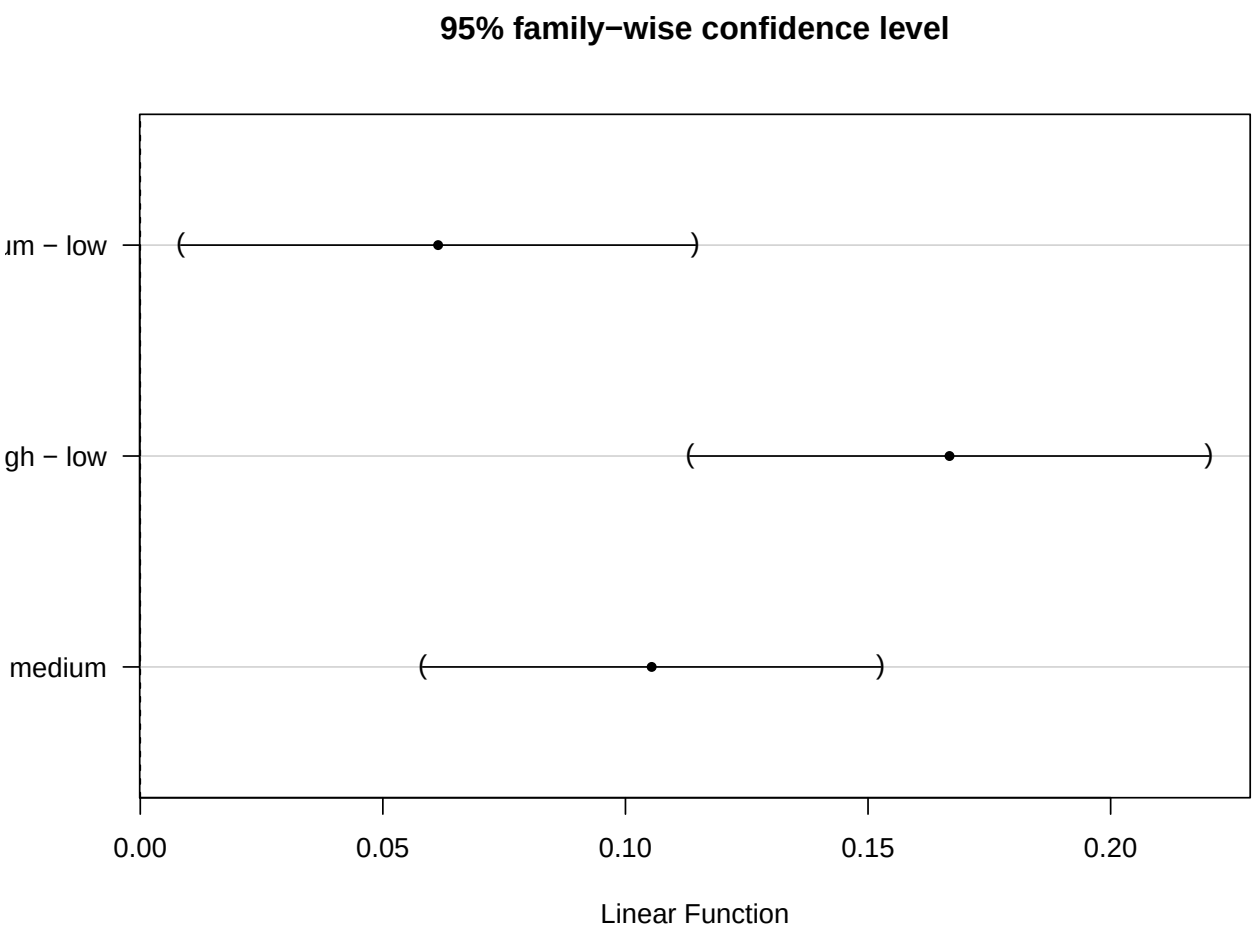
```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme4::lmer(formula = MPDC ~ Level + (1 + Level | Subject), data = Datos)
##
## Linear Hypotheses:
##                     Estimate Std. Error z value         Pr(>|z|)
## medium - low == 0    0.0614     0.0226    2.71           0.0067 **
## high - low == 0      0.1668     0.0228    7.30 0.00000000000085 ***
## high - medium == 0   0.1054     0.0202    5.23 0.00000033795045 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- holm method)
```

```
confint(contr)
```

```
##
##   Simultaneous Confidence Intervals
##
```

```
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme4::lmer(formula = MPDC ~ Level + (1 + Level | Subject), data = Datos)
##
## Quantile = 2.342
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##                      Estimate lwr     upr
## medium - low == 0  0.06139  0.00836 0.11442
## high - low == 0    0.16679  0.11330 0.22029
## high - medium == 0 0.10540  0.05821 0.15260
```

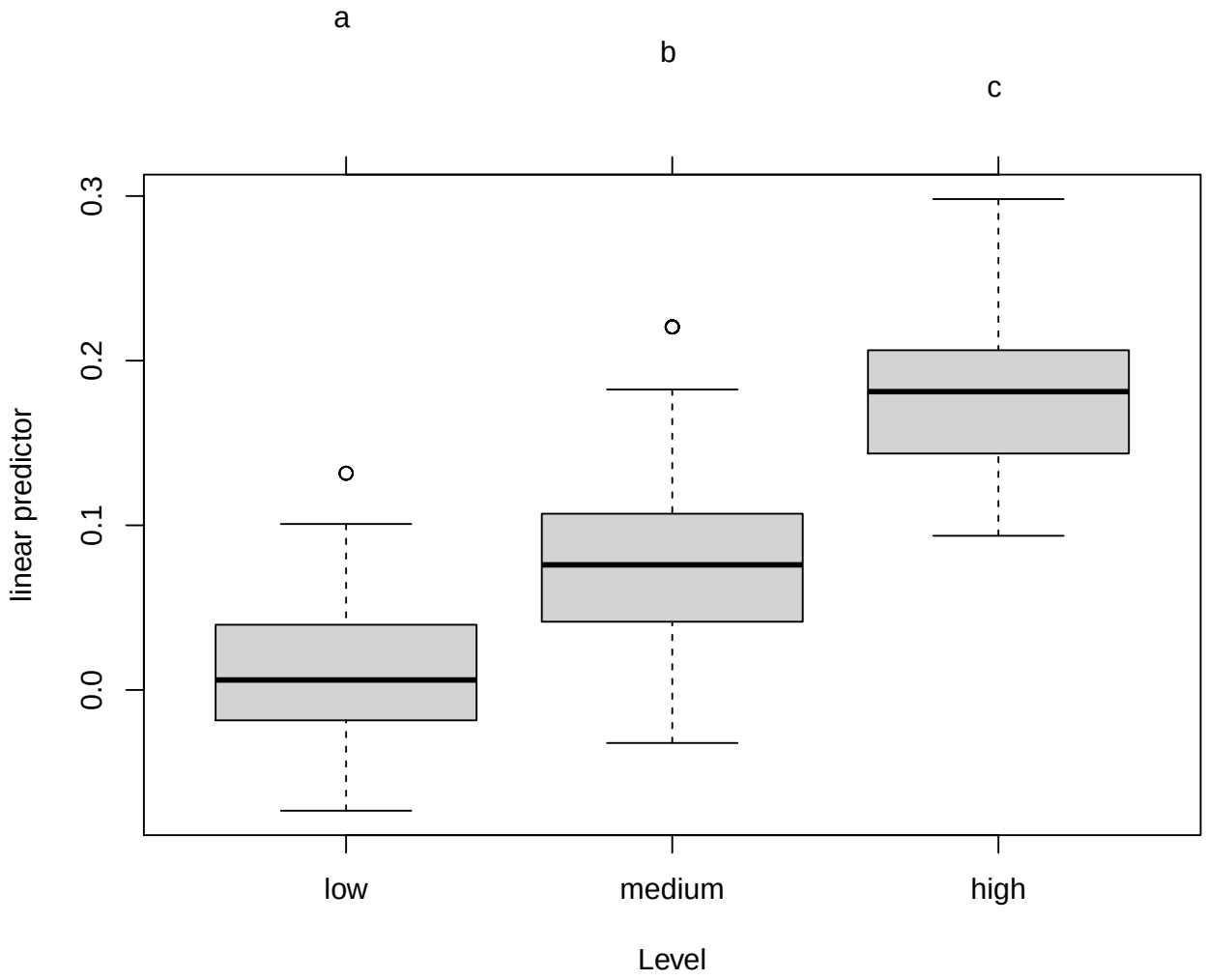
```
plot(confint(contr))
```




```
contr.cld <- cld(contr)
old.par <- par(mai=c(1,1,1.25,1), no.readonly = TRUE)
plot(contr.cld)
```

Subject	Name	Training	Nivel	BLPS	MPDC	APCPS	PD	Entropy	TTP	PDS	SequenceMemory	SMN	id	Level	res	fit
s10	SequenceMemory_r24	FALSE	3	2.744	1.1568	0.4215	1.6368	-1.974	7176806	0	r24	24	14	medium	0.9363	0.2205
s11	SequenceMemory_r22	FALSE	6	3.611	0.5424	0.1502	0.8607	-2.056	6139927	0	r22	22		high	0.3360	0.2063
s03	SequenceMemory_r21	FALSE	3	3.775	0.4244	0.1124	0.6202	-2.071	3588223	0	r21	21	12	medium	0.3285	0.0960
s09	SequenceMemory_r02	FALSE	3	3.637	0.4854	0.1335	0.9587	-2.046	5641562	0	r02	2	1	medium	0.3161	0.1693
s11	SequenceMemory_r10	FALSE	6	3.505	0.5068	0.1446	0.9528	-2.008	7335892	0	r10	10	11	high	0.3005	0.2063
s10	SequenceMemory_r05	FALSE	1	3.083	0.4308	0.1397	0.8144	-1.820	8153300	0	r05	5	5	low	0.2992	0.1316

Subject	Name	Training	Nivel	BLPS	MPDC	APCPS	PD	Entropy	TTP	PDS	SequenceMemory	SMN	id	Level	res	fit
s13	SequenceMemory_r15	FALSE	3	4.255	-0.3314	-0.0779	-0.0439	-1.973	2033376	0	r15	15	8	medium	-0.2991	-0.0322
s02	SequenceMemory_r30	FALSE	6	4.359	-0.1703	-0.0391	0.1220	-2.067	7794586	0	r30	30	17	high	-0.3120	0.1416
s14	SequenceMemory_r25	FALSE	3	3.697	-0.2953	-0.0799	0.0867	-1.767	6080251	0	r25	25	12	medium	-0.3248	0.0295
s01	SequenceMemory_r17	FALSE	6	4.023	-0.2674	-0.0665	0.1099	-1.911	2292544	0	r17	17	7	high	-0.3636	0.0962
s03	SequenceMemory_r03	FALSE	6	4.688	-0.2167	-0.0462	0.2240	-2.162	8432543	0	r03	3	1	high	-0.4141	0.1974
s16	SequenceMemory_r28	FALSE	6	4.849	-0.3060	-0.0631	-0.0414	-2.184	14432973	0	r28	28	15	high	-0.4704	0.1644




```
par(old.par)
```

```
Datos2 = Datos
Datos2$res = residuals(MPDC_mixed_lme4,type="pearson")
Datos2$fit = fitted(MPDC_mixed_lme4,type="pearson")

Datos2 %>% arrange(desc(res)) %>% head() %>% kable() %>%
  kable_styling(latex_options = c("striped", "scale_down"))
```

```
Datos2 %>% arrange(desc(res)) %>% tail() %>% kable() %>%
  kable_styling(latex_options = c("striped", "scale_down"))
```

```r
shapiro.test(Datos2$res)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Datos2$res
## W = 0.95, p-value = 0.00000002
```

```r
goftest::ad.test(Datos2$res,null="pnorm",mean=mean(Datos2$res), sd=sd(Datos2$res), estimated=TRUE)
```

```
##
##  Anderson-Darling test of goodness-of-fit
##  Braun's adjustment using 17 groups
##  Null hypothesis: Normal distribution
##  with parameters mean = 0.000000000000000006172011589524l1, sd =
##  0.14486365280981
##  Parameters assumed to have been estimated from data
##
## data:  Datos2$res
## Anmax = 3.2, p-value = 0.3
```

```r
rstatix::levene_test(data=ungroup(Datos2),res~Level)
```

```
## # A tibble: 1 x 4
##     df1   df2 statistic     p
##   <int> <int>     <dbl> <dbl>
## 1     2   299     0.552 0.576
```

The same model without the outlier

We repeat the analysis without the outlier

```r
# we exclude the outlier
Datos <- Datos %>% filter(!(Subject=="s10"&SMN==24))
```

```r
MPDC_mixed_lme4 <- lme4::lmer(MPDC ~ Level + (1+Level|Subject),data=Datos)
summary(MPDC_mixed_lme4)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: MPDC ~ Level + (1 + Level | Subject)
##    Data: Datos
##
## REML criterion at convergence: -297.5
##
## Scaled residuals:
##    Min     1Q Median     3Q    Max
## -3.425 -0.604  0.051  0.544  2.421
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Subject  (Intercept) 0.005351 0.0731
##           Levelmedium 0.003135 0.0560   -0.41
##           Levelhigh   0.000637 0.0252   -0.43  1.00
##  Residual             0.018510 0.1361
## Number of obs: 301, groups:  Subject, 17
```

```
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   0.0144     0.0241    0.60
## Levelmedium   0.0537     0.0248    2.16
## Levelhigh     0.1659     0.0219    7.57
##
## Correlation of Fixed Effects:
##             (Intr) Lvlmdm
## Levelmedium -0.610
## Levelhigh   -0.591  0.648
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular
```

```
anova(MPDC_mixed_lme4)
```

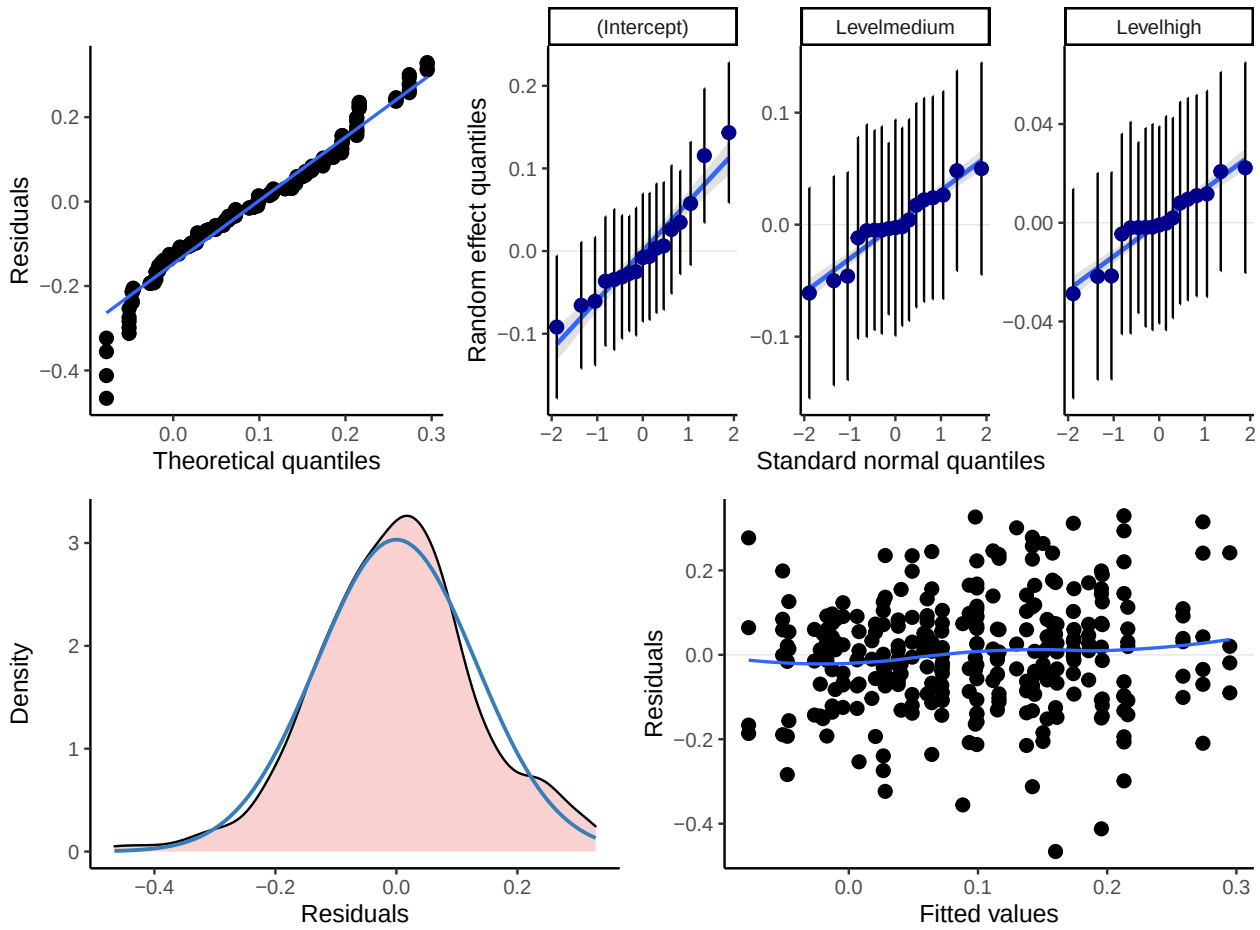
```
## Analysis of Variance Table
##       npar Sum Sq Mean Sq F value
## Level    2    1.3   0.651    35.1
```

```
coef(MPDC_mixed_lme4)
```

```
## $Subject
##     (Intercept) Levelmedium Levelhigh
## s01   -0.077515    0.050947    0.1658
## s02   -0.022066    0.048793    0.1642
## s03    0.017883    0.080068    0.1776
## s04    0.007974    0.103661    0.1882
## s05    0.040502    0.075793    0.1755
## s06   -0.010492    0.048766    0.1640
## s07    0.020634    0.051770    0.1650
## s08   -0.046355    0.041875    0.1614
## s09    0.071979    0.101828    0.1867
## s10    0.129896    0.007680    0.1442
## s11    0.049075    0.050054    0.1639
## s12    0.006411    0.057843    0.1679
## s13   -0.051164    0.003590    0.1444
## s14   -0.020043    0.048254    0.1639
## s15   -0.012833    0.071036    0.1740
## s16   -0.016916    0.077626    0.1769
## s17    0.157713   -0.007391    0.1372
##
## attr(,"class")
## [1] "coef.mer"
```

```
p<-plot_model(MPDC_mixed_lme4, type = "diag")

(q<-{p[[1]]+theme(plot.title=element_blank(),plot.subtitle=element_blank())+scale_x_continuous(name="Th
```




```
# muy importante Tukey para lme4.
contr <- glht(MPDC_mixed_lme4,linfct=mcp(Level="Tukey"))
summary(contr, test = adjusted("holm"))
```

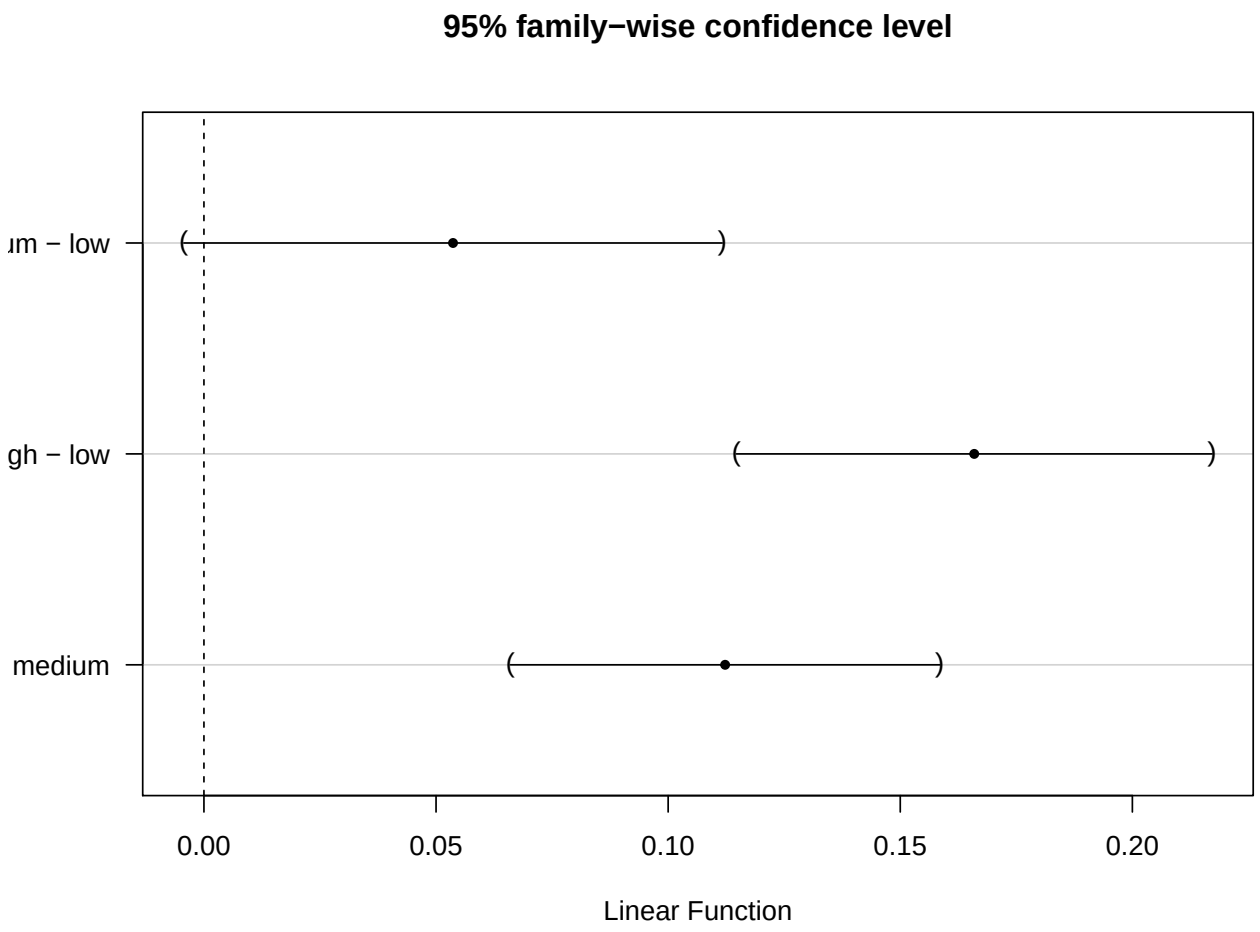
```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme4::lmer(formula = MPDC ~ Level + (1 + Level | Subject), data = Datos)
##
## Linear Hypotheses:
##                     Estimate Std. Error z value         Pr(>|z|)
## medium - low == 0    0.0537     0.0248    2.16             0.03 *
## high - low == 0      0.1659     0.0219    7.57 0.00000000000011 ***
## high - medium == 0   0.1123     0.0198    5.68 0.00000002746644 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- holm method)
```

```
confint(contr)
```

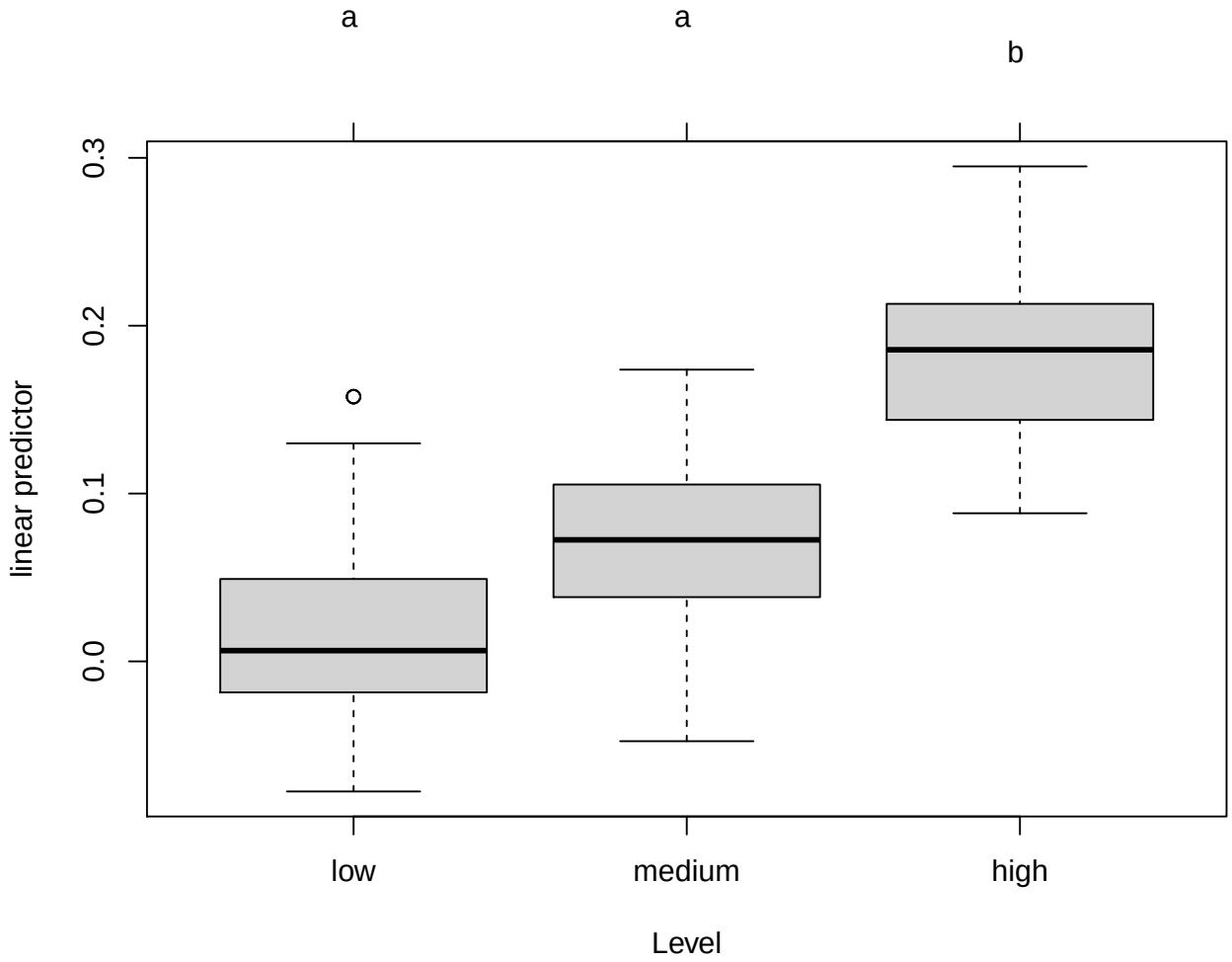
```
##
##   Simultaneous Confidence Intervals
```

```
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lme4::lmer(formula = MPDC ~ Level + (1 + Level | Subject), data = Datos)
##
## Quantile = 2.338
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##                     Estimate lwr      upr
## medium - low == 0   0.05366 -0.00433  0.11165
## high - low == 0     0.16593  0.11470  0.21717
## high - medium == 0  0.11227  0.06603  0.15852
```

```
plot(confint(contr))
```




```
contr.cld <- cld(contr)
### use sufficiently large upper margin
old.par <- par(mai=c(1,1,1.25,1), no.readonly = TRUE)
### plot
plot(contr.cld)
```




```
par(old.par)
```

```
Datos2=Datos
Datos2$res = residuals(MPDC_mixed_lme4,type="pearson")
Datos2$fit = fitted(MPDC_mixed_lme4,type="pearson")
shapiro.test(Datos2$res)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Datos2$res
## W = 0.99, p-value = 0.07
```

```
goftest::ad.test(Datos2$res,null="pnorm",mean=mean(Datos2$res), sd=sd(Datos2$res), estimated=TRUE)
```

```
##
##  Anderson-Darling test of goodness-of-fit
##  Braun's adjustment using 17 groups
##  Null hypothesis: Normal distribution
##  with parameters mean = 0.00000000000000009392430362989 85, sd =
##  0.131513289694428
##  Parameters assumed to have been estimated from data
##
## data:  Datos2$res
```

```
## Anmax = 1.5, p-value = 1
```

```
rstatix::levene_test(data=ungroup(Datos2),res~Level)
```

```
## # A tibble: 1 x 4
##     df1   df2 statistic     p
##   <int> <int>     <dbl> <dbl>
## 1     2   298      1.76 0.175
```

Non parametric tests

```
kruskal.test(MPDC ~ Level, data=Datos)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  MPDC by Level
## Kruskal-Wallis chi-squared = 59, df = 2, p-value = 0.0000000000002
```

```
PMCMR::posthoc.kruskal.nemenyi.test(data=Datos,MPDC~Level, dist="Tukey")
```

```
##
##  Pairwise comparisons using Tukey and Kramer (Nemenyi) test
##                    with Tukey-Dist approximation for independent samples
##
## data:  MPDC by Level
##
##        low              medium
## medium 0.011            -
## high   0.00000000000049 0.00000037308780
##
## P value adjustment method: none
```

```
PMCMRplus::tukeyTest(data=Datos,MPDC~Level)
```

```
##        low             medium
## medium 0.015           -
## high   0.0000000000031 0.0000002277642
```