Solución del Taller

1. Estimar los **efectos** de cada uno de los factores **principales** y los efectos de **interacción doble**. Grafique los efectos estimados en un gráfico de probabilidad normal, úselo para seleccionar un modelo tentativo.

   Solución:

   $\mathbf{Y}$: Resistividad de la Oblea hallada en los diferentes niveles de los **cuatro** factores considerados: **A, B, C, D**.

   $\mathbf{k} = \mathbf{4}$: Número de factores. $\mathbf{n} = \mathbf{1}$: Número de réplicas.

   Recuerde que el **estimador del efecto** de cada Factor es: El **contraste** asociado al Factor dividido por $\mathbf{n} * \mathbf{2^{k-1}} = \mathbf{1} \times \mathbf{2^3} = \mathbf{8}$.

Y	1,92	11,28	1,09	5,75	2,13	9,53	1,03	5,35	1,6	11,73	1,16	4,68	2,16	9,11	1,07	5,3
Signos de A	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1
	-1,92	11,28	-1,09	5,75	-2,13	9,53	-1,03	5,35	-1,6	11,73	-1,16	4,68	-2,16	9,11	-1,07	5,3
Contraste de A	50,57															
Efecto estimado de A	6,32125															

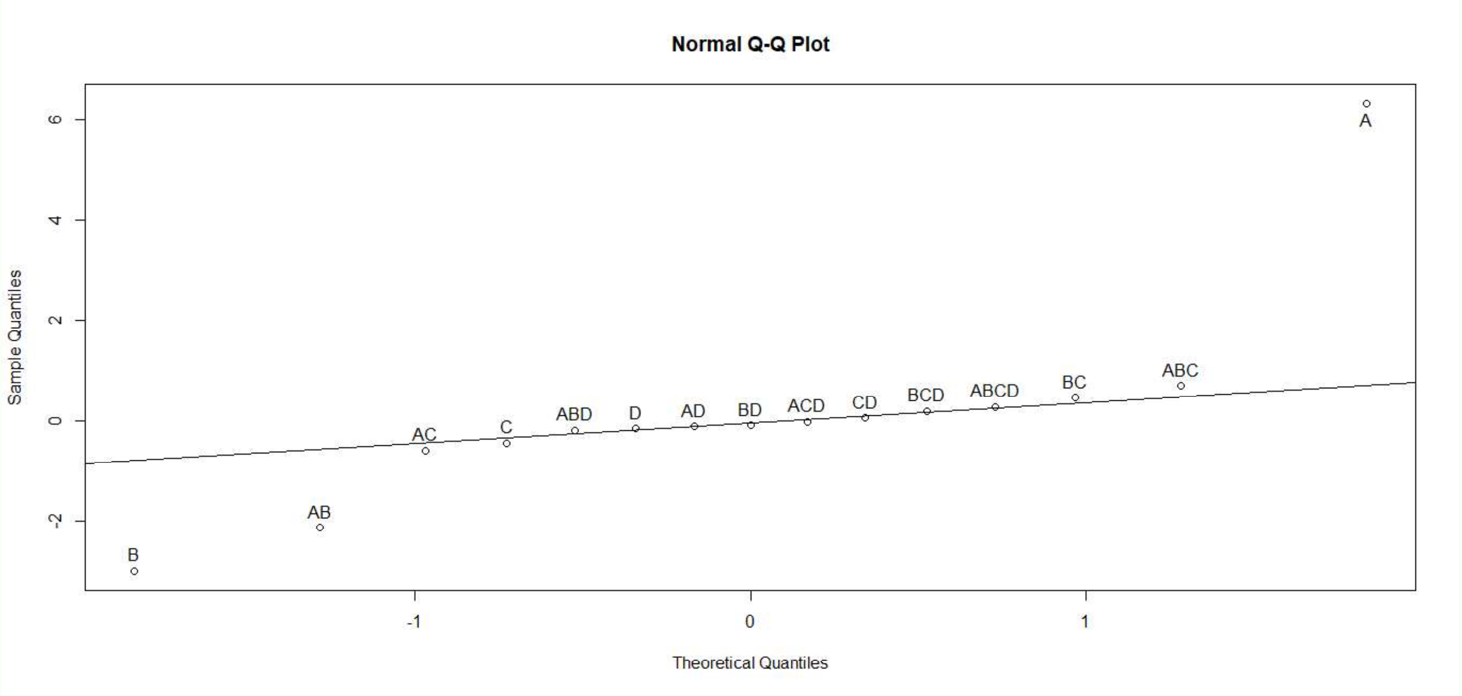
En forma análoga se hallan los otros **efectos**, a partir de la multiplicación de la **TABLA** de signos por el respectivo valor de $\mathbf{Y}$ y luego sumarlos y dividirlos por $\mathbf{8}$:

Resistividad	A	B	C	D	AB	AC	AD	BC	BD	CD	ABC	ABD	BCD	ACD	ABCD
1,92	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1
11,28	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	-1	1	-1
1,09	-1	1	-1	-1	-1	1	1	-1	-1	1	1	1	1	-1	-1
5,75	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	1
2,13	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1
9,53	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	1	-1	1
1,03	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	-1	1	1
5,35	1	1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	-1
1,6	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	1	1	1	-1
11,73	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	1	-1	1
1,16	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	-1	1	1
4,68	1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	-1
2,16	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	1
9,11	1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	-1	1	-1
1,07	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	1	-1	-1
5,3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

También, los estimadores de los **efectos** se pueden hallar en $\mathbf{R}$, a partir de la función $\mathbf{lm}$

```
datos<- matrix( c( 1.92,-1,-1,-1,-1, 11.28,1,-1,-1,-1, 1.09,-1,1,-1,-1,
5.75,1,1,-1,-1, 2.13,-1,-1,1,-1, 9.53,1,-1,1,-1, 1.03,-1,1,1,-1, 5.35,1,1,1,-1,
1.6,-1,-1,-1,1, 11.73,1,-1,-1,1, 1.16,-1,1,-1,1, 4.68,1,1,-1,1, 2.16,-1,-1,1,1,
9.11,1,-1,1,1, 1.07,-1,1,1,1, 5.3,1,1,1,1 ), ncol=5, byrow=T )

y      <-  datos[,1]; A <- datos[,2]; B <- datos[,3]; C <- datos[,4];D <- datos[,5]

mod1   <-  lm(y~A*B*C*D)

summary(mod1)

coef   <- mod1$coefficients
efectos <- 2*mod1$coefficients
efectos <-efectos[-1]
efectos
#    A        B        C        D       A:B       A:C       B:C       A:D       B:D
# 6.32125 -3.00375 -0.44125 -0.15875 -2.13875 -0.59625  0.45875 -0.11375 -0.09375
#    C:D     A:B:C    A:B:D    A:C:D     B:C:D     A:B:C:D
# 0.05875  0.68875 -0.19375 -0.02125   0.18875    0.28375
```

```
identify(qqnorm(efectos[1:14]),labels=c("A","B","C","D","AB","AC","BC","AD","BD",
        "CD","ABC","ABD","ACD","BCD","ABCD"))
qqline(efectos)
```



Modelo Tentativo:

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, \; i = 1, 2, \; j = 1, 2, \; k = 1, 2, 3, 4, \; \varepsilon_{ijk} \sim NID(0, \sigma^2). \quad (3)$$

2. Ajuste el modelo obtenido en la parte **1**. y haga el análisis de **residuales** para la verificación de los supuestos. ¿Hay indicios de que el modelo **NO** es adecuado?

```
y <- datos[,1]; A <- as.factor(datos[,2]);
B <- as.factor(datos[,3]); C <- as.factor(datos[,4])
modelo1<-aov(y~A*B)
summary(modelo1)
####Salida del R#############################################################
Analysis of Variance Table
Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
A          1 159.833 159.833 333.088 4.049e-10 ***
B          1  36.090  36.090  75.211 1.630e-06 ***
A:B        1  18.297  18.297  38.130 4.763e-05 ***
Residuals 12   5.758   0.480
```

NOTA: Observe que tanto el efecto de interacción **AB** como el de los efectos principales **A** y **B** son todos altamente **significativos**, a un nivel de significancia del $5\,\%$.

Verificación de los supuestos del Modelo:

Normalidad:

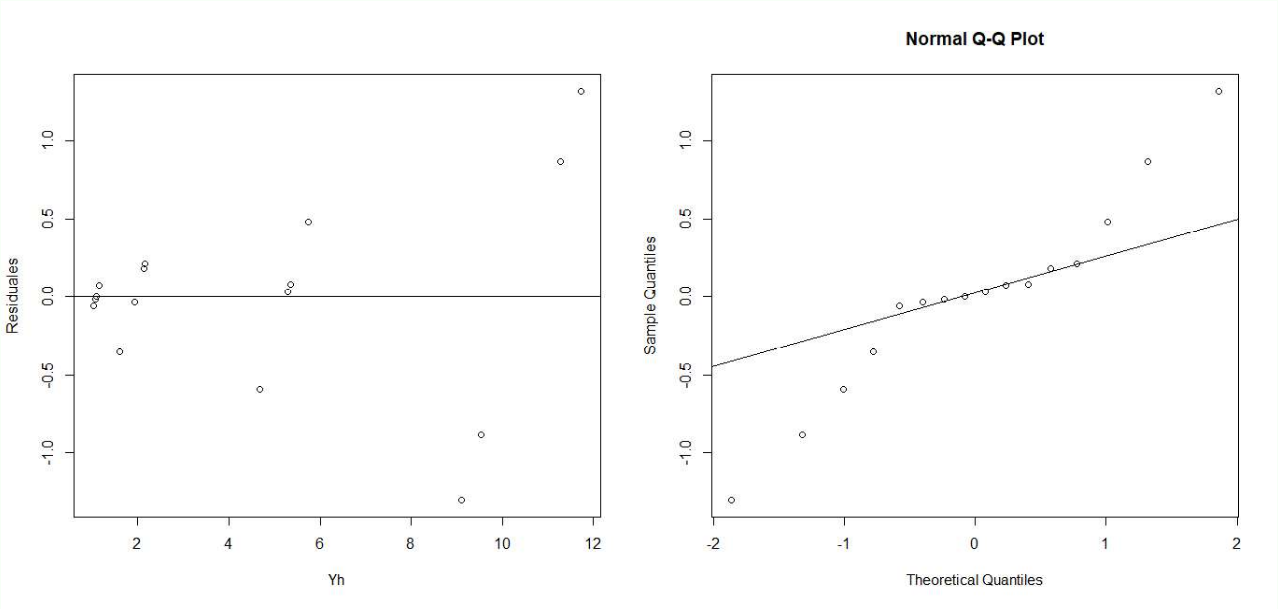
Shapiro-Wilk normality test

data: res

W = 0.94285, p-value = 0.3854

```
##########################################################################
res<-mod1$residuals
shapiro.test(res)
##########################################################################
```

```
par(mfrow=c(1,2))
plot(y,res,ylab="Residuales",xlab="Yh")
abline(h=0)
qqnorm(res)
qqline(res)
```



Note que al parecer el supuesto de **homogeneidad de varianza** **NO** se cumple, existe tendencia en forma de parlante.
En cuanto a la normalidad, este supuesto se verifica al usar la prueba de Shapiro-Wilks a un nivel de significancia del $5\,\%$.

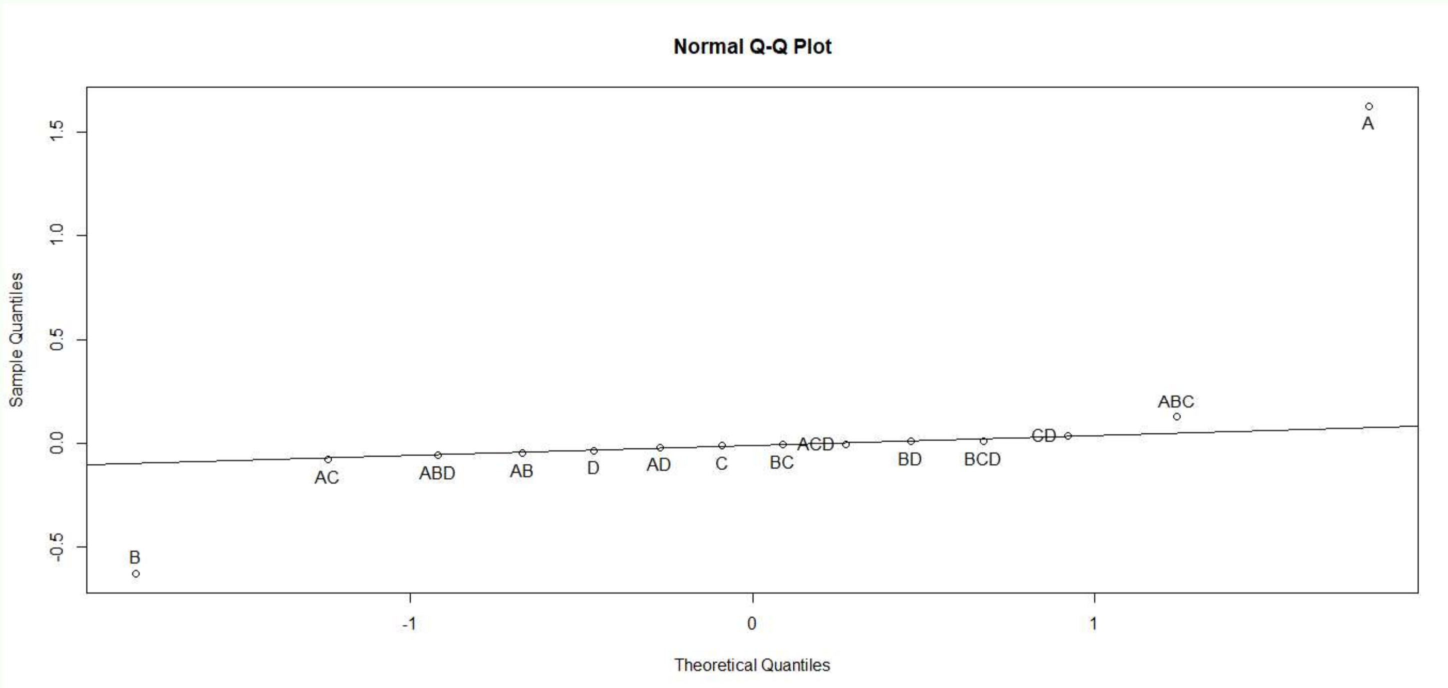
```
bartlett.test(res,A)
#        Bartlett test of homogeneity of variances
#data:  res and A
#Bartlett's K-squared = 12.857, df = 1,
#p-value = 0.0003363

bartlett.test(res,B)
#        Bartlett test of homogeneity of variances
#data:  res and B
#Bartlett's K-squared = 6.4881, df = 1,
#p-value = 0.01086
```

Conclusión Se obtiene una conclusión similar con la prueba de **Bartlett**, para cada factor individual **NO** hay **homogeneidad de varianza** en los errores del modelo.

...

3. Repita el análisis de las partes 1 y 2 con $\ln(\mathbf{y})$ como variable respuesta. ¿Existen indicios de que la transformación ha sido apropiada?

```
y<-log(y)
A <- datos[,2]; B <- datos[,3]; C <- datos[,4];D <- datos[,5]
mod1   <-  lm(y~A*B*C*D);summary(mod1)
coef   <- mod1$coefficients
efectos <- 2*mod1$coefficients
efectos <-efectos[-1]
efectos
########################################
identify(qqnorm(efectos[1:14]),labels=c("A","B","C","D","AB",
        "AC","BC","AD", "BD", "CD","ABC","ABD","ACD","BCD","ABCD"))
qqline(efectos)
```



Observe que sólo los factores principales $\mathbf{A}$ y $\mathbf{B}$ tienen efectos importantes en la resistividad de la Oblea (en escala logarítmica).

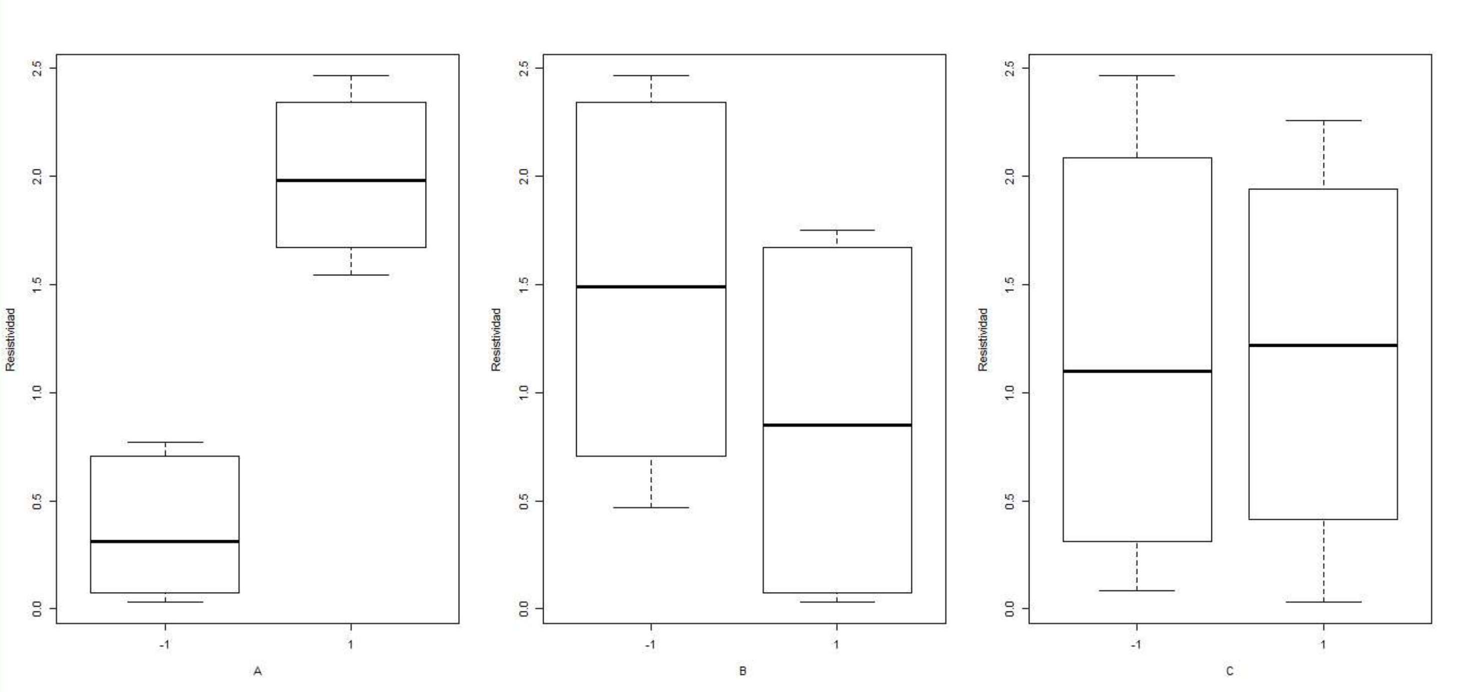
```
mod2<-aov(y~A*B)
summary(mod2)
             Df Sum Sq Mean Sq F value   Pr(>F)
A             1 10.572  10.572  954.04 8.33e-13 ***
B             1  1.580   1.580  142.61 5.10e-08 ***
A:B           1  0.010   0.010    0.88    0.367
Residuals    12  0.133   0.011
```

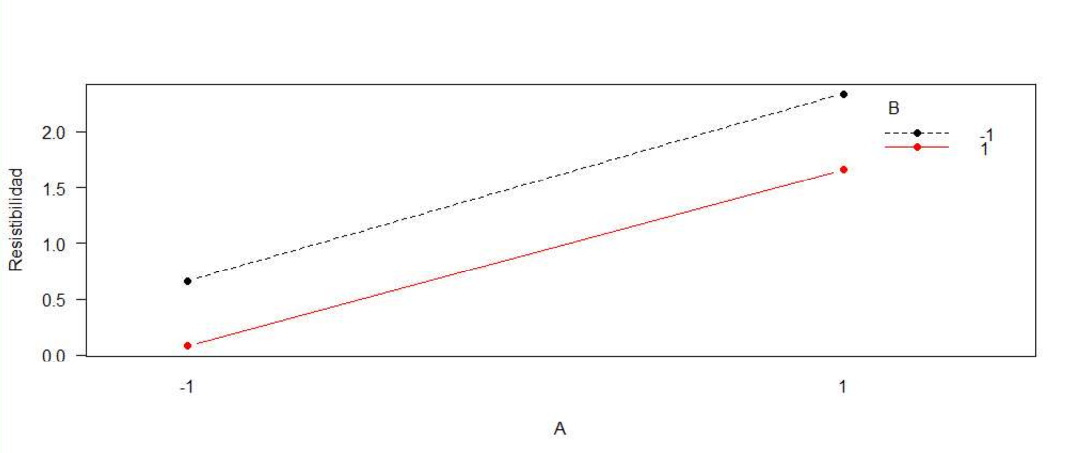
Note que la interacción entre los factores $\mathbf{A}$ y $\mathbf{B}$ no es significativa, a un nivel de significancia del $\mathbf{5}\%$.

Verificación de los Supuestos:

```
###Modelo sin Interacción######
mod3 <- aov(y~A+B)
summary(mod3)
res    <-  mod3$residuals
shapiro.test(res)
####   Shapiro-Wilk normality test
####   data:  res
####   W = 0.91145, p-value = 0.1228
bartlett.test(res,A)
#        Bartlett test of homogeneity of variances
#        data:  res and A
#        Bartlett's K-squared = 0.33025, df = 1, p-value = 0.5655
bartlett.test(res,B)
#        Bartlett test of homogeneity of variances
#        data:  res and B
#        Bartlett's K-squared = 0.24004, df = 1, p-value = 0.6242
```

```
par(mfrow=c(1,3))
boxplot(split(y,A),xlab='A',ylab='Resistividad')
boxplot(split(y,B),xlab='B',ylab='Resistividad')
boxplot(split(y,C),xlab='C',ylab='Resistividad')
```





4. Ajuste un modelo en términos de las variables codificadas que puede ser usado para predecir la resistividad.

```
A <- datos[,2]; B <- datos[,3]; C <- datos[,4];D <- datos[,5];y<-log(y)
modelo4<-lm(y~A+ B)
summary(modelo4)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.1854     0.0262   45.25 1.09e-15 ***
A             0.8129     0.0262   31.03 1.41e-13 ***
B            -0.3143     0.0262  -12.00 2.09e-08 ***
```

Modelo Ajustado:

$$\hat{\log}(\mathbf{y}) = \hat{\beta}_0 + \hat{\beta}_1\,\mathbf{A} + \hat{\beta}_2\,\mathbf{B}$$

$$\hat{\log}(\mathbf{y}) = \mathbf{1.1854} + \mathbf{0.8129\,A} - \mathbf{0.3143\,B},\ \ \mathbf{A} = \pm\mathbf{1},\ \mathbf{B} = \pm\mathbf{1}$$

5. Suponga que el experimentador ha realizado cuatro corridas al centro junto con las 16 corridas experimentales anteriores. Las medidas de resistividad en el centro fueron: $\mathbf{8.15}$, $\mathbf{7.63}$, $\mathbf{8.95}$ y $\mathbf{6.48}$. Analice el experimento otra vez con los puntos centrales.
¿Qué conclusiones se pueden obtener?

- **Ptos al Centro**: $\mathbf{8.15,\ 7.63,\ 8.9,\ 6.48}$ y $\mathbf{n_c = 4}$.

- **Media** en los puntos centrales usando el $\log$:
  $\mathbf{y_c^* = \log(y)}$ $\mathbf{\bar{y}_c^* = 2.04762}$.
- Media cuadrática del Error para probar **curvatura**:

$$\mathbf{MS_E = \frac{SS_E}{n_c - 1} = \frac{\sum_{\text{puntos centrales}} (y_i^* - \bar{y}_c^*)^2}{4 - 1} = 0.0185.}$$

- **Promedio** en los puntos factoriales: $\mathbf{\bar{y}_f^* = 1.1854}$
- Suma de cuadrados asociada a la **curvatura:**

$$\mathbf{SS_{curvatura} = \frac{n_f n_c \left(\bar{y}_f^* - \bar{y}_c^*\right)^2}{n_f + n_c} = \frac{16 * 4 * (1.1854 - 2.04762)^2}{16 + 4} = 2.3788}$$

Modelo:

$$\log(\mathbf{Y}) = \beta_0 + \sum_{\mathbf{j=1}}^{\mathbf{2}} \beta_{\mathbf{j}} \mathbf{x_j} + \sum\sum_{\mathbf{i<j}} \beta_{\mathbf{ij}} \mathbf{x_i x_j} + \sum_{\mathbf{j=1}}^{\mathbf{2}} \beta_{\mathbf{jj}} \mathbf{x_j^2} + \varepsilon,$$

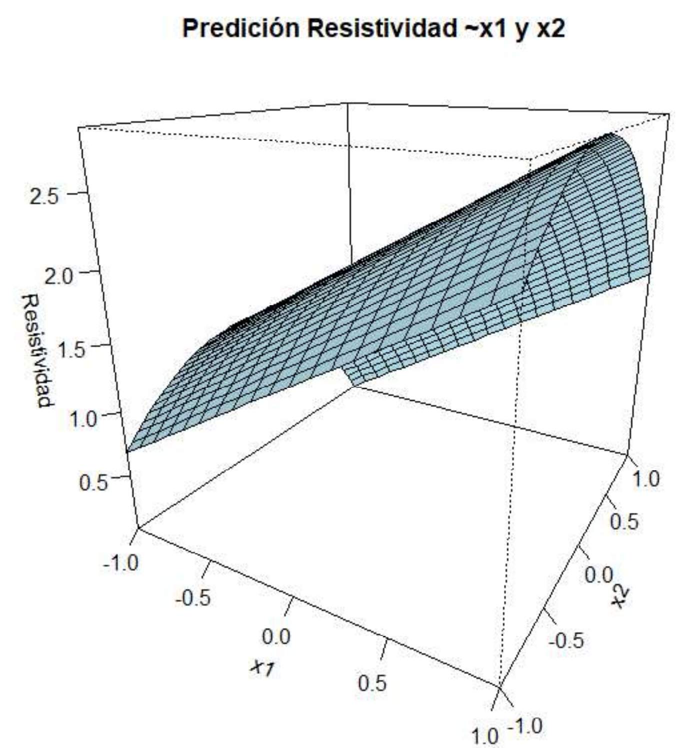
Hipótesis para curvatura:

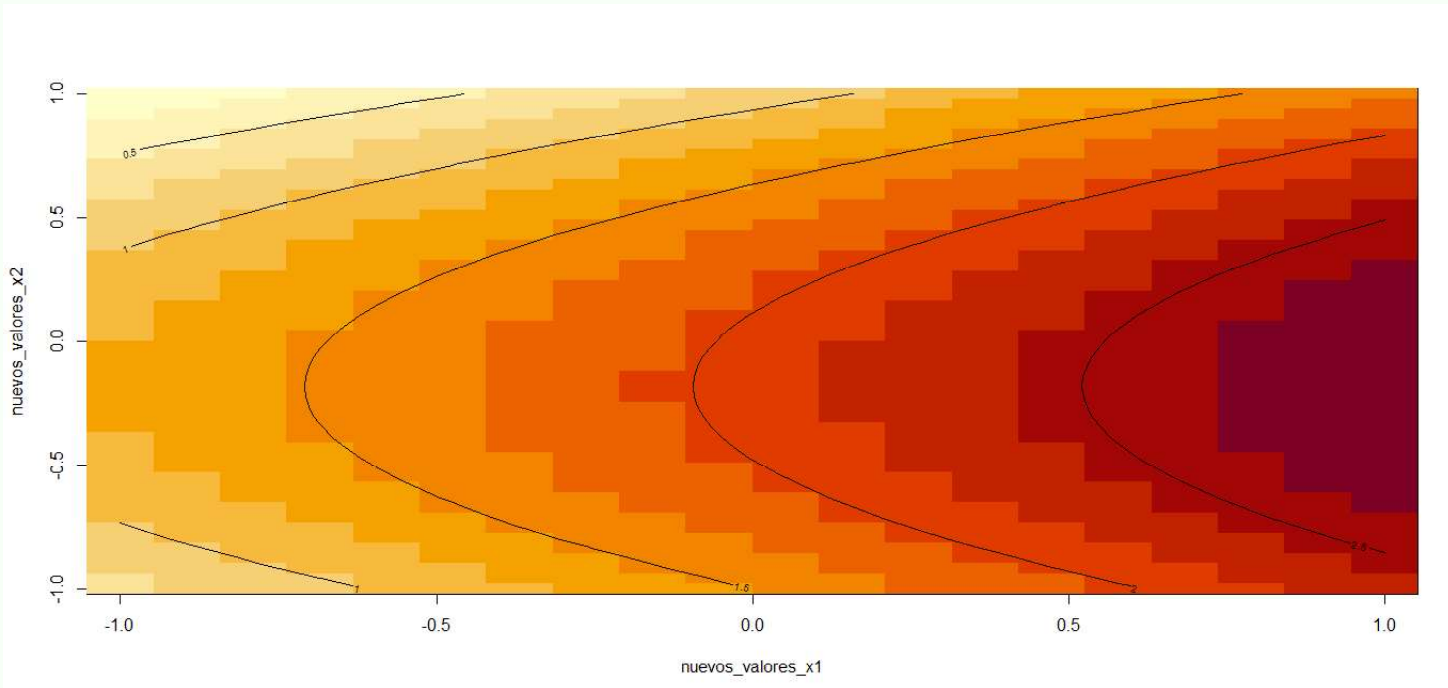
$$\mathbf{H_0} \; : \; \beta_{11} + \beta_{22} = \mathbf{0}$$

Estadístico de Prueba:

$$\mathbf{F} = \frac{\mathbf{SS_{curvatura}}}{\mathbf{MSE}} \sim \mathbf{F_{1,\, n_c - 1}} = \mathbf{F_{1,\, 3}}$$

Además $\mathbf{F_{cal}} = \frac{\mathbf{2.3788}}{\mathbf{0.0185}} = \mathbf{128.58}$, y como $\mathbf{F_{0.05,1,3}} = \mathbf{qf(0.95, 1, 3)} = \mathbf{10.12}$ se rechaza $\mathbf{H_0}$ indicando, con un nivel de significancia del $\mathbf{5}\,\%$ que existe **curvatura** en la superficie esperada.


Predición Resistividad ~x1 y x2
Resistividad
2.5
2.0
1.5
1.0
0.5
x1
-1.0
-0.5
0.0
0.5
1.0
x2
-1.0
-0.5
0.0
0.5
1.0



nuevos_valores_x2
nuevos_valores_x1
-1.0
-0.5
0.0
0.5
1.0