

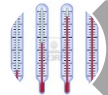
Sesión III

Diseño completamente aleatorio

Diseño de experimentos



Identificación del problema



Elección de factores, niveles y rangos



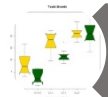
Selección de la variable respuesta



Selección del diseño experimental



Realización del experimento



Análisis estadístico de los datos



Conclusiones y recomendaciones

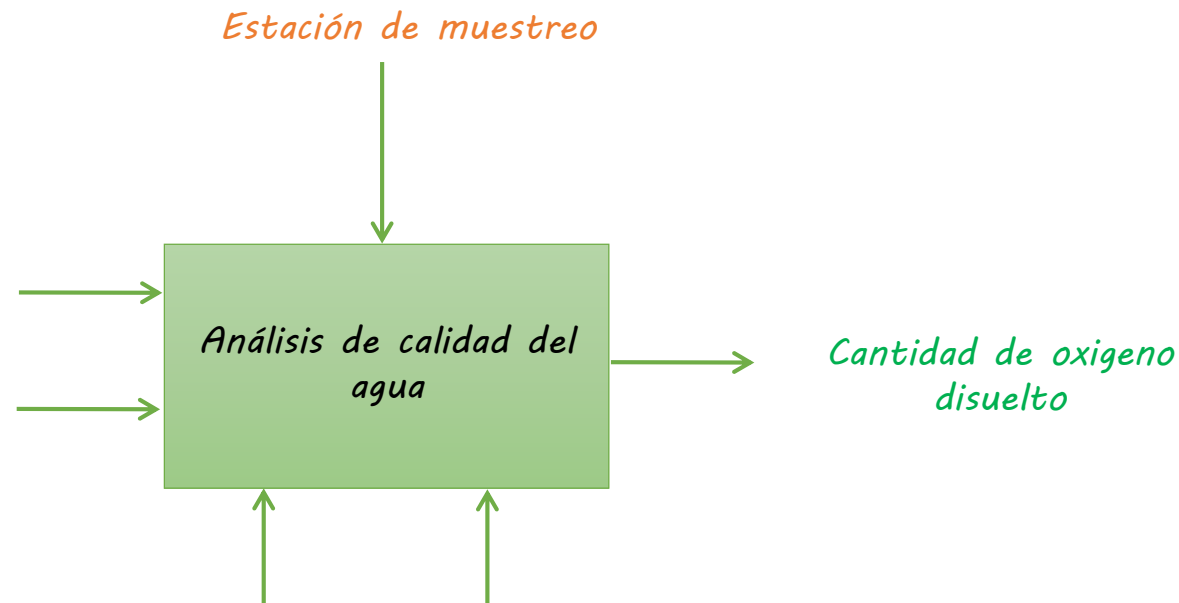
Impacto de la industria manufacturera en la calidad del agua del rio



Para estudiar el nivel de contaminación de un río, se recogen muestras de agua en cuatro estaciones distintas del río y se estudia la cantidad de oxígeno existente en el agua recogida en cada uno de ellos. Las dos primeras estaciones se escogen de manera que el río no ha pasado por una planta industrial existente en la zona. La tercera y cuarta estación se eligen después de pasar el río por la planta industrial, de manera que la tercera esta próximo a ésta y la cuarta más alejada. En cada estación se seleccionan cinco muestras de agua. Los datos vienen dados en la siguiente tabla

Estación	Contenido de oxígeno disuelto				
1	6.1	5.9	6.3	6.1	6.0
2	6.6	6.3	6.4	6.4	6.5
3	4.3	4.8	5.0	4.7	5.1
4	6.2	6.0	6.1	5.8	5.9

Objetivo: Investigar los efectos de la estación de recolección de una muestra de agua en la calidad del agua de un río.



$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, i = 1, 2, \dots, a \text{ y } j = 1, 2, \dots, n, \varepsilon_{ij} \sim N(0, \sigma^2), \varepsilon_{ij} \text{ ind}$$

Unidad experimental: Volumen de agua

Número de replicas



Estación de
medición 1



Estación de
medición 2



Estación de
medición 3



Estación de
medición 4

Lectura de datos

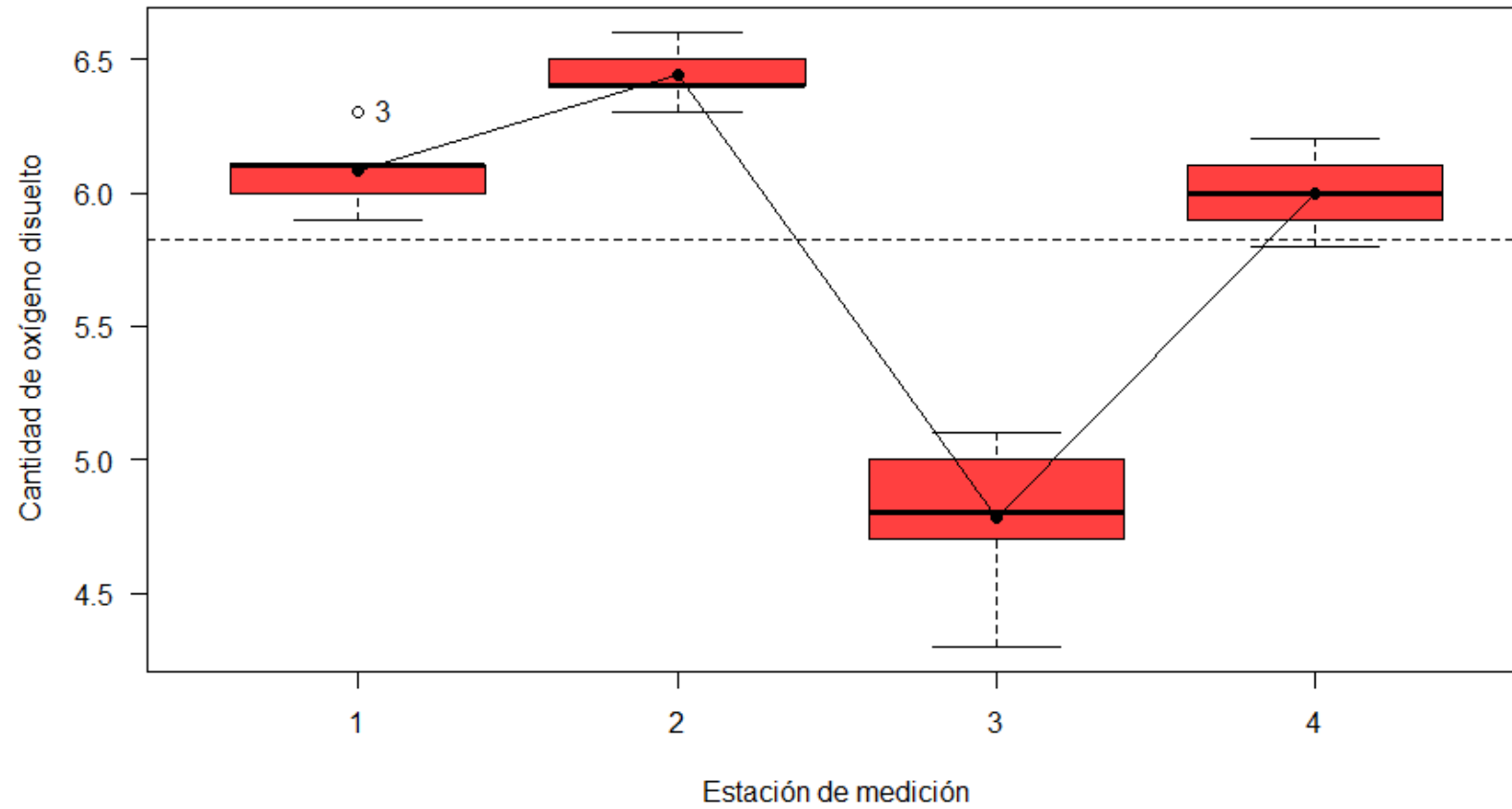
```
datos <- read.table(file="https://tinyurl.com/y9pmra95",  
                    header=T, sep="\t", dec=".")
```

En R

```
estacion.f <- factor(datos$estacion)
medias <- tapply(datos$oxigeno, estacion.f, mean)

require(car)
Boxplot(datos$oxigeno ~ estacion.f, xlab="Estación de medición",
        ylab="Cantidad de oxígeno disuelto",
        las=1, col="brown1")
points(medias, col="black", pch=16)
lines(medias, col="black", pch=16)
abline(h=mean(datos$oxigeno), col="black", lty=2)
```

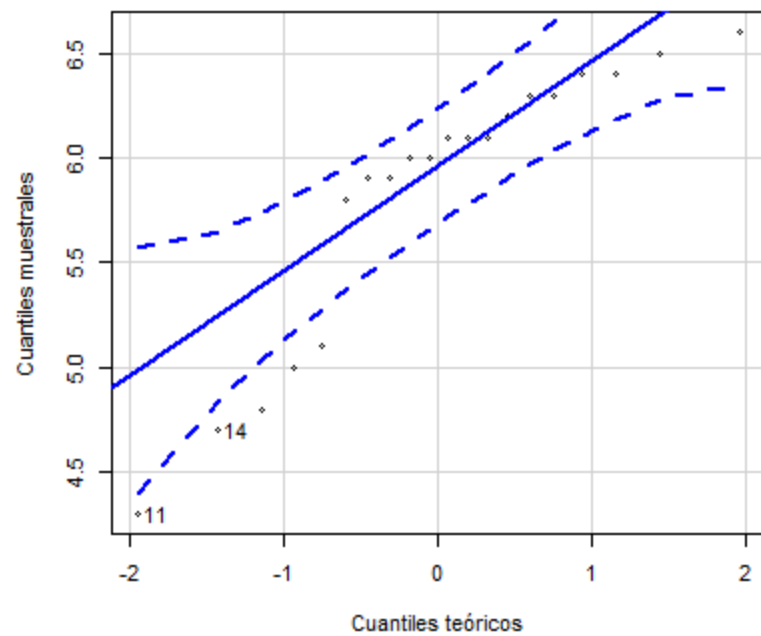

Boxplot y gráfico de medias



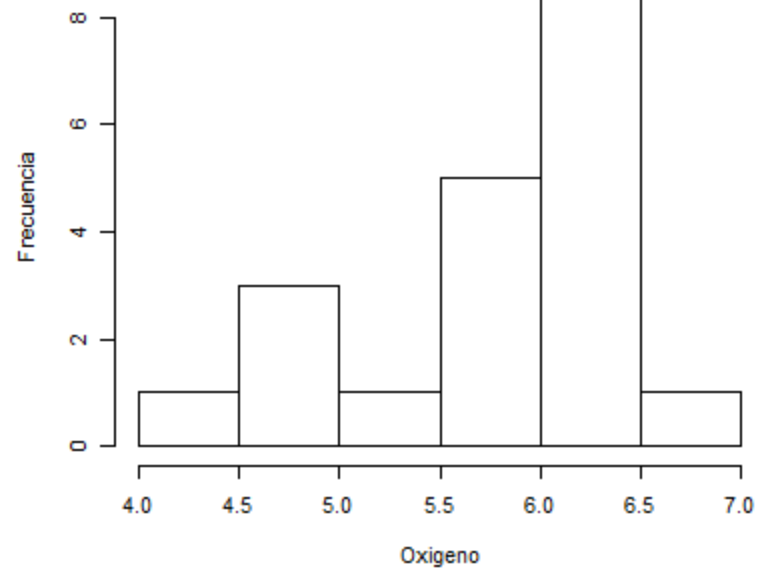
En R

```
require(car)
par(mfrow=c(1,3))
qqPlot(datos$oxigeno, xlab="Cuantiles teóricos", ylab="Cuantiles muestrales", main="Gráfico cuantil-cuantil")
hist(datos$oxigeno, xlab="Oxigeno", ylab="Frecuencia", main="Histograma")
boxplot(datos$oxigeno, xlab="Oxigeno", main="Boxplot")
```

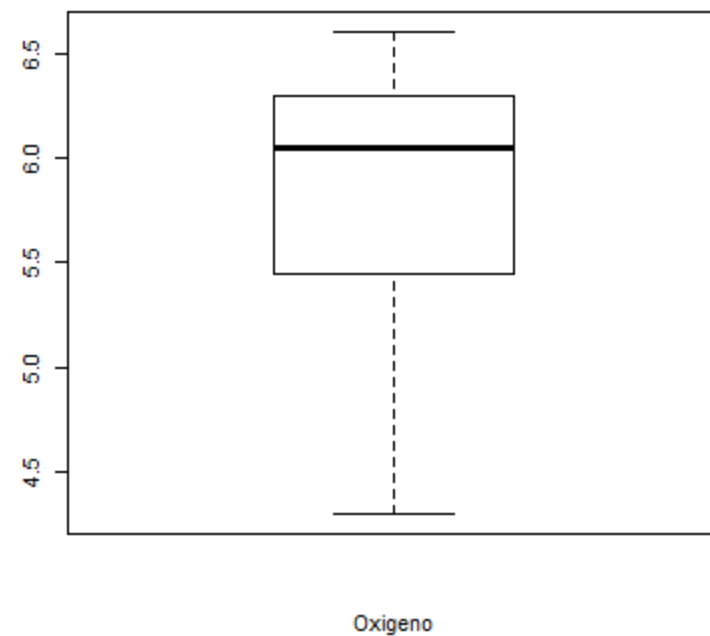
Gráfico cuantil-cuantil



Histograma



Boxplot



En R

```
Modelo<- lm(datos$oxigeno ~ estacion.f)  
summary(Modelo)
```

En R

Call:

```
lm(formula = datos$oxigeno ~ estacion.f)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.480	-0.085	0.010	0.115	0.320

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	6.0800	0.0886	68.623	< 2e-16	***
estacion.f2	0.3600	0.1253	2.873	0.011	*
estacion.f3	-1.3000	0.1253	-10.375	1.64e-08	***
estacion.f4	-0.0800	0.1253	-0.638	0.532	

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

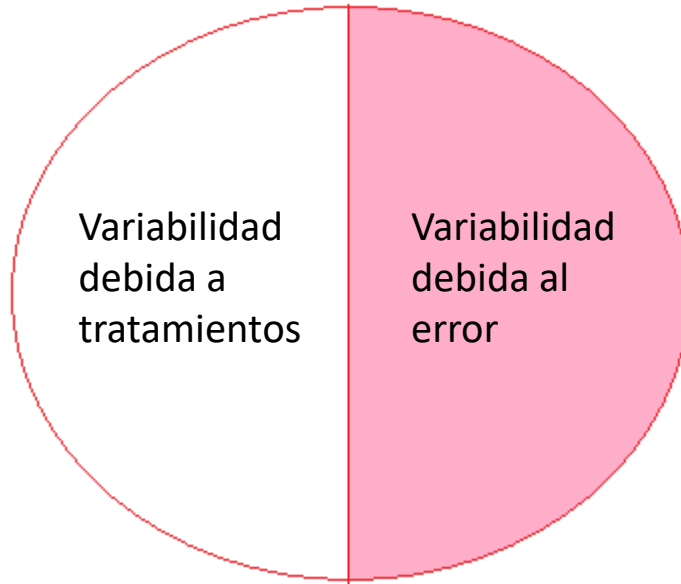
Residual standard error: 0.1981 on 16 degrees of freedom

Multiple R-squared: 0.9257, Adjusted R-squared: 0.9118

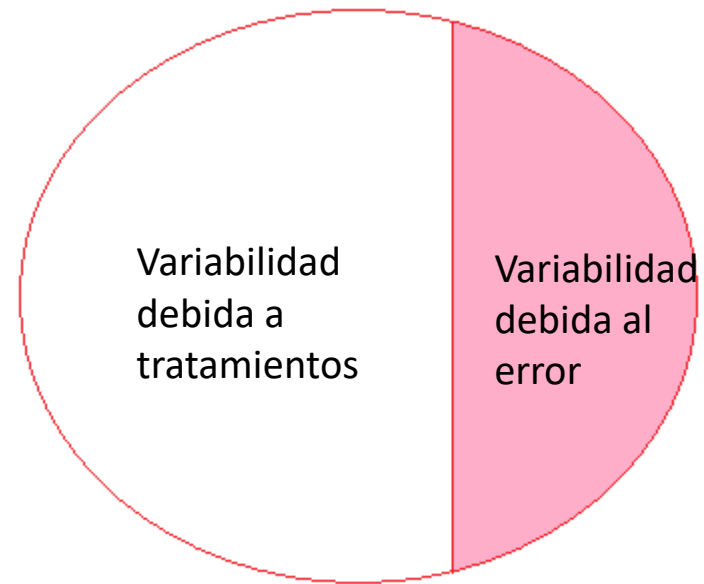
F-statistic: 66.49 on 3 and 16 DF, p-value: 2.982e-09

Análisis de varianza

No hay efecto de tratamientos



Hay efecto de tratamientos



En R

```
anova(Modelo)
```

```
Analysis of Variance Table
```

```
Response: datos$oxigeno
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
estacion.f	3	7.8295	2.60983	66.493	2.982e-09 ***
Residuals	16	0.6280	0.03925		

```
---
```

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

En R

```
Modelo1<-aov(datos$oxigeno ~ estacion.f)  
summary(Modelo1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
estacion.f	3	7.830	2.6098	66.49	2.98e-09 ***
Residuals	16	0.628	0.0392		

En R

```
##Comparación del Oxígeno y Estación
```

```
#Boxplot comparativo entre Oxígeno y Estación
```

```
par(mfrow=c(1,2))
```

```
Boxplot(datos$oxigeno ~ estacion.f, xlab="Estación de medición",  
        ylab="Cantidad de oxígeno disuelto", main="Valores reales",  
        col="brown1", las=2)
```

```
abline(h=mean(datos$oxigeno), col=1, lty=1)
```

```
##Comparación de Oxígeno promedio y factor
```

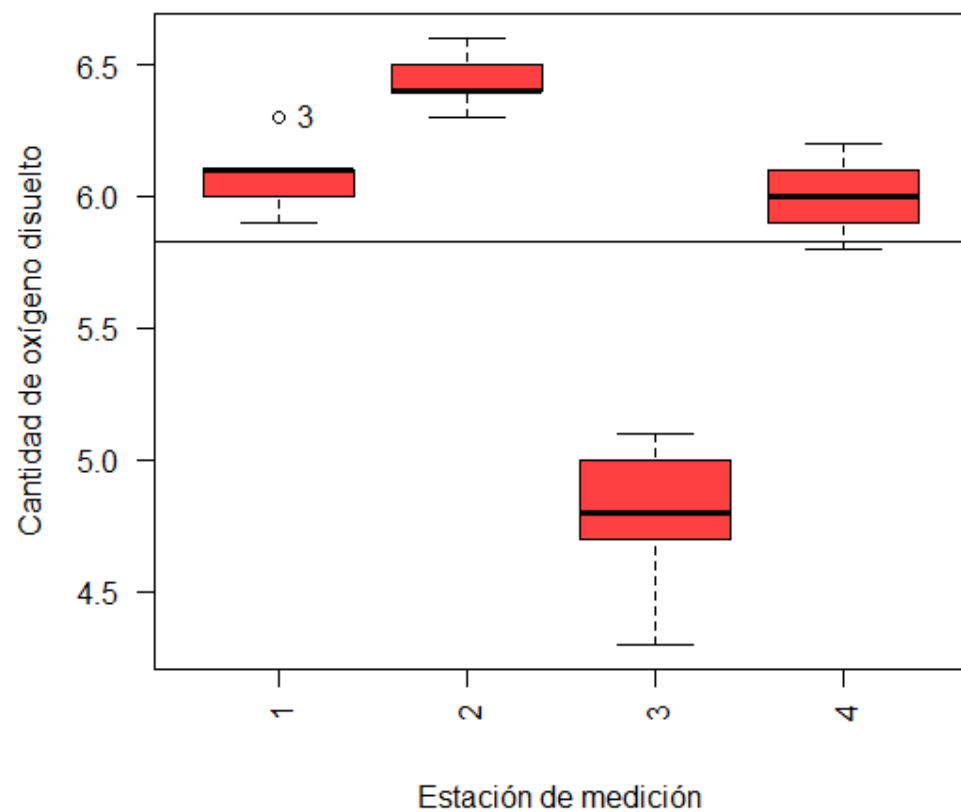
```
#Boxplot comparativo Oxígeno promedio y Estación
```

```
Boxplot(Modelo$fit ~ estación.f, xlab="Estación de medición",  
        ylab="Cantidad de oxígeno disuelto promedio",  
        main="Valores ajustados", las=2)
```

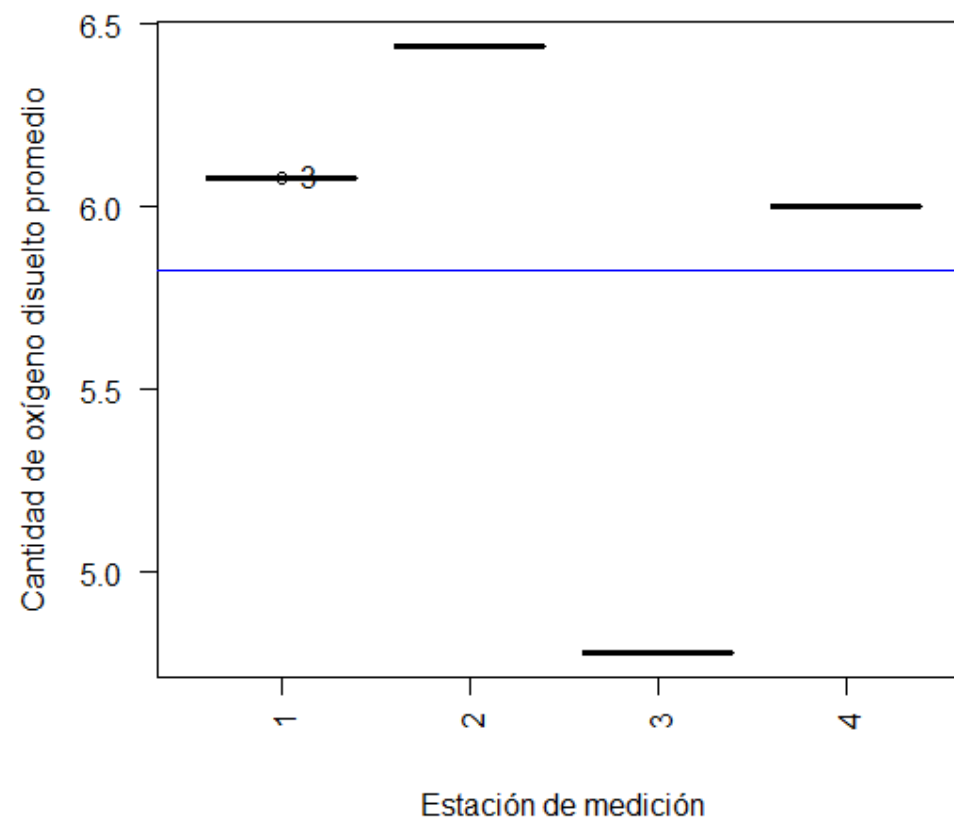
```
abline(h=mean(datos$oxigeno), col=4, lty=1)
```

En R

Valores reales



Valores ajustados



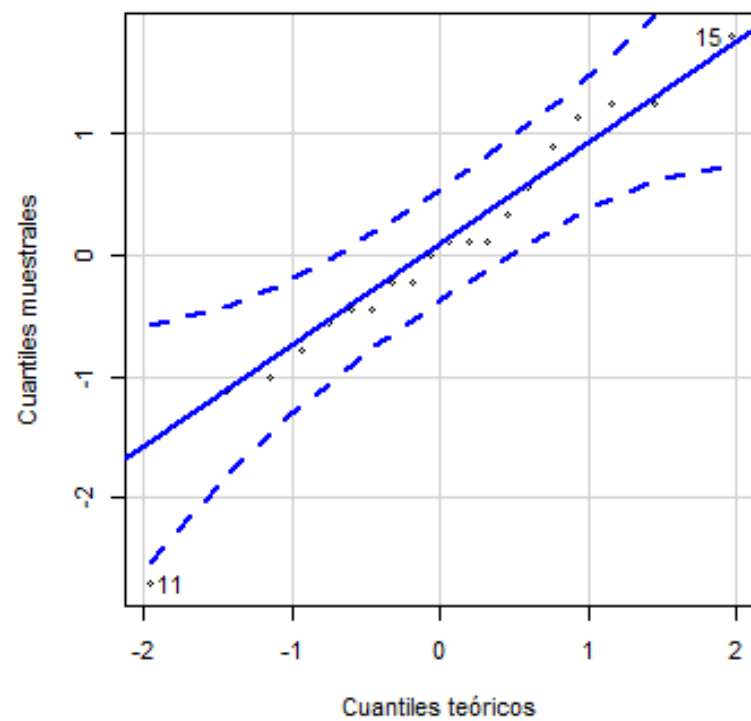
Validación del modelo en R

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, i = 1, 2, \dots, a \text{ y } j = 1, 2, \dots, n, \varepsilon_{ij} \sim N(0, \sigma^2), \varepsilon_{ij} \text{ ind}$$

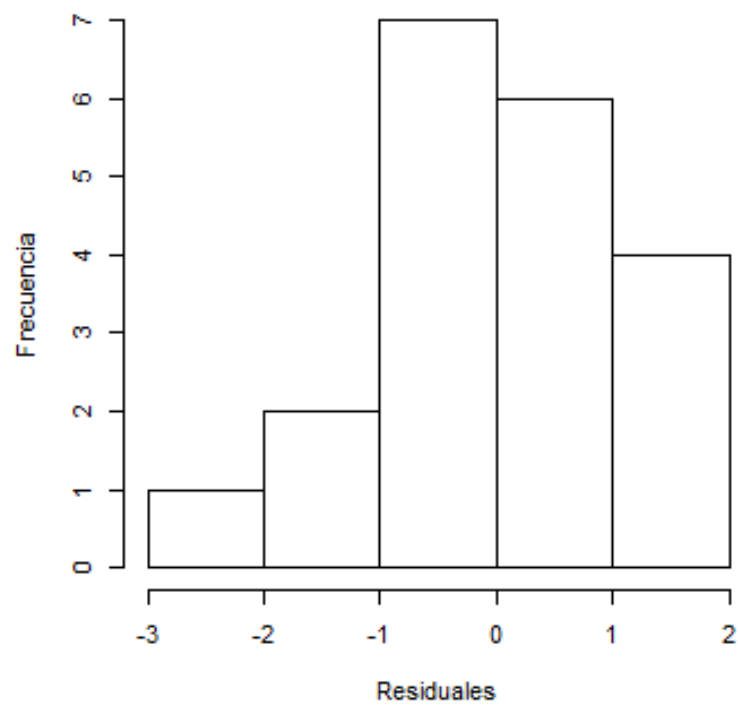
```
require(car)
par(mfrow=c(1,3))
residuales <- rstandard(Modelo)
qqPlot(residuales, xlab="Cuantiles teóricos", ylab="Cuantiles muestrales",
       main="Gráfico cuantil-cuantil")
hist(residuales, xlab="Residuales", ylab="Frecuencia", main="Histograma")
Boxplot(residuales, ylab="Residuales", main="Boxplot")
```

En R

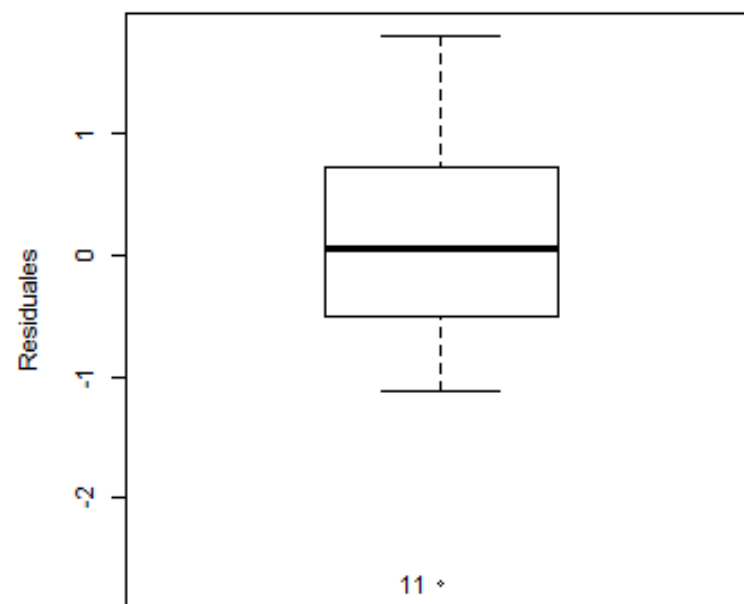
Gráfico cuantil-cuantil



Histograma



Boxplot



En R

```
require(nortest)
shapiro.test(residuales)
ad.test(residuales)
```

Shapiro-Wilk normality test

```
data:  residuales
W = 0.95915, p-value = 0.527
```

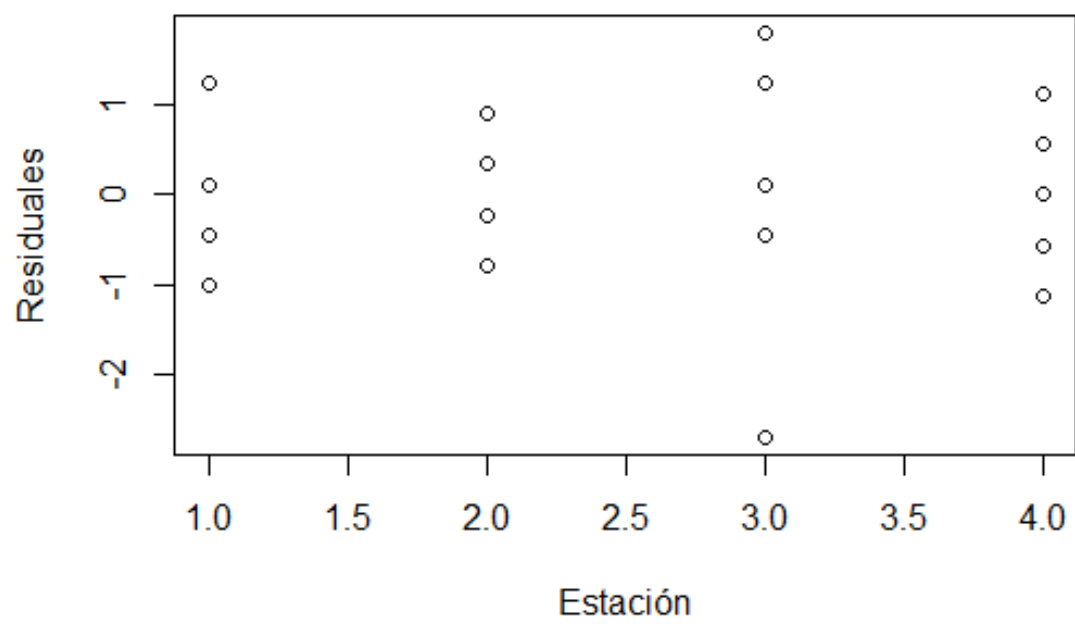
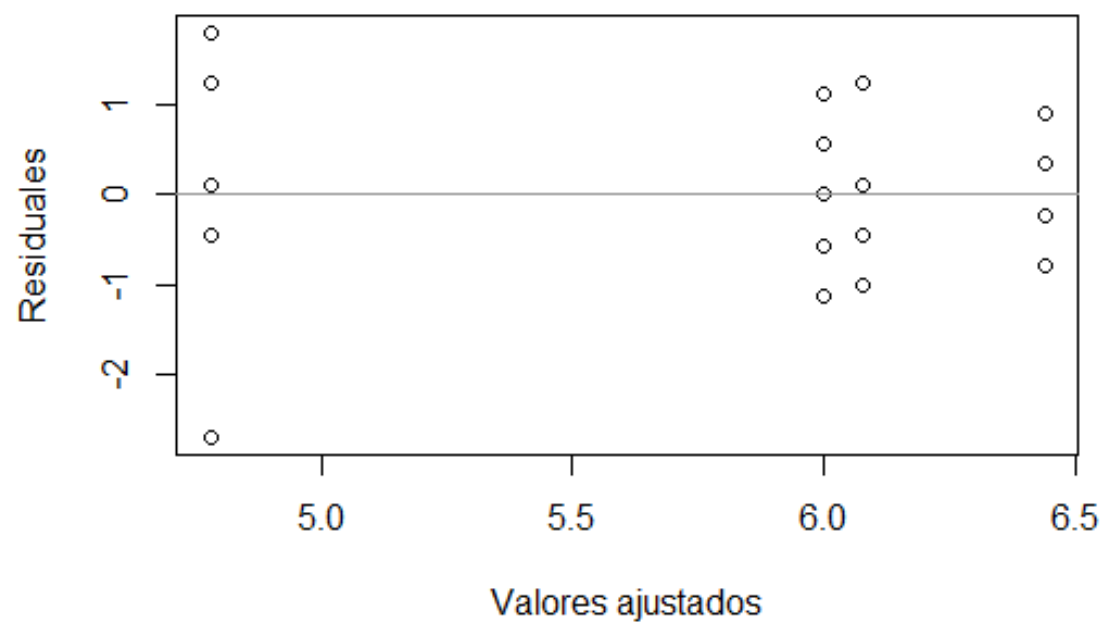
Anderson-Darling normality test

```
data:  residuales
A = 0.28448, p-value = 0.592
```

En R

```
valores_ajustados<-fitted(Modelo)
par(mfrow=c(1,2))
plot(valores_ajustados, residuales, xlab="Valores ajustados", ylab="Residuales")
abline(h=0, col = "gray60")
plot(as.numeric(estacion.f), residuales, xlab="Estación", ylab="Residuales")
abline(h=0, col = "gray60")
```

En R



En R

```
require(car)
leveneTest(datos$oxigeno ~estacion.f)
bartlett.test(datos$oxigeno ~estacion.f)

require(lmtest)
bptest(datos$oxigeno ~estacion.f)
```


En R

```
Levene's Test for Homogeneity of Variance (center = median)
      Df F value Pr(>F)
group  3  1.2678 0.3189
      16
```

Bartlett test of homogeneity of variances

```
data:  datos$oxigeno by estacion.f
Bartlett's K-squared = 4.4567, df = 3, p-value = 0.2162
```

studentized Breusch-Pagan test

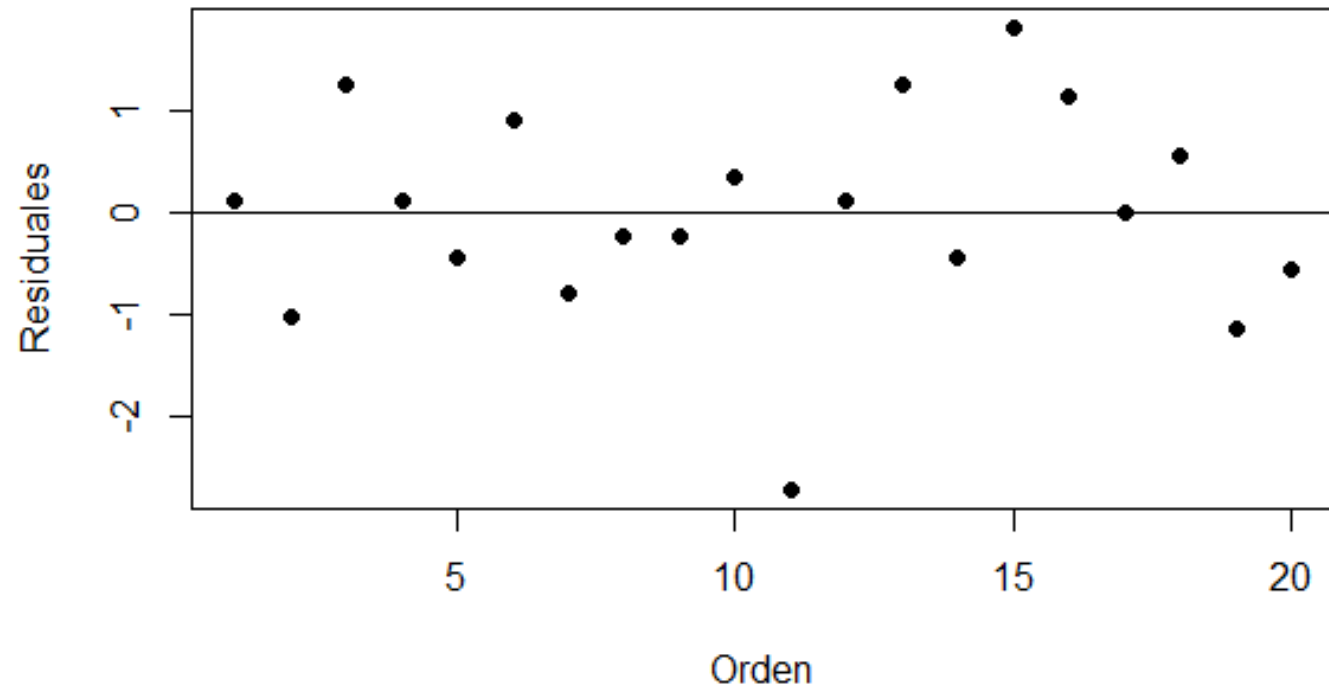
```
data:  oxigeno ~ estacion.f
BP = 5.3502, df = 3, p-value = 0.1479
```

En R

```
plot(residuales, pch=16, ylab="Residuales", xlab="Orden", main="Gráfico de Orden vs  
Residuales")  
abline(h=0)
```

En R

Gráfico de Orden vs Residuales



En R

```
require(lmtest)
dwtest(Modelo, alternative = "two.sided")
bgtest(Modelo)
Box.test(Modelo$res, type = "Ljung")
```

En R

Durbin-Watson test

```
data:  Modelo  
DW = 2.251, p-value = 0.8918  
alternative hypothesis: true autocorrelation is not 0
```

Breusch-Godfrey test for serial correlation of order up to 1

```
data:  Modelo  
LM test = 0.39843, df = 1, p-value = 0.5279
```

Box-Ljung test

```
data:  Modelo$res  
X-squared = 0.41432, df = 1, p-value = 0.5198
```

En R

```
require(agricolae)  
LSD.test(Modelo, "estacion.f", console=TRUE, group=FALSE)
```

Study: Modelo ~ "estacion.f"

LSD t Test for datos\$oxigeno

Mean Square Error: 0.03925

estacion.f, means and individual (95 %) CI

	datos.oxigeno		std r		LCL		UCL	Min	Max
1	6.08	0.1483240	5	5.892176	6.267824	5.9	6.3		
2	6.44	0.1140175	5	6.252176	6.627824	6.3	6.6		
3	4.78	0.3114482	5	4.592176	4.967824	4.3	5.1		
4	6.00	0.1581139	5	5.812176	6.187824	5.8	6.2		

Alpha: 0.05 ; DF Error: 16

Critical Value of t: 2.119905

Comparison between treatments means

	difference	pvalue	signif.	LCL	UCL
1 - 2	-0.36	0.0110	*	-0.6256234	-0.09437663
1 - 3	1.30	0.0000	***	1.0343766	1.56562337
1 - 4	0.08	0.5322		-0.1856234	0.34562337
2 - 3	1.66	0.0000	***	1.3943766	1.92562337
2 - 4	0.44	0.0029	**	0.1743766	0.70562337
3 - 4	-1.22	0.0000	***	-1.4856234	-0.95437663

En R

```
TukeyHSD(Modelo1, which = "estacion.f")
```


Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = datos\$oxigeno ~ estacion.f)

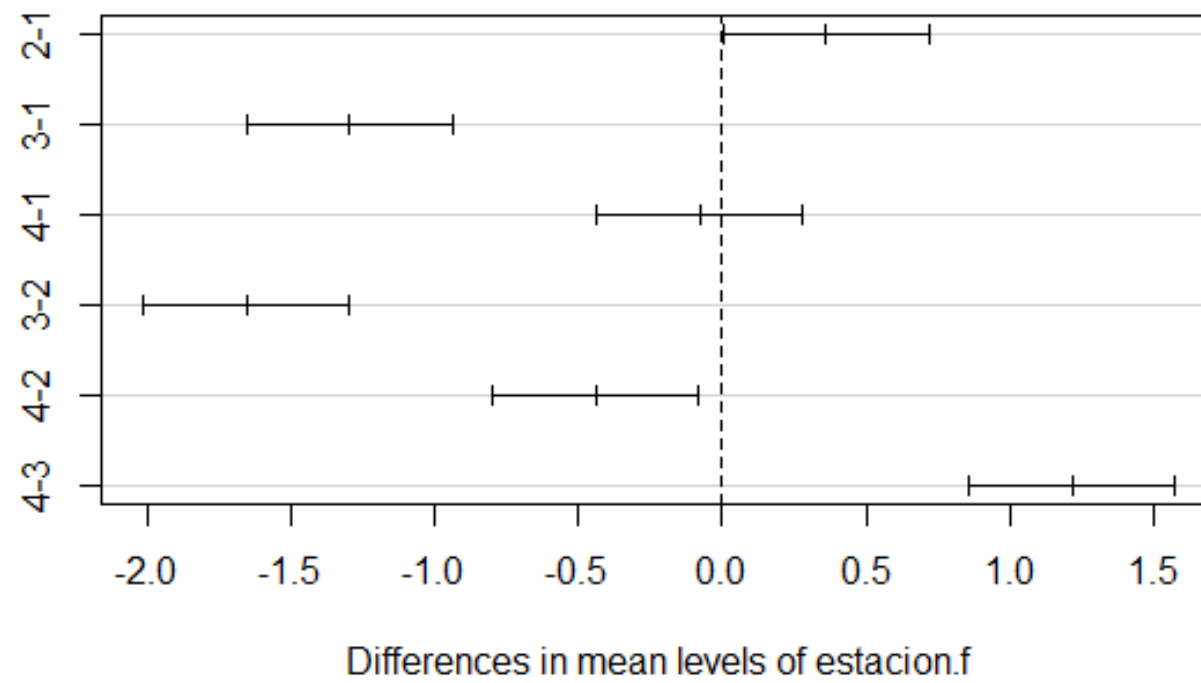
	diff	lwr	upr	p adj
2-1	0.36	0.001515244	0.71848476	0.0488533
3-1	-1.30	-1.658484756	-0.94151524	0.0000001
4-1	-0.08	-0.438484756	0.27848476	0.9179993
3-2	-1.66	-2.018484756	-1.30151524	0.0000000
4-2	-0.44	-0.798484756	-0.08151524	0.0138156
4-3	1.22	0.861515244	1.57848476	0.0000002

En R

```
plot(TukeyHSD(Modelo1, which = "estacion.f"))
```

En R

95% family-wise confidence level



Taller

<https://tinyurl.com/y8uukezj>

Referencias

Fox, J. and Weisberg, S. (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage. URL: <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>.

Gross, J. and Ligges, U. (2015). nortest: Tests for Normality. R package version 1.0-4. <https://CRAN.R-project.org/package=nortest>.

Gutiérrez, P.H. & De La Vara, S. R. (2012). *Análisis y diseño de experimentos*. McGraw-Hill. Edición 3.

Referencias

Mendiburu, F. (2017). agricolae: Statistical Procedures for Agricultural Research. R package version 1.2-8. <https://CRAN.R-project.org/package=agricolae>.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

RStudio Team (2018). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>.

Zeileis, A. & Hothorn, T. (2002). Diagnostic Checking in Regression Relationships. R News 2(3), 7-10. URL <https://CRAN.R-project.org/doc/Rnews/>.