

# Análisis exploratorio de datos de fitoplancton encontrado en canales Baker y Martinez, bajo CIMAR Fiordo N°20

Duncan Rosales Schulz

30 November, 2021

## #CONSIDERACIONES INICIALES

## # A tibble: 1,001 x 6

##	ESPECIE	TIPO	PROFUNDIDAD	CANAL	ESTACIÓN	ABUNDANCIA
##	<chr>	<chr>	<dbl>	<chr>	<chr>	<dbl>
##	1 Cerataulina bergonii	DIATOMEAS	0	BAKER	6	600
##	2 Cerataulina bergonii	DIATOMEAS	0	BAKER	7	1500
##	3 Cerataulina bergonii	DIATOMEAS	0	BAKER	8	1600
##	4 Cerataulina bergonii	DIATOMEAS	0	MARTÍNEZ	96	2000
##	5 Cerataulina bergonii	DIATOMEAS	0	MARTÍNEZ	97	300
##	6 Chaetoceros affinis	DIATOMEAS	0	BAKER	5p	3400
##	7 Chaetoceros affinis	DIATOMEAS	0	BAKER	5	2400
##	8 Chaetoceros affinis	DIATOMEAS	0	BAKER	6	2700
##	9 Chaetoceros affinis	DIATOMEAS	0	BAKER	7	900
##	10 Chaetoceros affinis	DIATOMEAS	0	BAKER	8	1600

## # ... with 991 more rows

## #Variables cuantitativas y factores

La variable cuantitativa, en este estudio, se trató de una cuantitativa de tipo discreta: ABUNDANCIA.

Los factores, que son varios, son los siguientes: ESPECIE, TIPO, PROFUNDIDAD, CANAL, ESTACION.

##	ESPECIE	TIPO	PROFUNDIDAD
##	Pseudo-nitzschia delicatissima: 65	DIATOMEAS	:699 0 :325
##	Pseudo-nitzschia subcurvata : 54	DINOFLAGELADOS	:301 5 :230
##	Thalassionema nitzschioides : 39	SILICOFLAGELADOS:	1 10:168
##	Rhizosolenia setigera : 36		20:123
##	Thalassiosira mendiolana : 34		30:104
##	Leptocylindrus danicus : 32		50: 51
##	(Other) :741		

##	CANAL	ESTACIÓN	ABUNDANCIA
##	BAKER :609	7 :134	Min. : 100
##	MARTÍNEZ:392	6 : 94	1st Qu.: 200
##		5 : 91	Median : 600
##		96 : 91	Mean : 10441
##		5p : 85	3rd Qu.: 2200
##		8 : 81	Max. :641200
##		(Other):425	

## # A tibble: 6 x 6

##	ESPECIE	TIPO	PROFUNDIDAD	CANAL	ESTACIÓN	ABUNDANCIA
##	<fct>	<fct>	<fct>	<fct>	<fct>	<dbl>

```
## 1 Cerataulina bergonii DIATOMEAS 0          BAKER      6          600
## 2 Cerataulina bergonii DIATOMEAS 0          BAKER      7          1500
## 3 Cerataulina bergonii DIATOMEAS 0          BAKER      8          1600
## 4 Cerataulina bergonii DIATOMEAS 0          MARTÍNEZ 96          2000
## 5 Cerataulina bergonii DIATOMEAS 0          MARTÍNEZ 97          300
## 6 Chaetoceros affinis  DIATOMEAS 0          BAKER      5p          3400

## tibble [1,001 x 6] (S3: tbl_df/tbl/data.frame)
## $ ESPECIE      : Factor w/ 90 levels "Centrodinium pavillardii",...: 3 3 3 3 3 11 11 11 11 11 ...
## $ TIPO         : Factor w/ 3 levels "DIATOMEAS","DINOFLAGELADOS",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ PROFUNDIDAD: Factor w/ 6 levels "0","5","10","20",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ CANAL        : Factor w/ 2 levels "BAKER","MARTÍNEZ": 1 1 1 2 2 1 1 1 1 1 ...
## $ ESTACIÓN     : Factor w/ 13 levels "10","11","12",...: 7 8 9 11 12 6 5 7 8 9 ...
## $ ABUNDANCIA  : num [1:1001] 600 1500 1600 2000 300 3400 2400 2700 900 1600 ...
```

## BALANCE DE DATOS Y VERIFICACIÓN DE DATOS

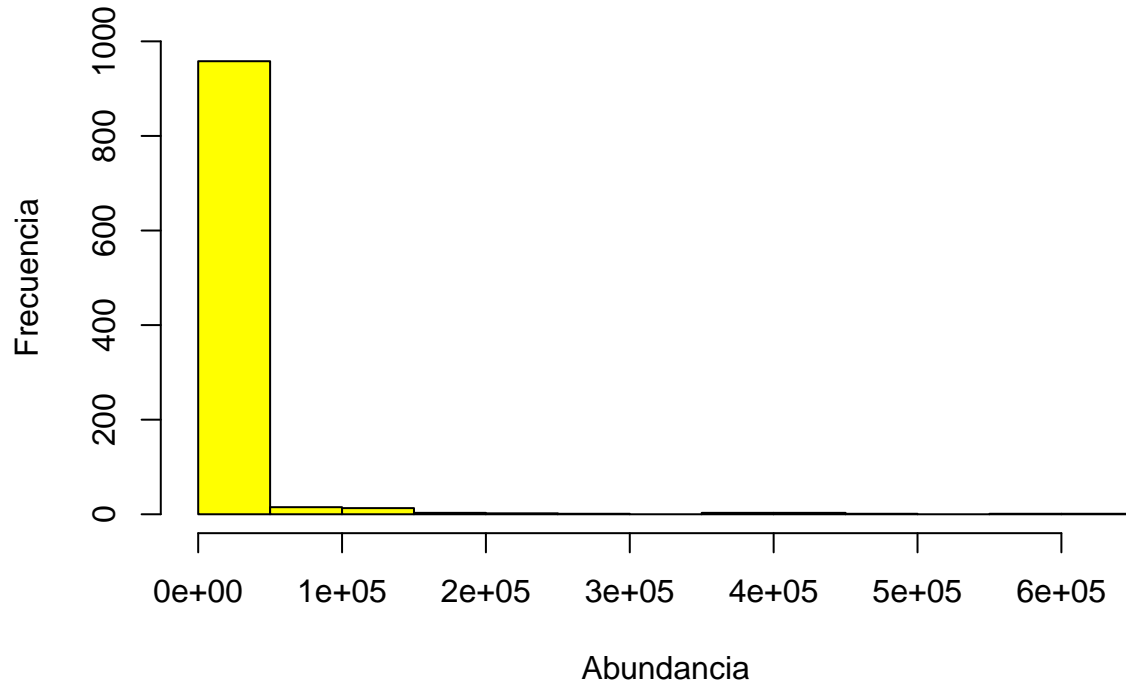
Tabla de frecuencia sobre factores de clasificacion con función table(), sobre la variable ABUNDANCIA.

```
##
##      100      200      300      400      500      600      700      800      900     1000     1100
##      187      113       89       57       41       44       31       30       29       17       18
##     1200     1300     1400     1500     1600     1700     1800     2000     2100     2200     2300
##       20       14        9        5       13        6        7       12        6        9        6
##     2400     2500     2600     2700     2800     2900     3000     3100     3200     3300     3400
##        8        7        5        7        1        5        4        6        2        3        4
##     3500     3600     3700     3800     3900     4000     4100     4200     4300     4400     4500
##        3        1        2        3        1        2        4        1        1        2        1
##     4600     4800     4900     5000     5100     5200     5300     5500     5600     5800     5900
##        3        2        1        1        1        1        1        2        1        4        2
##     6100     6300     6400     6600     6700     7000     7100     7200     7400     7600     8000
##        2        1        3        2        1        1        1        1        1        3        1
##     8200     8300     8500     8600     8700     8900     9000     9500     10000     10300     10500
##        3        1        1        1        2        1        1        1        1        1        2
##    10800    11000    11200    11400    11600    11700    12000    12200    12700    12800    13100
##        1        1        2        1        1        1        2        1        1        3        3
##    13400    13500    13700    14200    15000    15200    15800    16000    16200    16500    16600
##        2        3        1        1        1        1        1        1        2        1        1
##    16800    16900    17000    17100    17500    18100    18300    20600    20800    21300    21400
##        1        1        2        2        1        1        1        1        1        1        1
##    21600    22800    22900    23200    23800    24000    24200    24800    25900    27600    32500
##        1        2        1        1        1        1        2        2        1        1        1
##    32800    33100    33900    34800    36400    36500    36600    39400    40200    41200    44300
##        1        1        1        1        1        1        1        1        1        1        1
##    47900    50400    54000    55800    61400    64700    65000    67600    75600    79200    79400
##        1        1        1        1        1        1        1        1        1        1        1
##    91400    95800    96400    99600    1e+05    100900    101600    102700    106400    108200    110900
##        1        1        1        1        1        1        1        1        1        1        1
##   111000    115700    126200    140700    142800    148200    161500    161800    162200    204800    238400
##        1        1        1        2        1        1        1        1        1        1        1
##  298400    354600    369800    384600    407676    423800    437000    468200    569200    641200
##        1        1        1        1        1        1        1        1        1        1
```

## #HISTOGRAMA

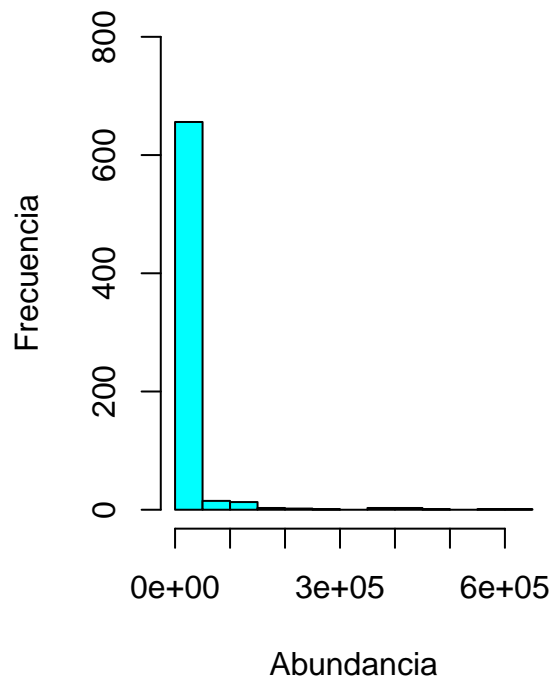
Como primera aproximación, se realizó un histograma de la distribución de frecuencia, de la variable ABUNDANCIA.

## Histograma

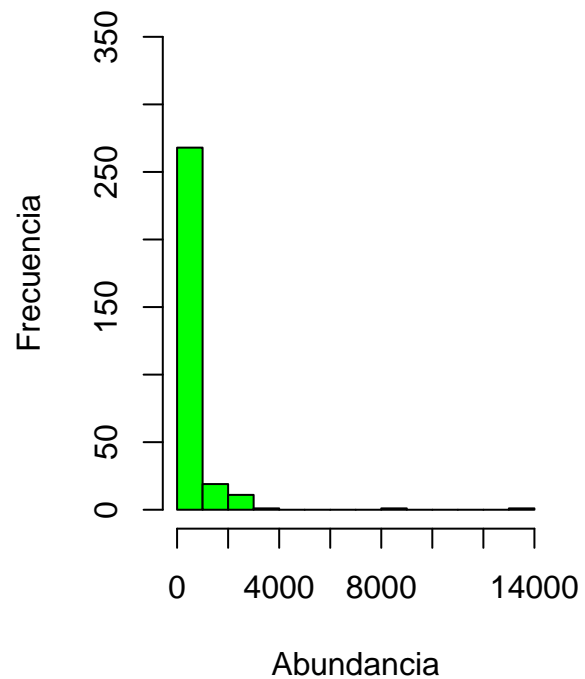


Se decidió mostrar la distribución de frecuencia de la variable ABUNDANCIA, separada entre DIATOMEAS y DINOFLAGELADOS.

### Hist Diatomeas



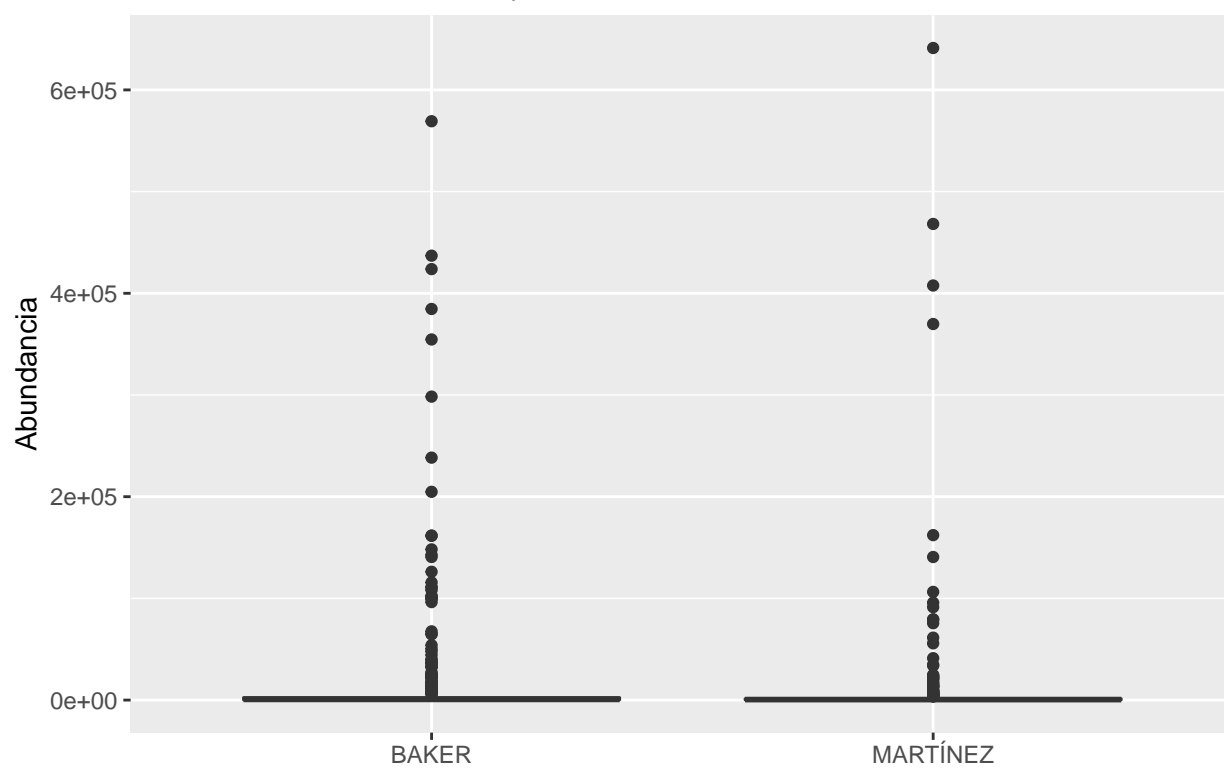
### Hist Dinoflagelados



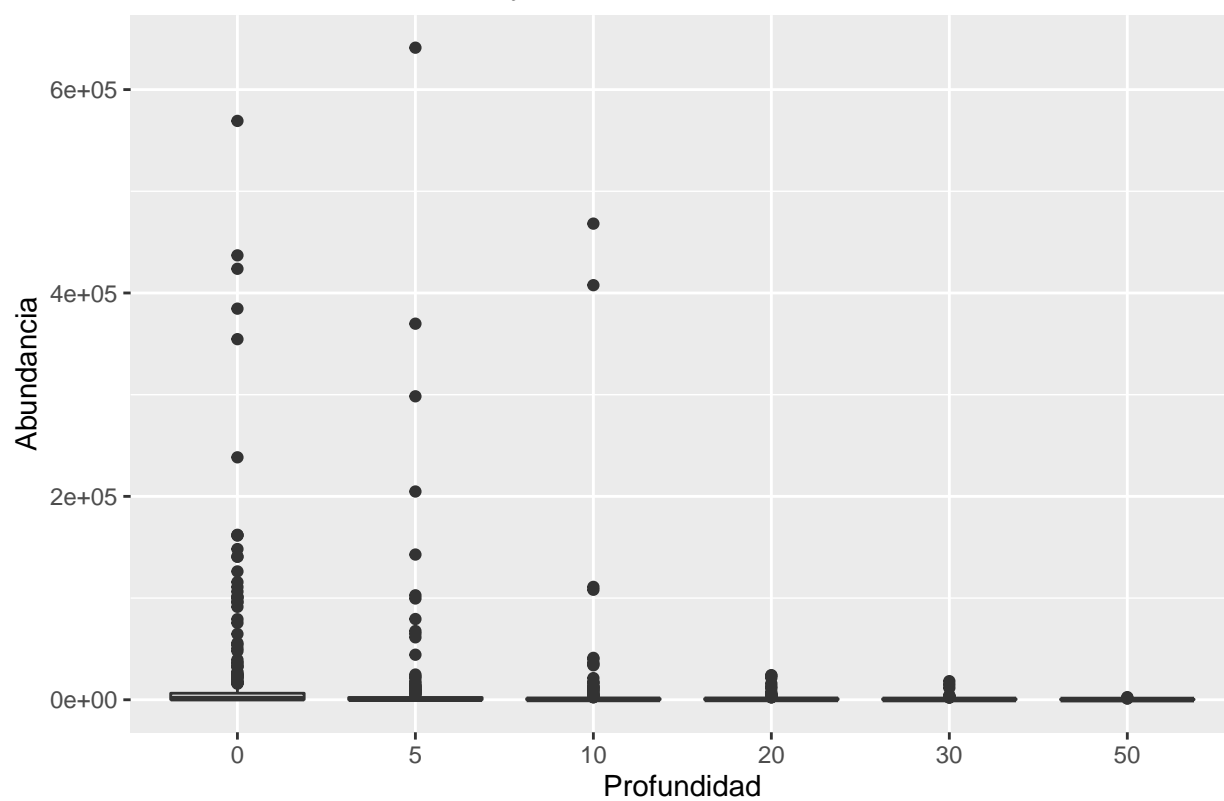
## #RELACIÓN ENTRE VARIABLES Y FACTORES

Se realizaron gráficos boxplot entre ABUNDANCIA, variable respuesta, y los factores o variables regresoras.

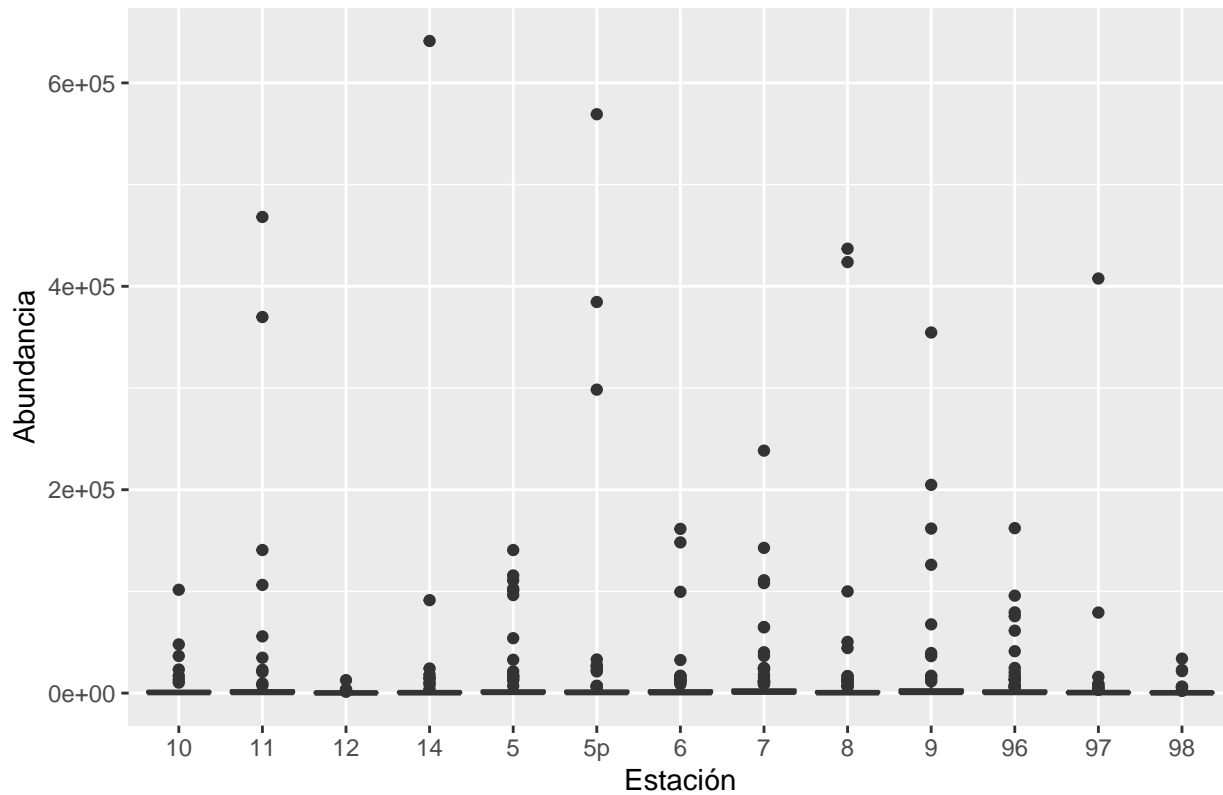
Relación ABUNDANCIA y CANAL



Relación ABUNDANCIA y PROFUNDIDAD



## Relación ABUNDANCIA y ESTACIÓN



## #TRANSFORMACIÓN

A continuación, se transforman los datos ABUNDANCIA; evalúa balance datos, generación histograma, boxplot de lo ya generado

```
##
##          2 2.30102999566398 2.47712125471966 2.60205999132796
##          187          113          89          57
## 2.69897000433602 2.77815125038364 2.84509804001426 2.90308998699194
##          41          44          31          30
## 2.95424250943932          3 3.04139268515822 3.07918124604762
##          29          17          18          20
## 3.11394335230684 3.14612803567824 3.17609125905568 3.20411998265593
##          14          9          5          13
## 3.23044892137827 3.25527250510331 3.30102999566398 3.32221929473392
##          6          7          12          6
## 3.34242268082221 3.36172783601759 3.38021124171161 3.39794000867204
##          9          6          8          7
## 3.41497334797082 3.43136376415899 3.44715803134222 3.46239799789896
##          5          7          1          5
## 3.47712125471966 3.49136169383427 3.50514997831991 3.51851393987789
##          4          6          2          3
## 3.53147891704225 3.54406804435028 3.55630250076729 3.56820172406699
##          4          3          1          2
## 3.57978359661681 3.5910646070265 3.60205999132796 3.61278385671974
##          3          1          2          4
## 3.6232492903979 3.63346845557959 3.64345267648619 3.65321251377534
##          1          1          2          1
```

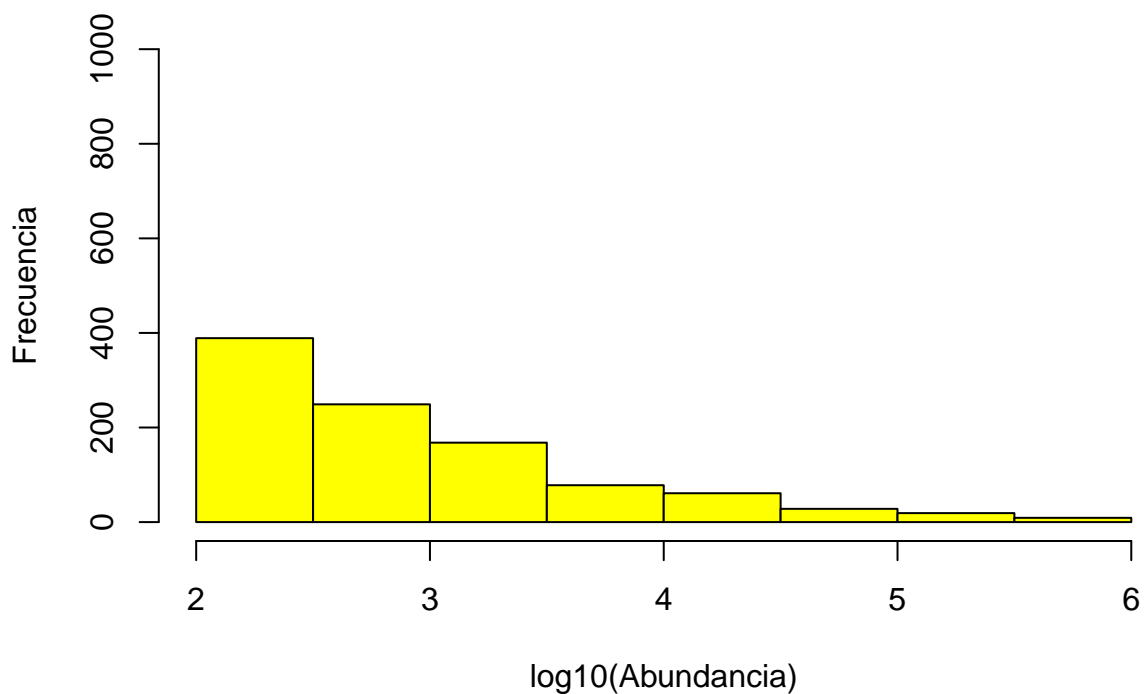
##	3.66275783168157	3.68124123737559	3.69019608002851	3.69897000433602
##	3	2	1	1
##	3.70757017609794	3.7160033436348	3.72427586960079	3.74036268949424
##	1	1	1	2
##	3.7481880270062	3.76342799356294	3.77085201164214	3.78532983501077
##	1	4	2	2
##	3.79934054945358	3.80617997398389	3.81954393554187	3.82607480270083
##	1	3	2	1
##	3.84509804001426	3.85125834871908	3.85733249643127	3.86923171973098
##	1	1	1	1
##	3.88081359228079	3.90308998699194	3.91381385238372	3.91907809237607
##	3	1	3	1
##	3.92941892571429	3.93449845124357	3.93951925261862	3.94939000664491
##	1	1	2	1
##	3.95424250943932	3.97772360528885	4	4.01283722470517
##	1	1	1	1
##	4.02118929906994	4.03342375548695	4.04139268515823	4.04921802267018
##	2	1	1	2
##	4.05690485133647	4.06445798922692	4.06818586174616	4.07918124604763
##	1	1	1	2
##	4.08635983067475	4.10380372095596	4.10720996964787	4.11727129565576
##	1	1	3	3
##	4.12710479836481	4.13033376849501	4.13672056715641	4.15228834438306
##	2	3	1	1
##	4.17609125905568	4.18184358794477	4.19865708695442	4.20411998265593
##	1	1	1	1
##	4.20951501454263	4.21748394421391	4.22010808804005	4.22530928172586
##	2	1	1	1
##	4.22788670461367	4.23044892137827	4.23299611039215	4.24303804868629
##	1	2	2	1
##	4.25767857486918	4.26245108973043	4.31386722036915	4.31806333496276
##	1	1	1	1
##	4.32837960343874	4.33041377334919	4.33445375115093	4.35793484700045
##	1	1	1	2
##	4.35983548233989	4.3654879848909	4.37657695705651	4.38021124171161
##	1	1	1	1
##	4.38381536598043	4.39445168082622	4.41329976408125	4.44090908206522
##	2	2	1	1
##	4.51188336097887	4.51587384371168	4.51982799377572	4.53019969820308
##	1	1	1	1
##	4.54157924394658	4.56110138364906	4.56229286445647	4.56348108539441
##	1	1	1	1
##	4.59549622182557	4.60422605308447	4.61489721603313	4.64640372622307
##	1	1	1	1
##	4.68033551341456	4.70243053644553	4.73239375982297	4.74663419893758
##	1	1	1	1
##	4.78816837114117	4.8109042806687	4.81291335664286	4.82994669594164
##	1	1	1	1
##	4.87852179550121	4.89872518158949	4.8998205024271	4.96094619573383
##	1	1	1	1
##	4.98136550907854	4.98407703390283	4.9982593384237	5
##	1	1	1	1
##	5.00389116623691	5.0068937079479	5.01157044359728	5.02694162795903
##	1	1	1	1

```

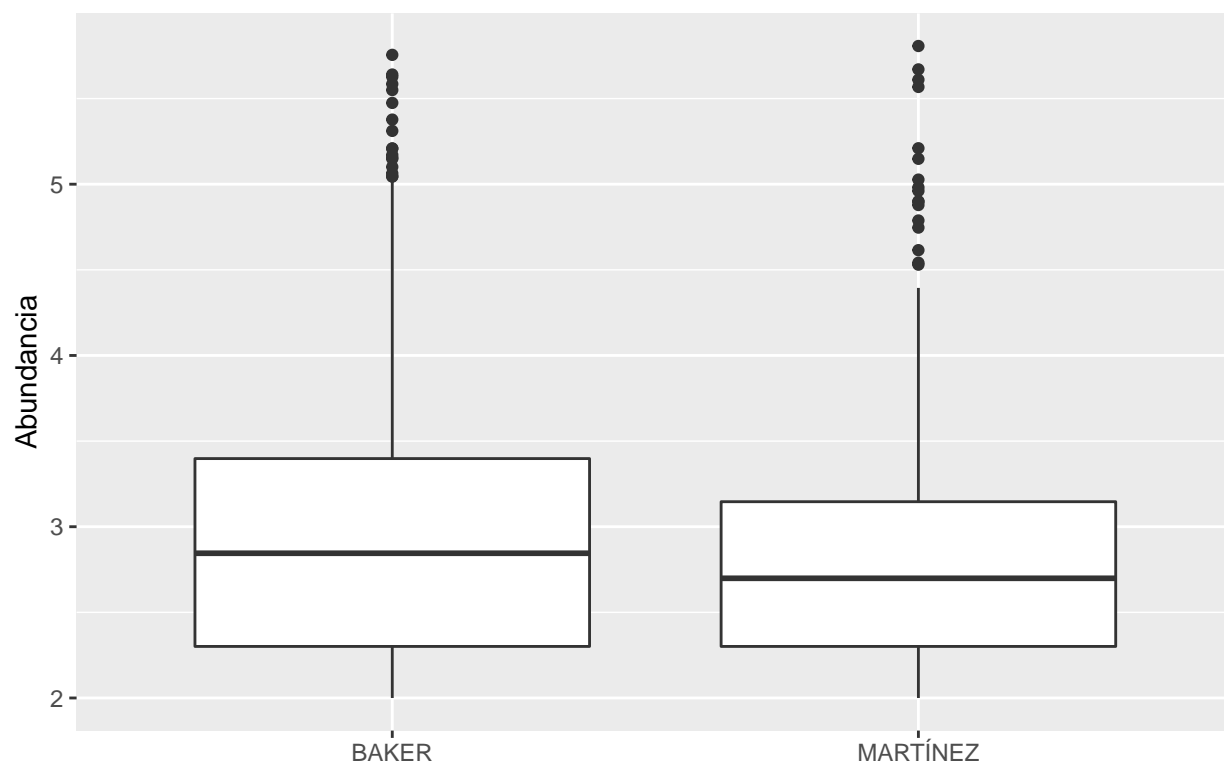
## 5.03422726077055 5.04493154614916 5.04532297878666 5.06333335895175
##          1          1          1          1
## 5.10105935490812 5.14829409743475 5.15472820744016 5.17084820364331
##          1          2          1          1
## 5.20817252666712 5.20897851727625 5.21005084987514 5.31132995230379
##          1          1          1          1
## 5.3773062510682 5.47479881880063 5.5497387312649 5.56796690682315
##          1          1          1          1
## 5.58500927990246 5.61031514516723 5.62716095237478 5.64048143697042
##          1          1          1          1
## 5.67043140936061 5.75526489141225 5.80699351368211
##          1          1          1

```

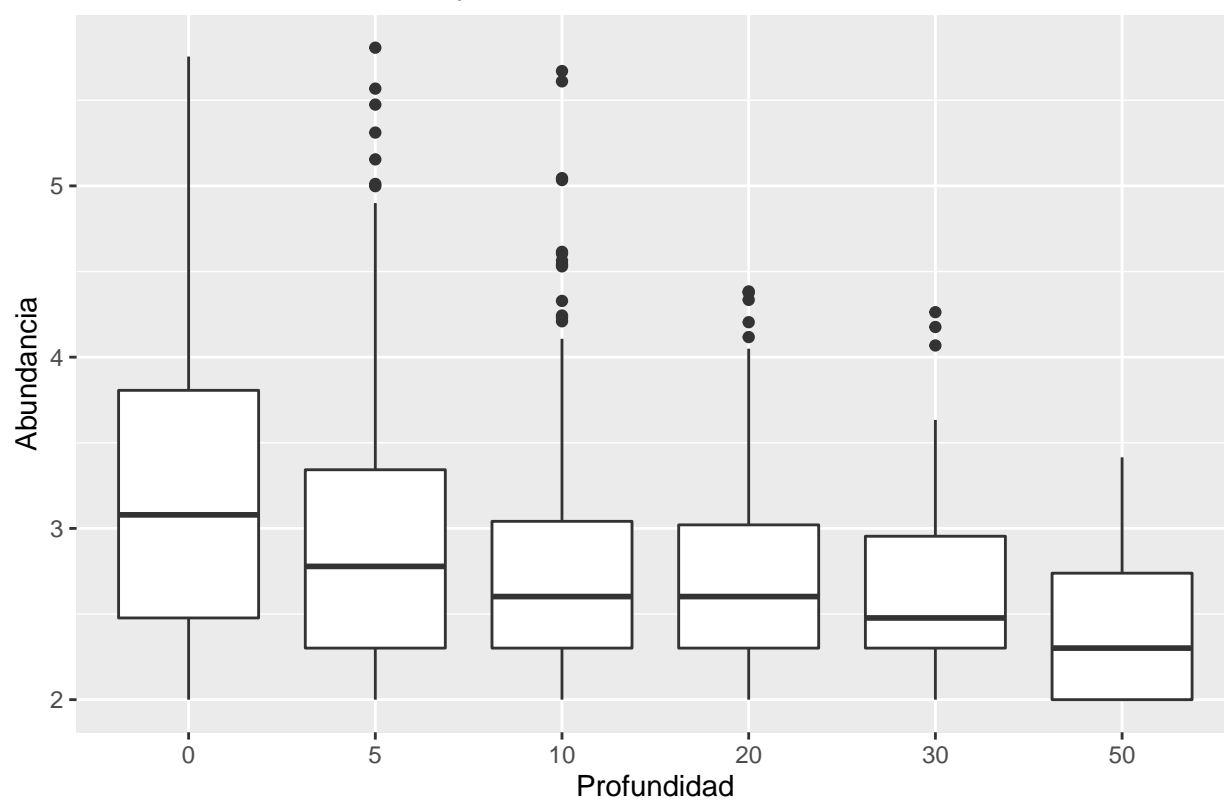
**Histograma**



Relación ABUNDANCIA y CANAL

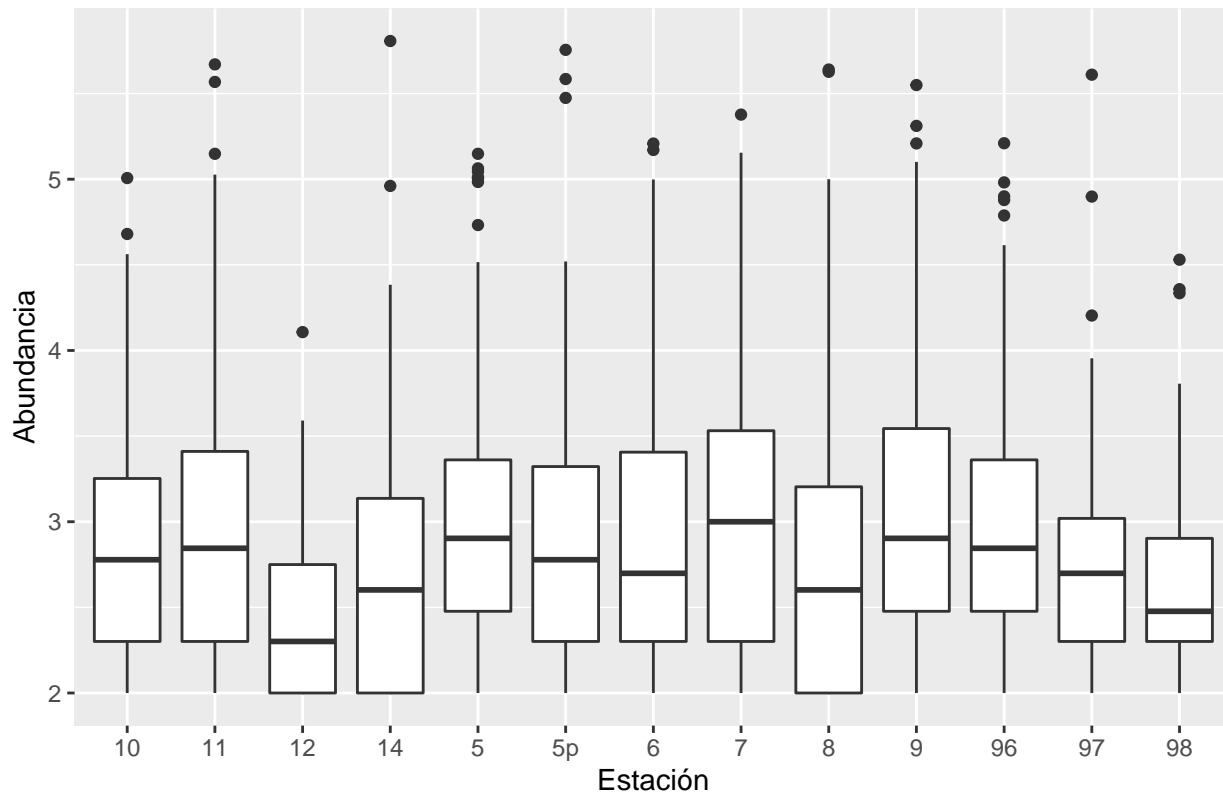


Relación ABUNDANCIA y PROFUNDIDAD





## Relación ABUNDANCIA y ESTACIÓN



### #DÓCIMA DE HIPOTESIS

Podría existir una asociación entre la variable respuesta ABUNDANCIA, y el factor ESTACION. Para ello, se presenta la siguiente H0 e H1:

**H<sub>0</sub>:**

$$\mu_{A(10)} \cap \mu_{A(11)} \cap \mu_{A(12)} \cap \mu_{A(14)} \cap \mu_{A(5)} \cap \mu_{A(5p)} \cap \mu_{A(6)} \cap \mu_{A(7)} \cap \mu_{A(8)} \cap \mu_{A(9)} \cap \mu_{A(96)} \cap \mu_{A(97)} \cap \mu_{A(98)} = \emptyset$$

**H<sub>1</sub>:**

$$\mu_{A(10)} \cap \mu_{A(11)} \cap \mu_{A(12)} \cap \mu_{A(14)} \cap \mu_{A(5)} \cap \mu_{A(5p)} \cap \mu_{A(6)} \cap \mu_{A(7)} \cap \mu_{A(8)} \cap \mu_{A(9)} \cap \mu_{A(96)} \cap \mu_{A(97)} \cap \mu_{A(98)} \neq \emptyset$$

### #ESTADÍSTICA PARAMÉTRICA

Se realizan pruebas para establecer si los datos de ABUNDANCIA cumplen con los supuestos de normalidad, para analizar con estadística paramétrica. Se evaluó independencia, homogeneidad de las varianzas, normalidad, usando un modelo lineal.

**## Call:**

**##** aov(formula = lm.aov)

**##**

**## Terms:**

**##** PROFUNDIDAD Residuals

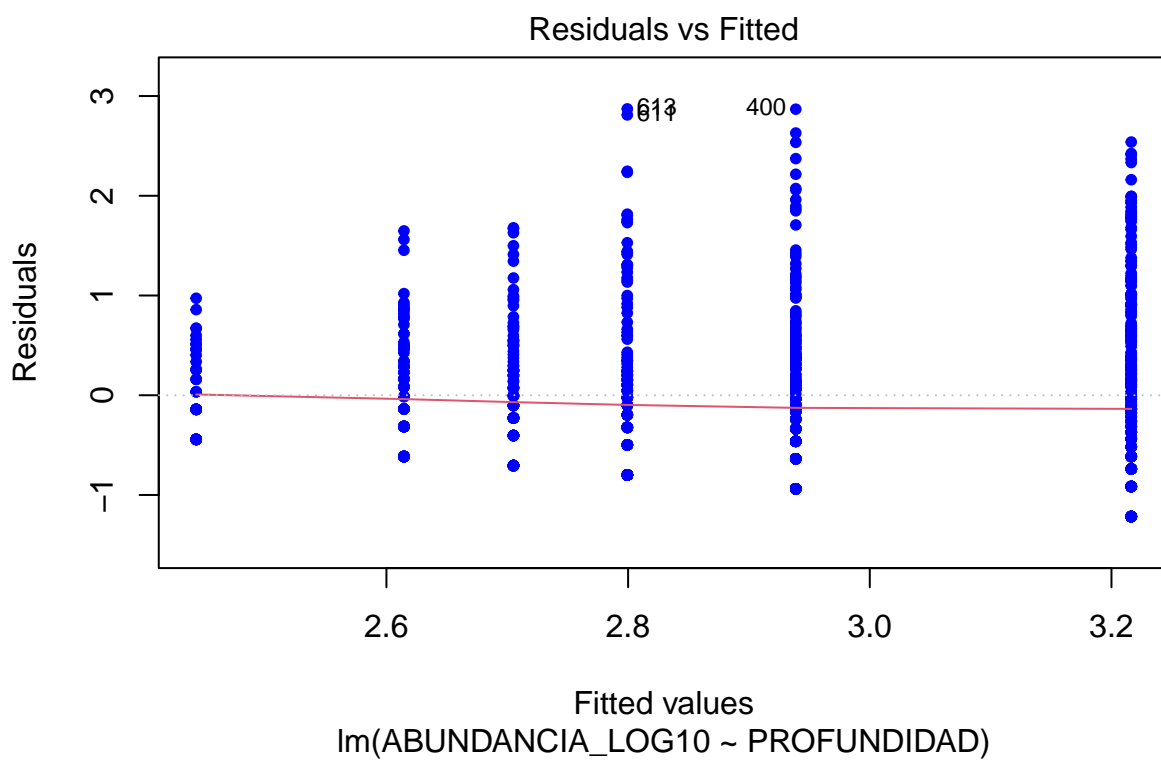
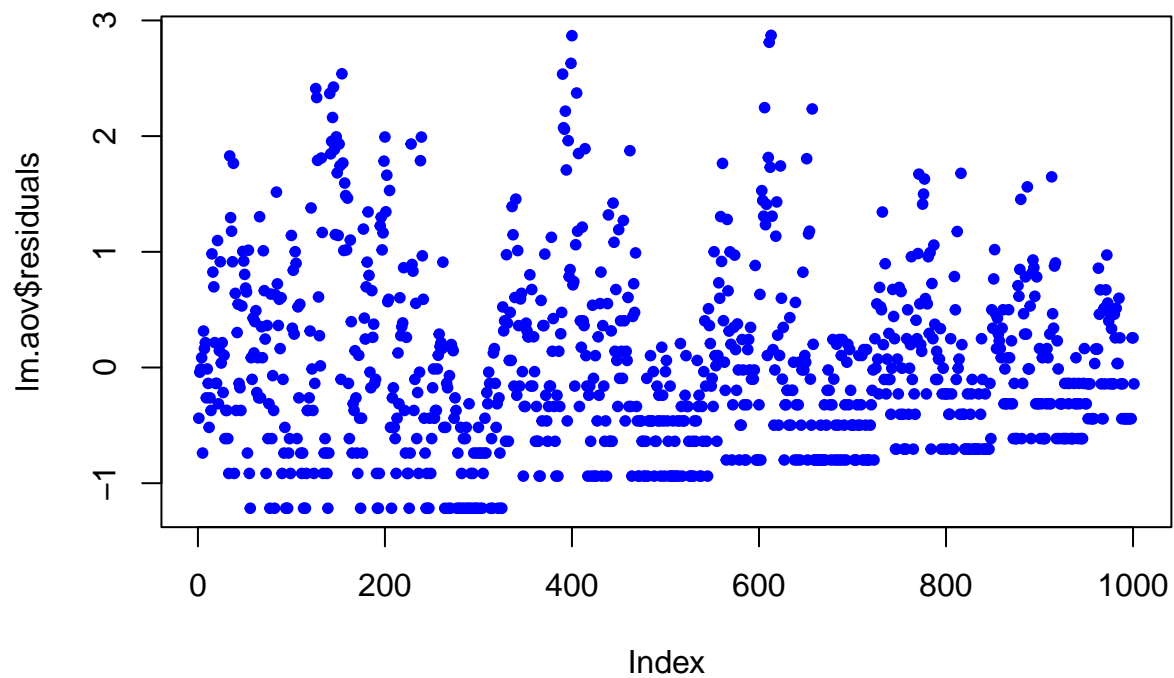
**##** Sum of Squares 58.0846 598.3078

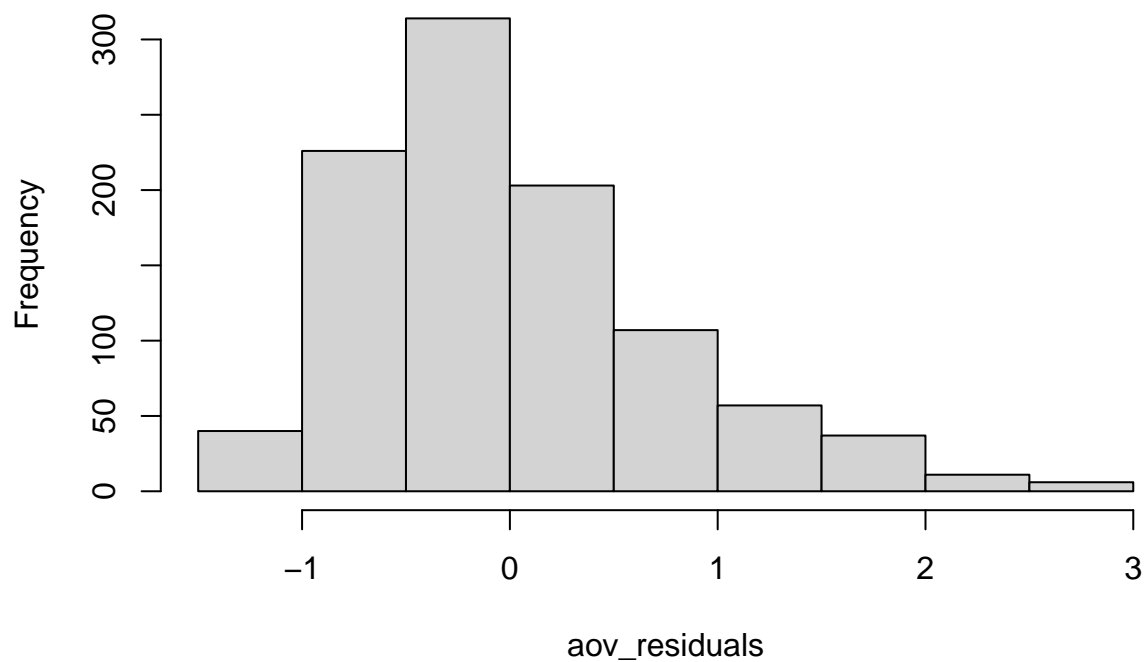
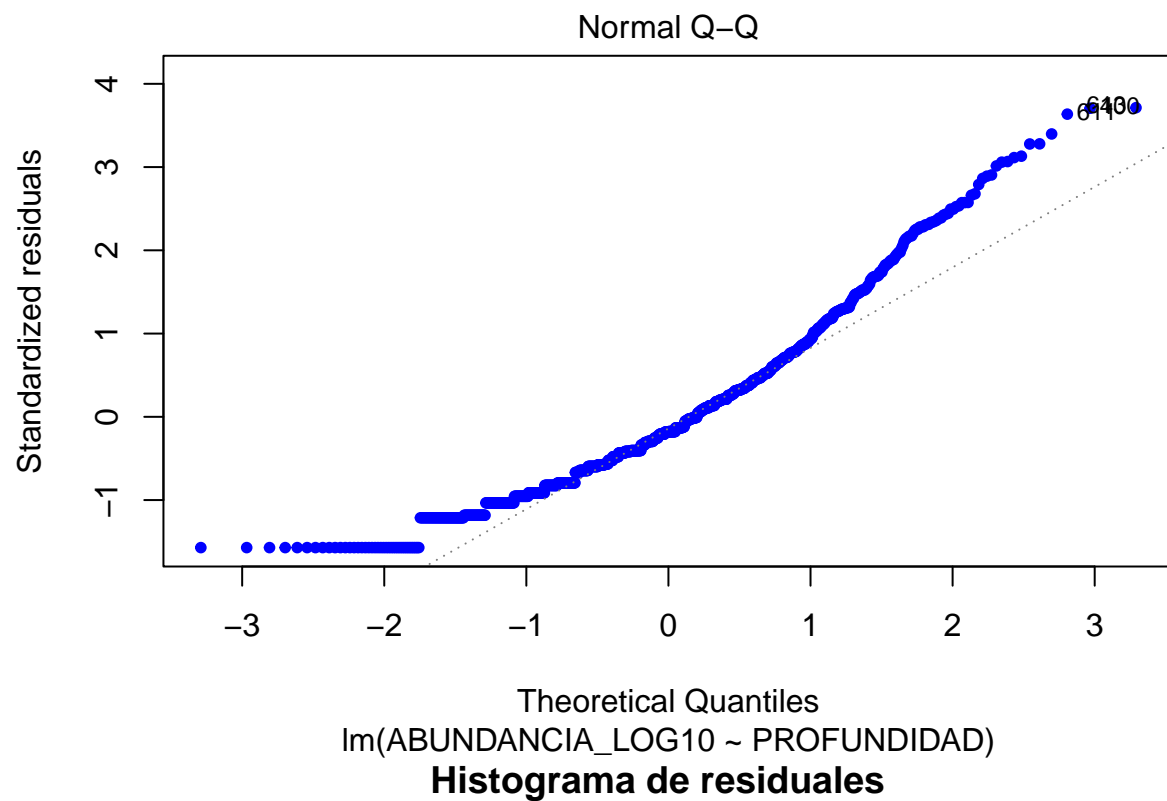
**##** Deg. of Freedom 5 995

**##**

**##** Residual standard error: 0.7754446

**##** Estimated effects may be unbalanced





```
##
## Shapiro-Wilk normality test
##
## data:  aov_residuales
## W = 0.94287, p-value < 2.2e-16
```

Interpretando los gráficos logrados, y el test de normalidad de Shapiro-Wilk, con un p-value  $< 2.2\text{e-}16$ , se establece que los datos estandarizados, no cumplen con los supuestos de la estadística paramétrica, por lo que habría de usarse estadística no paramétrica para evaluarlos.

#REGRESIÓN LINEAL MIXTA

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	3.181	0.1691	18.81	4.749e-67
ESTACIÓN11	0.1521	0.2528	0.6017	0.5475
ESTACIÓN12	-0.6651	0.3382	-1.966	0.04956
ESTACIÓN14	-0.271	0.2572	-1.054	0.2923
ESTACIÓN5	0.2233	0.2117	1.055	0.2918
ESTACIÓN5p	0.1624	0.2255	0.72	0.4717
ESTACIÓN6	0.07554	0.2107	0.3585	0.7201
ESTACIÓN7	0.1081	0.2139	0.5054	0.6134
ESTACIÓN8	0.08689	0.2117	0.4104	0.6816
ESTACIÓN9	0.28	0.2294	1.22	0.2226
ESTACIÓN96	0.1505	0.2205	0.6826	0.495
ESTACIÓN97	-0.3381	0.2674	-1.265	0.2063
ESTACIÓN98	-0.5344	0.2392	-2.235	0.02568
PROFUNDIDAD5	-0.2614	0.3587	-0.7287	0.4664
PROFUNDIDAD10	-0.562	0.2804	-2.004	0.04537
PROFUNDIDAD20	-0.7957	0.3856	-2.063	0.03937
PROFUNDIDAD30	-0.7914	0.5735	-1.38	0.1679
PROFUNDIDAD50	-0.8795	0.7932	-1.109	0.2678
ESTACIÓN11:PROFUNDIDAD5	0.01754	0.4465	0.03928	0.9687
ESTACIÓN12:PROFUNDIDAD5	0.6179	0.5785	1.068	0.2857
ESTACIÓN14:PROFUNDIDAD5	0.327	0.4362	0.7496	0.4537
ESTACIÓN5:PROFUNDIDAD5	-0.1676	0.4202	-0.3989	0.69
ESTACIÓN5p:PROFUNDIDAD5	-0.2099	0.437	-0.4803	0.6312
ESTACIÓN6:PROFUNDIDAD5	-0.2559	0.4197	-0.6099	0.5421
ESTACIÓN7:PROFUNDIDAD5	0.06352	0.409	0.1553	0.8766
ESTACIÓN8:PROFUNDIDAD5	-0.3222	0.4528	-0.7116	0.4769
ESTACIÓN9:PROFUNDIDAD5	0.00204	0.4275	0.004773	0.9962
ESTACIÓN96:PROFUNDIDAD5	-0.03575	0.4116	-0.08685	0.9308
ESTACIÓN97:PROFUNDIDAD5	0.2535	0.4508	0.5623	0.5741
ESTACIÓN98:PROFUNDIDAD5	0.1873	0.4366	0.429	0.668
ESTACIÓN11:PROFUNDIDAD10	0.2131	0.3759	0.567	0.5708
ESTACIÓN12:PROFUNDIDAD10	0.5837	0.4738	1.232	0.2183
ESTACIÓN14:PROFUNDIDAD10	0.3712	0.5161	0.7194	0.4721
ESTACIÓN5:PROFUNDIDAD10	0.006533	0.402	0.01625	0.987
ESTACIÓN5p:PROFUNDIDAD10	-0.206	0.3885	-0.5303	0.596
ESTACIÓN6:PROFUNDIDAD10	-0.008994	0.375	-0.02398	0.9809
ESTACIÓN7:PROFUNDIDAD10	0.4359	0.3436	1.269	0.2048
ESTACIÓN8:PROFUNDIDAD10	-0.2248	0.3866	-0.5814	0.5611
ESTACIÓN9:PROFUNDIDAD10	-0.2647	0.4034	-0.6561	0.5119
ESTACIÓN96:PROFUNDIDAD10	-0.02478	0.3915	-0.06329	0.9496
ESTACIÓN97:PROFUNDIDAD10	0.7228	0.402	1.798	0.07249
ESTACIÓN98:PROFUNDIDAD10	0.5235	0.3966	1.32	0.1872
ESTACIÓN11:PROFUNDIDAD20	0.192	0.5195	0.3696	0.7118
ESTACIÓN12:PROFUNDIDAD20	0.6869	0.5955	1.153	0.249
ESTACIÓN14:PROFUNDIDAD20	-0.01351	0.6216	-0.02173	0.9827

	Estimate	Std. Error	t value	Pr(> t )
ESTACIÓN5:PROFUNDIDAD20	0.1398	0.4686	0.2985	0.7654
ESTACIÓN5p:PROFUNDIDAD20	0.2403	0.4701	0.5112	0.6093
ESTACIÓN6:PROFUNDIDAD20	0.09417	0.4632	0.2033	0.839
ESTACIÓN7:PROFUNDIDAD20	0.384	0.4395	0.8737	0.3825
ESTACIÓN8:PROFUNDIDAD20	0.03514	0.4637	0.07578	0.9396
ESTACIÓN9:PROFUNDIDAD20	0.009425	0.4561	0.02066	0.9835
ESTACIÓN96:PROFUNDIDAD20	0.6882	0.5185	1.327	0.1847
ESTACIÓN97:PROFUNDIDAD20	0.6333	0.5267	1.202	0.2295
ESTACIÓN98:PROFUNDIDAD20	1.174	0.5722	2.052	0.04049
ESTACIÓN11:PROFUNDIDAD30	-0.09874	0.6565	-0.1504	0.8805
ESTACIÓN12:PROFUNDIDAD30	0.577	0.7074	0.8156	0.4149
ESTACIÓN14:PROFUNDIDAD30	0.4978	0.6975	0.7136	0.4756
ESTACIÓN5:PROFUNDIDAD30	0.1774	0.6482	0.2737	0.7844
ESTACIÓN5p:PROFUNDIDAD30	0.1814	0.6334	0.2864	0.7746
ESTACIÓN6:PROFUNDIDAD30	0.1049	0.6479	0.1619	0.8715
ESTACIÓN7:PROFUNDIDAD30	0.4311	0.6237	0.6913	0.4895
ESTACIÓN8:PROFUNDIDAD30	-0.1315	0.6418	-0.2049	0.8377
ESTACIÓN9:PROFUNDIDAD30	0.06953	0.8082	0.08603	0.9315
ESTACIÓN96:PROFUNDIDAD30	0.04474	0.6316	0.07084	0.9435
ESTACIÓN97:PROFUNDIDAD30	0.6325	0.6869	0.9207	0.3575
ESTACIÓN98:PROFUNDIDAD30	0.7062	0.6462	1.093	0.2748
ESTACIÓN11:PROFUNDIDAD50	-0.3027	0.9822	-0.3081	0.7581
ESTACIÓN12:PROFUNDIDAD50	1.08	1.008	1.072	0.2842
ESTACIÓN14:PROFUNDIDAD50	-0.03003	0.9311	-0.03225	0.9743
ESTACIÓN5:PROFUNDIDAD50	-0.04445	0.8551	-0.05198	0.9586
ESTACIÓN5p:PROFUNDIDAD50	0.02351	0.8586	0.02738	0.9782
ESTACIÓN6:PROFUNDIDAD50	0.3024	0.8917	0.3392	0.7346
ESTACIÓN7:PROFUNDIDAD50	0.07737	0.8494	0.09109	0.9274
ESTACIÓN8:PROFUNDIDAD50	-0.03844	0.9725	-0.03953	0.9685
ESTACIÓN9:PROFUNDIDAD50	-0.1989	0.9238	-0.2154	0.8295
ESTACIÓN96:PROFUNDIDAD50	0.08247	0.9216	0.08948	0.9287
ESTACIÓN97:PROFUNDIDAD50	0.4438	0.8901	0.4986	0.6182
ESTACIÓN98:PROFUNDIDAD50	0.6537	0.8988	0.7273	0.4672

Table 2: Modelo de efectos fijos con interacción, de los datos df2, transformados  $\log_{10}(\text{ABUNDANCIA})$

Observations	Residual Std. Error	$R^2$	Adjusted $R^2$
1001	0.775	0.1555	0.08503

Con un  $r^2$  de 0.155, se puede establecer que no existe relación estadísticamente significativa entre las relaciones establecidas en el modelo de regresión lineal mixta.