

QTL-SRS

Marcos Mancilla

30 November 2021

Realizando análisis exploratorio de datos

```
##      Fecha                Centro                Unidad
## Min.      :2019-02-09 00:00:00 Length:12825      Length:12825
## 1st Qu.:2019-06-01 00:00:00 Class :character Class :character
## Median :2019-07-26 00:00:00 Mode  :character Mode  :character
## Mean    :2019-07-25 16:53:46
## 3rd Qu.:2019-09-20 00:00:00
## Max.    :2019-12-09 00:00:00
##      Subgrupo      N_peces      Biomasa_kg      Mortalidad
## Length:12825      Min.      : 0      Min.      : 0      Min.      : 0.000
## Class :character  1st Qu.:16308  1st Qu.: 23766  1st Qu.: 0.000
## Mode  :character  Median :28340  Median : 42046  Median : 1.000
##                      Mean  :24143  Mean  : 44360  Mean  : 9.889
##                      3rd Qu.:29028  3rd Qu.: 59946  3rd Qu.: 8.000
##                      Max.   :29209  Max.   :114599  Max.   :1982.000
## Mortalidad_Biomasa_kg N_cosecha      Biomasa_cosecha      Alimento_kg
## Min.      : 0.000      Min.      : 0.0      Min.      : 0.0      Min.      : 0.0
## 1st Qu.: 0.000      1st Qu.: 0.0      1st Qu.: 0.0      1st Qu.: 285.0
## Median : 2.015      Median : 0.0      Median : 0.0      Median : 446.0
## Mean    : 27.549      Mean    : 76.9      Mean    : 270.1      Mean    : 440.6
## 3rd Qu.: 17.845      3rd Qu.: 0.0      3rd Qu.: 0.0      3rd Qu.: 604.0
## Max.    :7671.965      Max.    :22602.0      Max.    :78009.3      Max.    :1560.0
##      Temperatura
## Min.      : 0.00
## 1st Qu.:10.79
## Median :11.20
## Mean    :11.15
## 3rd Qu.:11.70
## Max.    :15.00

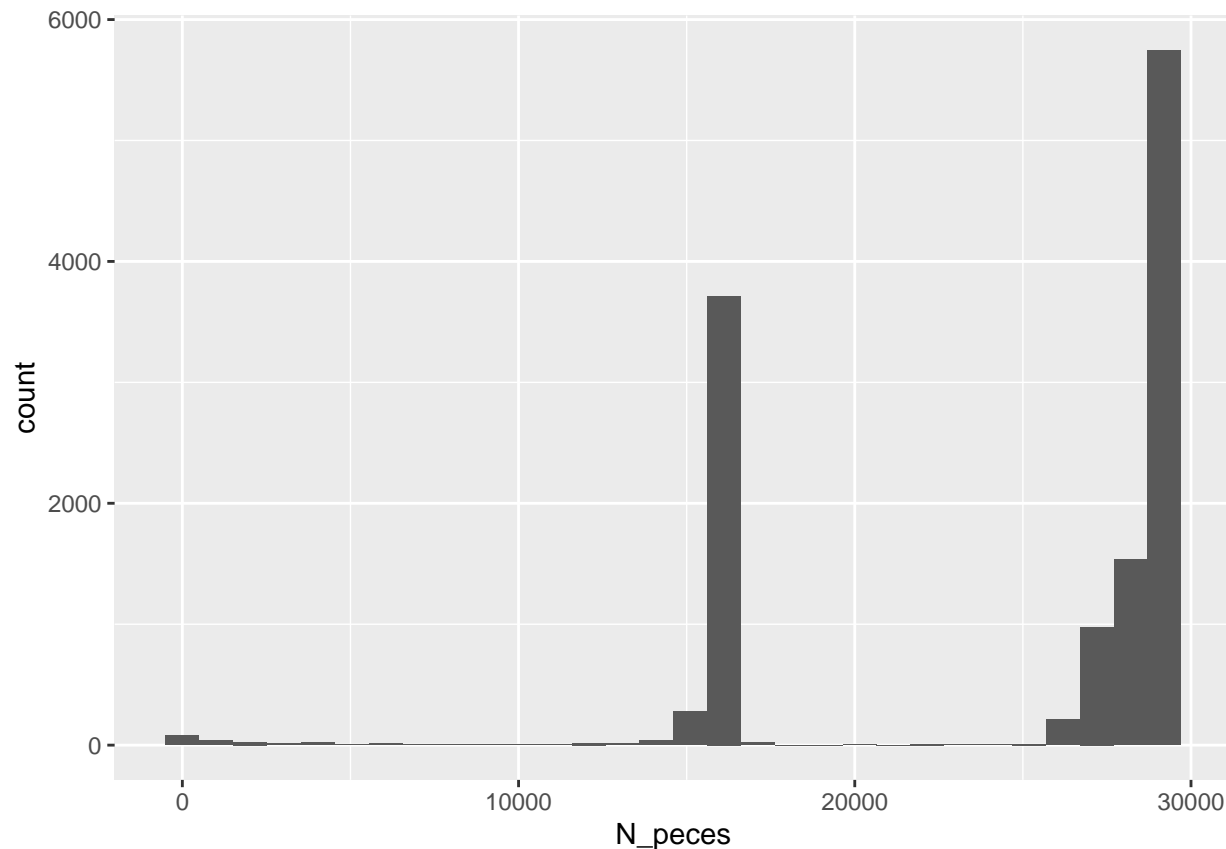
## # A tibble: 6 x 12
##      Fecha                Centro Unidad Subgrupo N_peces Biomasa_kg Mortalidad
##      <dtm>                <chr> <chr> <chr>      <dbl>      <dbl>      <dbl>
## 1 2019-08-24 00:00:00 A      205 QTL2      16302      35625.      0
## 2 2019-08-25 00:00:00 A      205 QTL2      16302      35879.      0
## 3 2019-08-26 00:00:00 A      205 QTL2      16302      36261.      0
## 4 2019-08-27 00:00:00 A      205 QTL2      16302      36684.      0
## 5 2019-08-28 00:00:00 A      205 QTL2      16302      37100.      0
## 6 2019-08-29 00:00:00 A      205 QTL2      16302      37435.      0
## # ... with 5 more variables: Mortalidad_Biomasa_kg <dbl>, N_cosecha <dbl>,
## # Biomasa_cosecha <dbl>, Alimento_kg <dbl>, Temperatura <dbl>
```

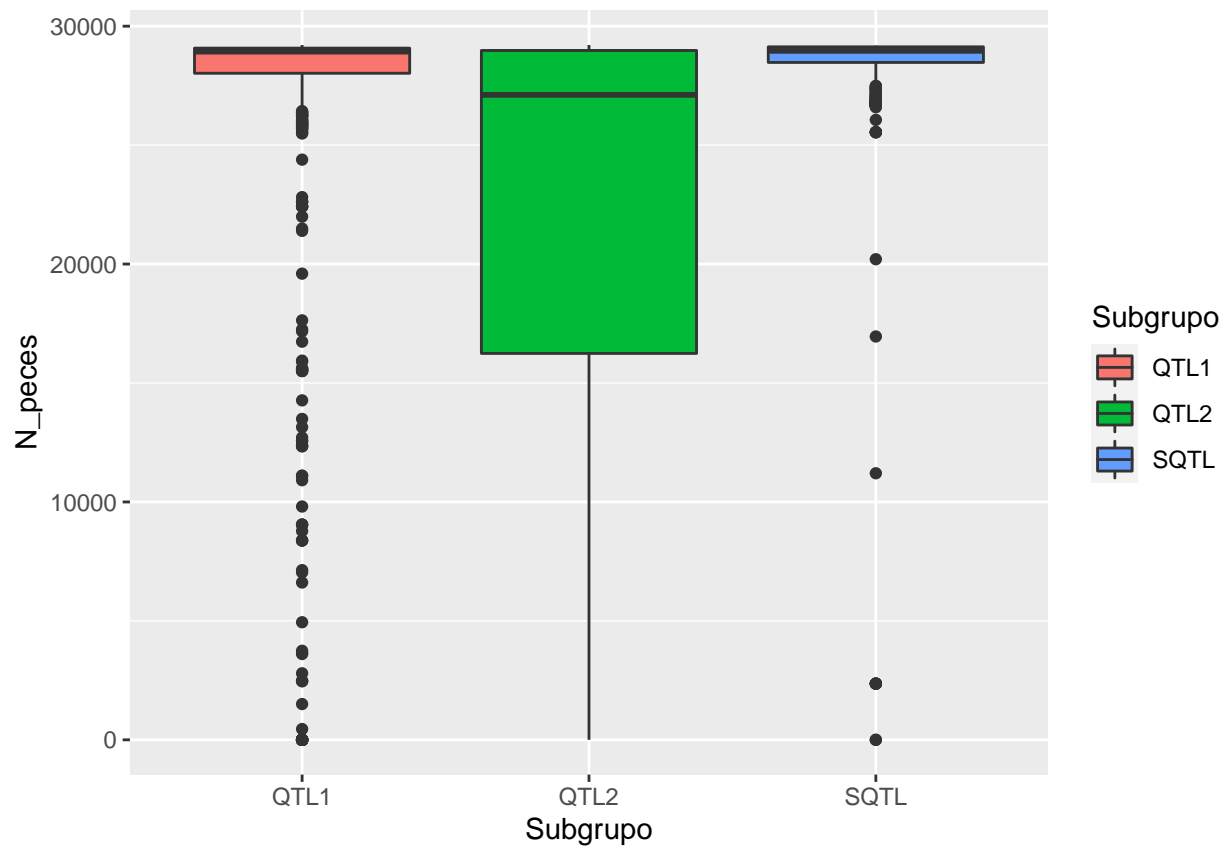
```
## tibble [12,825 x 12] (S3: tbl_df/tbl/data.frame)
## $ Fecha          : POSIXct[1:12825], format: "2019-08-24" "2019-08-25" ...
## $ Centro         : chr [1:12825] "A" "A" "A" "A" ...
## $ Unidad         : chr [1:12825] "205" "205" "205" "205" ...
## $ Subgrupo       : chr [1:12825] "QTL2" "QTL2" "QTL2" "QTL2" ...
## $ N_peces        : num [1:12825] 16302 16302 16302 16302 16302 ...
## $ Biomasa_kg     : num [1:12825] 35625 35879 36261 36684 37100 ...
## $ Mortalidad     : num [1:12825] 0 0 0 0 0 0 0 1 2 2 ...
## $ Mortalidad_Biomasa_kg: num [1:12825] 0 0 0 0 0 ...
## $ N_cosecha      : num [1:12825] 0 0 0 0 0 0 0 0 0 0 ...
## $ Biomasa_cosecha : num [1:12825] 0 0 0 0 0 0 0 0 0 0 ...
## $ Alimento_kg    : num [1:12825] 455 300 450 500 490 396 472 562 435 556 ...
## $ Temperatura    : num [1:12825] 11.3 11.1 11.1 11.3 10.3 ...
```

Observamos que el set de datos tiene 12.825 observaciones de 12 variables. Existe una variable fecha, 3 variables como texto (Centro, Unidad y Subgrupo) y 8 variables numéricas. Nos se aprecian datos faltantes.

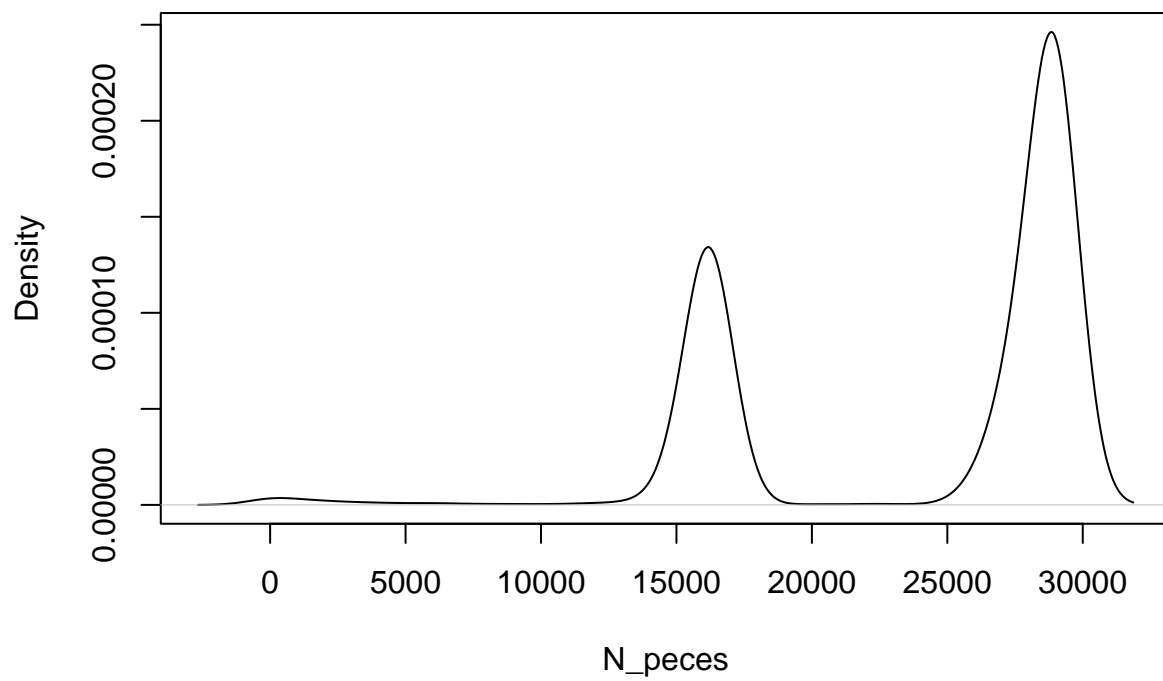
Graficamos variables de interés (histogramas, boxplots, densidad y densidad acumulada).

Peces por subgrupo

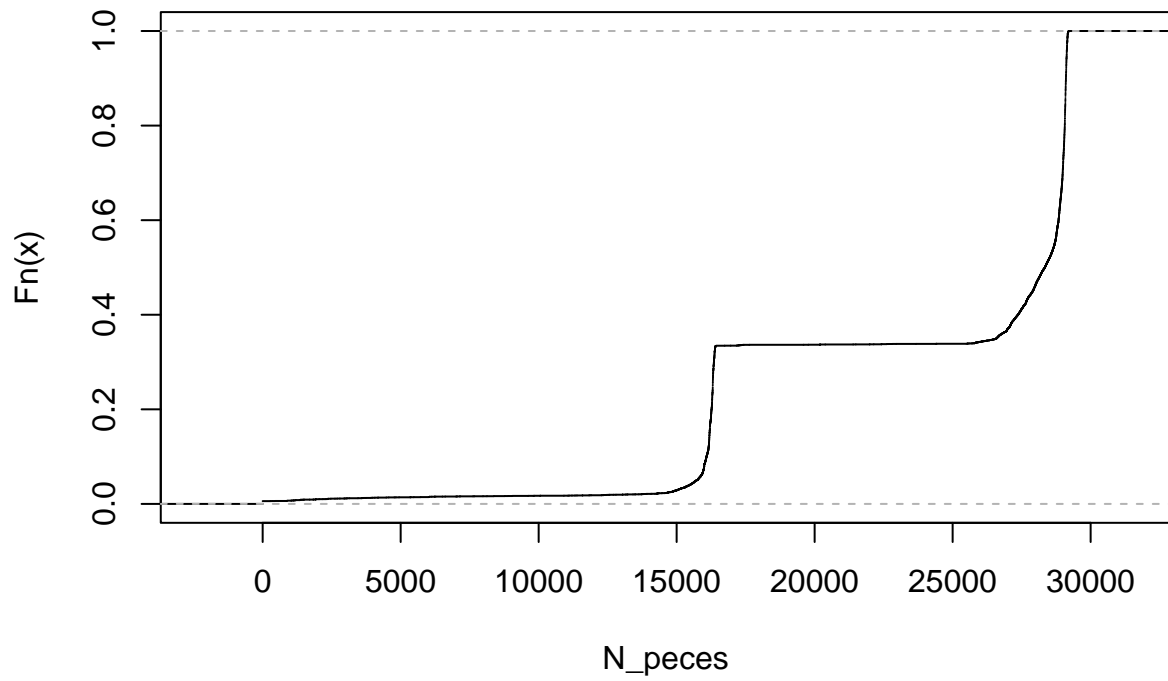




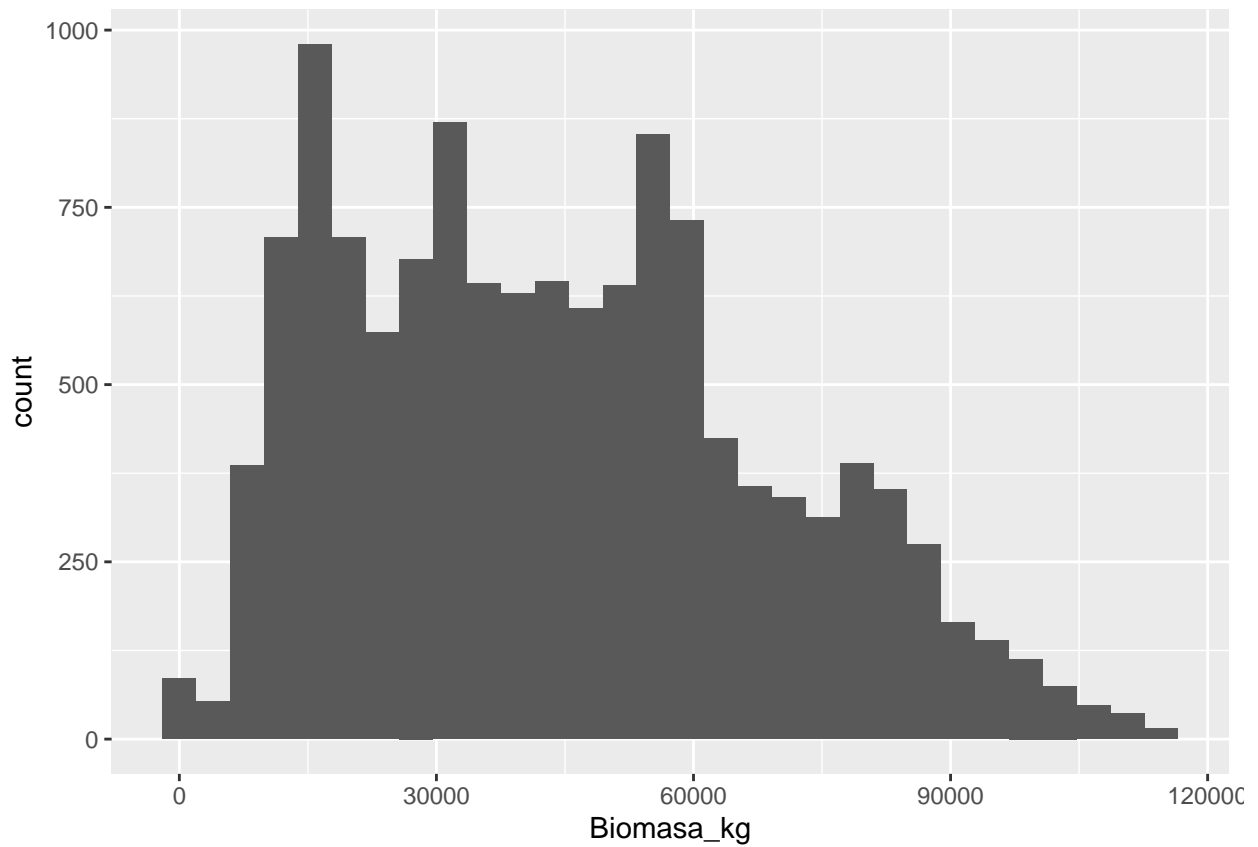
Densidad empírica

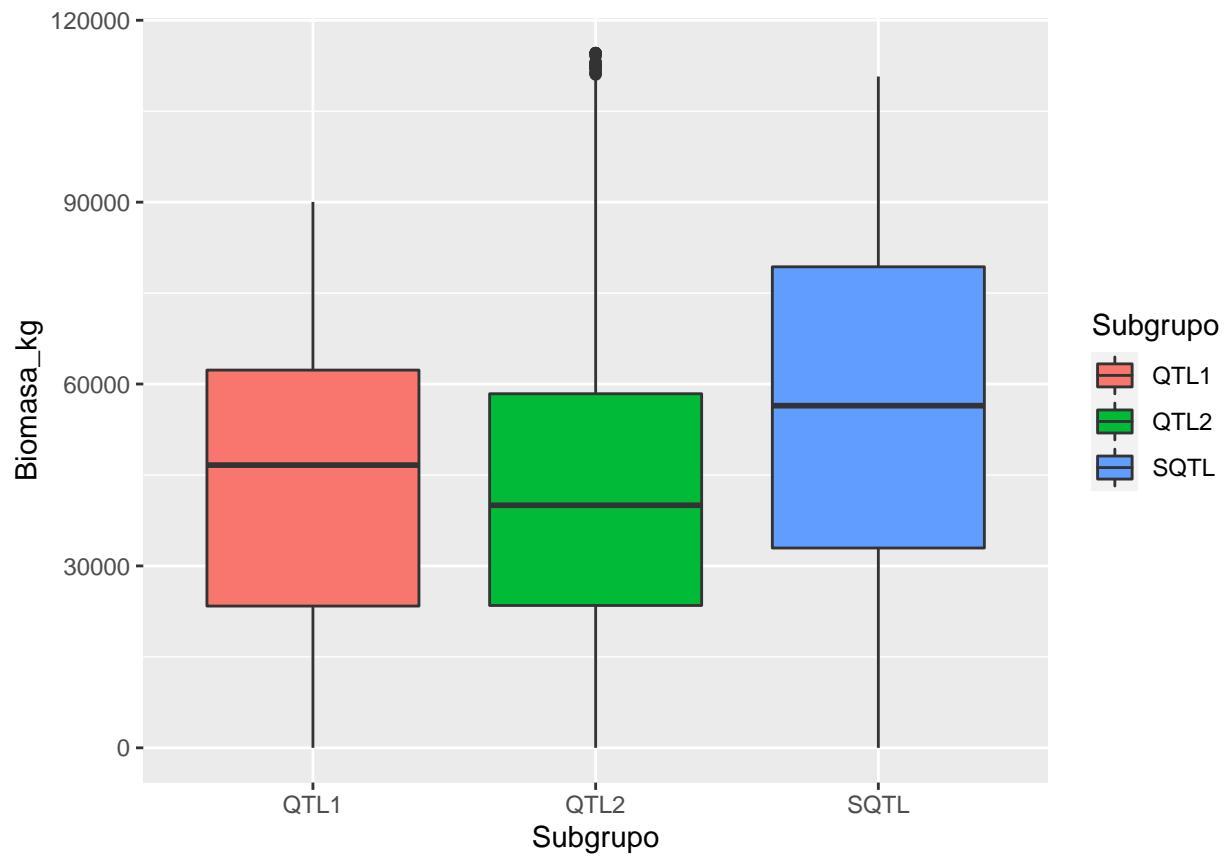


Distribución acumulada empírica

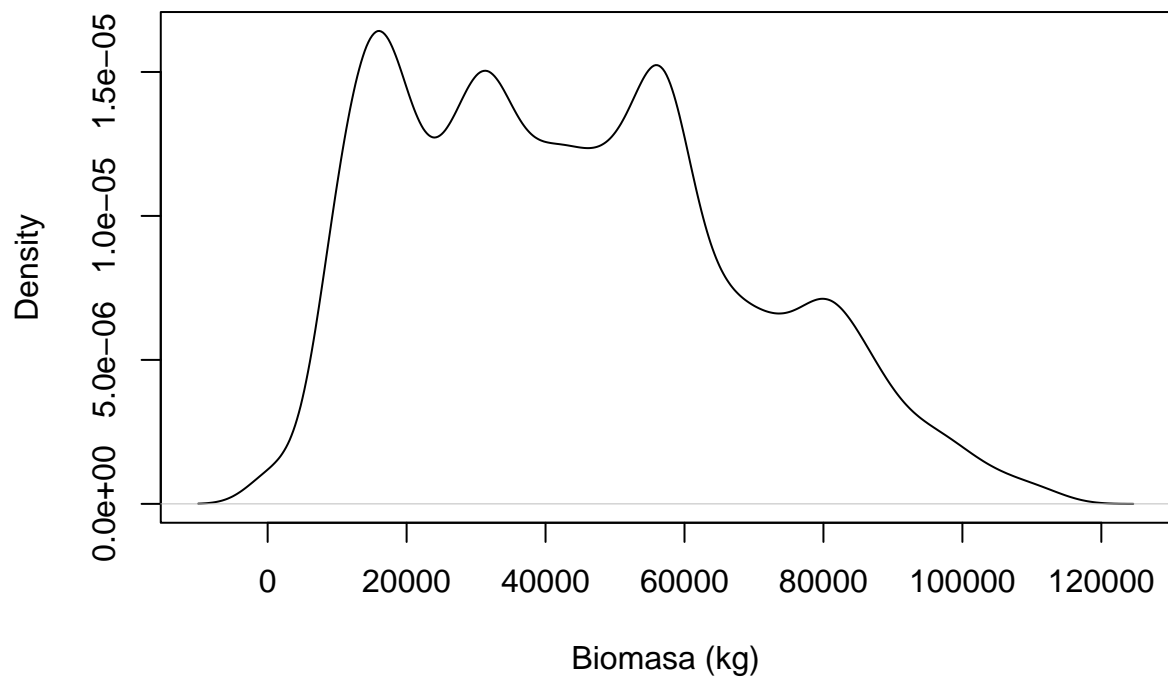


Biomasa viva por subgrupo

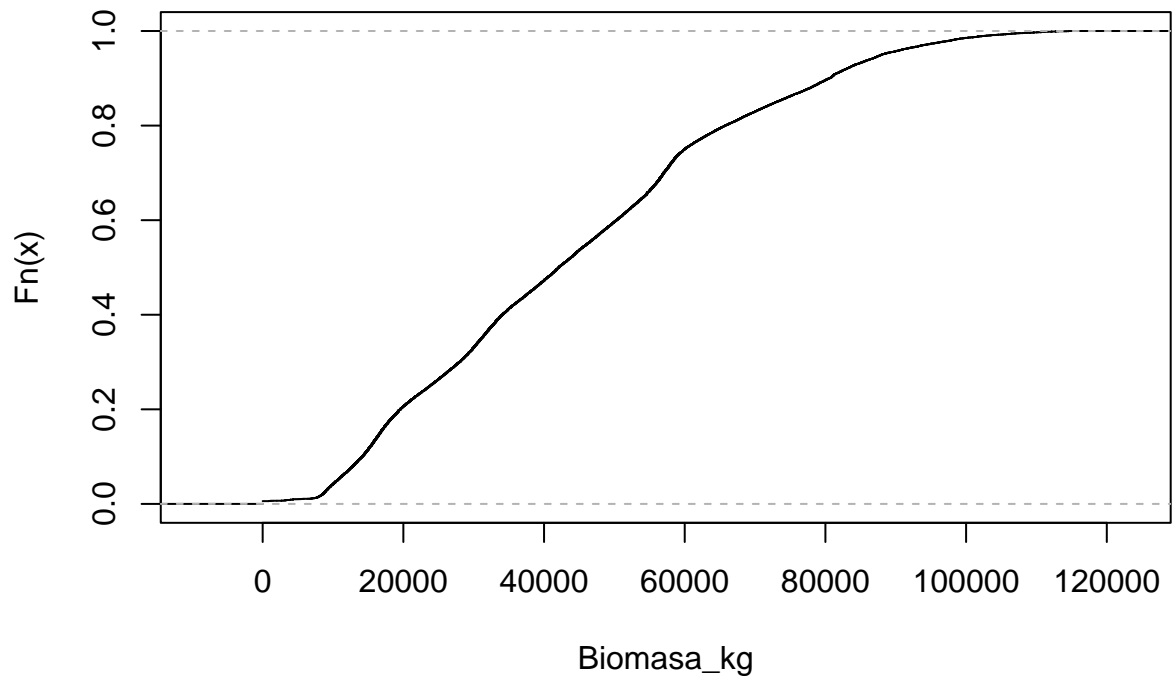




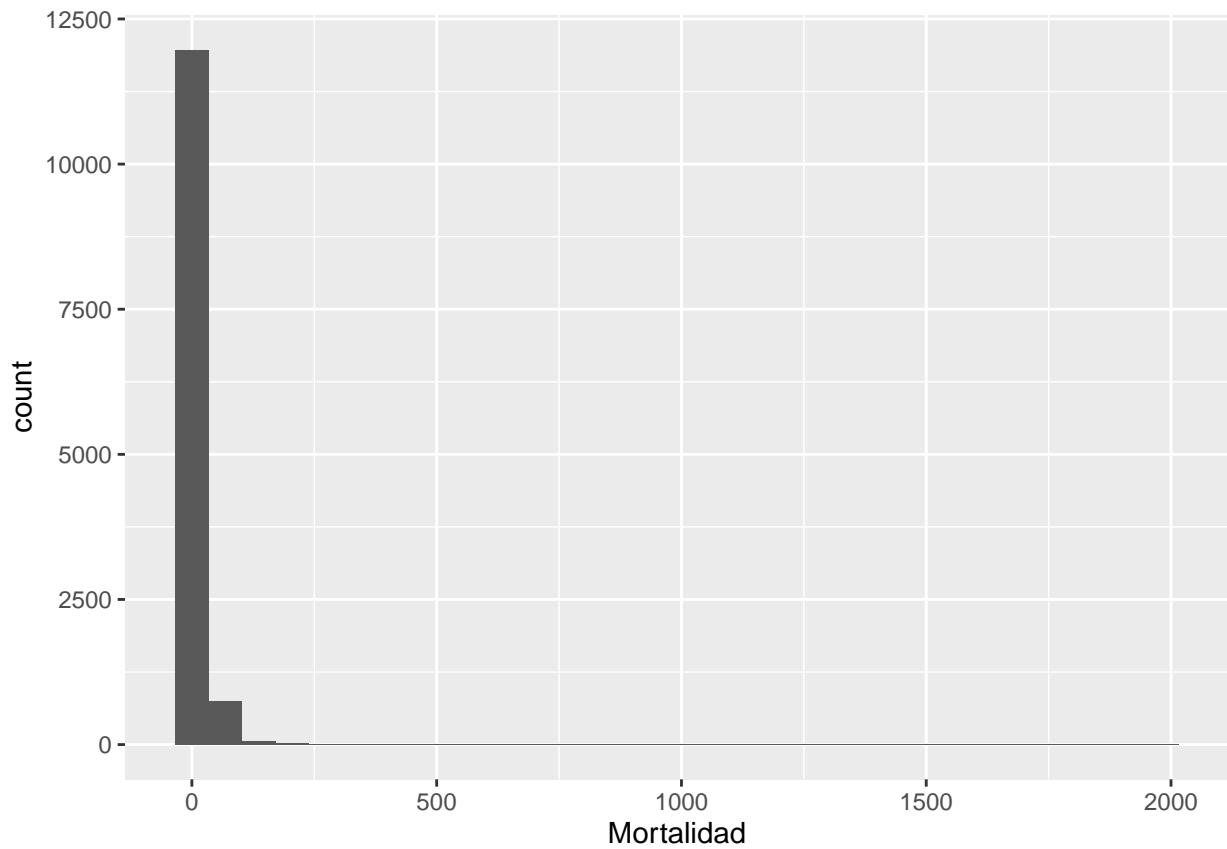
Densidad empírica



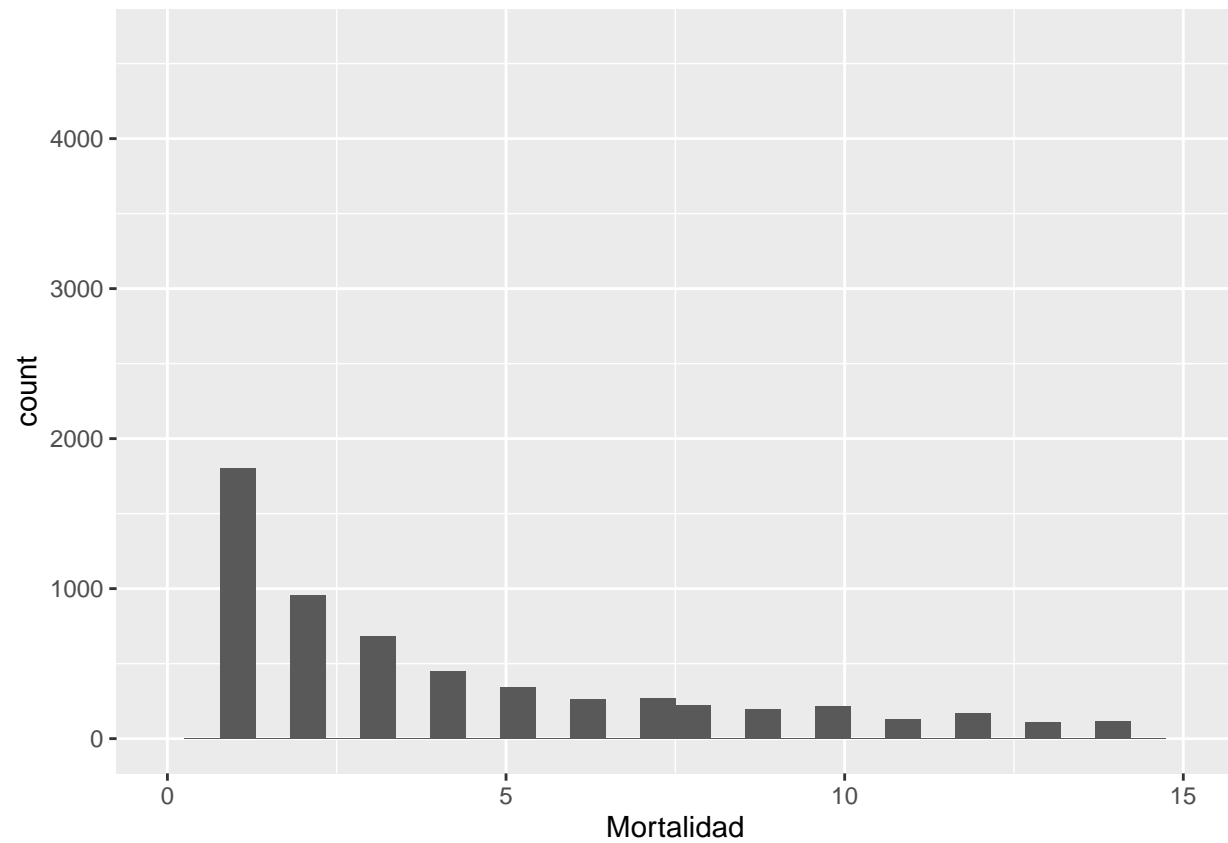
Distribución acumulada empírica

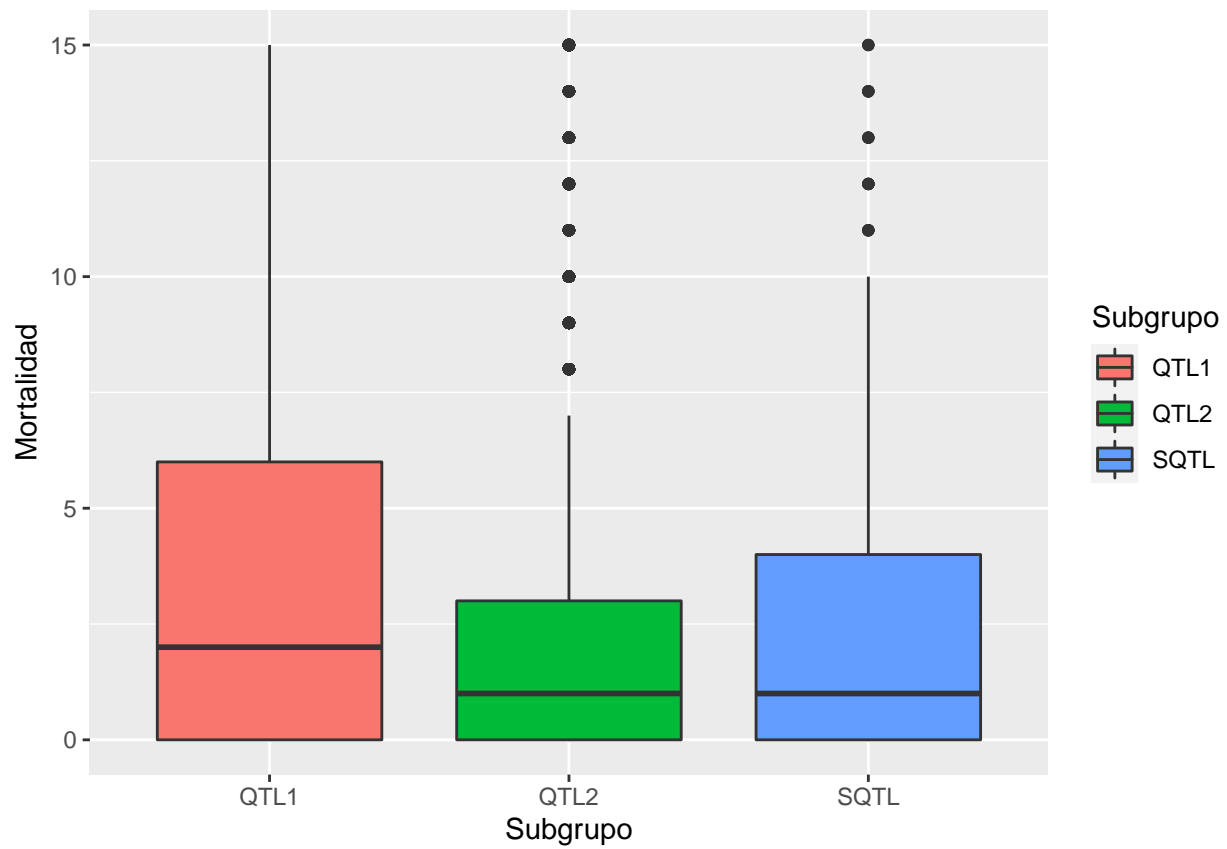


Mortalidad (n° peces) por subgrupo

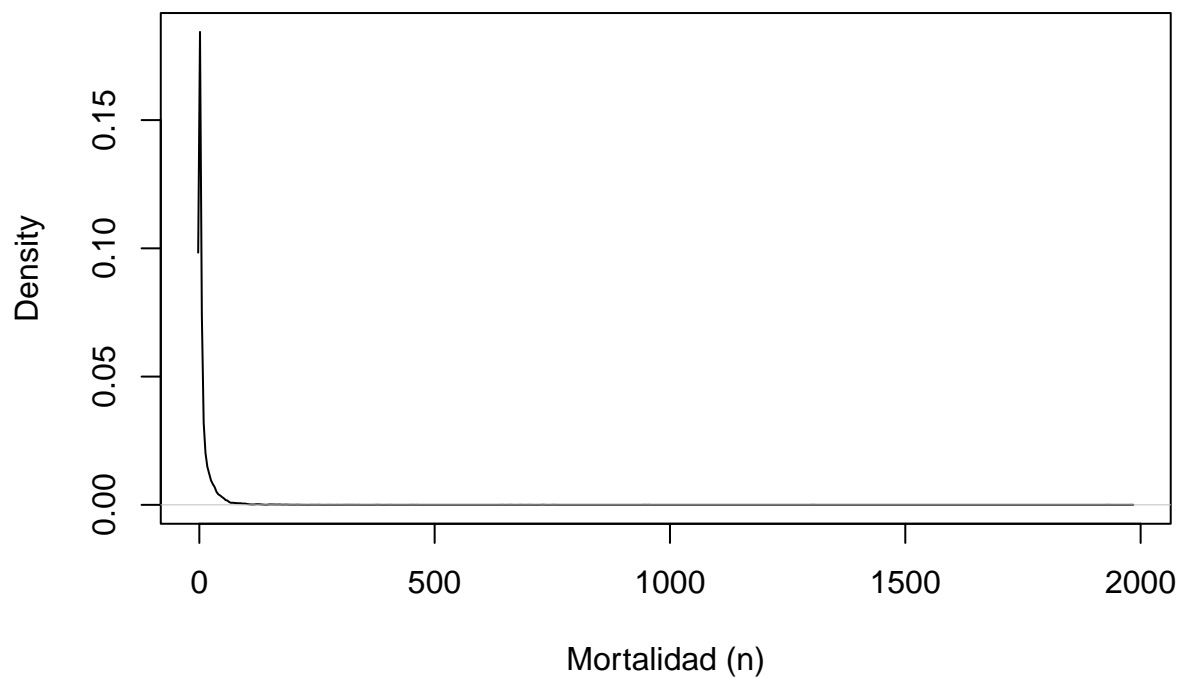


Para apreciar mejor como se comporta esta variable, haremos un histograma ajustando el eje x

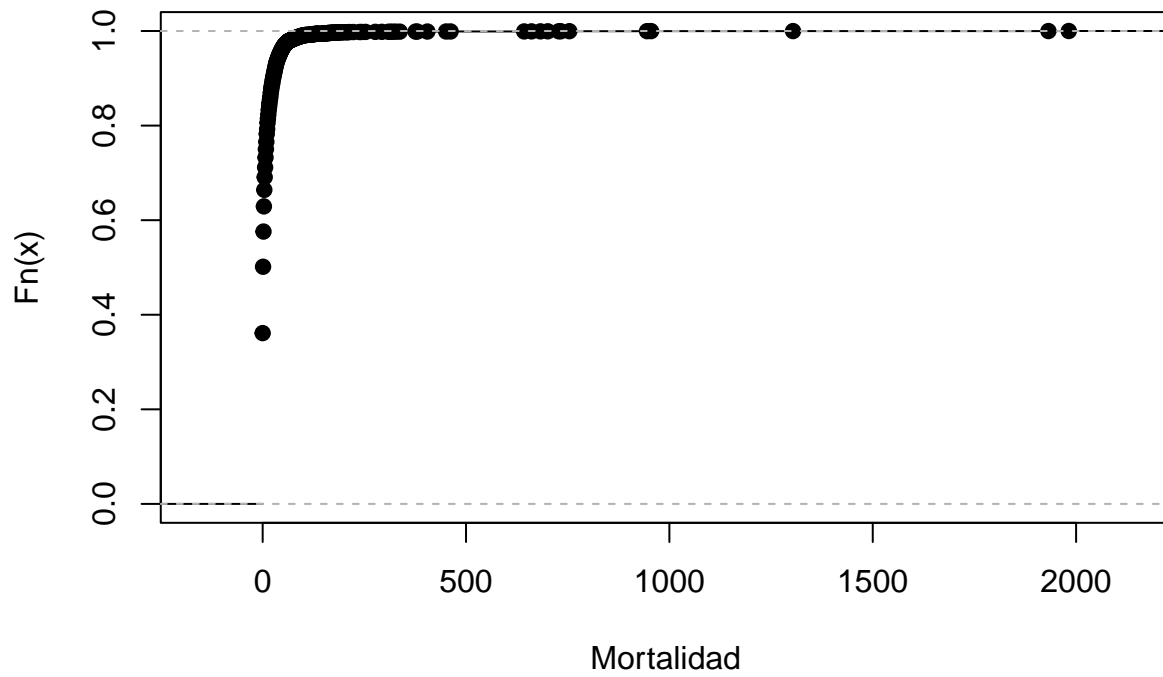




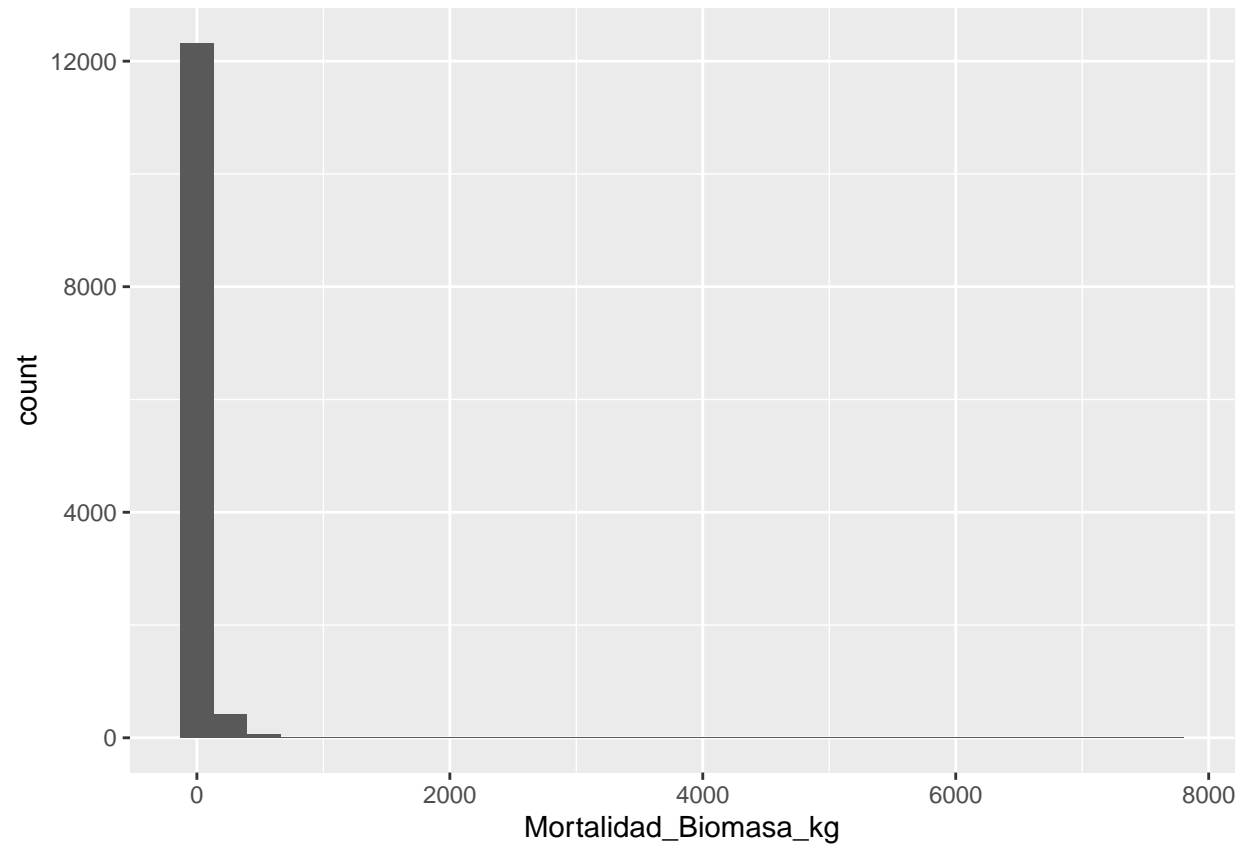
Densidad empírica

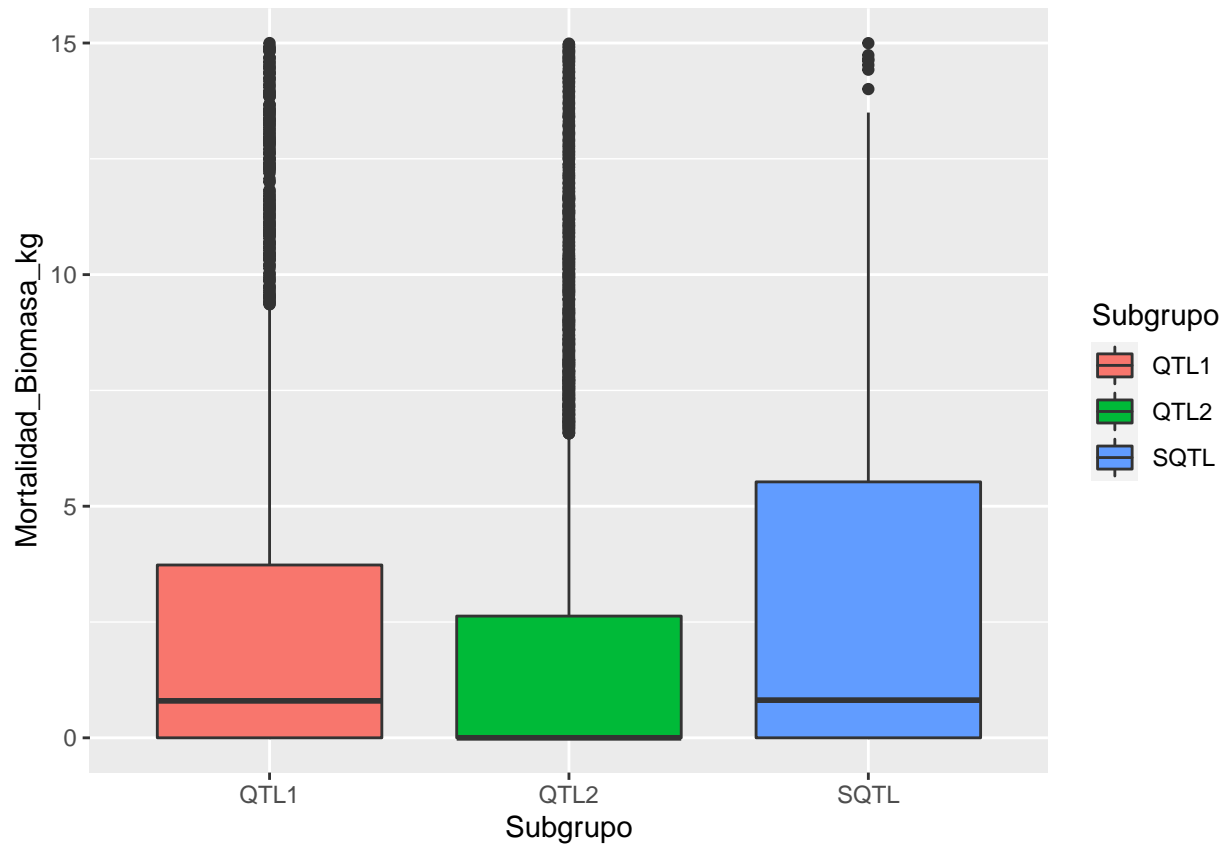


Distribución acumulada empírica

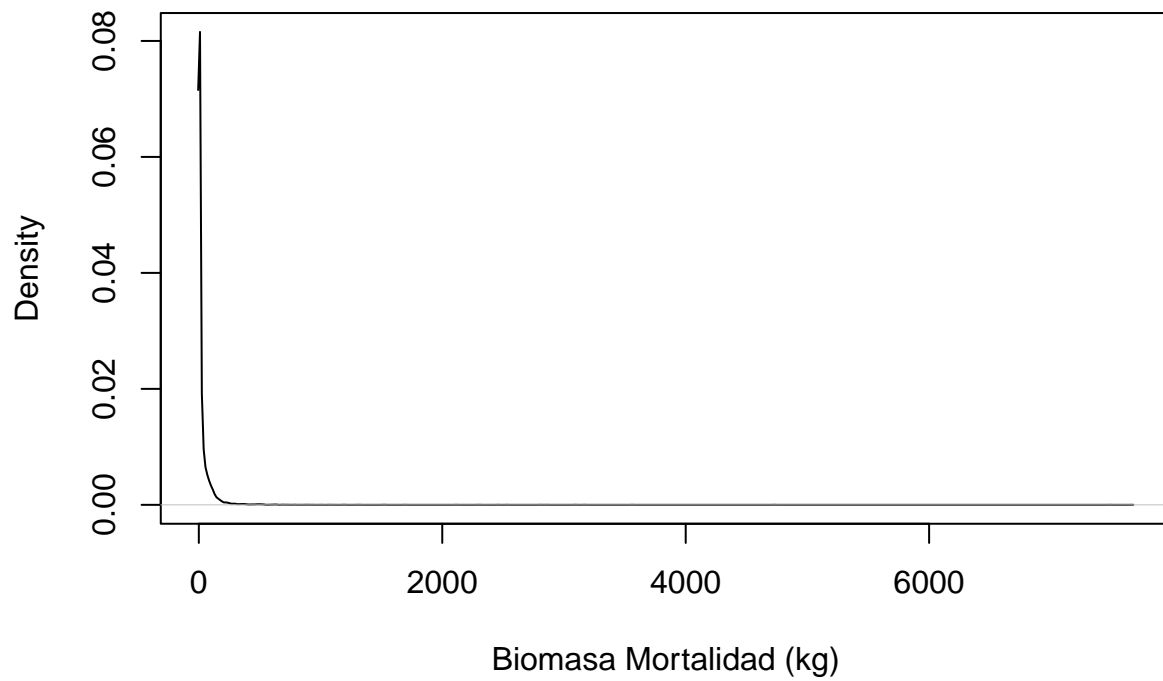


Biomasa de la mortalidad por subgrupo

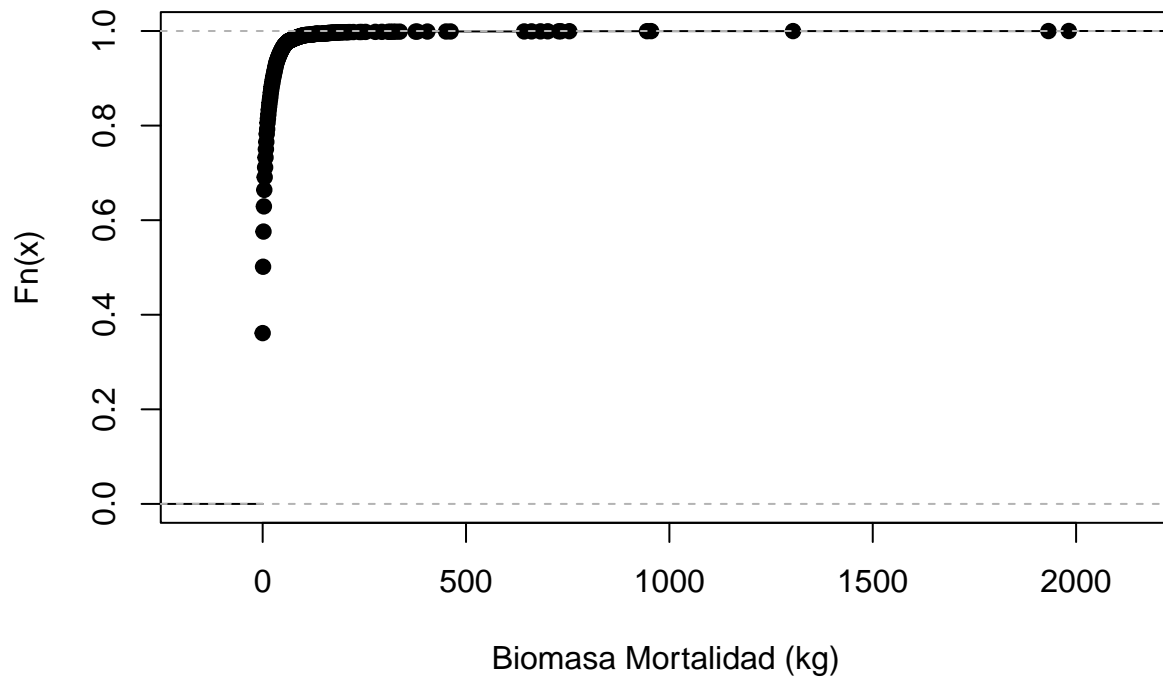




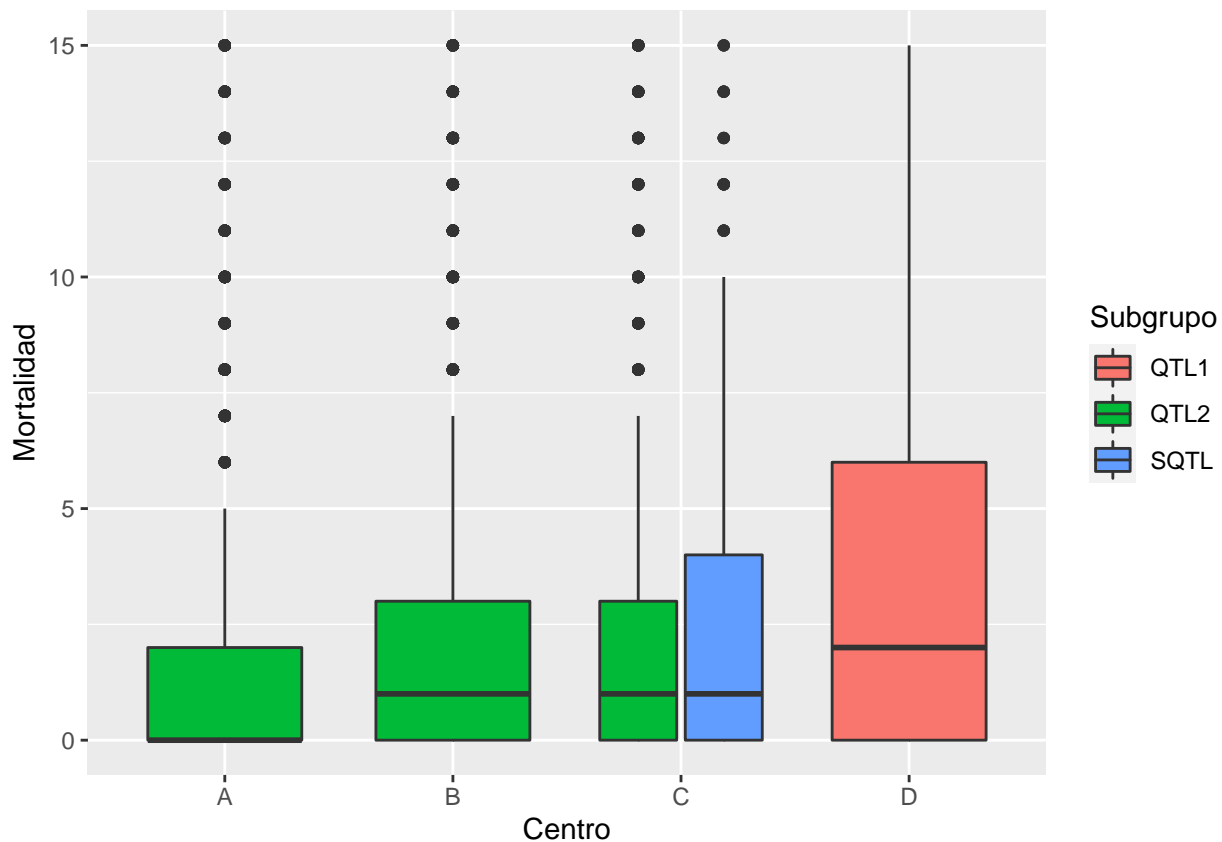
Densidad empírica



Distribución acumulada empírica



Hacemos boxplot de variable respuesta en función de otras variables de interés.



Transformaciones de variables a factor

```
##
## QTL1 QTL2 SQTL
## 3256 9078 491

##
##   A    B    C    D
## 4130 3222 2217 3256

## tibble [12,825 x 12] (S3: tbl_df/tbl/data.frame)
##  $ Fecha           : POSIXct[1:12825], format: "2019-08-24" "2019-08-25" ...
##  $ Centro           : Factor w/ 4 levels "A","B","C","D": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Unidad            : Factor w/ 24 levels "101","102","103",...: 19 19 19 19 19 19 19 19 19 19 ...
##  $ Subgrupo          : Factor w/ 3 levels "QTL1","QTL2",...: 2 2 2 2 2 2 2 2 2 2 ...
##  $ N_peces           : num [1:12825] 16302 16302 16302 16302 16302 ...
##  $ Biomasa_kg         : num [1:12825] 35625 35879 36261 36684 37100 ...
##  $ Mortalidad         : num [1:12825] 0 0 0 0 0 0 0 1 2 2 ...
##  $ Mortalidad_Biomasa_kg: num [1:12825] 0 0 0 0 0 ...
##  $ N_cosecha          : num [1:12825] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Biomasa_cosecha     : num [1:12825] 0 0 0 0 0 0 0 0 0 0 ...
##  $ Alimento_kg        : num [1:12825] 455 300 450 500 490 396 472 562 435 556 ...
##  $ Temperatura        : num [1:12825] 11.3 11.1 11.1 11.3 10.3 ...
```

```

##      Fecha                Centro      Unidad      Subgrupo
## Min.    :2019-02-09 00:00:00 A:4130    105      : 909    QTL1:3256
## 1st Qu.:2019-06-01 00:00:00 B:3222    107      : 902    QTL2:9078
## Median :2019-07-26 00:00:00 C:2217    104      : 900    SQTL: 491
## Mean   :2019-07-25 16:53:46 D:3256    110      : 893
## 3rd Qu.:2019-09-20 00:00:00      106      : 890
## Max.    :2019-12-09 00:00:00      109      : 889
##                                     (Other):7442
##      N_peces      Biomasa_kg      Mortalidad      Mortalidad_Biomasa_kg
## Min.    :    0      Min.    :    0      Min.    :    0.000      Min.    :    0.000
## 1st Qu.:16308      1st Qu.: 23766      1st Qu.:    0.000      1st Qu.:    0.000
## Median :28340      Median : 42046      Median :    1.000      Median :    2.015
## Mean   :24143      Mean   : 44360      Mean   :    9.889      Mean   :   27.549
## 3rd Qu.:29028      3rd Qu.: 59946      3rd Qu.:    8.000      3rd Qu.:   17.845
## Max.    :29209      Max.    :114599      Max.    :1982.000      Max.    :7671.965
##
##      N_cosecha      Biomasa_cosecha      Alimento_kg      Temperatura
## Min.    :    0.0      Min.    :    0.0      Min.    :    0.0      Min.    :    0.00
## 1st Qu.:    0.0      1st Qu.:    0.0      1st Qu.: 285.0      1st Qu.:10.79
## Median :    0.0      Median :    0.0      Median : 446.0      Median :11.20
## Mean   :   76.9      Mean   : 270.1      Mean   : 440.6      Mean   :11.15
## 3rd Qu.:    0.0      3rd Qu.:    0.0      3rd Qu.: 604.0      3rd Qu.:11.70
## Max.    :22602.0      Max.    :78009.3      Max.    :1560.0      Max.    :15.00
##

```

Determinar si los datos están balanceados

Calculamos el número de observaciones por Subgrupo. Luego expresamos la cifra como proporción.

```

##
##      A      B      C      D
## QTL1    0    0    0 3256
## QTL2 4130 3222 1726    0
## SQTL    0    0 491    0
##
##      Centro
## Subgrupo      A      B      C      D
## QTL1 0.0000000 0.0000000 0.0000000 0.2538791
## QTL2 0.3220273 0.2512281 0.1345809 0.0000000
## SQTL 0.0000000 0.0000000 0.0382846 0.0000000

```

Table 1: Proporción datos por Subgrupo de peces en diferentes Centros

	A	B	C	D
QTL1	0.0000000	0.0000000	0.0000000	0.2538791
QTL2	0.3220273	0.2512281	0.1345809	0.0000000
SQTL	0.0000000	0.0000000	0.0382846	0.0000000

Los datos no están balanceados. Hay más observaciones para el Subgrupo QTL2 (70,7%) que para otros subgrupos. El Subgrupo en menor proporción es el SQTL que está presente solo en Centro C.

Según lo sugerido en la evaluación 1, se agregan valores de estadística descriptiva al set de datos en formato de tabla.

Table 2: Media y varianza Mortalidad

Media_M	Var_M	sd	n
9.889123	1590.286	39.87839	12825

Table 3: Media y varianza Mortalidad por Subgrupo

Subgrupo	Media_M	Var_M	sd	n
QTL1	11.124386	880.1255	29.66691	3256
QTL2	9.599251	1923.0556	43.85266	9078
SQTL	7.057027	130.1478	11.40823	491

Table 4: Media y varianza Mortalidad por Centro y Subgrupo

Centro	Subgrupo	Media_M	Var_M	sd	n
A	QTL2	9.394431	3231.8437	56.84931	4130
B	QTL2	8.299193	403.1877	20.07953	3222
C	QTL2	12.516222	1618.7368	40.23353	1726
C	SQTL	7.057027	130.1478	11.40823	491
D	QTL1	11.124386	880.1255	29.66691	3256

FIN DE CURSO. ANALISIS ESTADISTICO

Se agrega una nueva variable al dataset (Days), la cual da cuenta en el tiempo del comportamiento de la variable respuesta (Mortalidad). Cargamos dicho dataset.

```
##      Fecha                Days                Centro
## Min.   :2019-02-09 00:00:00 Min.   : 1.0 Length:12817
## 1st Qu.:2019-06-01 00:00:00 1st Qu.: 61.0 Class :character
## Median :2019-07-26 00:00:00 Median :119.0 Mode  :character
## Mean   :2019-07-25 18:14:44 Mean   :116.5
## 3rd Qu.:2019-09-20 00:00:00 3rd Qu.:172.0
## Max.   :2019-12-09 00:00:00 Max.   :247.0
##      Unidad      Subgrupo      N_peces      Biomasa_kg
## Length:12817      Length:12817      Min.   :    0      Min.   :    0
## Class :character  Class :character  1st Qu.:16308  1st Qu.: 23857
```

```
## Mode :character Mode :character Median :28345 Median : 42067
## Mean :24153 Mean : 44384
## 3rd Qu.:29029 3rd Qu.: 59960
## Max. :29209 Max. :114599
## Mortalidad Mortalidad_Biomasa_kg N_cosecha Biomasa_cosecha
## Min. : 0.000 Min. : 0.000 Min. : 0.00 Min. : 0.0
## 1st Qu.: 0.000 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.0
## Median : 1.000 Median : 2.017 Median : 0.00 Median : 0.0
## Mean : 9.895 Mean : 27.566 Mean : 76.95 Mean : 270.3
## 3rd Qu.: 9.000 3rd Qu.: 17.866 3rd Qu.: 0.00 3rd Qu.: 0.0
## Max. :1982.000 Max. :7671.965 Max. :22602.00 Max. :78009.3
## Alimento_kg Temperatura
## Min. : 0.0 Min. : 0.00
## 1st Qu.: 285.0 1st Qu.:10.79
## Median : 447.0 Median :11.20
## Mean : 440.8 Mean :11.16
## 3rd Qu.: 604.0 3rd Qu.:11.70
## Max. :1560.0 Max. :15.00
## tibble [12,817 x 13] (S3: tbl_df/tbl/data.frame)
## $ Fecha : POSIXct[1:12817], format: "2019-08-24" "2019-08-25" ...
## $ Days : num [1:12817] 1 2 3 4 5 6 7 8 9 10 ...
## $ Centro : chr [1:12817] "A" "A" "A" "A" ...
## $ Unidad : chr [1:12817] "205" "205" "205" "205" ...
## $ Subgrupo : chr [1:12817] "QTL2" "QTL2" "QTL2" "QTL2" ...
## $ N_peces : num [1:12817] 16302 16302 16302 16302 16302 ...
## $ Biomasa_kg : num [1:12817] 35625 35879 36261 36684 37100 ...
## $ Mortalidad : num [1:12817] 0 0 0 0 0 0 0 1 2 2 ...
## $ Mortalidad_Biomasa_kg: num [1:12817] 0 0 0 0 0 ...
## $ N_cosecha : num [1:12817] 0 0 0 0 0 0 0 0 0 0 ...
## $ Biomasa_cosecha : num [1:12817] 0 0 0 0 0 0 0 0 0 0 ...
## $ Alimento_kg : num [1:12817] 455 300 450 500 490 396 472 562 435 556 ...
## $ Temperatura : num [1:12817] 11.3 11.1 11.1 11.3 10.3 ...
## - attr(*, "na.action")= 'omit' Named int [1:4] 3257 3258 3259 3260
## ..- attr(*, "names")= chr [1:4] "3257" "3258" "3259" "3260"
```

Transformamos a factor lo que haya que transformar.

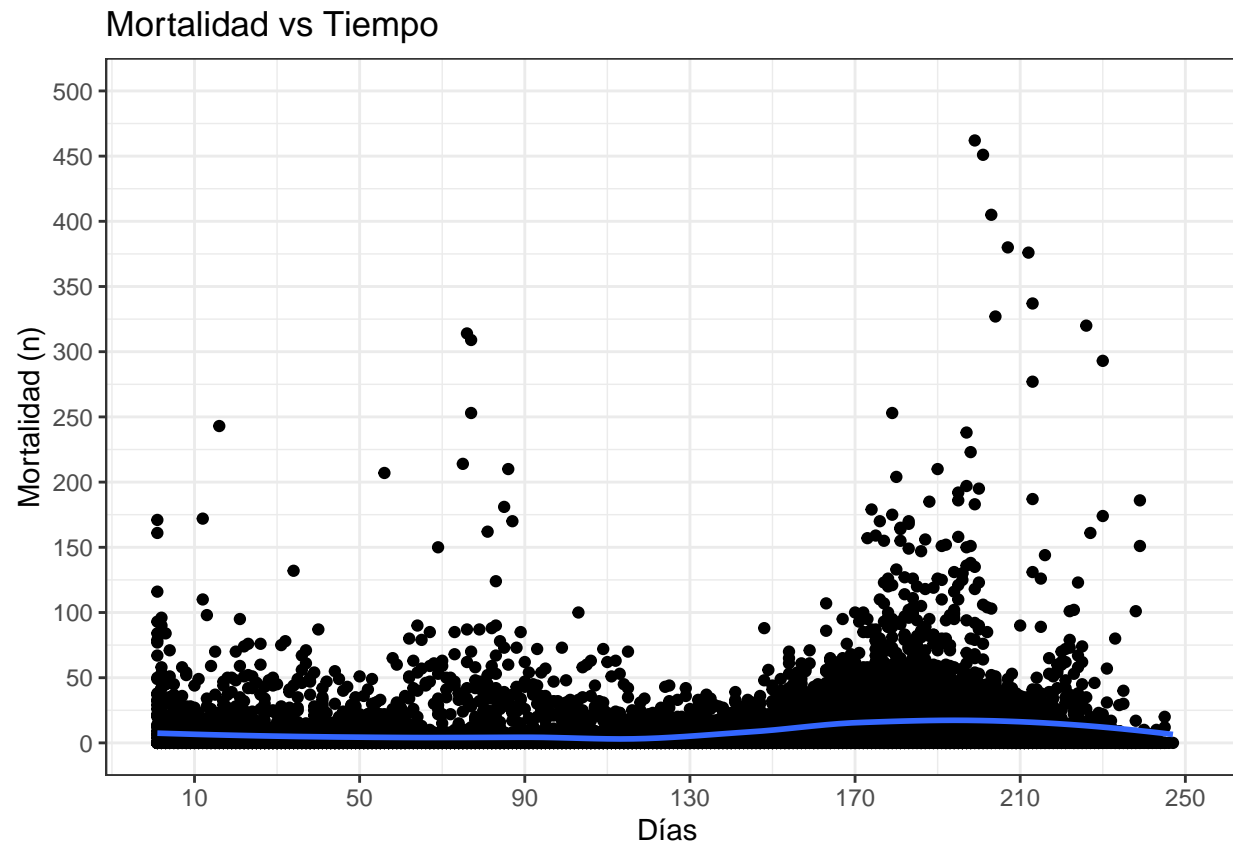
```
## tibble [12,817 x 13] (S3: tbl_df/tbl/data.frame)
## $ Fecha : POSIXct[1:12817], format: "2019-08-24" "2019-08-25" ...
## $ Days : num [1:12817] 1 2 3 4 5 6 7 8 9 10 ...
## $ Centro : Factor w/ 4 levels "A","B","C","D": 1 1 1 1 1 1 1 1 1 1 ...
## $ Unidad : Factor w/ 24 levels "101","102","103",...: 19 19 19 19 19 19 19 19 19 19 ...
## $ Subgrupo : Factor w/ 3 levels "QTL1","QTL2",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ N_peces : num [1:12817] 16302 16302 16302 16302 16302 ...
## $ Biomasa_kg : num [1:12817] 35625 35879 36261 36684 37100 ...
## $ Mortalidad : num [1:12817] 0 0 0 0 0 0 0 1 2 2 ...
## $ Mortalidad_Biomasa_kg: num [1:12817] 0 0 0 0 0 ...
## $ N_cosecha : num [1:12817] 0 0 0 0 0 0 0 0 0 0 ...
## $ Biomasa_cosecha : num [1:12817] 0 0 0 0 0 0 0 0 0 0 ...
## $ Alimento_kg : num [1:12817] 455 300 450 500 490 396 472 562 435 556 ...
## $ Temperatura : num [1:12817] 11.3 11.1 11.1 11.3 10.3 ...
## - attr(*, "na.action")= 'omit' Named int [1:4] 3257 3258 3259 3260
## ..- attr(*, "names")= chr [1:4] "3257" "3258" "3259" "3260"
```

```

##      Fecha              Days      Centro      Unidad
## Min.   :2019-02-09 00:00:00 Min.   : 1.0 A:4124 105 : 909
## 1st Qu.:2019-06-01 00:00:00 1st Qu.: 61.0 B:3221 107 : 902
## Median :2019-07-26 00:00:00 Median :119.0 C:2217 104 : 900
## Mean   :2019-07-25 18:14:44 Mean   :116.5 D:3255 110 : 892
## 3rd Qu.:2019-09-20 00:00:00 3rd Qu.:172.0      106 : 889
## Max.   :2019-12-09 00:00:00 Max.   :247.0      109 : 889
##                                     (Other):7436
## Subgrupo      N_peces      Biomasa_kg      Mortalidad
## QTL1:3255 Min.   : 0 Min.   : 0 Min.   : 0.000
## QTL2:9071 1st Qu.:16308 1st Qu.: 23857 1st Qu.: 0.000
## SQTL: 491 Median :28345 Median : 42067 Median : 1.000
##          Mean  :24153 Mean  : 44384 Mean  : 9.895
##          3rd Qu.:29029 3rd Qu.: 59960 3rd Qu.: 9.000
##          Max.   :29209 Max.   :114599 Max.   :1982.000
##
## Mortalidad_Biomasa_kg N_cosecha      Biomasa_cosecha      Alimento_kg
## Min.   : 0.000 Min.   : 0.00 Min.   : 0.0 Min.   : 0.0
## 1st Qu.: 0.000 1st Qu.: 0.00 1st Qu.: 0.0 1st Qu.: 285.0
## Median : 2.017 Median : 0.00 Median : 0.0 Median : 447.0
## Mean   : 27.566 Mean   : 76.95 Mean   : 270.3 Mean   : 440.8
## 3rd Qu.: 17.866 3rd Qu.: 0.00 3rd Qu.: 0.0 3rd Qu.: 604.0
## Max.   :7671.965 Max.   :22602.00 Max.   :78009.3 Max.   :1560.0
##
##      Temperatura
## Min.   : 0.00
## 1st Qu.:10.79
## Median :11.20
## Mean   :11.16
## 3rd Qu.:11.70
## Max.   :15.00
##

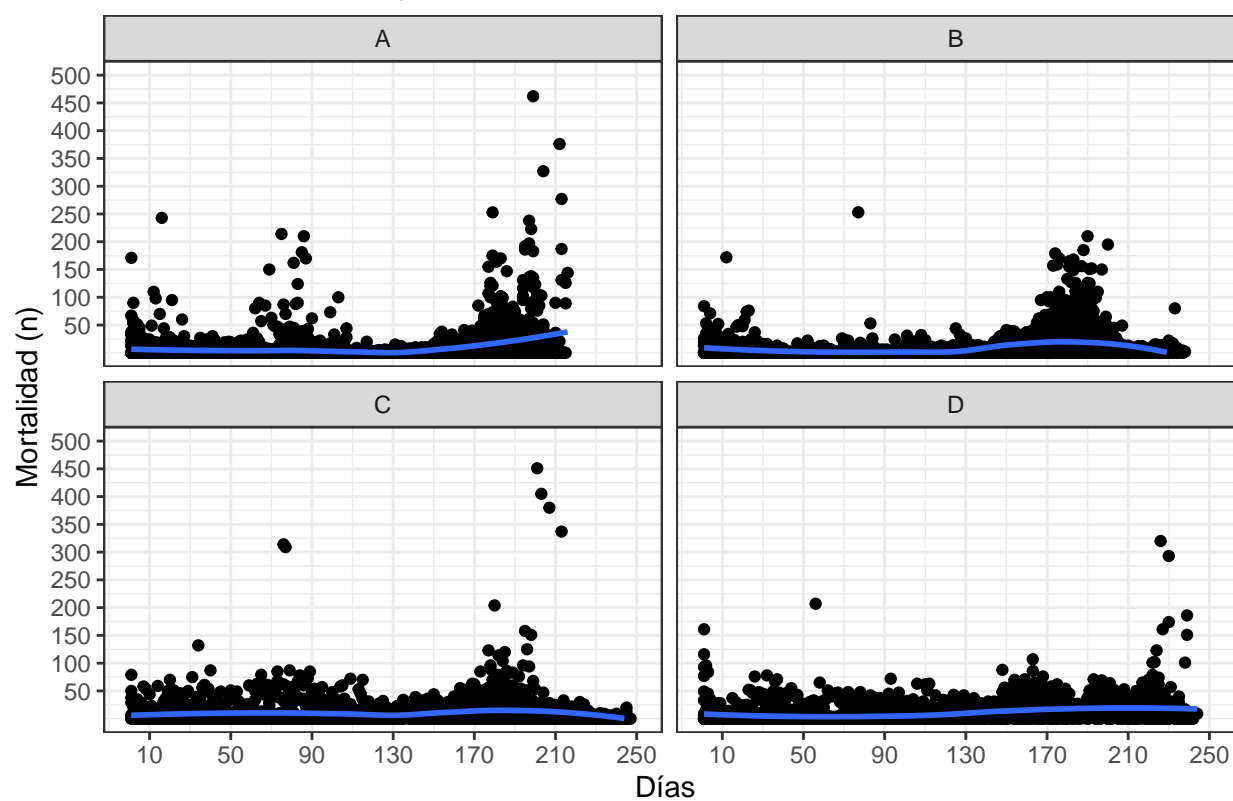
```


Exploramos el comportamiento de la variable respuesta en función del tiempo (Days)

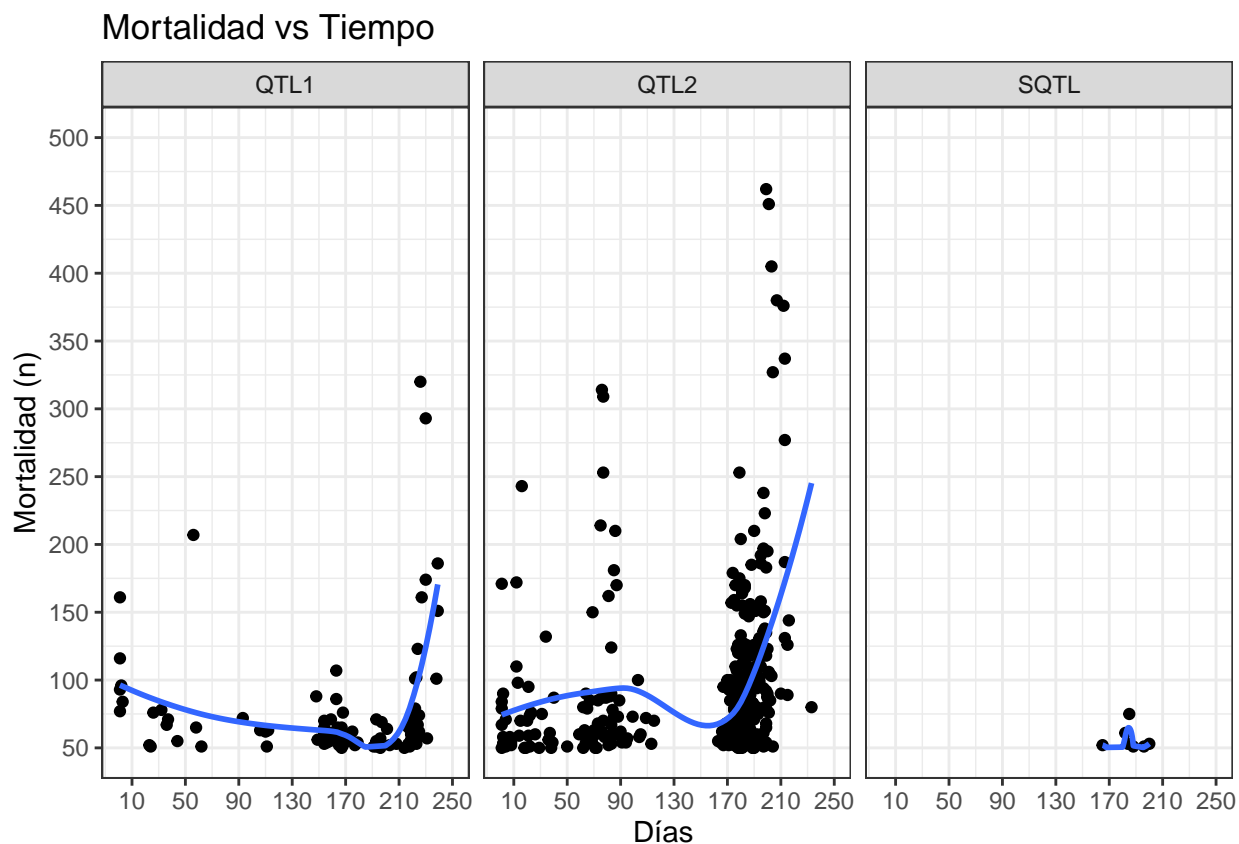


En el exploratorio se observa un comportamiento no lineal de la variable respuesta en los distintos centros.

Mortalidad vs Tiempo



Vemos la distribución de la variable respuesta vs tiempo en función del Subgrupo.



Agrupando datos por “Subgrupo”, se observa un comportamiento no lineal de la variable respuesta en los distintos centros.

Dada la distribución de datos, aplicamos el modelo lineal que se ajuste a una distribución de Poisson.

```
##
## Call:
## glm(formula = Mortalidad ~ Days, family = poisson, data = dat2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -7.373  -3.703  -2.656  -0.420  122.184
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.079e+00  7.798e-03  138.4   <2e-16 ***
## Days         9.001e-03  4.794e-05  187.7   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
```

```

##      Null deviance: 386860  on 12816  degrees of freedom
## Residual deviance: 348019  on 12815  degrees of freedom
## AIC: 377753
##
## Number of Fisher Scoring iterations: 6
##
## Call:
## glm(formula = Mortalidad ~ Days + Centro, family = poisson(),
##      data = dat2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -7.621   -3.671   -2.626   -0.468   121.069
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.122e+00  8.399e-03 133.625 < 2e-16 ***
## Days         9.134e-03  4.872e-05 187.467 < 2e-16 ***
## CentroB     -2.976e-01  8.016e-03 -37.120 < 2e-16 ***
## CentroC     -1.200e-02  8.185e-03  -1.466   0.143
## CentroD      3.601e-02  7.361e-03   4.892 9.98e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 386860  on 12816  degrees of freedom
## Residual deviance: 345874  on 12812  degrees of freedom
## AIC: 375615
##
## Number of Fisher Scoring iterations: 7
##
## Call:
## glm(formula = Mortalidad ~ Days + Subgrupo + Centro, family = poisson(),
##      data = dat2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -7.970   -3.658   -2.596   -0.432   121.035
##
## Coefficients: (1 not defined because of singularities)
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.153e+00  9.255e-03 124.551 < 2e-16 ***
## Days         9.170e-03  4.888e-05 187.604 < 2e-16 ***
## SubgrupoQTL2 -3.535e-02  7.362e-03  -4.802 1.57e-06 ***
## SubgrupoSQTL -6.335e-01  1.971e-02 -32.137 < 2e-16 ***
## CentroB     -2.983e-01  8.016e-03 -37.213 < 2e-16 ***
## CentroC      9.437e-02  8.558e-03  11.027 < 2e-16 ***
## CentroD              NA           NA      NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)

```

```
##
## Null deviance: 386860 on 12816 degrees of freedom
## Residual deviance: 344645 on 12811 degrees of freedom
## AIC: 374387
##
## Number of Fisher Scoring iterations: 7
```

Se observa que el set de datos tiene muchos ceros. Por lo tanto, aplicamos un Modelo Poisson inflado con ceros. Se habilita paquete pscl previamente.

```
##
## Call:
## zeroinfl(formula = Mortalidad ~ Days, data = dat2, dist = "poisson")
##
## Pearson residuals:
##      Min      1Q   Median      3Q      Max
## -1.9220 -1.1503 -0.8904 -0.1791 205.4033
##
## Count model coefficients (poisson with log link):
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.8116478  0.0078852   229.8  <2e-16 ***
## Days        0.0066502  0.0000484   137.4  <2e-16 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.1770205  0.0372533    4.752 2.02e-06 ***
## Days       -0.0066360  0.0002951  -22.490 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 1
## Log-likelihood: -1.493e+05 on 4 Df
##
## Call:
## zeroinfl(formula = Mortalidad ~ Days + Subgrupo, data = dat2, dist = "poisson")
##
## Pearson residuals:
##      Min      1Q   Median      3Q      Max
## -2.4673 -1.1349 -0.8656 -0.1874 194.6595
##
## Count model coefficients (poisson with log link):
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.733e+00  9.129e-03  189.80  <2e-16 ***
## Days        6.763e-03  4.869e-05  138.89  <2e-16 ***
## SubgrupoQTL2 1.174e-01  6.255e-03   18.77  <2e-16 ***
## SubgrupoSRTL -4.664e-01  1.786e-02  -26.12  <2e-16 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.4263917  0.0518695   -8.220 < 2e-16 ***
## Days       -0.0068131  0.0003008  -22.650 < 2e-16 ***
```

```
## SubgrupoQTL2  0.8291136  0.0474317  17.480  < 2e-16 ***
## SubgrupoSQL  0.4179369  0.1091594   3.829  0.000129 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 1
## Log-likelihood: -1.484e+05 on 8 Df

##
## Call:
## zeroinfl(formula = Mortalidad ~ Days + Centro, data = dat2, dist = "poisson")
##
## Pearson residuals:
##      Min       1Q   Median       3Q      Max
## -2.4519 -1.1300 -0.8612 -0.1850 174.7783
##
## Count model coefficients (poisson with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.934e+00  8.621e-03  224.39  < 2e-16 ***
## Days         7.071e-03  4.953e-05  142.77  < 2e-16 ***
## CentroB     -4.262e-01  8.049e-03  -52.94  < 2e-16 ***
## CentroC     -6.257e-02  8.126e-03   -7.70  1.36e-14 ***
## CentroD     -2.491e-01  7.340e-03  -33.93  < 2e-16 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.5026647  0.0449004  11.195  < 2e-16 ***
## Days        -0.0066303  0.0003023 -21.932  < 2e-16 ***
## CentroB     -0.2161890  0.0491345  -4.400  1.08e-05 ***
## CentroC     -0.3081023  0.0556512  -5.536  3.09e-08 ***
## CentroD     -0.9496128  0.0527367 -18.007  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 10
## Log-likelihood: -1.475e+05 on 10 Df
```

No se puede correr Modelo Poisson inflado con ceros usando los factores Subgrupo y Centro al mismo tiempo.

Para comparar los modelos, usamos la función AIC.

Table 5: Comparación de modelos

	df	AIC
modelo0	2	377752.6
modelo1	5	375614.6
modelo2	6	374387.0
modelo3	4	298668.6
modelo4	8	296784.1
modelo5	10	294955.8

De todos los modelos generados, el modelo 5 tiene menor valor AIC. Por lo tanto, este es el modelo que mejor se ajusta a los datos.

Conclusiones: En el modelo 5, todas las variables regresoras aportan significativamente al modelo (se rechaza H_0 que no hay efecto del coeficiente). Cuando no se aplica un modelo con distribución de Poisson inflado con ceros, la variable de clasificación “Centro C” (ver modelo 2) o “Centro D” (modelo 3) no aportan significativamente al modelo.