

Nuevas Exploraciones

Juan C. Correa

3/18/2021

1. Introducción

Pantaleon ha hecho un gran esfuerzo por aumentar la producción de caña y azúcar en el último año y los datos nos alegran al mostrar el comportamiento estadístico logrado en la gestión de los últimos 12 meses. En este informe, nos concentramos en presentar algunos hechos relevantes de la producción en Pantaleon a partir del siguiente pool de variables:

- TCH (Toneladas de Caña por Hectárea)
- TERCIO
- rendimiento
- Madurante
- Cuadrante
- grup_var
- Premadurante

```
library(readxl)
AOV <- read_excel("Documents/GitHub/Pantaleon/PantaleonData4.xlsx")
```

```
## New names:
## * TOTAL -> TOTAL...23
## * CONTAR -> CONTAR...24
## * edad -> edad...33
## * edad -> edad...39
## * TOTAL -> TOTAL...88
## * ...
```

```
DATOS <- AOV[c(7, 10, 19, 42, 43, 54, 67, 68)]
variable.names(DATOS)
```

```
## [1] "TCH"          "rendimiento"  "Cuadrante"    "grup_var"     "Status"
## [6] "TERCIO"       "Madurante"    "Premadurante"
```

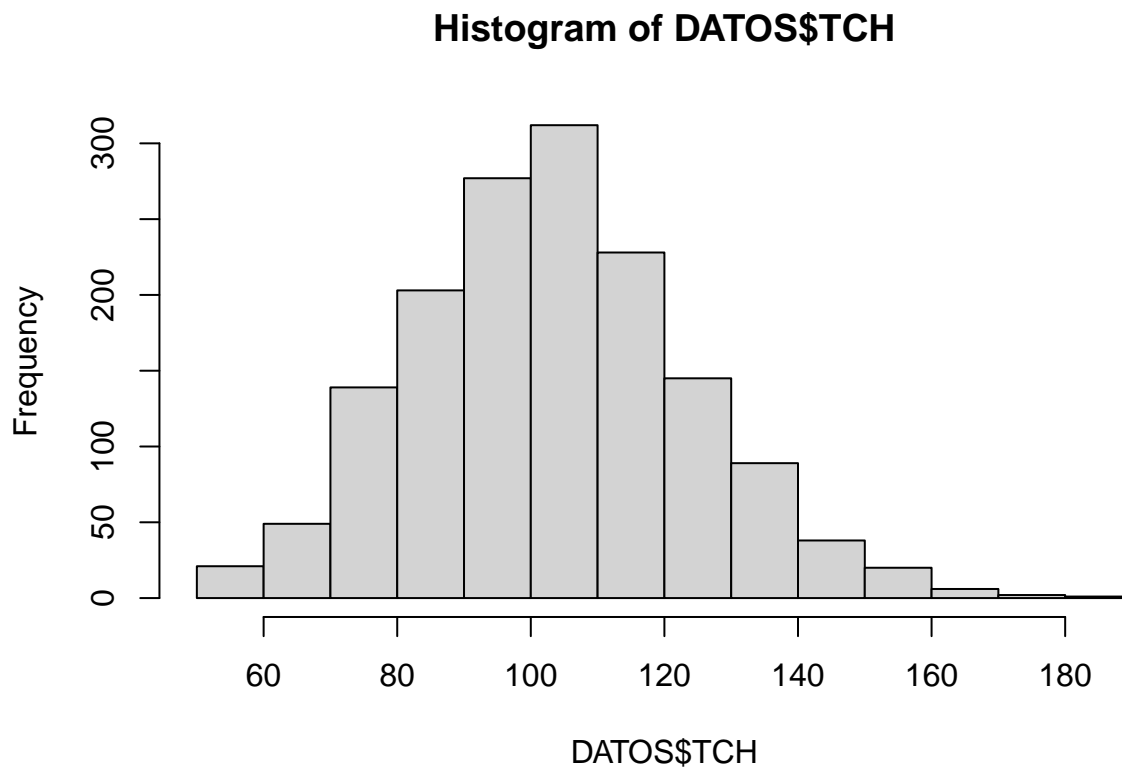
2. Toneladas de Caña por Hectárea (TCH)

Una de las variable clave de la producción en Pantaleon es el total de toneladas de caña por hectárea que se produce en cada fina o lote. Acá mostramos cómo fue el comportamiento estadístico de TCH en general. La producción mínima observada fue de 50.97, la máxima fue de 181.81 y la producción promedio alcanzó un total de 102.91 toneladas por hectárea.

```
summary(DATOS$TCH)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    50.97   88.69   102.50   102.91   116.83   181.81
```

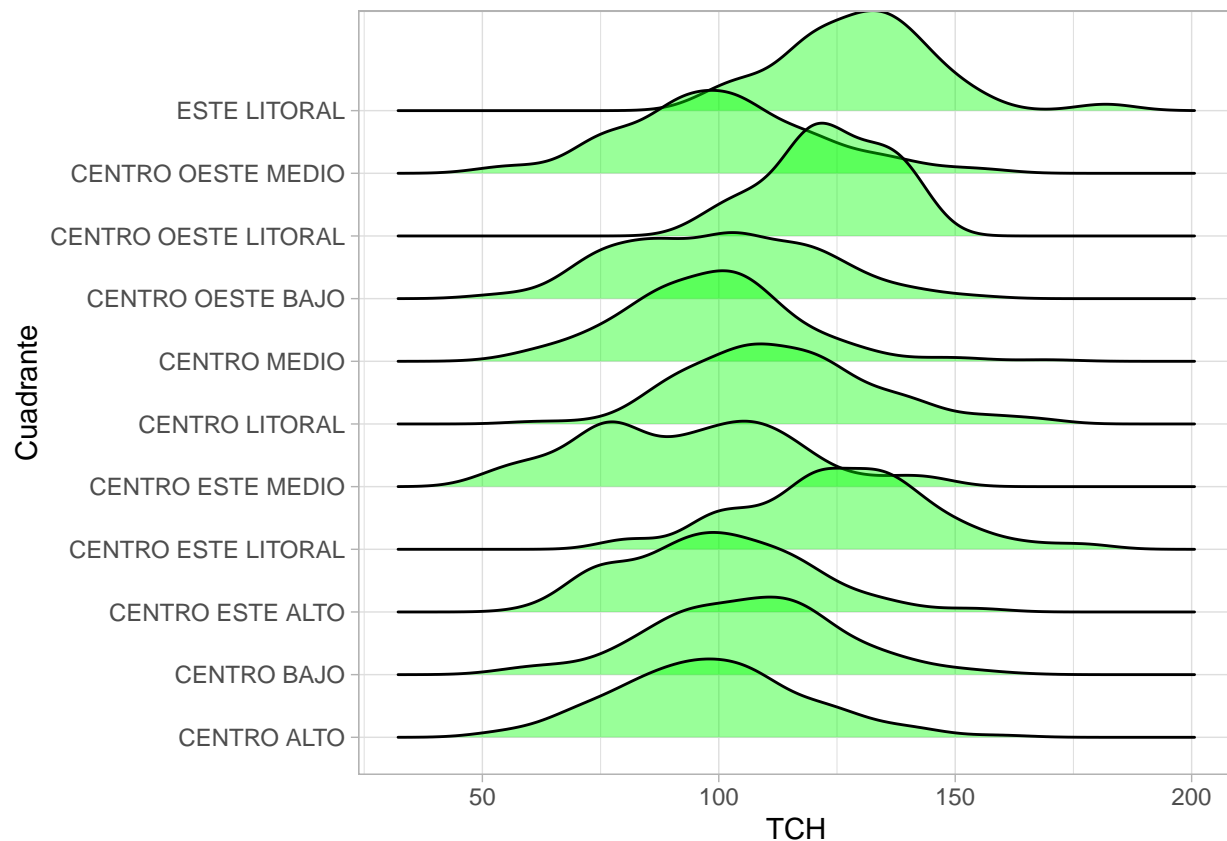
```
hist(DATOS$TCH)
```



La variable TCH se comportó de manera diferente según el cuadrante en el que se cosechó. La siguiente gráfica nos presenta este comportamiento diferencial.

```
library(ggplot2)
library(ggribes)
ggplot(DATOS,
  aes(x=TCH, y=Cuadrante)) +
  geom_density_ridges(fill="green", alpha = 0.4) +
  ylab("Cuadrante") + xlab("TCH") +
  theme(axis.text.y = element_text(family="Arial", face="bold", colour="black", size=rel(1))) +
  theme(axis.text.x = element_text(family="Arial", face="bold", colour="black", size=rel(1))) +
  theme_light()
```

```
## Picking joint bandwidth of 6.27
```



```
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
```

```
##
## %+%, alpha
```

```
describeBy(DATOS$TCH, group = DATOS$Cuadrante, mat = TRUE, digits = 0)
```

```
##      item      group1 vars  n mean sd median trimmed mad min max
## X11      1  CENTRO ALTO    1 229  98 20    98     98 18 52 161
## X12      2  CENTRO BAJO    1 296 106 19   106    106 19 53 162
## X13      3  CENTRO ESTE ALTO  1 136  99 19    99     99 19 61 156
## X14      4  CENTRO ESTE LITORAL  1  51 125 20   124    125 19 81 177
## X15      5  CENTRO ESTE MEDIO  1  65  93 22    93     93 25 51 145
## X16      6  CENTRO LITORAL    1 109 113 20   111    112 18 61 168
## X17      7  CENTRO MEDIO     1 337  98 18    98     98 16 52 170
## X18      8  CENTRO OESTE BAJO  1 159 100 21   100    100 23 51 155
## X19      9  CENTRO OESTE LITORAL  1  18 124 12   122    124 12 98 141
## X110     10  CENTRO OESTE MEDIO  1  91 100 20   100    100 17 52 157
## X111     11  ESTE LITORAL    1  39 130 16   130    129 16 97 182
##      range skew kurtosis se
## X11     109    0        0  1
## X12     109    0        0  1
## X13      95    0        0  2
## X14      96    0        0  3
```

```
## X15      94      0      -1  3
## X16     106      0       0  2
## X17     118      1       1  1
## X18     104      0      -1  2
## X19      42      0      -1  3
## X110    105      0       0  2
## X111     85      1       1  3
```

Resulta evidente que el total de TCH muestra una distribución estadística diferente según su cuadrante. Sin embargo, desde el punto de vista estadístico estas diferencias resultan ser estadísticamente significativas ($F = 25.77$, $df = 10$, $p < 2e-16$); dicho de otra manera, las diferencias observadas en el TCH por cuadrante son importantes o meritorias de mayor rigor en términos de calidad o volumen de producción.

```
modelo1 <- aov(DATOS$TCH ~ DATOS$Cuadrante)
summary(modelo1)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## DATOS$Cuadrante  10  96062    9606   25.77 <2e-16 ***
## Residuals      1519 566178     373
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(e1071)
skewness(modelo1$residuals)
```

```
## [1] 0.2450868
```

```
kurtosis(modelo1$residuals)
```

```
## [1] 0.2615575
```

```
skewness(DATOS$TCH)
```

```
## [1] 0.2511583
```