

Tabulación versus Graficación de Datos Estadísticos

Juan C. Correa

Material de uso exclusivo para
INGENIO PANTALEON, S.A.
Diagonal 6, 10-31, Zona 10

Ciudad de Guatemala



- 1 Tabulación de datos estadísticos
- 2 Tabulación versus Graficación: Cuándo usarlas
- 3 Ejemplo y Asignación



Tabulación de datos estadísticos

Los resultados de calcular las tres propiedades de una distribución:

- Tendencia Central (i.e., promedio, mediana, modo)
- Dispersión (i.e., desviación estándar, varianza, rango, rango intercuartilar)
- Forma (asimetría y kurtosis)

Usualmente deben disponerse en un formato que facilite la lectura y la comparación. Por ejemplo, para entender diferencias en la producción o el rendimiento de un lote o parcela en comparación con otro. Las Tablas cumplen ese propósito, al igual que lo hacen las gráficas.



Tabulación versus Graficación

Si **las tablas** y también **los gráficos** cumplen el mismo propósito de transmitir información, ¿pueden usarse indiscriminadamente?



Tabulación de datos estadísticos

Veamos qué ocurre cuando aplicamos la siguiente sintaxis

```
library(apaTables)
apa.1way.table(iv = Estacion,
               dv = TCH,
               data = PantaleonData,
               show.conf.interval = TRUE)
```



Tabulación de datos estadísticos

Descriptive statistics for TCH as a function of Estacion.

Estacion	M	M_95%_CI	SD
Amazonas	125.30	[119.22, 131.39]	20.02
Bonanza	119.98	[109.97, 129.99]	24.78
Bouganvilia	103.59	[99.33, 107.86]	19.41
Cengicana	101.85	[99.17, 104.52]	19.89
Cocales	99.23	[93.52, 104.93]	21.50
Concepción	98.29	[95.38, 101.21]	19.59
Costa Brava	93.07	[89.79, 96.35]	18.67
El Balsamo	97.05	[94.60, 99.51]	18.11
Irlanda	123.51	[117.43, 129.60]	12.24
La Giralda	100.41	[93.35, 107.47]	15.92
Naranjales	107.68	[93.57, 121.78]	15.25
Puyumate	100.97	[97.97, 103.96]	19.99
San Antonio EV	112.55	[101.54, 123.56]	15.39
San Rafael	128.72	[123.81, 133.63]	16.54
Tehuantepeq	105.93	[103.77, 108.08]	19.45

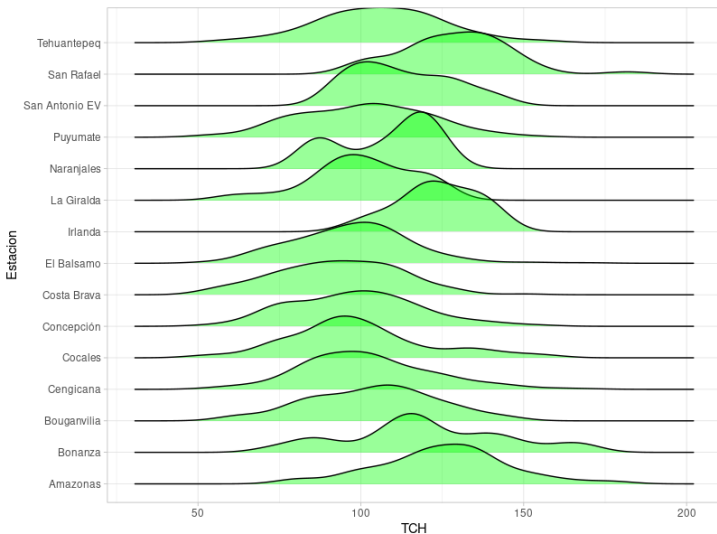


Graficación de datos estadísticos

```
library(ggplot2)
library(ggribes)
ggplot(PantaleonData,
  aes(x=TCH, y=Estacion)) +
  geom_density_ridges(fill="green", alpha = 0.4) +
  ylab("Estacion") + 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()
```



Graficación de datos estadísticos



Cuándo y cómo usar tabulación/graficación

- La tabulación enfatiza precisiones numéricas de los estadísticos reportados en la tabla que no son evidentes en las gráficas.
- La graficación suele ser más “amigable” a la vista de la audiencia en comparación con una fría tabla que muestra números.
- Si se quiere describir una variable, a veces es suficiente con la prosa (siendo innecesarias las tabulaciones y las gráficas). Con dos o más variables, hay que usar tabulación y graficación.
- La audiencia a la cual se dirige un informe debe ser el criterio más importante para decidir cómo reportar los resultados.
- La mejor práctica, sin embargo, es aquella que hace un uso inteligente de tabulaciones y graficaciones.



Veamos el siguiente artículo a modo de ejemplo

<https://www.sciencedirect.com/science/article/pii/S0969698918302339?via%3Dihub>, cuyo repositorio está disponible <https://github.com/jcorrean/WebMining-OFD> ¿Con los datos de

Pantaleon, sería posible preparar un informe de gestión en el que se combine de manera inteligente la tabulación y la graficación? De ser posible, ¿Cuáles serían las relaciones entre variables que más le interesarían y por qué?

