

# Matemáticas Computacionales

## Práctica 2: Estudio de una Base Datos

1904381 Itzel Guadalupe Vega Yañez  
Semestre Febrero - Junio 2021

03 de Marzo del 2021

### 1. Introducción

En esta práctica se estudiara una base de datos con la estadística descriptiva básica. Se analizara los atributos, tipos de datos, indicadores básicos de una y de dos variables. Se muestran algunos de los primeros pasos para luego implementar un preprocesamiento de datos para luego utilizarla en alguna experimentación o trabajo a tratar.

### 2. Base de datos: Tabaquismo, alcohol y (O)cáncer de esófago

La base de datos es un estudio de casos y controles de (o) cáncer de esófago en Ille-et-Vilaine, Francia (esoph), en Ille-et-Vilaine (Francia) se ha llevado a cabo un estudio retrospectivo de casos y controles. Los logaritmos de los riesgos relativos de desarrollar la enfermedad aumentan linealmente con el consumo diario de alcohol y tabaco de forma independiente. Además de que se discuten brevemente las implicaciones prácticas para fines de salud pública.[3]

La incidencia de cáncer de esófago es alta en Francia y particularmente en Bretaña. El alcohol y el tabaquismo son los principales factores etiológicos del carcinoma de células escamosas, el tipo de cáncer de esófago más frecuente en esa región. Esta combinación de alcohol y tabaquismo también predispone a la cirrosis, siendo frecuente la cirrosis alcohólica y el tabaquismo aumenta la gravedad de las enfermedades hepáticas[2]. Además del consumo de alcohol, el tabaquismo también es un factor etiológico importante generalmente aceptado para los cánceres del tracto digestivo superior. Los efectos multiplicadores y dependientes de la dosis del alcohol y el tabaco sobre el riesgo de ESCC se conocen desde hace décadas y se confirman en estudios y metanálisis[1].

#### 2.1. Descripción del conjunto de datos

El dataset esoph cuenta con un marco de datos con registro de 88 combinaciones de edad/alcohol/tabaco

1. Agegp: Es el grupo de edad a partir de los 25 años y se dividen en 6 grupos

Los 6 grupos se dividen en:

- 25-34
- 35-44
- 45-54
- 55-64
- 65-74
- 75+

2. Alcgp: Consumo de alcohol(g/dia) con 4 niveles

Los 4 niveles son:

- 0-39
- 40-79
- 80-119
- 120+

3. Tobgp: Consumo de tabaco(g/dia) con 4 niveles

Los 4 niveles son:

- 0-9
- 10-19
- 20-29
- 30+

4. "Ncases": Numero de casos

5. "Ncontrols": Numero de controles

## 2.2. Estadística descriptiva de una variable

Los datos de los grupos de edades, se pueden observar en la Figura(1).

Los datos del consumo de alcohol de las personas, se puede observar en la Figura(2).

Los datos del consumo de tabaco de las personas, se puede observar en la Figura(3).

Los datos de numero de casos tiene un mínimo de 0.00 y un máximo de 17.00, media de 2.273 y mediana de 1.00. La Figura(4)se muestra la grafica de densidad de los numeros de casos.

Los datos de los numeros controlados tiene un mínimo de 1.00 y un máximo de 60.00, media de 11.08 y mediana de 6.00. La Figura(5)se muestra la grafica de densidad de los numeros controlados.

Agremamos un histograma en la que se puede ver la frecuencia de cada observacion y ademas se muestra la distribucion de las ocurrencias en cada categoria.La Figura(6) se muestra la grafica, ademas de que tambien se realizo un bloxplots donde se puede ver los casos y controlados de cada categoria. La Figura(7) se aprecia la grafica de bloxplots.[4]

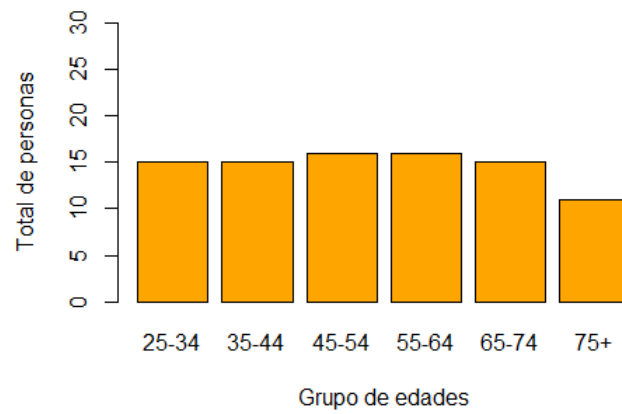


Figura 1: Gráfica de barra del grupos de edades

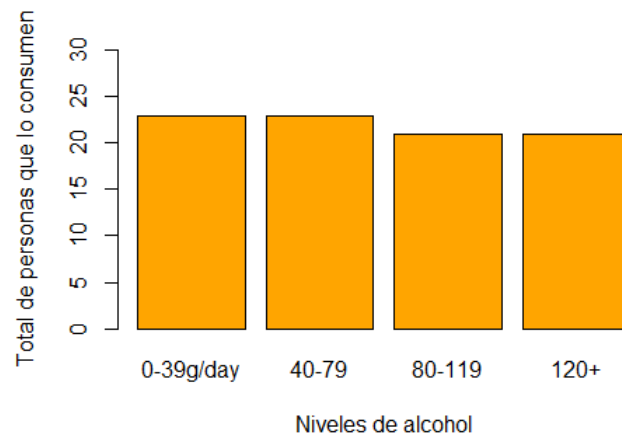


Figura 2: Gráfica de barra sobre el consumo del alcohol

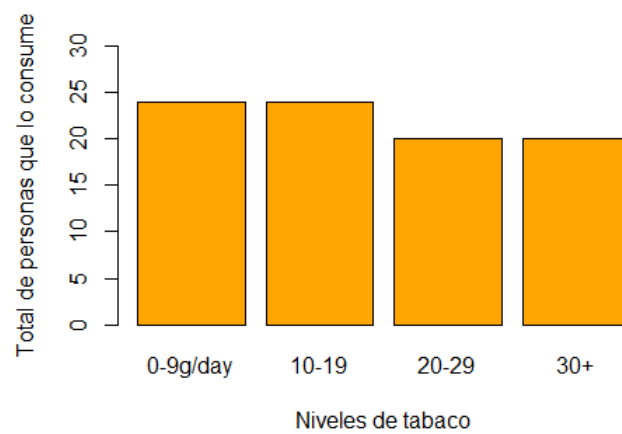


Figura 3: Gráfica de barra sobre el consumo del tabaco

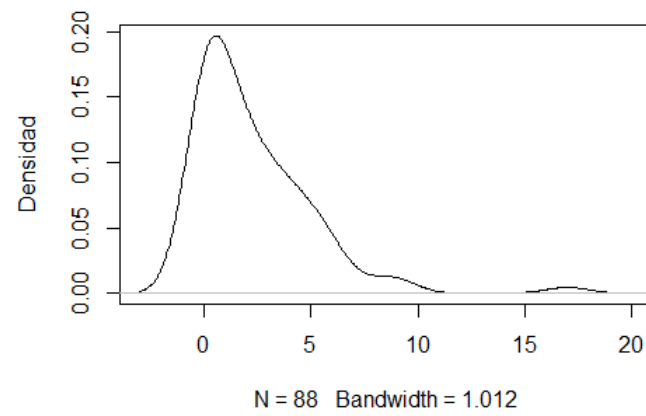


Figura 4: Gráficas de densidad de los numeros de casos

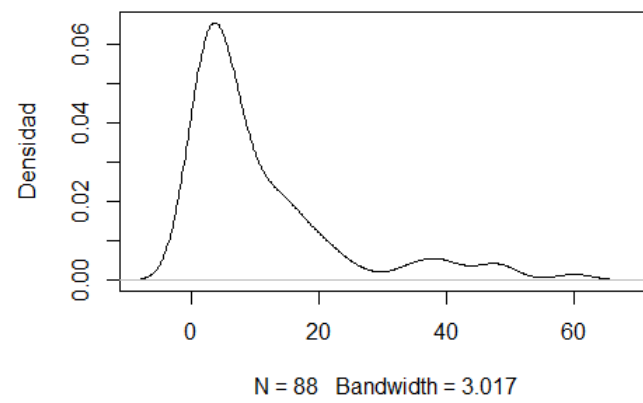


Figura 5: Gráficas de densidad de los numeros controlados

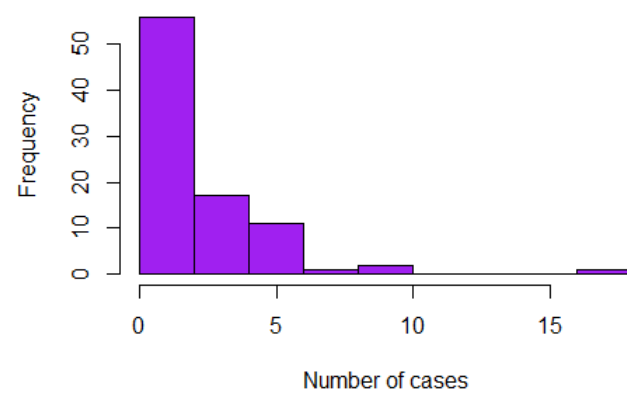


Figura 6: Gráficas de densidad de los numeros controlados

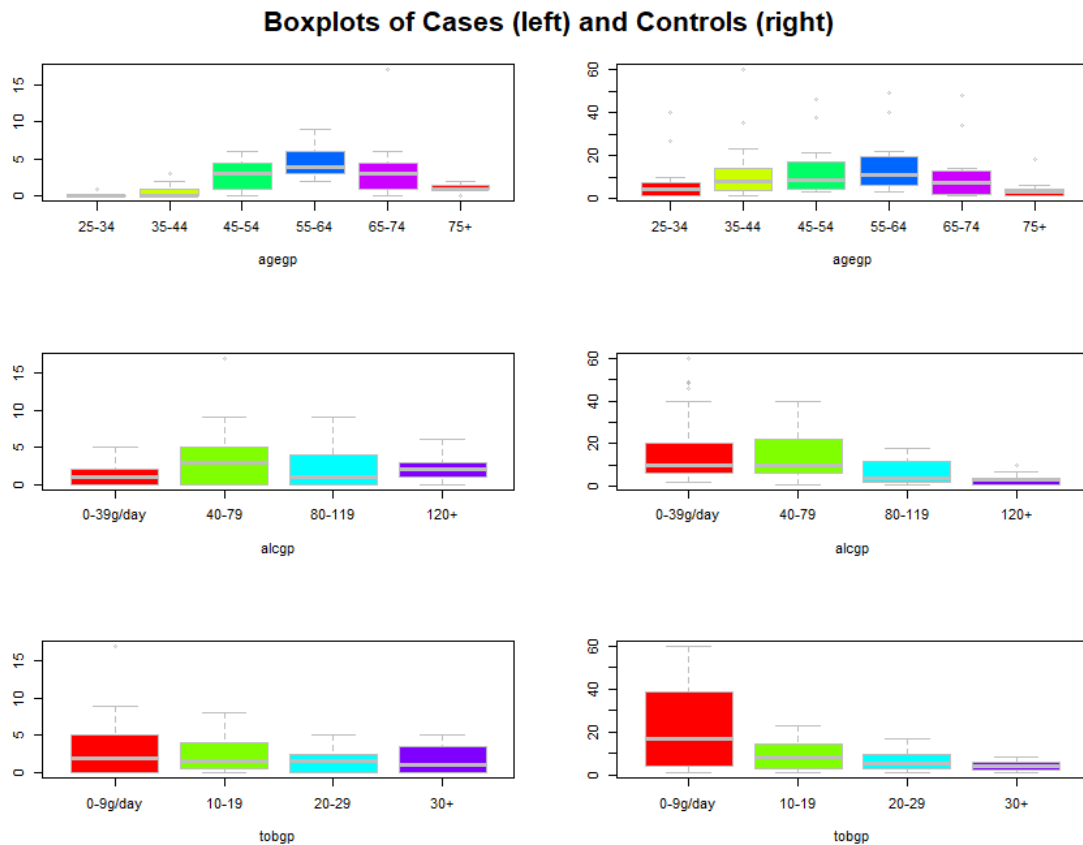


Figura 7: Gráficas de densidad de los numeros controlados

### 2.3. Estadística descriptiva de dos variables

En las siguientes figuras (8) y (9) se muestran la relacion entre los casos y controles

## Referencias

- [1] Andreas Helminen, Mikko Salaspuro, and S. Vakevainen. Aldh2 genotype has no significant effect on oral cavity's exposure to carcinogenic acetaldehyde after rinsing the mouth with ethanol. *Alcohol and Alcoholism*, 46:21–21, 09 2011.
- [2] Florence Trivin, Boucher Eveline, Elodie Vauléon, Isabelle Cumin, Elisabeth Pris , Odile Audrain, and Jean-Luc Raoul. Management of esophageal carcinoma associated with cirrhosis: A retrospective case-control analysis. *Journal of oncology*, 2009:173421, 12 2009. doi: 10.1155/2009/173421.
- [3] AJ Tuyns, G P quignot, and O Jensen. Esophageal cancer in ille-et-vilaine in relation to levels of alcohol and tobacco consumption. risks are multiplying. *Bulletin du cancer*, 64: 45–60, 02 1977.
- [4] Itzel Vega. Repositorio de Github. <https://github.com/ItzelVega>, 2021.

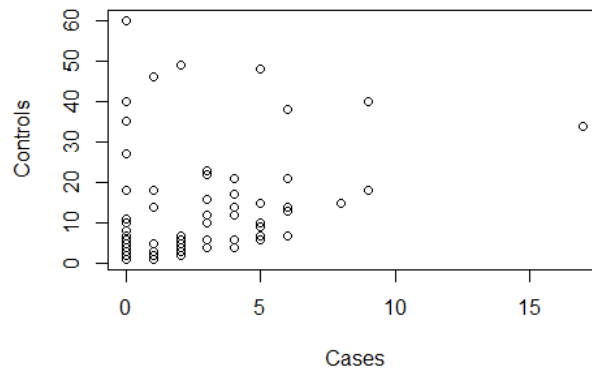


Figura 8: Gráficas de densidad de los numeros controlados

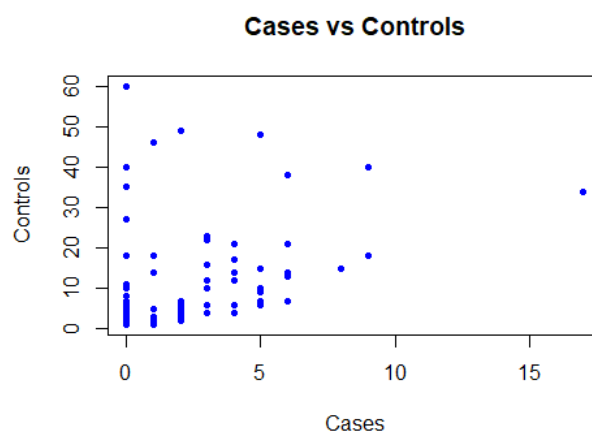


Figura 9: Gráficas de densidad de los numeros controlados