



Trabajo Práctico

El desperdicio de alimentos

Estadística Aplicada 11.67

Comisión C

Grupo 5

Integrantes:

Camilo Barbero

Ignacio Simón

Profesor: Pedro Camilo Cosatto Ammann

Ayudante: Ariel Waisburg

Índice

1. Búsqueda del set de datos	1
2. Análisis descriptivo	1
Bibliografía	5

1. Búsqueda del set de datos

Como tema se eligió el desperdicio de alimentos alrededor del mundo por lo que se seleccionó un dataset de corte transversal en el que se especifican las toneladas y kg por capita de desperdicios de comida que generan los diferentes actores de la sociedad (hogares, minoristas y servicios de comida) en cada uno de los países del mundo, entre otros territorios.

Al momento de escoger variables aleatorias para el análisis, se decidió por 3 variables aleatorias numéricas y una categoría, siendo esta última una estimación de nivel de confianza de las Naciones Unidas. La variable categórica puede tomar cualquiera de los 4 niveles de confianza: “High Confidence”, “Medium Confidence”, “Low Confidence” y “Very Low Confidence”. Estas variables se las define como:

$X = \text{Toneladas estimadas de desperdicio de comida de hogares del país por año (t)}$

$Y = \text{Toneladas estimadas de desperdicio de comida de minoristas del país por año (t)}$

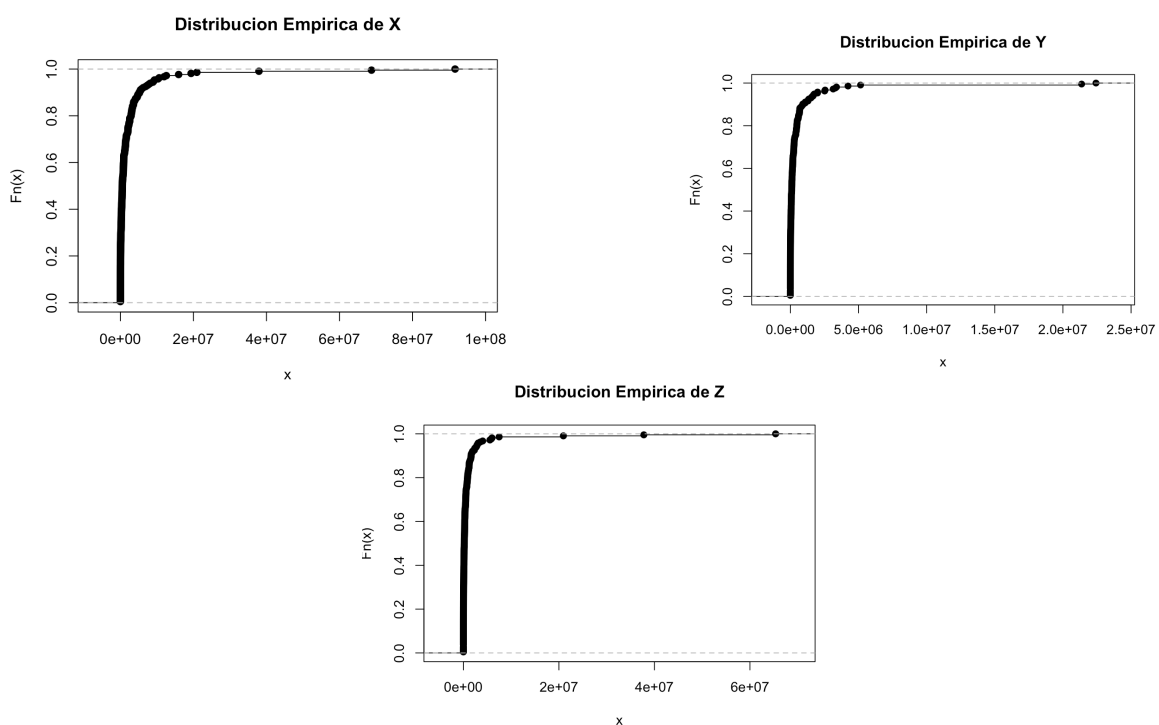
$Z = \text{Toneladas estimadas de desperdicio de comida de servicios de comida del país por año (t)}$

$W = \text{Estimación de confianza de los datos del país}$

2. Análisis descriptivo

T2.1)

Se realizó un gráfico de la función de distribución acumulada empírica para cada una de las variables numéricas.



Se puede observar que a partir de la muestra obtenida en el año 2021, las variables aleatorias numéricas al tratarse las tres de toneladas estimadas de desperdicio de comida de los diferentes actores societarios, tiende a concentrar sus valores en el principio de su rango, pero se pueden observar ciertos outliers. Esto es más reconocible al realizar el diagrama de caja y bigotes o boxplot más adelante.

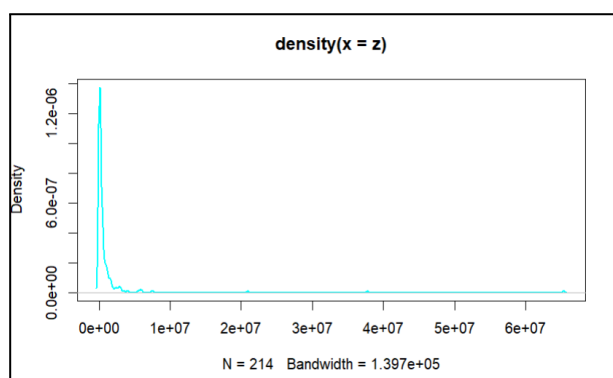
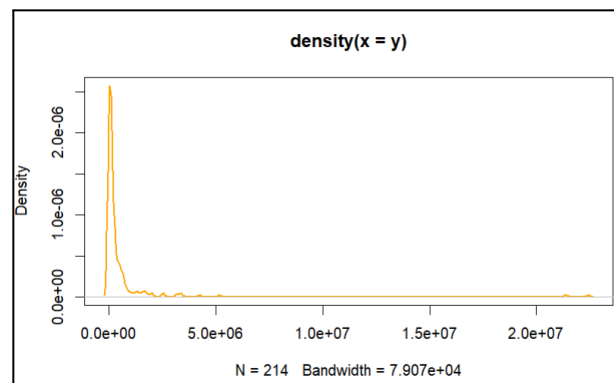
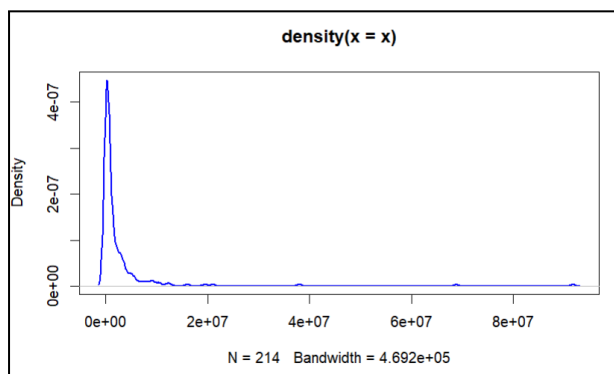


Diagrama de caja y bigotes de X

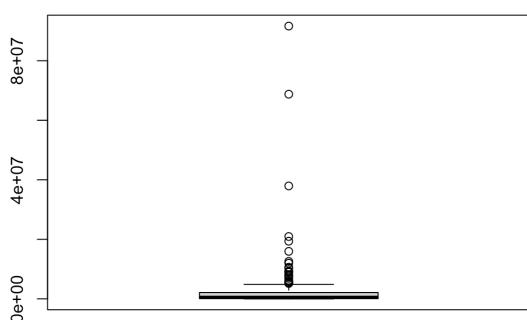


Diagrama de caja y bigotes de Y

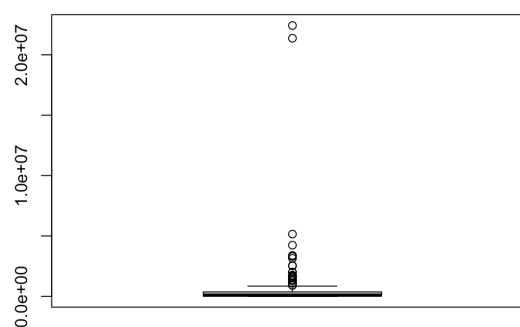


Diagrama de caja y bigotes de Z

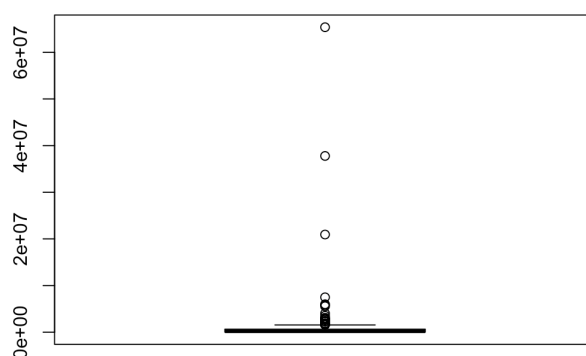
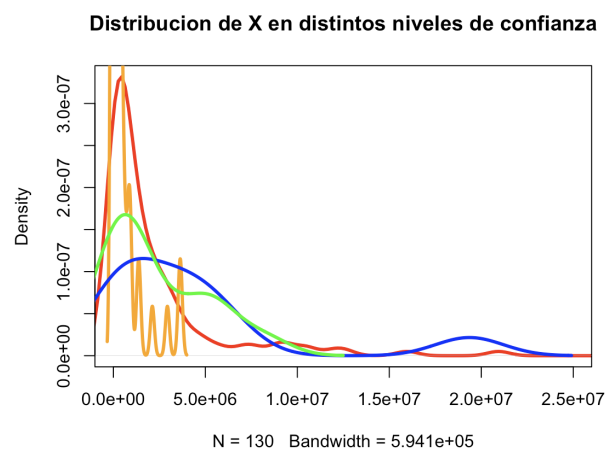
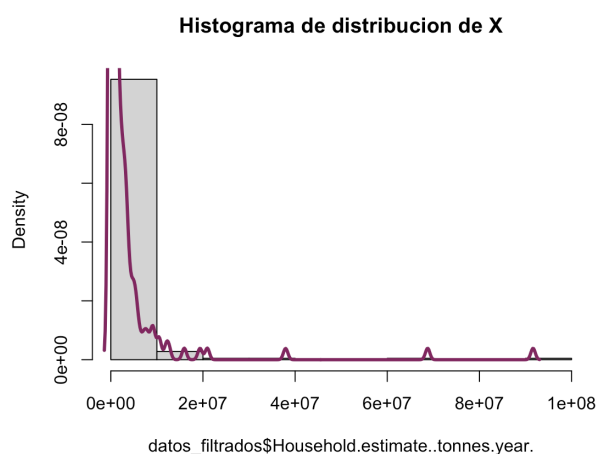


Tabla de datos relevantes del punto 2, T2.2):

Mediciones/Variables	X	Y	Z
Media	2658896	552045.4	1138859
Mediana	520508	100650	188466
Cuantil 0.25	67385	12243	22013
Cuantil 0.75	2119455	364585	644496
Rango inter-cuartilico	2042999	344306.5	608370.5
Desviación Estándar	8596906	2195578	5380459
Coefficiente de Variación	3.233262	3.97717	4.724429
Coefficiente de Asimetría	7.743123	8.700989	9.675113
Coefficiente de Curtosis	71.52448	83.97481	105.8755

Se puede interpretar que como para las tres variables la mediana se encuentra bastante más a la izquierda de la media, las variables tienen una asimetría a la derecha que se especifica en los histogramas de distribución. Es posible darse cuenta de esto asimismo, viendo que el coeficiente de asimetría es positivo en todos los casos. Asimismo la desviación estándar es grande, lo cual indica que hay mucha distancia entre en los datos y la media, posiblemente debido a datos inflados como los outliers, que modifican sustancialmente el valor de la media.

T2.4)



En el punto 4, se confeccionó por un lado un histograma de la distribución de X superpuesto con su función de densidad, y por el otro lado se superpusieron cuatro funciones de densidad para X según los distintos valores posibles que podría tomar la variable categórica W .

Como se explicó previamente, la variable W se trata del nivel de confianza que tiene las Naciones Unidas en los datos proporcionados por el país. Tomando esto en cuenta, se utilizaron colores de frío a cálido para indicar los valores de “Very Low Confidence” a “High Confidence” respectivamente de W^1 . De esta manera, se puede observar como hay un mayor achatamiento en la curva de densidad a mayor nivel de confianza, y mayor apuntamiento a menor nivel de confianza. Finalmente, cabe aclarar que hay una similitud entre la distribución de X y la de X con los diferentes valores de W . En esta se puede ver como la distribución de la variable X con “Low Confidence” (naranja) es similar a la distribución general de X . Se podría decir que los valores de X con “Low Confidence” afectan en gran medida a la distribución X en general, generando ese apuntamiento alto al inicio.

¹ “Very Low Confidence”: rojo, “Low Confidence”: naranja, “Medium Confidence”: verde, “High Confidence”: azul.

Bibliografia

- ARVIDSSON, J. (2019, March 9). *Food Waste*. Kaggle. Retrieved August 20, 2023, from <https://www.kaggle.com/datasets/joebeachcapital/food-waste/code>
- Forbes, H., Quested, T., & O'Connor, C. (2021, March 4). *UNEP Food Waste Index Report 2021* | *UNEP*. UN Environment Programme. Retrieved August 20, 2023, from <https://www.unep.org/resources/report/unep-food-waste-index-report-2021>