

# Reporte 1

## Datos UANL

**Materia: MAA(Aprendizaje Automático)**  
**Profesor: José Anastacio Hernández Saldaña**  
**Grupo: 03**  
**Alumno: José de Jesús Almanza Trejo**

# Contenido

Introducción.....	3
Definiciones .....	4
Descripción de los datos .....	4
Distribución de número de empleados por Tipo.....	6
Distribución de número de dependencias por Tipo .....	7
Histograma de distribución de Sueldo Neto por Tipo .....	8
Distribución de Sueldo Neto por fecha.....	9
Prueba no paramétrica .....	10

## Introducción

A continuación, se presenta un análisis descriptivo de las entidades encontradas en la base de datos compartida por el profesor a partir de una extracción (web scrapping) desde la página de la UANL.

Se incluye gráficas de la información de interés tales como histogramas, diagrama de violín (caja y bigotes), gráfico de pie; así como pruebas no paramétricas sobre la sospecha de igualdad en la mediana de algunas distribuciones.

# Definiciones

## Descripción de los datos

La tabla de información contiene las siguientes variables:

Variable	clasificación	Tipo
Nombre	texto	cualitativa
Sueldo Neto	continua	cuantitativa
dependencia	texto	cualitativa
Fecha	discreta	temporalidad
Tipo	texto	cualitativa

A partir de esta información identificamos 3 variables de interés que podemos explorar:

1. Sueldo Neto
2. Número de empleados por dependencia
3. Número de dependencias por Tipo

De la totalidad de los datos independientemente de la temporalidad de los mismos tenemos los siguientes estadísticos por cada una de las variables de interés:

Estadístico	Sueldo Neto	Número empleados por Fecha/dependencia	Número de dependencias por Fecha/Tipo
Conteo	636,201	6,124	288
Promedio	\$14,241.68	103.88	21.26
Desviación estándar	\$9,578.44	223.37	14.72
Mínimo	\$175.41	1	1
Percentil 25	\$8,007.66	18	8
Percentil 50 (Mediana)	\$11,426.50	45	23.5
Percentil 75	\$17,654.63	117	28
Máximo	\$147,051.59	2,615	54

Obtenemos una descripción general de cómo se distribuye el Sueldo Neto promedio a nivel dependencia (sin importar la fecha de observación):

Dependencia	Sueldo Neto promedio	Veces desviaciones estándar de la media
Dirección de Unidades Académicas	\$48,706.90	7.43
Dirección de Investigación Educativa	\$30,192.22	3.35
Dirección de Responsabilidad Social Universitaria	\$29,785.79	3.26
Dirección Administrativa de la SIID	\$27,286.85	2.71
Dirección de Desarrollo de Modelos Tecnológicos	\$26,102.06	2.45

Se puede decir con un corte de 3 veces la desviación estándar a partir de la media, que tenemos 3 Direcciones dentro de la UANL que tienen un promedio de Sueldo Neto atípico (outliers), se muestra a continuación un gráfico que todas las dependencias:

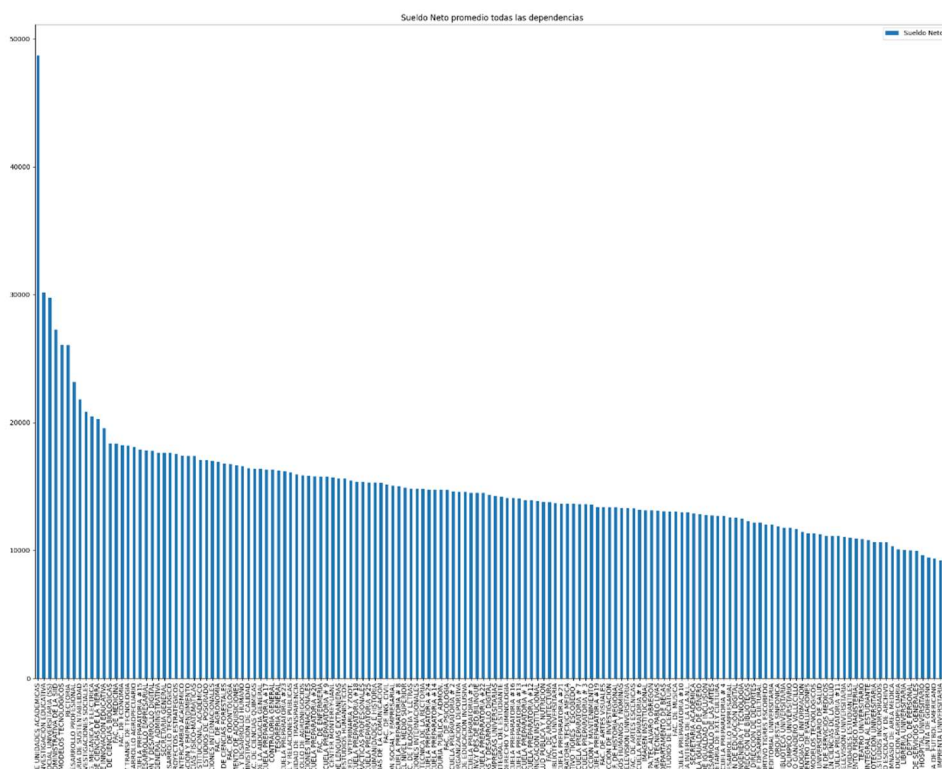


Fig 1.

## Distribución de número de empleados por Tipo

Observamos cómo se distribuye de manera general el porcentaje de empleados por el tipo de dependencia:

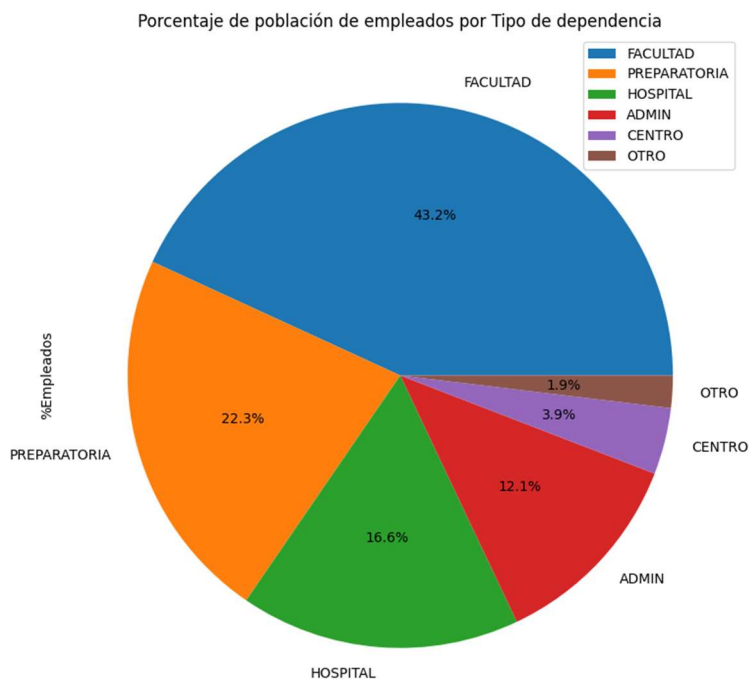


Fig 2.

Podemos decir de manera general que la mayoría de los empleados están concentrados en las Facultades, Preparatorias y el Hospital Universitario; sumando aproximadamente el 81% del total de la plantilla de empleados.

## Distribución de número de dependencias por Tipo

Se muestra a continuación por Tipo de dependencia, en promedio cuántas dependencias hay por cada tipo (sin contemplar temporalidad) en promedio:

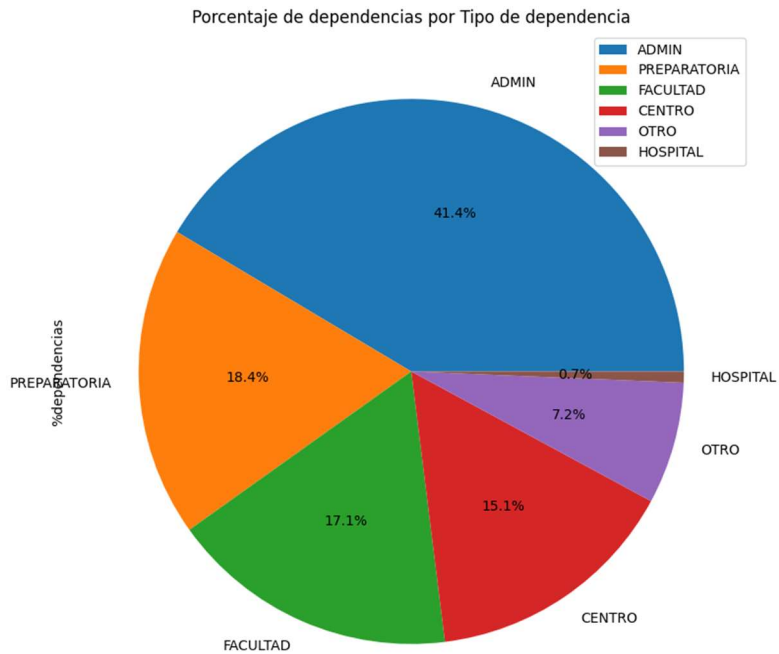


Fig 3.

De aquí podemos concluir que tenemos en promedio más departamentos de Administración que Preparatorias o Facultades, aún y cuando estos departamentos de Administración cuentan con muchos menos empleados que estas otras entidades.

## Histograma de distribución de Sueldo Neto por Tipo

Generamos un gráfico de violín para observar la distribución de los Sueldos Netos por Tipo de dependencia (sin importar temporalidad):

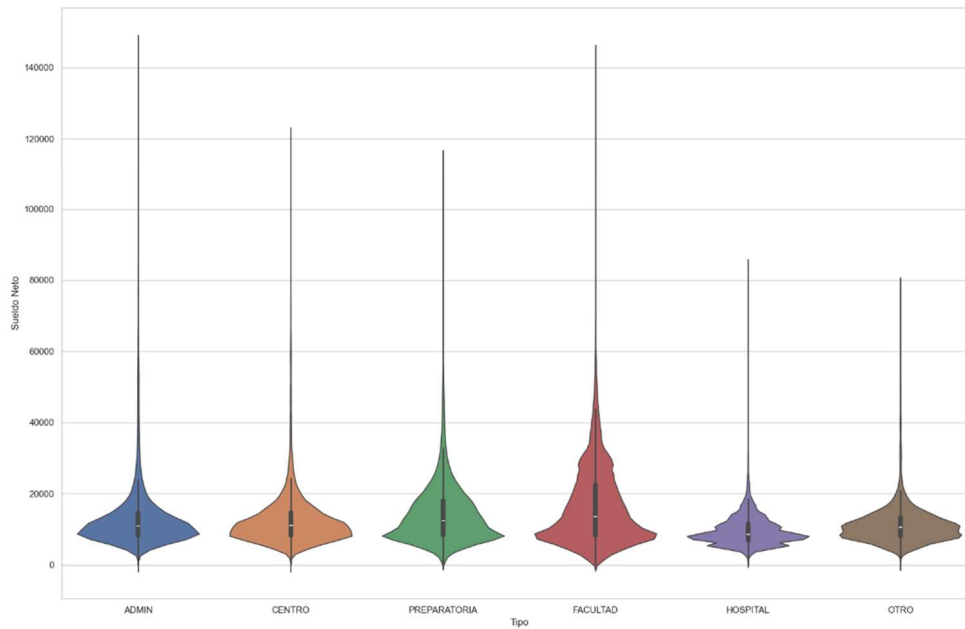


Fig 4.

Visualmente podemos observar una mayor desviación en los sueldos de Preparatoria y Facultad, en Administraciones y Facultades se observan algunos casos que muestran valores atípicos.

Podemos probar cuál de estos Tipos de dependencias tienen medianas iguales por medio de la prueba de Kruskal Wallis.



## Distribución de Sueldo Neto por fecha

Generamos el histograma de Sueldo Neto por Fecha donde podemos observar que los datos siguen probablemente una distribución gamma o beta por la forma sesgada de la distribución de estos datos, como ejemplo podemos observar la gráfica del mes de enero 2024:

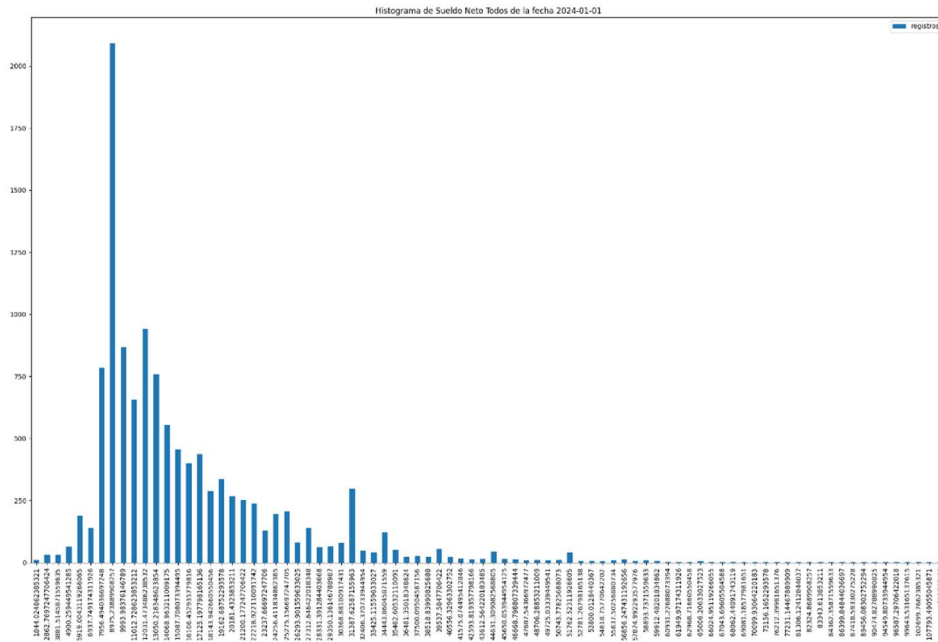


Fig 5.

## Prueba no paramétrica

De acuerdo con la figura 5 podemos decir que el Ingreso Neto no sigue una distribución Normal ya que tiene una forma sesgada. De acuerdo con la figura 4 podemos decir que al menos un par de Tipos de dependencia tienen la misma mediana por lo que podemos aplicar la prueba de Kruskal Wallis para determinar si es cierto para al menos un par de Tipos.

La prueba paramétrica consiste en probar que:

H0: Las medianas de las poblaciones son iguales

Vs

H1: Las medianas de las poblaciones no son iguales

Se rechaza H0 si el p-valor de la prueba estadística (Chi cuadrada) es menor que el nivel de significancia de 0.05 (usualmente este valor).

Prueba Kruskal Wallis de CENTRO vs ADMIN: KruskalResult(statistic=np.float64(0.187526305347088), pvalue=np.float64(0.6649834762718729))

Prueba Kruskal Wallis de PREPARATORIA vs ADMIN: KruskalResult(statistic=np.float64(2132.1352959752153), pvalue=np.float64(0.0))

Prueba Kruskal Wallis de PREPARATORIA vs CENTRO: KruskalResult(statistic=np.float64(1010.5698458593691), pvalue=np.float64(9.052621918241111e-222))

Prueba Kruskal Wallis de FACULTAD vs ADMIN: KruskalResult(statistic=np.float64(5372.859044412944), pvalue=np.float64(0.0))

Prueba Kruskal Wallis de FACULTAD vs CENTRO: KruskalResult(statistic=np.float64(2290.1228079969846), pvalue=np.float64(0.0))

Prueba Kruskal Wallis de FACULTAD vs PREPARATORIA: KruskalResult(statistic=np.float64(1878.740695734864), pvalue=np.float64(0.0))

Prueba Kruskal Wallis de HOSPITAL vs ADMIN: KruskalResult(statistic=np.float64(11820.252800633105), pvalue=np.float64(0.0))

Prueba Kruskal Wallis de HOSPITAL vs CENTRO: KruskalResult(statistic=np.float64(5288.056886213051), pvalue=np.float64(0.0))

Prueba Kruskal Wallis de HOSPITAL vs PREPARATORIA: KruskalResult(statistic=np.float64(25958.5604286773), pvalue=np.float64(0.0))

Prueba Kruskal Wallis de HOSPITAL vs FACULTAD: KruskalResult(statistic=np.float64(35790.9587934486), pvalue=np.float64(0.0))

Prueba Kruskal Wallis de OTRO vs ADMIN: KruskalResult(statistic=np.float64(233.94120781510193), pvalue=np.float64(8.238760581720604e-53))

Prueba Kruskal Wallis de OTRO vs CENTRO: KruskalResult(statistic=np.float64(183.77340960145042), pvalue=np.float64(7.270729388595677e-42))

Prueba Kruskal Wallis de OTRO vs PREPARATORIA: KruskalResult(statistic=np.float64(1468.2708333267856), pvalue=np.float64(0.0))

Prueba Kruskal Wallis de OTRO vs FACULTAD: KruskalResult(statistic=np.float64(2219.299779808321), pvalue=np.float64(0.0))

Prueba Kruskal Wallis de OTRO vs HOSPITAL: KruskalResult(statistic=np.float64(1672.1097407233672), pvalue=np.float64(0.0))

Dados los resultados anteriores podemos decir que con un nivel de significancia de 0.05 las medianas de las distribuciones de Sueldo Neto de las dependencias de CENTRO y ADMIN muestran el mismo valor de la mediana; y de la misma forma con un nivel de significancia de 0.05 podemos decir que ningún otro par de Tipos de dependencia tienen la misma mediana de la distribución de Sueldo Neto.