

Historico de Bitácoras Onomásticas del SEMEFO-DF

Análisis de datos

Las Históricas de Bitácoras Onomásticas del Servicio Médico Forense del Distrito Federal (SEMEFO-DF), han sido sistematizada por la Comisión Nacional de Búsqueda (CNB). Este conjunto de datos se obtuvo a través de una solicitud de acceso a la información disponible en el Portal de Transparencia del Instituto Nacional de acceso a la Información (INA) en noviembre de 2019 y existe una copia en el portal ciudadano [data.mx](#) de Codepend Mexico.

El SEMEFO-DF (Servicio Médico Forense Distrito Federal) ahora INCIFCO-CdMx es una institución oficial encargada de realizar autopsias y estudios forenses para esclarecer las causas de muerte, principalmente en casos de muertes violentas, sospechosas o relacionadas con hechos delictivos. Las **Bitácoras Onomásticas (BO)** son registros administrativos principales utilizados por el SEMEFO para documentar el ingreso de cadáveres o restos humanos a sus instalaciones. Estas bitácoras contienen información detallada sobre cada ingreso, como el nombre de la persona fallecida, su edad, sexo, fecha de ingreso, institución de procedencia y diagnóstico preliminar de la causa de muerte. En muchos casos, estos registros incluyen datos clave para identificar patrones históricos de mortalidad, enfermedades comunes y factores de riesgo, lo que las convierte en una fuente valiosa para estudios demográficos, epidemiológicos y de salud pública.

Podemos acceder a esta información con la [Plataforma de Transparencia](#) o en [datamx](#).

Código de color

1. ■ Cadáver conocido
2. ■ Feto
3. ■ Recién nacido
4. ■ Miembros
5. ■ Restos óseos
6. ■ Cadáver desconocido
7. ■ Recién nacido
8. ■ Fetal
9. ■ Miembros
10. ■ Restos óseos
11. ■ Múltiples
12. ■ Restos óseos

```
In [1]: from joblib import load
import pandas as pd
import numpy as np
from numpy import isnan
from helper_pandas import *
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud

from scipy.stats import chi2_contingency
import statsmodels.api as sm
warnings.filterwarnings('ignore')

in_file = 'data_clean/HBO_clean.pkl'
with open(in_file, 'rb') as f:
    df = load(f)

#Out[1]: DataFrame: 96825 rows & 12 columns
df.dtypes[1].head(2)
df[df['Identificación'].str.startswith('S-D')]
display(df.shape)
df.tail(1)
```

```
Out[1]:      Número_progresivo_transcripción Nombre_completo_transcripción Primer_apellido Segundo_apellido Nombres_propios Fecha_transcripción Fecha_estándar
ID
```

ID	Número_progresivo_transcripción	Nombre_completo_transcripción	Primer_apellido	Segundo_apellido	Nombres_propios	Fecha_transcripción	Fecha_estándar
BO_1982_07489		S-D			placenta	NaN	NaN
BO_1982_07490		S-D			5 dedos del pie derecho de un feto	NaN	NaN
BO_1982_07491		S-D			dedo de desconocido	NaN	NaN
BO_1982_07492		S-D			4 dedos de desconocido	NaN	NaN
BO_1982_07493		S-D			osamenta de desconocido	NaN	NaN

5 rows x 31 columns

```
<matplotlib.figure.Figure at 0x10000000>
```

```
In [2]: from matplotlib.colors import LinearSegmentedColormap
dic_color = {'Población':'#0504aa', 'S-D':'#999999', 'S-D (DF)':'#d9d9d9', 'Masculino':'#37eb8b', 'Masculino (DF)':'#80bd31',
            'Femenino':'#98e4a3', 'Femenino (DF)':'#bebad4',
            'Cada vez conocido':'#04d0b2', 'Cada vez desconocido':'#ffdf52',
            'Recién nacido':'#87542f',
            'Feto':'#f77f00',
            'Miembros':'#7781bf',
            'Restos óseos':'#fb5a33'}
```

```
gray_to_green = LinearSegmentedColormap.from_list("gray_to_green", ['#d9d9d9', '#2a7e19'])
```

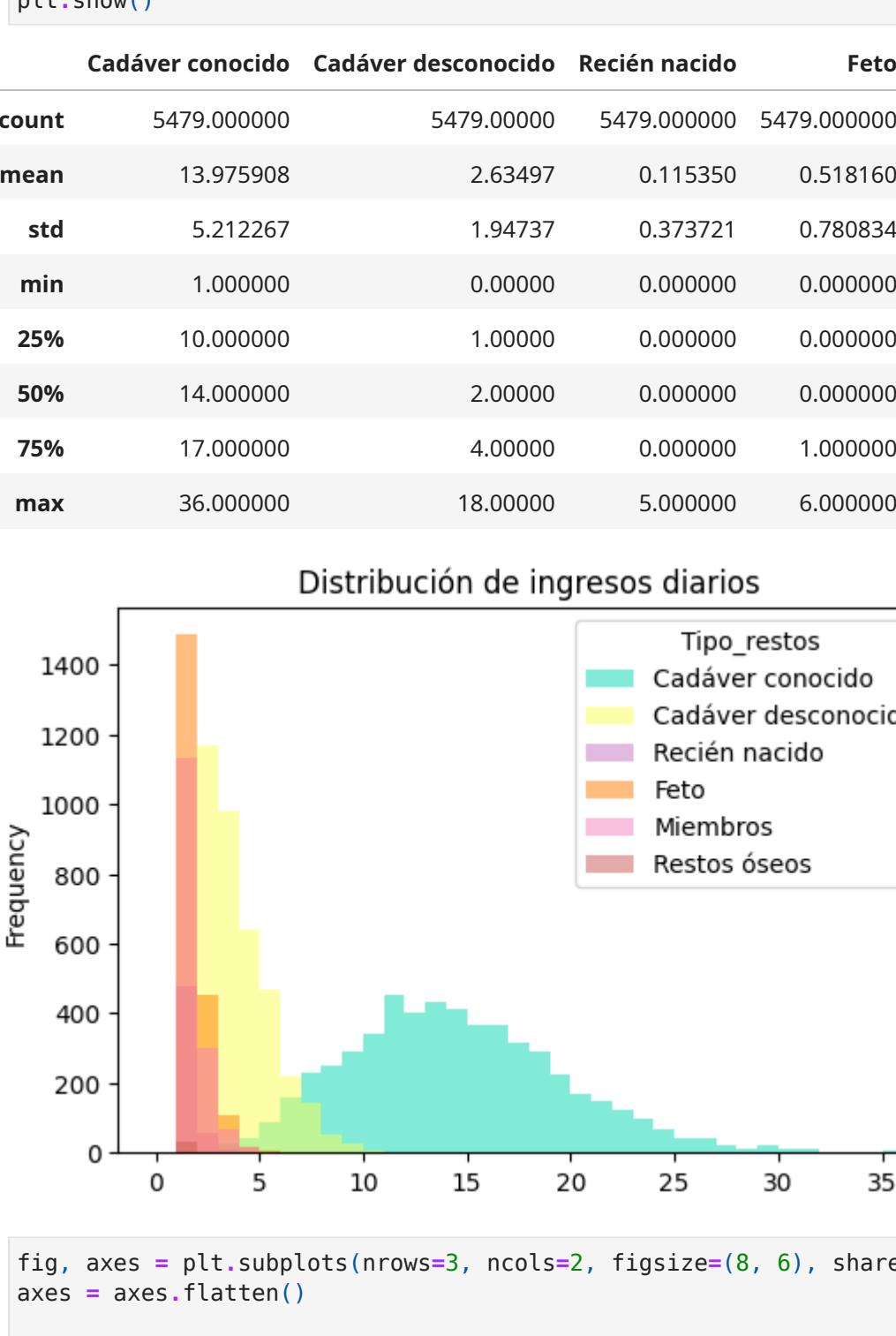
Fecha

'Fecha_transcripción', 'Fecha_estándar', 'Fecha_año', 'Fecha_mes', 'Fecha_díasemana',

Ingresos anuales y diarios

```
In [3]: df_año_tipo = df.pivot_table(index='Fecha_año', columns='Tipo_restos', aggfunc='size')
df_año_tipo = df_año_tipodf['Tipo_restos'].cat.categories.iloc[:, ::-1]
display(df_año_tipo)

fig, ax = plt.subplots(figsize=(6, 4))
df_año_tipo.plot.area(ax=ax, color=dic_color, linewidth=0, alpha=0.5)
ax.set_xlabel('Año')
ax.set_title('Ingresos anuales por tipo de restos')
ax.set_xticklabels([x.get_text().replace('-', '') for x in ax.get_xticklabels()]) # Remove negative sign from tick labels
plt.tight_layout()
plt.savefig('Imagenes/B0_IngresosRestosSexo.svg')
plt.show()
```



Prueba de independencia χ^2

Supuestos:

- Observaciones Independientes:
- Frecuencias Esperadas Suficientes: guitar Restos óseos
- Muestra Aleatoria Simple

Resultado: Muestras no independientes: Chi-square: 721.48 Estadísticamente significativa: P-value: 2.36e-116

```
In [4]: data = df_año_tipo.drop(columns=['Restos óseos', 'Recién nacido'])

# Perform Chi-Square test
chi2, p_value, dof, expected = chi2_contingency(data)
print("Chi-Square statistic: (%.2f)" % chi2)
print("Degrees of freedom: (%d)" % dof)
print("P-value: (%.2f)" % p_value)

# visualize residuals
thr_resid = 5
resid = (data.values - expected) / np.sqrt(expected)
resid = pd.DataFrame(resid, index=data.index, columns=data.columns)
resid = np.abs(resid) > thr_resid
#display(resid)

Chi-square statistic: 5.301962484517
Degrees of freedom: 42
P-value: 1.7391436192219905e-99
```

Regresión Lineal

Variable independiente: año (X)

Variable dependiente: decesos/ingresos (Y)

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Supuestos:

- Linealidad
- Independencia
- Homocedasticidad: La varianza de los residuos es constante
- Normalidad residuos

Regresión Cadáver conocido anual R-squared: 0.957 Pendiente β_1 : 193.043

Regresión Cadáver desconocido anual R-squared: 0.825 Pendiente β_1 : 25.882

Regresión Miembros anual R-squared: 0.687 Pendiente β_1 : 8.968

Regresión Feto anual R-squared: 0.556 Pendiente β_1 : -8.918

Regresión Restos óseos anual R-squared: 0.210 NO Pendiente β_1 : -0.439

Regresión Recién nacido anual R-squared: 0.058 NO Pendiente β_1 : -0.893

```
In [5]: X = sm.add_constant([i for i in data.index])

for col in data.columns:
    print(f'Regresión {col} anual')
    y1 = data[col]
    model1 = sm.OLS(y1, X).fit()
    print(f'R-squared: {model1.rsquared:.3f} | Pendiente $|\beta_1|{model1.params.x1:.3f}|n)
```

Regresión Miembros anual

OLS Regression Results

Dep. Variable: Miembros R-squared: 0.687

Model: OLS Adj. R-squared: 0.663

Method: Least Squares F-statistic: 93.23

Date: Mon, 27 Oct 2025 Prob (F-statistic): 0.000133

Time: 13:17:56 Log-likelihood: -70.228

No. Observations: 15 AIC: 144.5

Df Model: 1 BIC: 145.9

Df Residuals: 13 Cond. No.: 145.9

Covariance Type: nonrobust

coef std err t P>|t| [0.025 0.975]

const -1.75e+04 3312.433 -5.306 0.000 -2.47e+04 -1.04e+04

x1 8.9679 1.677 5.347 0.000 5.345 12.591

Omnibus: 2.569 Durbin-Watson: 2.169

Prob(Omnibus): 0.954 Jarque-Bera (JB): 0.849

Skew: 0.516 Prob(JB): 0.654

Kurtosis: 3.542 Cond. No.: 9.03e+05

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 9.03e+05. This might indicate that there are strong multicollinearity or other numerical problems.

Regression Cadáver conocido anual

OLS Regression Results

Dep. Variable: Feto R-squared: 0.556

Model: OLS Adj. R-squared: 0.521

Method: Least Squares F-statistic: 16.25

Date: Mon, 27 Oct 2025 Prob (F-statistic): 0.000143

Time: 13:17:56 Log-likelihood: -74.828

No. Observations: 15 AIC: 152.8

Df Model: 1 BIC: 154.2

Df Residuals: 13 Cond. No.: 164.8

Covariance Type: nonrobust

coef std err t P>|t| [0.025 0.975]

const -5.015e+04 6533.298 -7.676 0.000 -6.43e+04 -3.6e+04

x1 25.8821 3.308 7.823 0.000 18.735 33.029

Omnibus: 1.881 Durbin-Watson: 2.034

Prob(Omnibus): 0.707 Jarque-Bera (JB): 1.178

Skew: 0.672 Prob(JB): 0.555

Kurtosis: 2.723 Cond. No.: 9.03e+05

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 9.03e+05. This might indicate that there are strong multicollinearity or other numerical problems.

Regression Cadáver desconocido anual

OLS Regression Results

Dep. Variable: Cadáver desconocido R-squared: 0.825

Model: OLS Adj. R-squared: 0.811

Method: Least Squares F-statistic: 2.85e-06

Date: Mon, 27 Oct 2025 Prob (F-statistic): 0.8818

Time: 13:17:56 Log-likelihood: -78.418

No. Observations: 15 AIC: 160.8

Df Model: 1 BIC: 162.8

Df Residuals: 13 Cond. No.: 166.3

Covariance Type: nonrobust

coef std err t P>|t| [0.025 0.975]

const -3.762e+05 2.24e+04 -16.812 0.000 -4.24e+05 -3.28e+05

x1 193.0429 11.329 17.042 0.000 -13.691 197.039

Omnibus: 1.981 Durbin-Watson: 2.173

Prob(Omnibus): 0.905 Jarque-Bera (JB): 0.918

Skew: 0.652 Prob(JB): 0.591

Kurtosis: 3.519 Cond. No.: 9.03e+05

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 9.03e+05. This might indicate that there are strong multicollinearity or other numerical problems.

Regression Recién nacido anual

OLS Regression Results

Dep. Variable: Recién nacido R-squared: 0.556

Model: OLS Adj. R-squared: 0.531

Method: Least Squares F-statistic: 2.05e-06

Date: Mon, 27 Oct 2025 Prob (F-statistic): 0.9028

Time: 13:17:56 Log-likelihood: -74.082

No. Observations: 15 AIC: 150.0

Df Model: 1 BIC: 152.0