

Reporte: Estadísticas de Divorcio (INEGI, 2023)

Carlos Hernández Márquez

“Doy mi palabra que he realizado esta actividad con integridad académica”

I. Caracterización de la base de datos y fundamentación metodológica

- **Fuente:** INEGI – Estadística de Divorcios 2023.
- **Cobertura:** México, periodo 2023 (163,587 registros).
- **Naturaleza:** Base administrativa, derivada de registros civiles y juzgados.
- **Estructura:** 64 variables iniciales → sociodemográficas, legales y de contexto.

Diccionario de variables: Estadísticas de Divorcios 2023

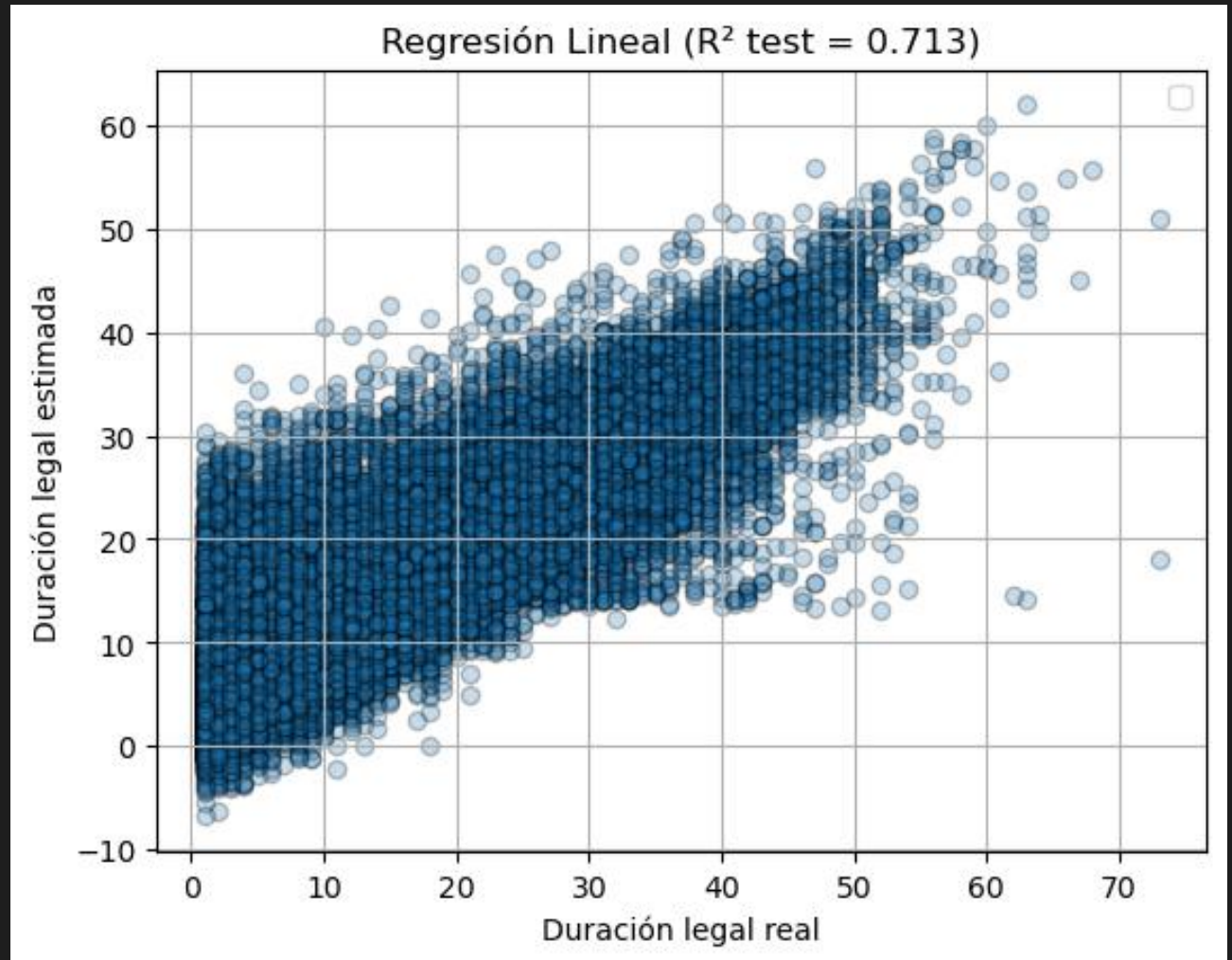
	Variable	Nombre	Catalogo	Tipo
0	anio_eje	Año de ejecutoria del divorcio.	año_ejecutoria	N
1	anio_mat	Año de registro del matrimonio.	año_registro	N
2	anio_reg	Año de registro de la demanda.	año_registro	N
3	anio_sen	Año de la sentencia.	año_sentencia	N
4	causa	Causas de divorcio.	causas_divorcio	N
5	con_acdiv1	Condición de actividad económica del primer divorciante.	condicion_actividad_economica	N
6	con_acdiv2	Condición de actividad económica del segundo divorciante.	condicion_actividad_economica	N
7	cus_hij	Número de hijos en custodia.	numero_de_hijos	N
8	custodia	Persona a quien se otorga la custodia.	custodia	N
9	dedic_div1	A qué se dedica el primer divorciante.	dedicado_a	N
10	dedic_div2	A qué se dedica el segundo divorciante.	dedicado_a	N
11	dia_eje	Día de ejecutoria del divorcio.	dia	N
12	dia_mat	Día de registro del matrimonio.	dia	N
13	dia_reg	Día de registro de la demanda.	dia	N
14	dia_sen	Día de la sentencia.	dia	N
15	dis_reoax	Distritos de registro de Oaxaca.	entidad_municipio_localidad_2022	N
16	dura_leg	Duración legal del matrimonio.	duracion_matrimonio	N

II. Corrección de inconsistencias y tratamiento de problemas en los datos

Problema	Acción aplicada	Resultado
Huecos (NaN)	Identificación de códigos especiales (INEGI) + imputación (Moda / Media-Mediana / KNN)	Base limpia sin valores faltantes
Outliers	Detección con Tukey (IQR) en variables cuantitativas	Se eliminaron pocos casos atípicos
Inconsistencias	$\text{edad_div} \approx \text{edad_mdiv} + \text{dura_leg}$	Mayor consistencia entre variables
Colinealidad	Eliminación de variables redundantes	Menor multicolinealidad
Interacciones	Creación de indicadores binarios (ej. <i>mismo sexo</i> , <i>misma entidad</i>) y diferencias de edad	Variables más interpretables
Variables categóricas	One-hot encoding en variables multicategoría	26 \rightarrow 155 columnas, listo para selección

III. Construcción y validación de modelos lineales y no lineales de predicción

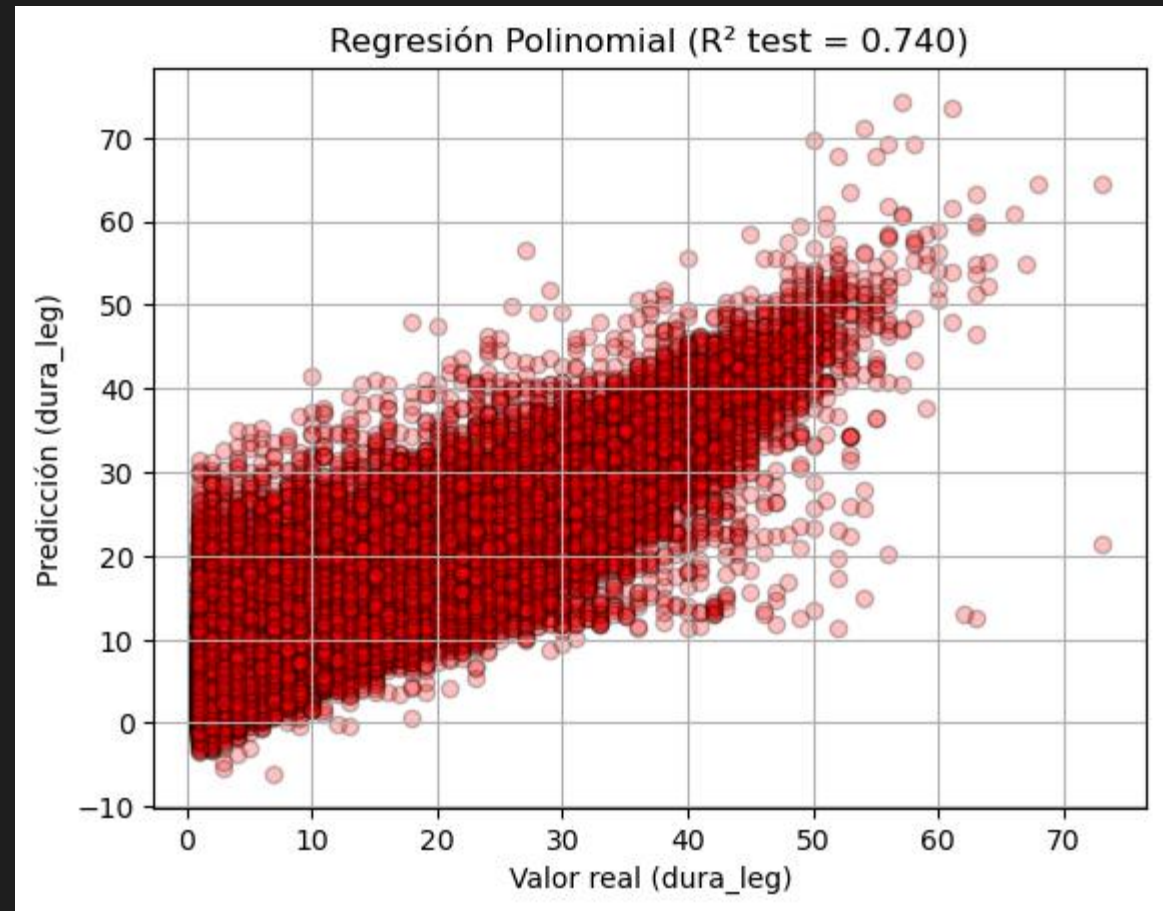
- **Selección de características:** Se aplicó una selección rápida hacia adelante
- Se quedó con 56 variables.
- R2 train: 0.7075 | R2 test: 0.7126
- El modelo logra explicar alrededor del **71% de la variabilidad** de la duración legal del matrimonio.
- Existe dispersión en valores altos de duración, indicando pérdida de ajuste en los extremos.



III. Construcción y validación de modelos lineales y no lineales de predicción

- Esto representa un incremento de 2.7 puntos porcentuales respecto al modelo lineal.
- En la gráfica, la nube de puntos se ajusta mejor a la diagonal, especialmente en valores altos de `dura_leg`.

Modelo	MSE	R ²
Regresión Lineal	37.17	0.713
Regresión Polinomial	33.64	0.740



IV. Aplicación del modelo óptimo para inferencia estadística

Para este análisis se tomaron en cuenta únicamente **variables numéricas continuas o discretas**, excluyendo las booleanas y categóricas codificadas en dummies.

- El modelo explica aproximadamente **71.3% de la variabilidad** de la duración legal del matrimonio.
- **F-statistic (p < 0.001)**: El modelo en su conjunto es altamente significativo.
- **Coeficientes significativos**: Prácticamente todas las variables y sus interacciones resultaron estadísticamente significativas (p < 0.05), excepto `edad_div1*hijos`.

OLS Regression Results							
Dep. Variable:	dura_leg		R-squared:	0.713			
Model:	OLS		Adj. R-squared:	0.713			
Method:	Least Squares		F-statistic:	1.363e+04			
Date:	Mon, 08 Sep 2025		Prob (F-statistic):	0.00			
Time:	05:49:04		Log-Likelihood:	-3.5370e+05			
No. Observations:	109668		AIC:	7.074e+05			
Df Residuals:	109647		BIC:	7.076e+05			
Df Model:	20						
Covariance Type:	nonrobust						
		coef	std err	t	P> t	[0.025	0.975]
	const	16.2157	0.035	459.345	0.000	16.146	16.285
	edad_div1	4.9159	0.054	90.232	0.000	4.809	5.023
	edad_div2	2.8444	0.052	54.623	0.000	2.742	2.946
	hijos	3.6644	0.037	98.394	0.000	3.591	3.737
	cus_hij	-1.9604	0.036	-54.266	0.000	-2.031	-1.890
	dif_edad_matrimonio	-2.1822	0.031	-69.332	0.000	-2.244	-2.121
	edad_div1^2	3.4999	0.055	63.664	0.000	3.392	3.608
	edad_div1 edad_div2	-6.0417	0.106	-57.052	0.000	-6.249	-5.834
	edad_div1 hijos	-0.0294	0.041	-0.717	0.473	-0.110	0.051
	edad_div1 cus_hij	-0.5222	0.047	-11.170	0.000	-0.614	-0.431
	edad_div1 dif_edad_matrimonio	-0.3370	0.024	-13.783	0.000	-0.385	-0.289
	edad_div2^2	2.9462	0.058	50.567	0.000	2.832	3.060
	edad_div2 hijos	0.1309	0.040	3.295	0.001	0.053	0.209
	edad_div2 cus_hij	-0.4165	0.045	-9.303	0.000	-0.504	-0.329
	edad_div2 dif_edad_matrimonio	0.1557	0.024	6.480	0.000	0.109	0.203
	hijos^2	-0.4803	0.020	-23.759	0.000	-0.520	-0.441
	hijos cus_hij	0.9047	0.047	19.414	0.000	0.813	0.996
	hijos dif_edad_matrimonio	-0.0897	0.024	-3.703	0.000	-0.137	-0.042
	cus_hij^2	-0.4852	0.034	-14.180	0.000	-0.552	-0.418
	cus_hij dif_edad_matrimonio	0.2949	0.025	11.941	0.000	0.247	0.343
	dif_edad_matrimonio^2	-0.6040	0.013	-44.910	0.000	-0.630	-0.578

**A continuación inferencias tomando en
cuenta un intervalo de confianza al
95%**

Por variable

1. Edad de los cónyuges al divorcio

- `edad_div1`: $\beta = 4.92$ [4.81, 5.02]
- `edad_div2`: $\beta = 2.84$ [2.74, 2.95]

Cada año adicional de edad se asocia con un **aumento de entre 2.7 y 5 años en la duración del matrimonio**, dependiendo de cuál cónyuge se considere.

Esto sugiere que los divorcios ocurren en edades más tardías tras matrimonios más prolongados.

2. Número de hijos

- `hijos`: $\beta = 3.66$ [3.59, 3.74]

Cada hijo adicional aumenta la duración legal esperada en **aprox. 3.6 años**.

Intervalo de confianza estrecho → efecto robusto.

Implica que la presencia de hijos actúa como un factor estabilizador del matrimonio.

3. Custodia de hijos

- `cus_hij`: $\beta = -1.96$ [-2.03, -1.89]

Tener hijos bajo custodia reduce la duración legal estimada en **~2 años por hijo**.

Señala un posible efecto de **disolución acelerada en matrimonios con custodia disputada**.

4. Diferencia de edad al casarse

- `dif_edad_matrimonio`: $\beta = -2.18$ [-2.24, -2.12]

Cada año adicional de diferencia de edad al casarse se asocia con **~2.2 años menos de duración**.

Relación negativa clara, consistente con estudios que muestran que grandes diferencias de edad reducen la estabilidad matrimonial.

No lineales

- `edad_div1^2`: $\beta = 3.50$ [3.39, 3.61]
- `edad_div2^2`: $\beta = 2.95$ [2.83, 3.06]

Existe **curvatura positiva**: a edades mayores, el efecto sobre la duración **se intensifica**.
Ejemplo: la diferencia entre casarse a los 20 vs 30 años no impacta tanto como la diferencia entre 40 vs 50.

- `hijos^2`: $\beta = -0.48$ [-0.52, -0.44]
- `cus_hij^2`: $\beta = -0.49$ [-0.55, -0.42]

Se observan **rendimientos decrecientes**: aunque tener más hijos inicialmente alarga la duración, el efecto se atenúa e incluso se revierte en familias muy numerosas o con muchos hijos en custodia.

- `dif_edad_matrimonio^2`: $\beta = -0.60$ [-0.63, -0.58]

Refuerza el impacto negativo de las grandes diferencias de edad: matrimonios con **brechas muy amplias se disuelven más rápido**.

Por interacción

- `edad_div1 * edad_div2`: $\beta = -6.04$ [-6.25, -5.83]
Interacción negativa: si ambos cónyuges tienen edades altas, la duración tiende a ser menor a lo esperado.
- `hijos * cus_hij`: $\beta = 0.90$ [0.81, 1.00]
El efecto negativo de la custodia se modera cuando hay más hijos en común: **la carga se reparte y reduce el impacto.**
- `cus_hij * dif_edad_matrimonio`: $\beta = 0.29$ [0.25, 0.34]
Contrarresta parcialmente el efecto negativo de la diferencia de edad cuando hay hijos en custodia.

Resumen general (lo más relevante)

- El modelo confirma que la edad al divorcio y el número de hijos son los predictores más robustos de la duración matrimonial.
- La custodia de hijos y la diferencia de edad al casarse son factores que reducen significativamente la duración.
- Los efectos cuadráticos y de interacción muestran que las relaciones sociales son complejas:
 - Los hijos alargan la duración, pero en exceso su efecto se revierte.
 - Las diferencias de edad pequeñas son tolerables, pero diferencias grandes reducen drásticamente la duración.
- El margen de error de los coeficientes es bajo, lo que aporta alta confianza estadística en las inferencias.



~~denuncia~~