

# Informe Final BIO4022

*Martina Alamo R*

*11/10/2019*

## Introducción

### Diabetes

La diabetes es una enfermedad crónica donde el páncreas no presenta la capacidad de secretar los niveles de insulina necesarios para la captación de la glucosa en la sangre. Existen dos tipos de diabetes:

- Diabetes mellitus 1: la cual se genera por una ingesta calórica elevada y constante en el tiempo que tiene como efecto un alza en los niveles de azúcar en la sangre, el cual no es capaz de ser regulado por los niveles de insulina secretado por el páncreas.
- Diabetes tipo 2: sus antecedentes tienen un componente genético, donde las células del páncreas secretan poco o nada de insulina.

En EEUU durante los años 1990 y 2009 hubo un alza en el número de casos diagnosticados con diabetes, lo que es de gran preocupación a problemas de salud pública’.

## Metodología y Resultados

### Análisis de los Datos

En este trabajo nos centraremos en el estudio de la base de datos *Diabetes 130 US hospitals for years 1999-2008* que concentra la información de varios pacientes ingresados a la clínica y que fueron diagnosticados con diabetes y la siguiente información: Edad, Raza, Media\_Tiempo\_en\_clínica, Número\_procedimientos\_lab, Número\_procedimientos, Número\_medicamentos (Strack et al. 2014).

Si realizamos un gráfico poniendo en el eje de la abscisa la Edad y en el eje de la ordenada el Tiempo en la clínica agrupado de acuerdo a la Raza, se obtiene un gráfico que no nos indica mucha información (Wickham 2014).

En este gráfico (Figura 2) no se observa ninguna tendencia de los datos que nos pueda indicar un comportamiento que pueda ser ajustado a un modelo. Sin embargo, lo interesante es la visualización gráfica de cómo se representan los datos en la Figura 2 ya que nos permite ver el tiempo de permanencia en clínica de cada paciente de acuerdo a su Edad y Raza.

### Datos agrupados

A partir del gráfico anterior (Figura 2) nos podemos percatar que es necesario realizar una segunda manipulación de los datos. Para ello se realizó una agrupación de los datos de los pacientes de acuerdo a su edad (definido en intervalos de 10 años) y agrupados respecto a la Raza que los determina para ver su permanencia en la clínica (Tabla 1).

Por otro lado tenemos los datos de los pacientes con diabetes de acuerdo a su edad y la media del número de procedimientos de laboratorio y medicamentos administrados durante su permanencia en la clínica (Tabla 2)

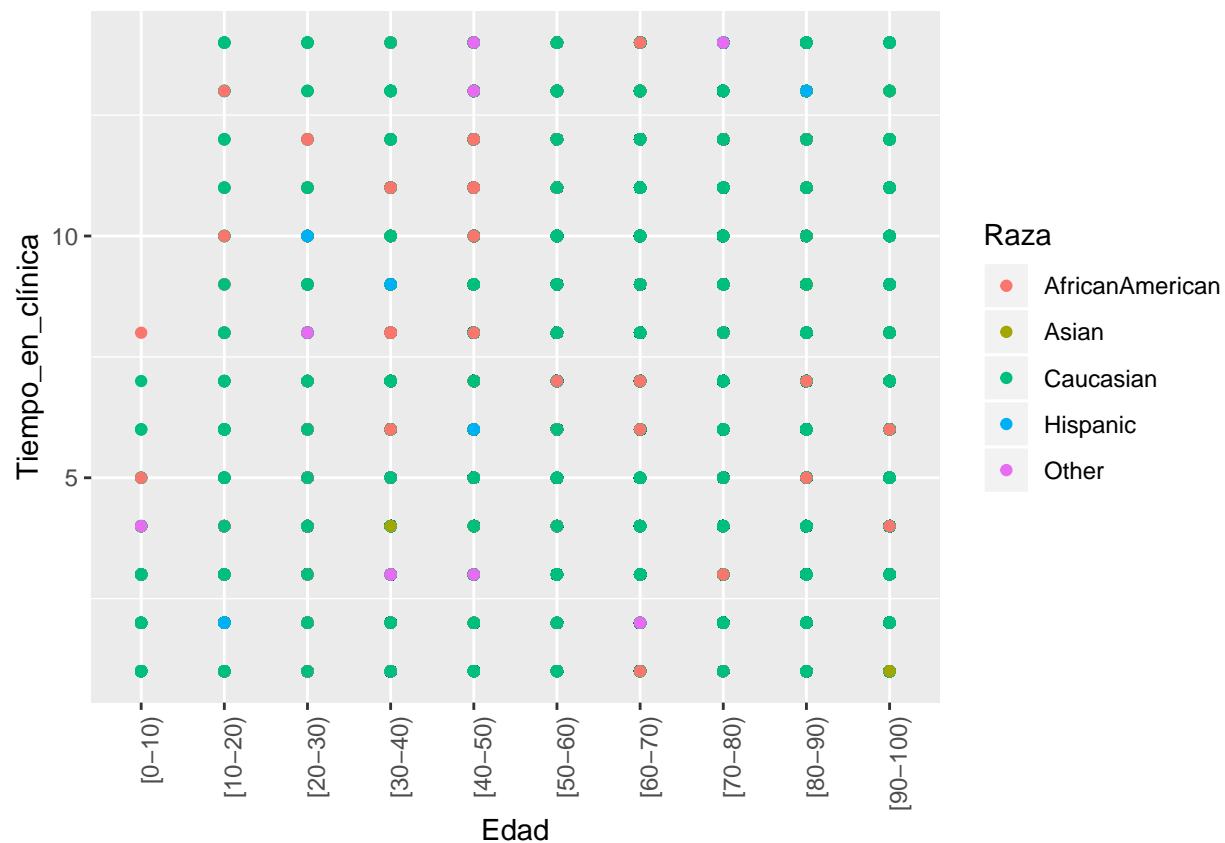


Figure 1: Primera aproximación de la visualización gráfica de los datos. En este gráfico se muestran los pacientes de acuerdo a su edad (en intervalos de 10 años) versus su tiempo de permanencia en la clínica (días) de acuerdo a la Raza representada.

Table 1: Pacientes con diabetes de acuerdo a su Raza y la media del tiempo de permanencia en clínica.

Raza	Media_Tiempo_en_clínica
AfricanAmerican	4.17
Asian	3.49
Caucasian	3.94
Hispanic	3.92
Other	3.85

Table 2: Pacientes con diabetes de acuerdo a su Raza y la media del número de medicamentos y el número de procedimientos realizados durante la permanencia en clínica.

Raza	Media_Número_procedimientos_lab	Media_Número_medicamentos
AfricanAmerican	43.60	13.08
Asian	40.61	11.33
Caucasian	42.81	13.72
Hispanic	41.74	12.20
Other	44.58	12.79

### Recopilación visual de los datos.

Si realizamos otro tipo de representación gráfica de los datos, de acuerdo a la tabla anterior (Tabla1), se obtiene el siguiente gráfico (Figura 3)

A partir de este gráfico se observa una tendencia de que la Media del Tiempo de permanencia en clínica aumenta a medida que progresa la edad de los paciente. Sin embargo, no se observa alguna tendencia en la permanencia del tiempo en clínica de los pacientes de acuerdo a su raza.

### Modelo

De acuerdo a la representación de los datos anteriores podemos sugerir los siguientes modelos

De acuerdo a los valores de AIC de cada modelo descritos en la Tabla 2, el modelo que mejor explica los resultados corresponde al modelo con un AIC de 32.811781134631, Fit2.

### Evaluación del modelo

Para evaluar si el modelo con el mejor AIC (AIC de menor valor) es capaz de ajustarse a los datos, se construyó una nueva base de datos a partir de los datos proporcionados por *Diabetes 130 US hospitals for years 1999-2008*. Estos “nuevos datos” incluyen el tiempo de permanencia en la clínica (días) versus la edad de los pacientes (en intervalos de 10 años).

Como se puede observar en la Figura 3, el modelo con el menor AIC, el Fit2 presenta una distribución de los datos similar al gráfico de la Figura 2. Esto nos sugiere que este modelo es capaz de explicar en gran medida el comportamiento de los datos.

### Discusión y Conclusión

La diabetes es una enfermedad crónica de alta incidencia en el mundo, y en este estudio de gran relevancia es su alta incidencia en EEUU. Durante la época de 1990 a 2009, hubo un gran interés por estudiar la admisión a clínica por pacientes diagnosticados con diabetes.

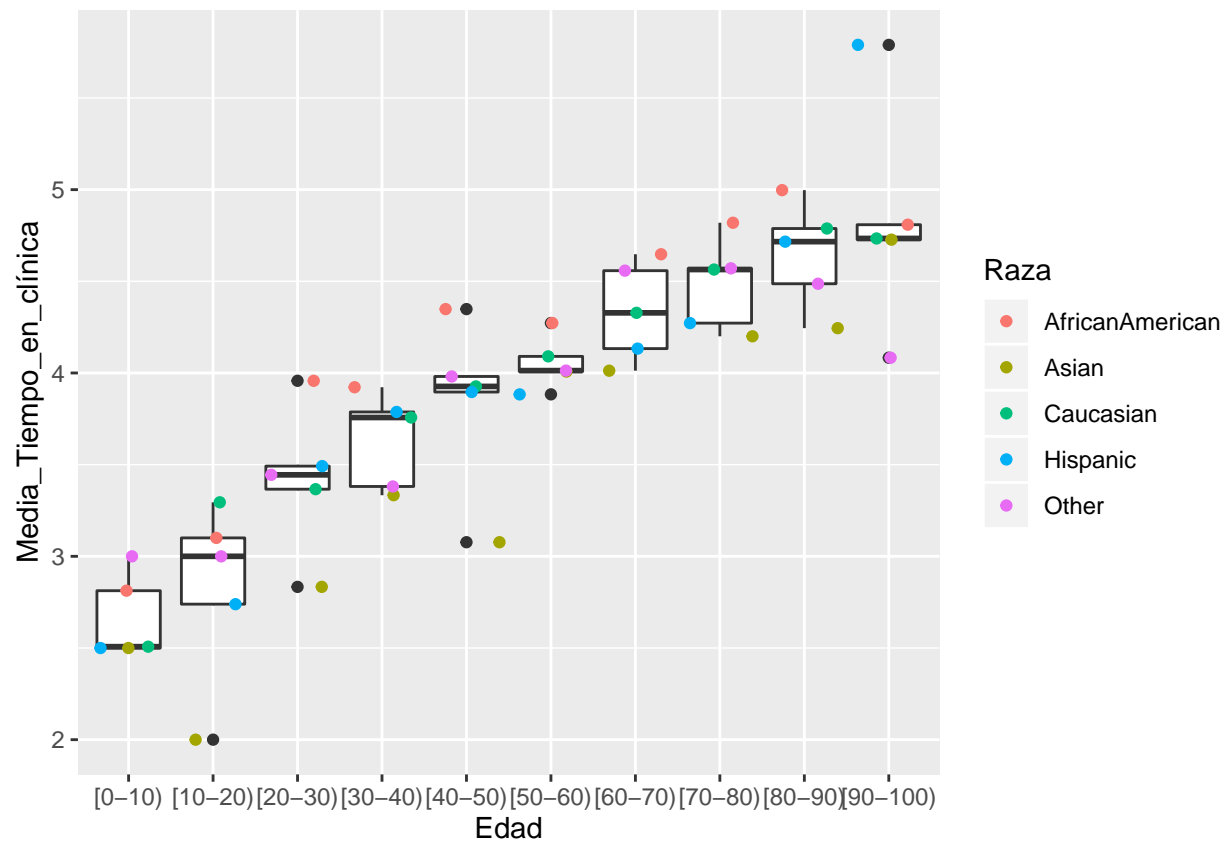


Figure 2: Representación final de los datos. En este gráfico se muestra el Tiempo de permanencia en clínica (días) de los pacientes de acuerdo a su edad (intervalo de 10 años) y definidos por su raza.

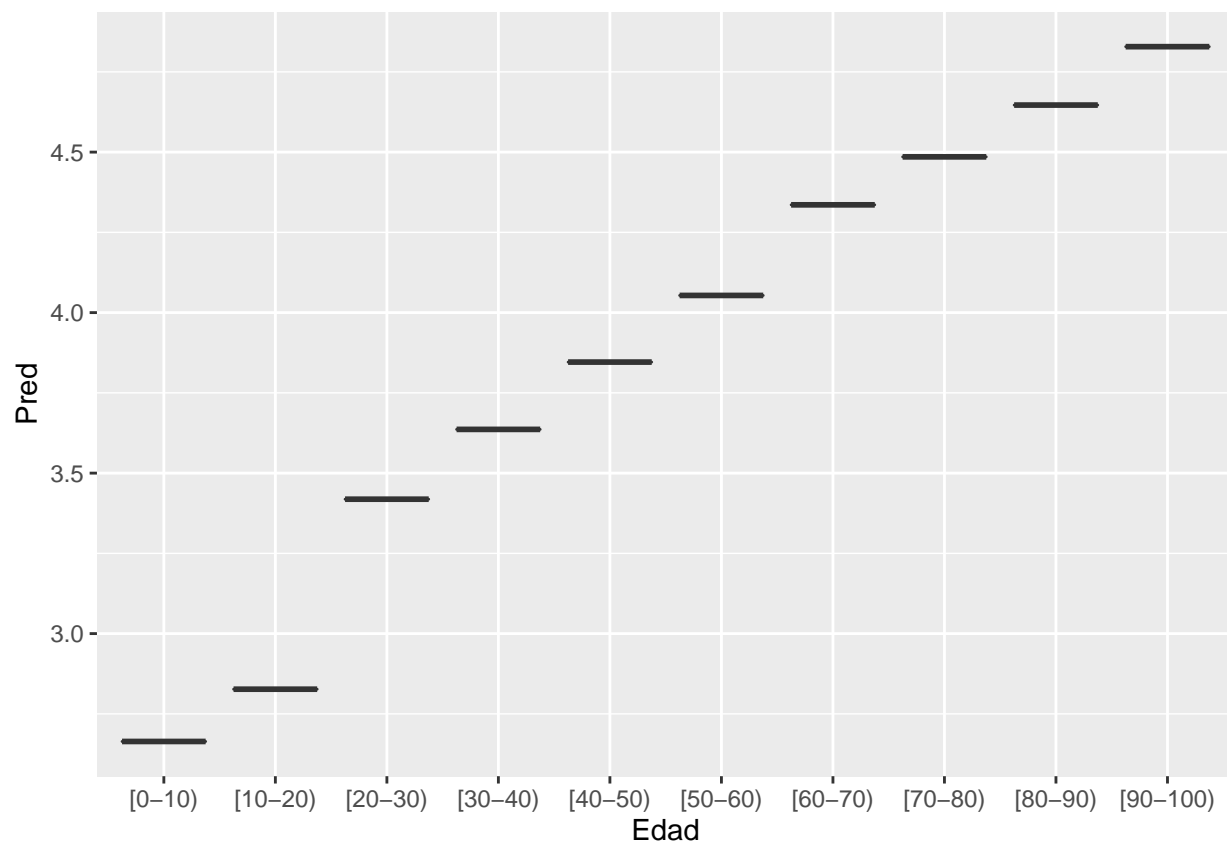


Figure 3: Evaluación del modelo: en este gráfico se muestra como el modelo lineal Fit2 se ajusta y explica la permanencia del tiempo en clínica de los pacientes con diabetes de acuerdo a su edad y raza.

Table 3: Análisis de los modelos propuestos (modelos lineales).

r.squared	AIC	p.value	Modelo	Ecuación
0.90	32.81	0	Fit2	Edad + Raza
0.82	53.32	0	Fit1	Edad

En este trabajo se realizó un análisis explorativo de la base de datos *Diabetes 130 US hospitals for years 1999-2008* que proporciona una diversa cantidad de información respecto a pacientes admitidos en clínica que presentan diabetes.

Del análisis realizado se determinó que la media del tiempo de permanencia en clínica de los pacientes diagnosticados con diabetes aumenta de acuerdo a los años de los pacientes admitidos, a mayor edad es más probable que su tiempo de permanencia en la clínica sea alto. Esto se puede deber al carácter crónico de la enfermedad que a medida que transcurre el tiempo hay un deterioro de los órganos, tales como el hígado, páncreas, corazón, entre otros.

A partir de la visualización gráfica de los datos (Figura 2) se propusieron dos modelos lineales que pudiesen explicar los datos, o que se ajusten a ellos. El primero modelo (Fit1) define que la media de permanencia en clínica es explicado por la Edad, mientras que el segundo modelo (Fit2) define que la media de permanencia en clínica es explicado por la Edad más la Raza. El análisis de estos dos modelos determinó que el modelo Fit2 es el que mejor explica los datos de los pacientes con diabetes, ya que presenta un valor de AIC menor.

La evaluación del modelo Fit2 (Figura 3) nos sugiere que el modelo es capaz de ajustarse a los datos al presentar una gráfica similar a la representación gráfica de los datos mostrados en la Figura 2. Además, se muestra que el valor de cuadrado de r es mayor en el modelo Fit2 en comparación al Fit1. Esto nos permite concluir que el modelo Fit2 es el mejor modelo que explica los datos de *Diabetes 130 US hospitals for years 1999-2008*.

## Referencias

- Strack, Beata, Jonathan P. Deshazo, Chris Gennings, Juan L. Olmo, Sebastian Ventura, Krzysztof J. Cios, and John N. Clore. 2014. "Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records." *BioMed Research International*. <https://doi.org/10.1155/2014/781670>.
- Wickham, Hadley. 2014. "Tidy data." *Journal of Statistical Software*. <https://doi.org/10.18637/jss.v059.i10>.