

PRESENTACION



Nombre: Bradigson Nuñez

Matricula: 2018-6114

Profesor: Keyn Tejada

Trabajo: Informe y reporte del trabajo de Regresion Lineal

Link Video: <https://youtu.be/XLguJjTEIzI>

Introducción

La Inteligencia Artificial a pesar de ser una nueva ciencia es bastante densa y profunda, por lo que consta de muchas ramas, una de estas es el Machine Learning o Aprendizaje Automático.

Este se caracteriza por ser de los más esenciales, ya que trata que la maquina aprenda, que es una de las bases de la Inteligencia Artificial. Por lo que es uno de los más enseñados al momento de introducirse a este mundo.

De todos los temas que tiene el Machine Learning con el que se suele comenzar es la Regresión Lineal. En este trabajo veremos que es la Regresión Lineal y estudiaremos por medio de una práctica y a través de análisis qué papel juega la Regresión Lineal y si esta es en verdad lo suficientemente eficiente para ser llamada Inteligencia Artificial.







Para profundizar un poco mas en el tema, empezare definiendo el machine learning y algunos de sus tipos.

¿Qué es Machine Learning?

El Machine Learning, o Aprendizaje Automático, es un subcampo de las ciencias de la computación y una rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que las computadoras aprendan. Los investigadores del aprendizaje de máquinas buscan algoritmos para convertir muestras de datos en programas de computadora, sin tener que escribir los últimos explícitamente.

Tipos

Son muchos los algoritmos utilizados con el fin de crear Machine Learning, por lo que estos se clasifican en grupos en base a la salida de los mismos. Estos son:

-  Aprendizaje supervisado
-  Aprendizaje no supervisado
-  Aprendizaje semisupervisado
-  Aprendizaje por refuerzo
-  Transducción
-  Aprendizaje multitareas

Mediante la lectura e investigación pude observar que de estos tipos los más famosos son el supervisado y el no supervisado. Pero el que trabajaremos será el Aprendizaje Supervisado.

Cabe destacar que esto es una porción de lo ya visto en clases .

Habalemos un poco de la regresion linea.

La Regresión Lineal es un modelo matemático usado para aproximar la relación de dependencia entre una variable dependiente (Y), las variables independientes (X) y un termino aleatorio.

En este proceso, la variable independiente serían las características que queremos usar para predecir algún valor dado de (Y). Y la variable dependiente es la característica que estamos tratando de predecir.

En este trabajo se le mostrara un ejemplo de Regresión Lineal y se realizara un análisis de los resultados.

Practica de Regresión Lineal Simple

He realizado una porción de código en **R** en el cual, utilizando distintas librerías, puse a prueba la efectividad de la Regresión Lineal con con una tabla de datos libres gob RD, correspondiente a partos del año 2019.

Las librerías que se utilizaron para este proyecto son:

- **library(readxl)**
Para la leer los datos exportados desde la tabla de excel, descargada desde la base de datos asignada (Datos libres Gob RD).
- **library(tidyverse)**
Esta fue utilizada para la manipulacion de datos de la tabla y su informacion.
- **library(dplyr)**
Esta la utilice para la ceacion y manipulacion de las graficas.

El codigo es el siguiente:



```
1 library(readxl)
2 library(tidyverse)
3 library(dplyr)
4 file.choose()
5 datos_partos <- "C:\\users\\Music\\Downloads\\PARTOS EN ADOLESCENTES ENERO-DICIEMBRE 2019.xlsx"
6 partos <- read_excel(datos_partos)
```

En la imagen del codigo podemos observar las librerias y la extracion de la tabla.

```

8 partos1 <- select(partos, PESO, EDAD)
9 pairs(partos1)
10 cor(partos1)
11
12
13
14
15 lm(PESO ~ EDAD, data = partos_datos)
16
17
18 partos_datos %>%
19   ggplot(aes(x = EDAD,
20             y = PESO)) +
21   geom_point() +
22   geom_abline(intercept = 3.43526 ,
23              slope = -0.01876 ,
24              col = 'blue') +
25   geom_vline(xintercept = 16, col = 'red' )
26

```

Análisis a Regresión Lineal con Base de Datos 1.

La tabal descargada contiene un conjunto de datos que contienen la Cantidad de Partos y gestaciones de jóvenes menores de edad de la Republica Dominicana entre enero y diciembre del año 2019.

Los campos utilizados de esta Tabla son la columna **Edad** y **PESO**. Para el entrenamiento del algoritmo utilizaremos un 100% de los datos de los campos o columnas mencionadas.

Luego de que el algoritmo eligiera el mejor conjunto para su entrenamiento y se efectuara el mismo pudimos conseguir los siguientes datos:

- El valor del coeficiente **a** es: 3.43526.
- Es decir, que el valor de la pendiente de la función lineal es de 3.43.
- El valor del coeficiente **b** es: -0.01876.
- Es decir, que la función tiene su intersección en el punto -0.01.
- De esto podemos determinar que la ecuación de la función de regresión lineal es $y = 3.43x + -0.01$.

Como todos sabemos, lo que la regresión lineal hace es trazar una línea recta la cual se ajuste lo más posible a los datos que ingresamos en el entrenamiento, tratando de trackear los valores mas óptimos. La información mostrada son los datos arrojado de dicha recta.

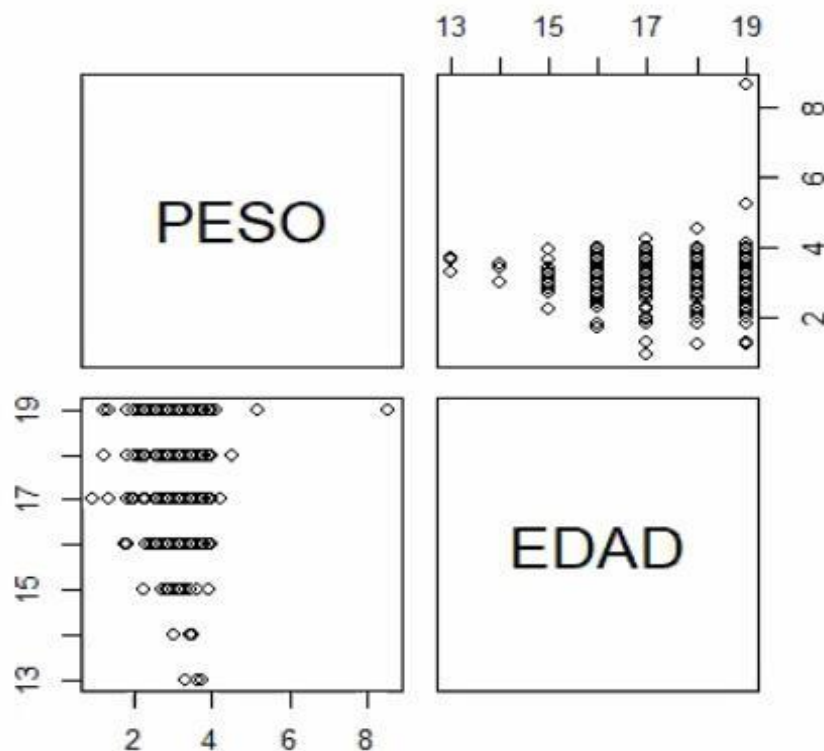
La grafica obtenidas son las siguiente:

Esta es la primera gráfica, esta grafica se obtuvo al tratar de encontrar las correlaciones entre la variable PESO y EDAD.

Utilizando la función `pairs()`, que busca hacer el análisis exploratorio de los datos para realizar las correlaciones.

En dicha grafica podemos ver como no encontramos correlaciones entre las dos variables debido a que la correlación vale cero.

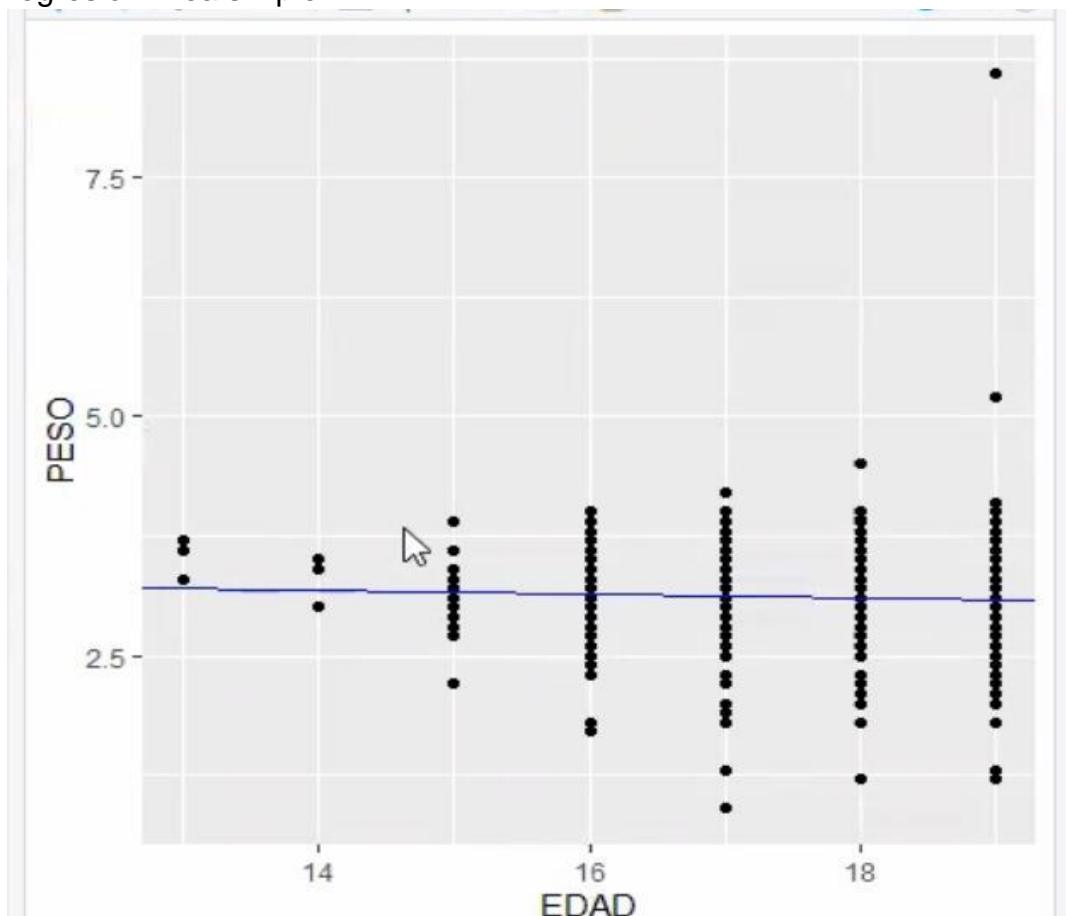
Cuanto más cerca del cero esté el coeficiente de correlación, más débil será la tendencia, es decir, habrá más dispersión en la nube de puntos. Si la correlación vale 1 o -1 diremos que la correlación es "perfecta", si la correlación vale 0 diremos que las variables no están correlacionadas.



Ahora bien si queremos ver estos datos de correlación más detallados o de una forma más óptima podemos utilizar la función `cor()`, que nos permite mostrar los valores a manera exploratoria con los análisis de correlación, lo mismo que en la gráfica, pero en análisis numérico.

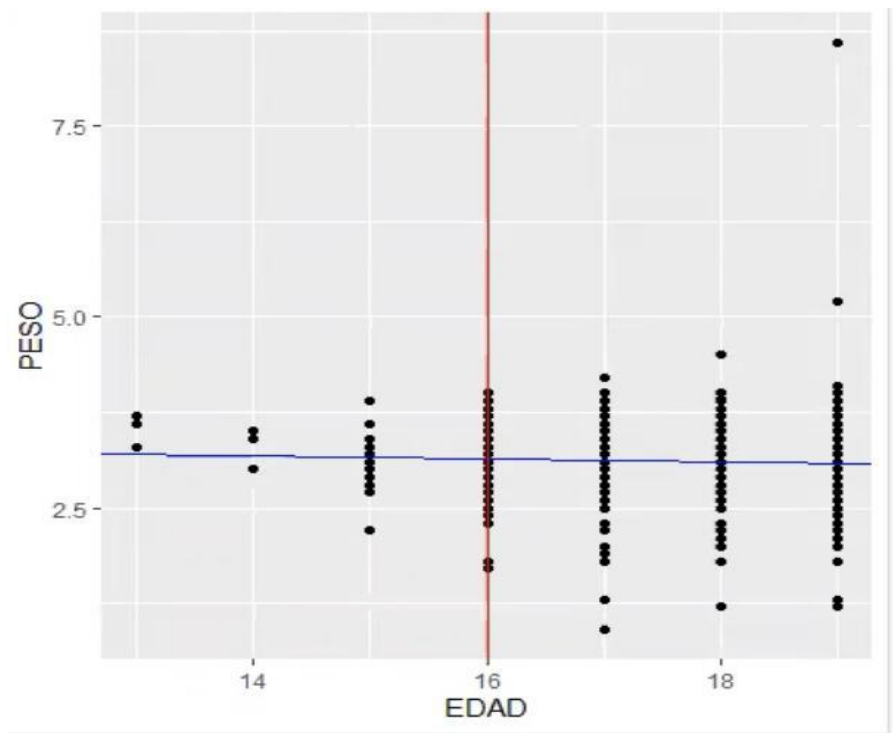
```
> cor(partos1)
           PESO      EDAD
PESO  1.00000000 -0.04192963
EDAD -0.04192963  1.00000000
> |
```

Ahora echemosle un vistazo al resultado mostrado en grafica de nuestra regresion linea simple:



Como podemos ver, luego de obtener los valores mas optimo la recta se encuentra lo más próxima posible a los puntos donde se encuentran los datos. Cada punto está determinado por un dato de entrada (el eje x), y un dato de salida (el eje y). El punto en el plano se encontraría de la forma siguiente (entrada, salida).

También se tomó una edad como referencia para trazar una línea para observar su punto de corte y de donde está cerca.



En la gráfica podemos observar como se toma la edad de 16 años y se traza una línea vertical donde con la línea azul se puede observar su punto de corte, el mismo aparenta estar entre 2,8, 2.9 o 3, pero si queremos saber con exactitud el valor del punto de corte, ese valor exacto con el que corta la línea, podemos usar el siguiente código.

```
26  
27  
28 edad <- 16  
29 medida <- -0.01876 * edad + 3.43526  
30 print(medida)  
17:1 (Top Level) ↕
```

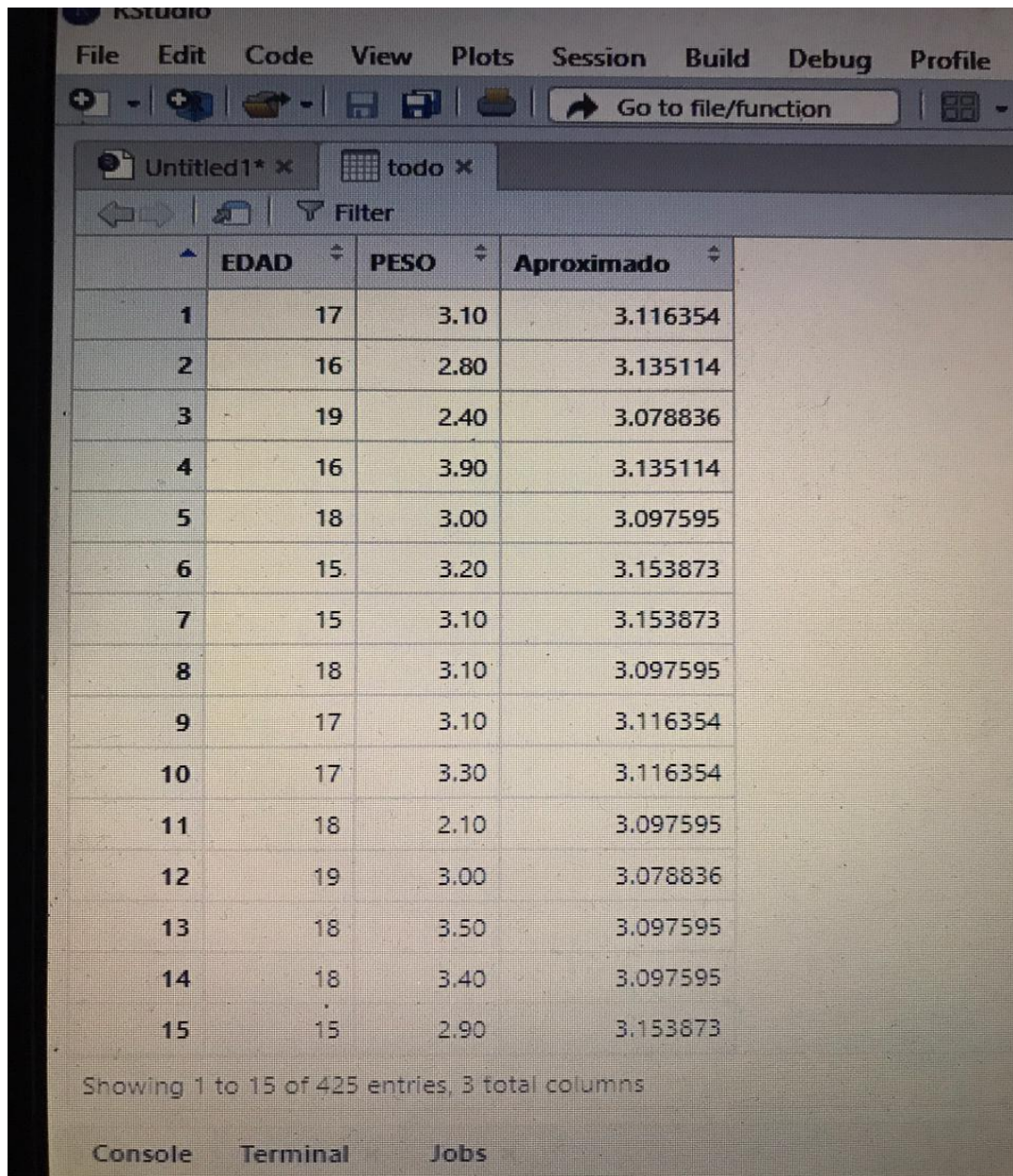
Con el siguiente código podremos observar el valor exacto con el que corta la línea roja.

El resultado sería:

```
> edad <- 16  
> medida <- -0.01876 * edad + 3.43526  
> print(medida)  
[1] 3.1351  
> |
```

Aca podemos observar que el valor exacto sería un [3.1351](#).

Ahora bien si extraemos los campos **PESO** y **EDAD** podremos ver que en muchos de los casos los valores son casi exactos con lo de nuestra regresión lineal.



The screenshot shows the RStudio interface with a data table displayed. The table has four columns: an index column, 'EDAD' (Age), 'PESO' (Weight), and 'Aproximado' (Approximate). The data is sorted by index, showing 15 rows. The 'Aproximado' column contains values that appear to be predictions from a linear regression model. The status bar at the bottom indicates 'Showing 1 to 15 of 425 entries, 3 total columns'.

	EDAD	PESO	Aproximado
1	17	3.10	3.116354
2	16	2.80	3.135114
3	19	2.40	3.078836
4	16	3.90	3.135114
5	18	3.00	3.097595
6	15	3.20	3.153873
7	15	3.10	3.153873
8	18	3.10	3.097595
9	17	3.10	3.116354
10	17	3.30	3.116354
11	18	2.10	3.097595
12	19	3.00	3.078836
13	18	3.50	3.097595
14	18	3.40	3.097595
15	15	2.90	3.153873

Ojo: Tube ue tirarle fotos ya que no queria hacer el screen shot la pc

	EDAD	PESO	Aproximado
16	18	2.70	3.097595
17	17	3.50	3.116354
18	17	3.50	3.116354
19	16	2.70	3.135114
20	17	3.50	3.116354
21	18	3.20	3.097595
22	17	3.40	3.116354
23	18	2.10	3.116354
24	17	3.50	3.135114
25	17	3.00	3.078836
26	17	2.80	3.135114
27	15	2.80	3.097595
28	18	3.00	3.153873
29	17	3.60	3.153873
30	17	3.40	3.097595

Showing 16 to 30 of 425 entries, 3 total columns

Console Terminal Jobs

En la grafica podemos observar como muchos de los valores se aproximan como en la fila **23, 25**. Y que otros no, como en la fiala **16 19**.

Esto sucede por:

- Que como dijimos, la regresión lineal es una línea que se acerca lo más posible al resultado, por lo que puntos tan alejados de la recta como los primeros tendrán un resultado totalmente erróneo.

Reporte De La Regresion Lineal Simple en R

En este reporte encontraremos los campos obtenidos de la tabla Embarazos, columna **PESO** y **EDAD**, con el resultados de la regresion lineal simple a su lado (Columna Aproximacion). La misma contiene una cnatidad de 425 filas con sus respectivos valores. (Solo muestro Una pocion de los 425 datos).

EDAD	PESO	Aproximado
17	3.1	3.11635423
16	2.8	3.13511356
19	2.4	3.07883559
16	3.9	3.13511356
18	3	3.09759491
15	3.2	3.15387288
15	3.1	3.15387288
18	3.1	3.09759491
17	3.1	3.11635423
17	3.3	3.11635423
18	2.1	3.09759491
19	3	3.07883559
18	3.5	3.09759491
18	3.4	3.09759491
15	2.9	3.15387288
18	2.7	3.09759491
17	3.5	3.11635423
17	3.5	3.11635423
16	2.7	3.13511356
17	3.5	3.11635423
18	3.2	3.09759491
17	3.4	3.11635423
18	2.1	3.11635423
17	3.5	3.13511356
17	3	3.07883559
17	2.8	3.13511356
15	2.8	3.09759491
18	3	3.15387288
17	3.6	3.15387288
17	3.4	3.09759491
19	3.4	3.11635423
19	3.8	3.11635423
15	3.4	3.09759491
17	3.3	3.07883559
18	3.4	3.09759491
19	2.6	3.09759491
18	2.5	3.15387288
16	3.5	3.09759491
18	4	3.11635423
16	2.4	3.11635423
19	3.4	3.13511356
19	2.7	3.11635423
19	2.2	3.09759491
19	3.5	3.11635423
19	3.5	3.09759491
18	3.6	3.11635423
19	1.8	3.11635423
18	3.4	3.11635423
17	3.1	3.15387288

18	2.7	3.09759491
16	2.8	3.11635423
17	3.5	3.11635423
18	2.7	3.07883559
19	2.8	3.07883559
15	3.3	3.15387288
17	2.2	3.11635423
18	3.4	3.09759491
18	3.2	3.07883559
15	3.9	3.09759491
18	3.1	3.13511356
19	3.5	3.09759491
17	3.3	3.13511356
17	2.5	3.07883559
16	2.5	3.07883559
16	3.2	3.07883559
17	3.4	3.07883559
13	3.6	3.07883559
17	3.7	3.09759491
19	3.5	3.07883559
18	2.2	3.09759491
16	1.8	3.11635423
19	1.2	3.09759491
17	3.1	3.13511356
18	3.9	3.11635423

16	3.7	3.09759491
18	2.2	3.07883559
15	3.2	3.15387288
18	3.5	3.11635423
16	2.9	3.09759491
18	3.4	3.09759491
19	3	3.15387288
18	2	3.09759491
15	3	3.07883559
16	3.4	3.11635423
19	3.3	3.11635423
19	1.3	3.13511356
19	3.3	3.13511356
16	3.5	3.11635423
17	3.4	3.19139153
16	3.1	3.11635423
17	3.6	3.07883559
17	3.6	3.09759491
19	2.3	3.13511356
18	2.8	3.07883559
16	3.1	3.11635423
17	2.9	3.09759491
16	3.9	3.13511356
18	3.93	3.09759491
16	4	3.15387288

18	3.4	3.09759491
19	3.3	3.13511356
19	3.4	3.09759491
19	2.2	3.11635423
19	2.8	3.13511356
17	1.9	3.07883559
18	3.2	3.13511356
17	3	3.09759491
17	3	3.15387288
18	3	3.15387288
18	2.9	3.09759491
18	3.8	3.11635423
17	3.4	3.11635423
17	3.6	3.09759491
19	2.6	3.07883559
15	3.2	3.09759491
17	3	3.09759491
17	2	3.15387288
19	5.2	3.09759491
19	3.4	3.11635423
14	3.4	3.11635423
17	3.1	3.13511356
18	3.5	3.11635423
19	3.3	3.09759491
18	3.3	3.11635423

18	3.4	3.09759491
18	3.2	3.11635423
17	3.3	3.11635423
18	3.9	3.11635423
18	3.1	3.15387288
16	3.5	3.09759491
17	2.8	3.11635423
19	3	3.11635423
19	2.3	3.07883559
17	2.9	3.07883559
19	3.1	3.15387288
19	3.1	3.11635423
15	3.1	3.09759491
18	3.1	3.07883559
19	3	3.09759491
16	3	3.13511356
19	3.2	3.09759491
16	3.8	3.13511356
16	3.2	3.07883559
18	3.1	3.07883559
19	2.9	3.07883559
19	2.7	3.07883559
17	4	3.07883559
16	3	3.09759491
16	3.4	3.07883559

19	2.5	3.09759491	16	3.3	3.11635423
16	3.4	3.11635423	19	3.4	3.09759491
19	3.5	3.09759491	19	3.1	3.13511356
16	3	3.13511356	17	2.6	3.09759491
17	3	3.11635423	17	3.7	3.15387288
17	2.8	3.09759491	16	3.3	3.09759491
17	3.5	3.07883559	17	3	3.13511356
19	3	3.15387288	17	3.6	3.09759491
17	0.9	3.11635423	17	3.8	3.11635423
18	3.1	3.09759491	19	2.8	3.13511356
19	2	3.09759491	19	3.5	3.07883559
16	2.8	3.15387288	18	3.5	3.13511356
17	3.6	3.09759491	18	2.2	3.09759491
15	3.4	3.07883559	17	3.5	3.15387288
19	3	3.11635423	19	3.2	3.15387288
19	2.5	3.11635423	19	3.2	3.09759491
15	3.6	3.13511356	14	3.5	3.11635423
19	2.4	3.13511356	19	3.1	3.11635423
17	3.6	3.11635423	18	3.1	3.09759491
18	3	3.19139153	18	3.6	3.07883559
17	3.4	3.11635423	18	2.7	3.09759491
19	3.7	3.07883559	17	3.8	3.09759491
19	2.8	3.09759491	17	2.8	3.15387288
17	2.3	3.13511356	17	3.2	3.09759491
16	3.3	3.07883559	18	2.5	3.11635423

15	3.2	3.11635423	18	4.5	3.09759491
17	3	3.13511356	15	2.7	3.11635423
16	3.7	3.11635423	18	2.5	3.09759491
16	2.9	3.09759491	17	3.9	3.11635423
17	3.7	3.11635423	17	3	3.11635423
19	3.4	3.11635423	16	3.6	3.11635423
18	3	3.13511356	18	3.2	3.15387288
19	3	3.07883559	16	3.2	3.09759491
15	3.9	3.13511356	19	2.9	3.11635423
18	3.5	3.09759491	17	3.3	3.11635423
17	2.2	3.15387288	17	3	3.07883559
19	1.8	3.15387288	16	3.4	3.07883559
19	2.2	3.09759491	18	2.5	3.15387288
19	2.2	3.11635423	16	2.9	3.11635423
17	3	3.11635423	18	3.6	3.09759491
17	3	3.09759491	18	3.3	3.07883559
19	3.2	3.07883559	18	3.1	3.09759491
17	3.1	3.09759491	19	2.8	3.13511356
19	3.5	3.09759491	16	2.3	3.09759491
18	3.2	3.15387288	18	2.7	3.13511356
18	2.8	3.09759491	18	2.7	3.07883559
16	2.8	3.11635423	16	3	3.07883559
17	2.3	3.11635423	14	3	3.07883559
18	2.6	3.13511356	17	3.1	3.07883559
13	3.3	3.11635423	19	3.9	3.07883559

17	3.1	3.09759491
18	3.1	3.11635423
18	3.5	3.11635423
17	2	3.13511356
16	2.8	3.11635423
18	3	3.09759491
18	3.7	3.11635423
19	2.7	3.09759491
18	3.8	3.11635423
19	3.5	3.11635423
19	3.8	3.11635423
19	3.5	3.15387288
18	3.4	3.09759491
18	2.7	3.11635423
18	3.1	3.11635423
17	1.3	3.07883559
19	3	3.07883559
18	3	3.15387288
17	3.4	3.11635423
19	3	3.09759491
17	3.7	3.07883559
18	2.8	3.09759491
17	3.2	3.13511356
16	3.6	3.09759491
18	2.9	3.13511356

17	4	3.07883559
16	3	3.07883559
18	2.9	3.07883559
18	3.8	3.07883559
16	3.2	3.07883559
16	3.1	3.09759491
18	1.2	3.07883559
19	2.9	3.09759491
19	3.3	3.11635423
17	3.8	3.09759491
17	4	3.13511356
16	3	3.11635423
16	3.9	3.09759491
16	3.1	3.07883559
17	2.7	3.15387288
18	3.1	3.11635423
17	3	3.09759491
14	3.5	3.09759491
19	2.9	3.15387288
17	2.9	3.09759491
19	3	3.07883559
19	3.2	3.11635423
15	3.3	3.11635423
16	3	3.13511356
19	3.2	3.13511356