

Universidad de Chile

Facultad de Economía y Negocios

# EL MUNDO DE LA COMPUTACIÓN Y LAS CIENCIAS DE LAS MATEMÁTICAS

*Trabajo de Curso*

Autores:

Ignacio Morales Mi y Alejandro Cerda Oyarzún

Verano 2023, Métodos Cuantitativos 1

Índice

|  |           |
|--|-----------|
| <b>1. Introducción</b>   | <b>2</b>  |
| <b>2. Parte I</b>  | <b>3</b>  |
| 2.1. Introducción y metodología . . . . .                        | 3         |
| 2.2. Datos: Estadísticas descriptivas . . . . .                  | 3         |
| 2.3. Modelos . . . . .   | 4         |
| 2.3.1. Modelo base: Regresión Lineal . . . . .                   | 4         |
| 2.3.2. Modelo extendido: Regresión Lineal . . . . .              | 5         |
| 2.3.3. Modelo extendido: Variables condicionadas . . . . .       | 6         |
| 2.4. Conclusión y observaciones . . . . .                        | 6         |
| <b>3. Parte II</b>   | <b>7</b>  |
| 3.1. Introducción y metodología . . . . .                        | 7         |
| 3.2. Datos: Estadísticas descriptivas . . . . .                  | 7         |
| 3.3. Modelos . . . . .   | 9         |
| 3.3.1. Modelo base: Regresión de Intervalo Lineal . . . . .      | 9         |
| 3.3.2. Modelo extendido: Regresión de Intervalo Lineal . . . . . | 10        |
| 3.4. Observaciones y conclusión . . . . .                        | 10        |
| <b>4. Conclusión final</b>                                       | <b>11</b> |
| <b>5. Librerías y códigos</b>                                    | <b>12</b> |
| <b>Referencias</b>   | <b>12</b> |

# 1. Introducción

Durante las últimas décadas los avances tecnológicos relacionados a la computación han sido exponenciales, creando nuevas formas de ver el mundo, y así, creando nuevas variables que pueden predecir resultados en diferentes áreas. Por eso, hemos pensado en como se relacionan algunos aspectos del mundo de la computación con los resultados académicos de alumnos universitarios.

Los primeros tópicos que se nos vinieron a la mente fueron la programación, como una disciplina fundamental en el avance de la tecnología, y los videojuegos online, uno de los entretenimientos que más se ha desarrollado en los últimos años. Debido a nuestro interés en estos dos temas, decidimos dividir nuestro trabajo en dos partes, una relacionado a la programación y otra a los videojuegos. En ambos, nuestro objetivo es predecir resultados académicos con variables explicativas.

Espero que disfruten este trabajo como nosotros disfrutamos hacerlo.

*Ignacio Morales Mi y Alejandro Cerda Oyarzún*

## 2. Parte I

### 2.1. Introducción y metodología

Durante la primera parte, quisimos relacionar las notas del ramo Programación y Algoritmo con las notas de los ramos previos de la malla curricular. Para esto, conseguimos las nóminas finales de todas las secciones de los ramos: Programación y Algoritmo, Estadística I, Métodos Matemáticos III y Contabilidad II. Por motivos de sencillez, los ramos pueden ser referenciados como programación, estadística, matemáticas y contabilidad.

Cabe destacar que los 4 ramos antes mencionados son obligatorios para alumnos de IICG, y pertenecen al 3r (Est. I, MM. III y Cont. II) y 4to semestre (Programación y Algoritmo).

Luego, con las nóminas finales, creamos un archivo de Excel con las notas de Programación, y con ayuda de la función buscarv, conseguimos crear una base de datos donde cada rut tenía asignada su nota final en los 3 ramos previos que consideramos en la investigación. En total, el tamaño de la muestra es de 76.

Quisimos ir más allá y encontrar nuevas variables relacionadas al alumno correspondiente. Por eso, escribimos un código de Python que nos pudo entregar el género y la residencia del estudiante, pudiendo obtener una base de datos más completa y con más variables.

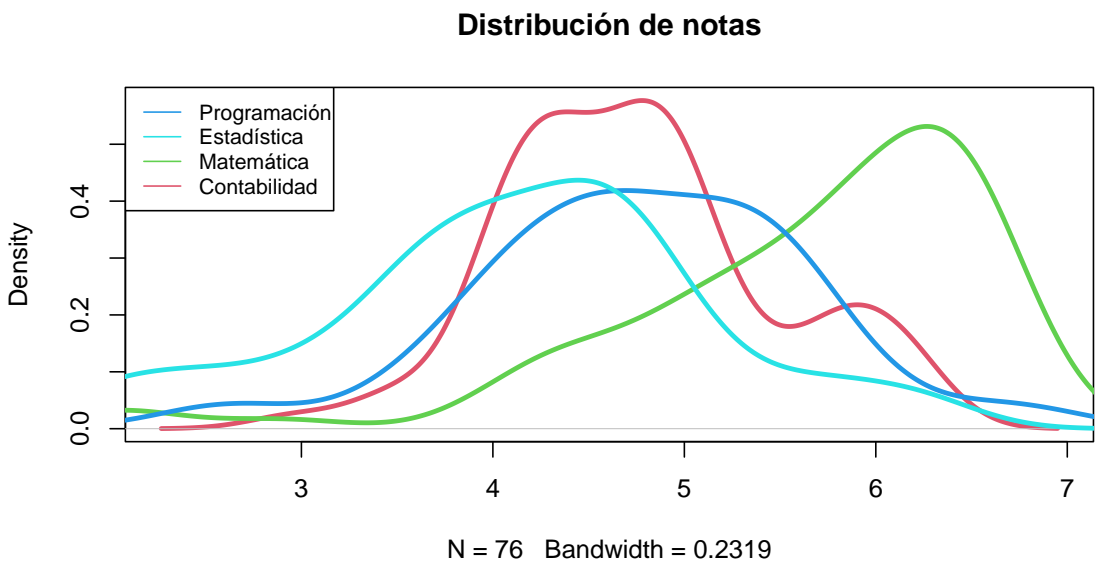
Las librerías usadas en Python fueron: Selenium, para web scrapping, y openpyxl, para leer, escribir y guardar archivos de Excel. Los datos reunidos son solo para la investigación y no serán difundidos.

### 2.2. Datos: Estadísticas descriptivas

La distribución de notas por ramo está representada por la siguiente tabla:

| Programación  | Estadística   | Matemática    | Contabilidad  |
|---------------|---------------|---------------|---------------|
| Min. :1.000   | Min. :1.798   | Min. :2.000   | Min. :2.964   |
| 1st Qu.:4.100 | 1st Qu.:3.570 | 1st Qu.:5.175 | 1st Qu.:4.251 |
| Median :4.709 | Median :4.160 | Median :5.900 | Median :4.686 |
| Mean :4.630   | Mean :4.121   | Mean :5.624   | Mean :4.754   |
| 3rd Qu.:5.341 | 3rd Qu.:4.673 | 3rd Qu.:6.325 | 3rd Qu.:5.072 |
| Max. :6.900   | Max. :6.279   | Max. :6.800   | Max. :6.254   |

Además, el 59 % de los alumnos son hombres y el 71 % vive en la capital, Santiago. El siguiente gráfico muestra la densidad de notas según el ramo:



2.3. Modelos

2.3.1. Modelo base: Regresión Lineal

Para nuestro modelo base, solo consideramos las notas de los 3 ramos previos, Estadística, Matemática y Contabilidad. Empezamos realizando una regresión lineal normal, para ver con que datos estamos tratando y si podemos ya encontrar una relación entre variables.

|               | Estimate | Std. Error | t value | Pr(> t ) |
|---------------|----------|------------|---------|----------|
| (Intercept)   | 0.9193   | 0.7879     | 1.17    | 0.2471   |
| notas_estad   | 0.2223   | 0.1471     | 1.51    | 0.1350   |
| notas_mateIII | 0.3532   | 0.1563     | 2.26    | 0.0269   |
| notas_conta   | 0.1702   | 0.2000     | 0.85    | 0.3977   |

Multiple R-squared: 0.2908, Adjusted R-squared: 0.2613. F-statistic: 9.843 on 3 and 72 DF, p-value: 1.612e-05

Podemos notar que, de nuestros ramos previos, Matemáticas III influye significativamente en la nota de Programación, Estadística I también influye pero con un menor coeficiente y con un mayor P-value, y que Contabilidad I influye muy poco y con un P-value bastante grande.

Cabe destacar que estamos trabajando con pocas muestras, por lo que se nos es difícil probar otras opciones como método de Ridger/Lasso con validación cruzada. Aún así, lo intentamos, pero tras realizar muchos intentos, no pudimos sacar conclusiones, porque en algunas realizaciones bajaba a 0 ciertos coeficientes y luego otros diferentes.

Decidimos usar selección de variables hacia adelante y hacia atrás, para ver si contabilidad era descartada por estos métodos. Obtuvimos lo siguiente:

|               | Estimate | Std. Error | t value | Pr(> t ) |
|---------------|----------|------------|---------|----------|
| (Intercept)   | 1.3083   | 0.6405     | 2.04    | 0.0447   |
| notas_estad   | 0.2521   | 0.1426     | 1.77    | 0.0813   |
| notas_mateIII | 0.4060   | 0.1431     | 2.84    | 0.0059   |

Multiple R-squared: 0.2837, Adjusted R-squared: 0.2641. F-statistic: 14.46 on 2 and 73 DF, p-value: 5.139e-06

Esta regresión sin la variable de contabilidad tiene un mayor R cuadrado ajustado, y un menor p-value en el estadístico F.

Aún así, si realizamos una regresión simple con solo contabilidad como variable explicativa, obtenemos lo siguiente:

|             | Estimate | Std. Error | t value | Pr(> t ) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 1.6498   | 0.8090     | 2.04    | 0.0450   |
| notas_conta | 0.6270   | 0.1684     | 3.72    | 0.0004   |

Multiple R-squared: 0.1578, Adjusted R-squared: 0.1465. F-statistic: 13.87 on 1 and 74 DF, p-value: 0.0003802

Podemos notar que contabilidad es relevante cuando se trata de explicar el rendimiento, pero que se queda atrás en cuanto a la precisión.

Esto es porque la nota de contabilidad no añade información a la regresión de nuestro modelo final, ya que estadística y matemática pueden explicar mejor el rendimiento si solo consideramos estas dos variables.

2.3.2. Modelo extendido: Regresión Lineal

Para nuestro modelo extendido consideramos 2 nuevas variables: Sexo y residencia. Ambas son variables dummy, la primera toma el valor 1 cuando el alumno es hombre, y la segunda, cuando el alumno reside en Santiago.

Primero, realizamos una regresión lineal simple para ver con que datos estamos tratando:

|               | Estimate | Std. Error | t value | Pr(> t ) |
|---------------|----------|------------|---------|----------|
| (Intercept)   | 1.1363   | 0.8662     | 1.31    | 0.1939   |
| notas_estad   | 0.2343   | 0.1492     | 1.57    | 0.1209   |
| notas_mateIII | 0.3558   | 0.1612     | 2.21    | 0.0305   |
| notas_conta   | 0.1259   | 0.2118     | 0.59    | 0.5543   |
| sexo_var      | -0.1839  | 0.2359     | -0.78   | 0.4383   |
| r_santiago    | 0.0538   | 0.2488     | 0.22    | 0.8293   |

Multiple R-squared: 0.2977, Adjusted R-squared: 0.2475. F-statistic: 5.934 on 5 and 70 DF, p-value: 0.0001248

Podemos ver, que al igual que en nuestro modelo base, estadística y matemáticas son las variables más incidentes, y que las demás tienen un p-value bastante grande.

AL igual que en nuestro modelo base, usamos selección de variables hacia adelante y hacia atrás. Nuestro modelo, quedó como:

|               | Estimate | Std. Error | t value | Pr(> t ) |
|---------------|----------|------------|---------|----------|
| (Intercept)   | 1.3083   | 0.6405     | 2.04    | 0.0447   |
| notas_estad   | 0.2521   | 0.1426     | 1.77    | 0.0813   |
| notas_mateIII | 0.4060   | 0.1431     | 2.84    | 0.0059   |

Multiple R-squared: 0.2837, Adjusted R-squared: 0.2641. F-statistic: 14.46 on 2 and 73 DF, p-value: 5.139e-06

Cabe destacar, que en este modelo si aplicamos validación cruzada con Ridger y Lasso, resultando que en la gran mayoría de veces las variables de sexo y residencia eran descartadas (En el caso de Lasso).

Nuestro modelo quedó igual a nuestro modelo base, tanto usando el método de selección de variables hacia adelante como hacia atrás. En este caso, también usamos el método mixto, quedando en las mismas variables a considerar.

Quisimos, al igual que en la parte anterior, generar una regresión solo con las variables descartadas, en este caso:

|             | Estimate | Std. Error | t value | Pr(> t ) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 4.6594   | 0.2896     | 16.09   | 0.0000   |
| sexo_var    | -0.2687  | 0.2604     | -1.03   | 0.3056   |
| r_santiago  | 0.1833   | 0.2822     | 0.65    | 0.5181   |

Multiple R-squared: 0.02103, Adjusted R-squared: -0.005788. F-statistic: 0.7842 on 2 and 73 DF, p-value: 0.4603

Con un R cuadrado ajustado negativo, y un p-value del estadístico de prueba F de 0.46, estamos en presencia de uno de los peores modelos de regresión para explicar las notas de programación

2.3.3. Modelo extendido: Variables condicionadas

Dentro de nuestro modelo extendido, queríamos ver si tenía incidencia en la nota final ser hombre y vivir en Santiago, por lo que añadimos una multiplicación de variables al modelo:

|                     | Estimate | Std. Error | t value | Pr(> t ) |
|---------------------|----------|------------|---------|----------|
| (Intercept)         | 1.4236   | 0.9348     | 1.52    | 0.1323   |
| notas_estad         | 0.2378   | 0.1496     | 1.59    | 0.1166   |
| notas_mateIII       | 0.3394   | 0.1627     | 2.09    | 0.0407   |
| notas_conta         | 0.1222   | 0.2123     | 0.58    | 0.5667   |
| sexo_var            | -0.4921  | 0.4406     | -1.12   | 0.2679   |
| r_santiago          | -0.2032  | 0.3980     | -0.51   | 0.6112   |
| sexo_var:r_santiago | 0.4234   | 0.5108     | 0.83    | 0.4100   |

Multiple R-squared: 0.3046, Adjusted R-squared: 0.2442. F-statistic: 5.038 on 6 and 69 DF, p-value: 0.0002462

Nuevamente, matemática y estadística son las variables con un p-value más pequeño. Realizando una selección de variables con el modelo hacia adelante, hacia atrás y mixto llegamos al mismo resultado, donde solo quedan matemática y estadística:

|               | Estimate | Std. Error | t value | Pr(> t ) |
|---------------|----------|------------|---------|----------|
| (Intercept)   | 1.3083   | 0.6405     | 2.04    | 0.0447   |
| notas_estad   | 0.2521   | 0.1426     | 1.77    | 0.0813   |
| notas_mateIII | 0.4060   | 0.1431     | 2.84    | 0.0059   |

Multiple R-squared: 0.2837, Adjusted R-squared: 0.2641. F-statistic: 14.46 on 2 and 73 DF, p-value: 5.139e-06

Realizamos modelos exponenciales, pero debido a la baja disponibilidad de datos y la inviabilidad de la validación cruzada, decidimos dejarlos en el anexo.

2.4. Conclusión y observaciones

Tanto en nuestro modelo base como en el extendido, llegamos al mismo resultado. La regresión lineal que mejor predice los datos, con las variables que consideramos , es:

$$Pr_i = 1,3083 + 0,2521E_i + 0,4060M_i + \epsilon_i$$

Consideramos que son variables significativas, debido a entre muchas cosas, a su p-value por coeficiente.  $E_i$  y  $M_i$  son las notas de estadística y matemática respectivamente, y  $Pr_i$  es la nota de programación, variable dependiente.

Antes de comenzar nuestro trabajo, consultamos diferentes estudios de investigación relacionados al impacto del género el las notas de programación, y en algunos estudios se demuestra que los rendimientos de las mujeres son marginalmente diferente al de los hombres, y que a cierto nivel (Ramos introductorios-avanzados) uno de los sexos promedia mejor nota, mientras que en otro sucede lo contrario (Mora-Rivera y cols., 2017). Esto ha sido comprobado en esta investigación, debido a que la variable que denota el sexo ha sido descartado en todas las metodologías de selección de variables que consideramos.

Observaciones que queremos destacar: Primero, hay un 4to curso previo en la malla curricular a Programación y Algoritmo, Introducción a la Macroeconomía. Decidimos no incluirlo debido a la dificultad de conseguir las nóminas finales, ya que al ser un ramo común pero no coordinado habían 9 secciones. Además cada una tenía metodologías diferentes. Segundo, no consideramos ramos previos como AMD porque el semestre correspondiente fue online, y queríamos evitar sesgo, como ha sido demostrado por varios estudios, donde alumnos de ciertas comunas tuvieron mejor acceso a clases remotas (Baeza Martín, 2021) y ciertos alumnos pudieron tener rendimientos sobre o bajo su promedio normal, por causas como el estrés o la ansiedad

### 3. Parte II

#### 3.1. Introducción y metodología

Para la segunda parte, nos quisimos enfocar en los videojuegos online, debido a que nosotros mismos somos activos usuarios de estas plataformas, y varias veces sentimos que nuestro rendimiento académico estaba condicionado con las horas que dedicábamos a jugar.

Decidimos relacionar los videojuegos (Y su contenido) con las notas de los ramos de estadística, ya que lo consideramos uno de los cursos más complicados en cuanto a dificultad se refiere. Para esto, creamos una encuesta donde preguntamos los promedios de los ramos de estadística, y algunas variables relacionadas tanto a la persona como a los videojuegos.

Esta encuesta fue distribuida de manera no probabilística, entre estudiantes universitarios que están cursando carreras donde realizaron 1 o más ramos de estadística. También, fue distribuida entre diferentes facultades y diferentes universidades, con el objetivo de recolectar más muestras y realizar una comparación con los resultados obtenidos en FEN

Las variables no relacionadas a videojuegos que consideramos, y que usaremos en los siguientes modelos son: Sexo, dummy que toma el valor 1 cuando el encuestado es hombre, carrera actual, variable categórica con 5 opciones, IC MA, IC,ME, IICG, CA y Externo a FEN, y Notas estadística, una variable de intervalo donde se registra el promedio de los ramos de estadística que cursó el encuestado.

Las variables relacionadas a los videojuegos que consideramos son: Juega online, variable dummy que toma el valor 1 cuando el alumno juega videojuegos multijugador online al menos 1 vez a la semana, horas juego, una variable de intervalos donde se representa la cantidad semanal de horas que dedica a jugar videojuegos, horas stream, una variable similar pero ahora considerando el tiempo que usa en ver contenido de videojuegos, contenido, una variable categórica que representa que tipo de contenido visualiza y que toma 4 valores: Por diversión, Para mejorar en el juego, Ambas y No consumo stream. Y por último, la variable elo máximo, una variable categórica ordinal que representa el elo relativo máximo alcanzado por el alumno. Elo es un sistema estadístico para ver el rango de los jugadores, usado principalmente en el ajedrez, pero se usa para representar la habilidad en los juegos multijugador.

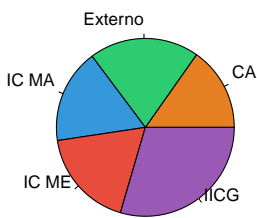
En total son 8 variables, y pudimos recolectar 105 muestras en total. De estas, 84 fueron respondidas por gente de FEN.

Durante esta encuesta realizamos muchos supuestos importantes, que serán detallados en el apartado de observaciones, al final de la parte II

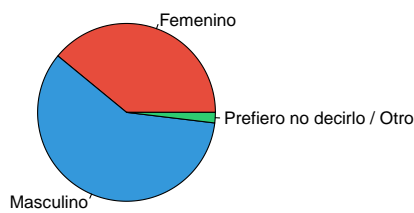
#### 3.2. Datos: Estadísticas descriptivas

La carrera y el sexo de las personas que contestaron la encuesta son los siguientes:

Carrera de los encuestados



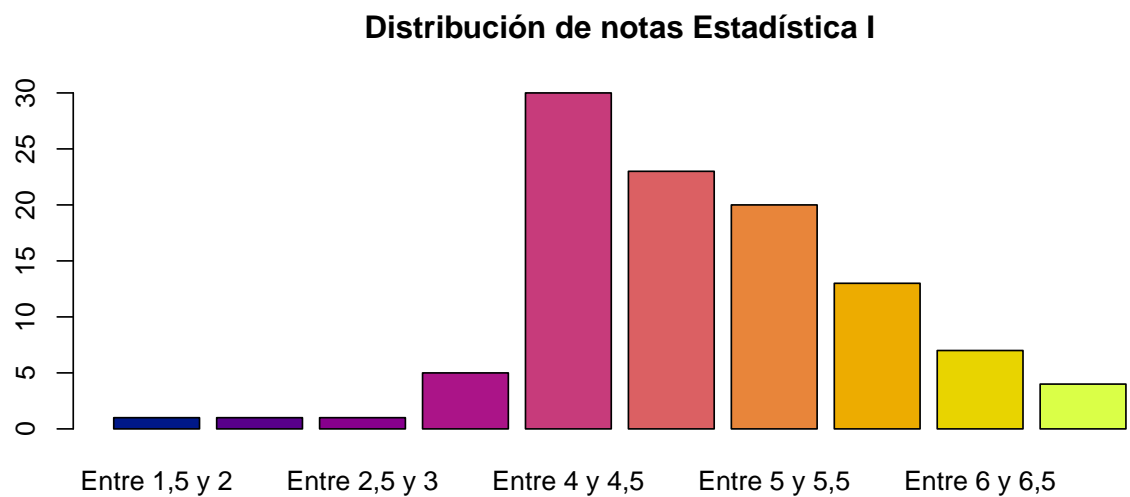
Género de los encuestados



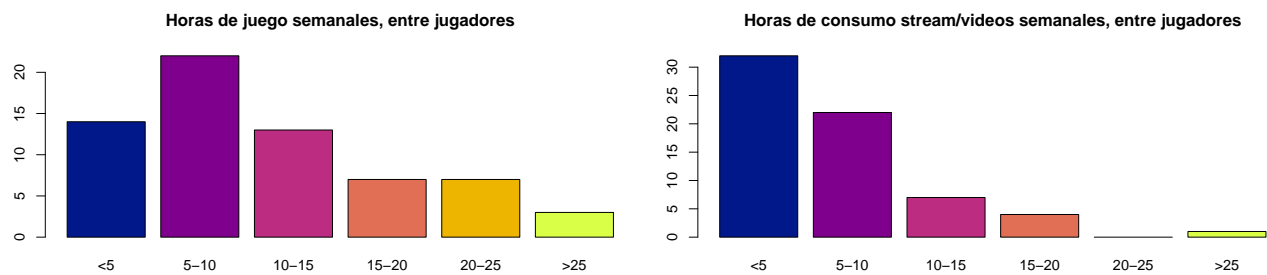
La mayoría de alumnos que contestaron la encuesta son de Ingeniería en Información y Control de Gestión, y el género que más respondió la encuesta es el masculino.



Las notas de estadística siguen esta distribución:

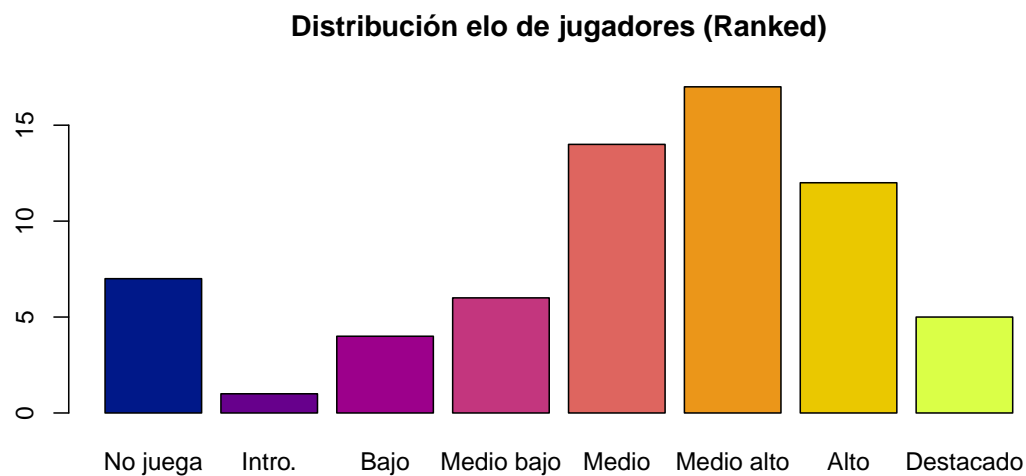


De las 105 personas que respondieron la encuesta, 66 declararon que jugaban videojuegos online frecuentemente. Dentro de las personas que jugaban, obtuvimos las siguientes estadísticas:



Además, dentro de las personas que consumían stream/videos de videojuegos, un 41 % veía contenido por diversión, un 7.5 % para mejorar en el juego (A través de guías, consejos, etc.), y otro 41 % lo hacía por ambas. El 10.5 % declaró no consumir contenido relacionado a los videojuegos.

El elo de los alumnos, en rangos de clasificación (ranked), sigue esta distribución:



Cabe destacar que los rangos están relativizados según la rareza, entre distintos juegos multijugador online.

3.3. Modelos

3.3.1. Modelo base: Regresión de Intervalo Lineal

Para nuestro modelo base consideramos las primeras variables de nuestra encuesta, es decir, las notas de estadística, carrera y género, y si juega o no videojuegos multijugador online 1 o más veces a la semana.

Para este modelo, utilizamos un método conocido como regresión de intervalo, ya que nuestra variable a explicar está en un intervalo. Para eso, usamos la librería survival en R. Dada la distribución de las notas de estadística (Véase pag.8, Estadísticas descriptivas), realizamos el supuesto de que las notas de distribuyen de manera normal (Campana de Gauss). Esto teniendo en cuenta que hay muy pocas notas bajo 4, ya que al reprobar esos cursos y darlos de nuevo no son registrados de manera directa en la encuesta.

Empezamos con una regresión con todas las variables, para ver con que datos estamos tratando:

|              | Estimate | Std. Error | z value | Pr(> z ) |
|--------------|----------|------------|---------|----------|
| (Intercept)  | 4.503    | 0.253      | 17.78   | ¡2e-16   |
| sexo_var     | 0.335    | 0.189      | 1.78    | 0.0757   |
| Externo_FEN  | 0.341    | 0.28       | 1.22    | 0.2239   |
| IC_MA        | 0.276    | 0.290      | 0.95    | 0.3416   |
| IC_ME        | 0.019    | 0.287      | 0.07    | 0.9473   |
| IICG         | 0.164    | 0.263      | 0.62    | 0.5338   |
| juega_online | 0.024    | 0.191      | 0.13    | 0.8998   |

Gaussian distribution. Scale= 0.832, Loglik(model)= -204. Chisq= 6.42 on 6 degrees of freedom, p= 0.38. AIC=424.0889 , BIC=445.3206

Podemos notar que la variable con un p-value más significativo es género, y la menos significativa es tanto juega online como ser de IC Mención economía. Podemos notar, además, que los únicos p-value que pueden ser considerablemente bajos, son los de sexo y Externo a FEN.

Debido a lo anterior, realizamos la siguiente regresión, con solo esas variables:

|             | Estimate | Std. Error | z value | Pr(> z ) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 4.644    | 0.134      | 34.67   | ¡2e-16   |
| sexo_var    | 0.330    | 0.169      | 1.95    | 0.051    |
| Externo_FEN | 0.219    | 0.208      | 1.05    | 0.293    |

Gaussian distribution. Scale= 0.838, Loglik(model)= -204.7 Chisq= 5.14 on 2 degrees of freedom, p= 0.077. AIC=417.375 , BIC=427.9909

Encontramos tanto un AIC como un BIC menor comparado con el primer modelo, pero aún así, definiendo externo como una dummy, el p-value sigue siendo significativo. Por eso, realizamos una regresión solo considerando una variable, el género:

|             | Estimate | Std. Error | z value | Pr(> z ) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 4.680    | 0.130      | 35.92   | ¡2e-16   |
| sexo_var    | 0.344    | 0.170      | 2.03    | 0.042    |

Gaussian distribution. Scale= 0.842, Loglik(model)= -205.2 .Chisq= 4.04 on 1 degree of freedom, p= 0.044. AIC=417.0184 , BIC=440.9041

Como resultado, este último modelo es el que mejor explica la nota de estadística, considerando solo las variables de nuestro modelo base. Esto podemos comprobarlo comparando el AIC de los modelos, además del p-value del test  $\chi^2$  . Nota: Usaremos el Critero de Akaike de aquí en adelante, ya que el Critero Bayesiano considera el intercepto como el mejor modelo para todos los posibles.

### 3.3.2. Modelo extendido: Regresión de Intervalo Lineal

Ahora, consideraremos todas las variables de nuestra encuesta. Realizamos una regresión base para ver con qué datos estamos tratando, pero al haber una gran cantidad de variables categóricas ordenadas (Horas de juego, de stream, y elo), insertar una tabla de grandes dimensiones solo haría que sea más confuso. Por eso, realizamos una selección de variables con el método hacia adelante y hacia atrás, primero, con las variables ordinales como categóricas normales, y luego como categóricas ordenadas.

Seleccionando variables, usando las categóricas ordenadas, obtuvimos lo siguiente:

|             | Estimate | Std. Error | z value | Pr(> z ) |
|-------------|----------|------------|---------|----------|
| (Intercept) | 4.680    | 0.130      | 35.92   | ¡2e-16   |
| sexo_var    | 0.344    | 0.170      | 2.03    | 0.042    |

Gaussian distribution. Scale= 0.842, Loglik(model)= -205.2 .Chisq= 4.04 on 1 degree of freedom, p= 0.044. AIC=417.0184 , BIC=440.9041

Es decir, solo el sexo es una variable significativa en este modelo.

Seleccionando variables, pero ahora con variables categóricas sin orden, obtuvimos que las 2 variables que minimizaban el AIC son, horas de visualización de stream y sexo:

|                            | Estimate | Std. Error | z value | Pr(> z ) |
|----------------------------|----------|------------|---------|----------|
| (Intercept)                | 5.5474   | 0.3429     | 16.18   | ¡2e-16   |
| sexo_var                   | 0.3197   | 0.1809     | 1.77    | 0.0772   |
| horas_stream entre 15 y 20 | -1.2425  | 0.5077     | -2.45   | 0.0144   |
| horas_stream entre 5 y 10  | -0.6889  | 0.3519     | -1.96   | 0.0503   |
| horas_stream más de 25     | -1.1171  | 0.8656     | -1.29   | 0.1968   |
| horas_stream menos de 5    | -1.0742  | 0.3386     | -3.17   | 0.0015   |
| horas_stream no consume    | -0.8701  | 0.3467     | -2.51   | 0.0121   |

Gaussian distribution. Scale=0.796 , Loglik(model)= -199.5. Chisq= 15.49 on 6 degrees of freedom, p= 0.017. AIC=415.0184 , BIC=436.2501

Este modelo tiene un menor AIC y un menor BIC comparado con el modelo de solo género, por lo que se acopla mejor a esta muestra de datos.

### 3.4. Observaciones y conclusión

Nuestro modelo final y extendido, queda como:

$$Es_i = 5,547 + 0,320Hom_i - 1,243E15y30_i - 0,689E5y10_i - 1,117M25_i - 1,074M5_i - 0,870NC_i + \epsilon_i$$

Donde las 6 últimas variables corresponden a variables dummy, que toman el valor 1 cuando las horas de consumo de stream están en ese rango, y  $Es_i$  corresponde a la nota de estadística.

Concluimos, que a esta escala y con estos datos, las horas de consumo de stream puede tener relación con las notas finales de estadística, como ha sido demostrado por estudios que dicen que tienen pequeñas correlaciones con jugar, ver TV y otras actividades de ocio relacionadas (Anand, 2007).

Realizamos muchos supuestos importantes en esta parte, empezando por la normalidad de las notas de estadística, como se mencionó anteriormente. También, realizamos el supuesto importante de que las personas no jugadoras no consumen contenido relacionado a estos videojuegos. Y por último, el supuesto de que las personas mantienen su tiempo de juego, consumo de contenido, y elo relativo durante el tiempo.

## 4. Conclusión final

A través de nuestro trabajo, pudimos encontrar relaciones entre el mundo de la computación y las ciencias exactas, como las matemáticas.

En la primera parte, pudimos comprobar que hay relación entre las notas de programación y las notas de ramos más lógicos, como estadística y matemática, descartando contabilidad que consideramos es un ramo que necesita menos cálculo y lógica. Esto ha sido también demostrado por estudios como el de Trejos Buriticá (2014), donde se establece que una base lógica puede ayudar a entender la programación de una mejor forma, que solo memorizando los comandos.

En la segunda parte, logramos encontrar una pequeña relación entre el consumo de contenido relacionado a videojuegos y las notas finales de estadística. Consideramos, que para la cantidad baja de muestras ( $n=105$ ), pudimos sacar alguna conclusión significativa.

Consideramos que, una conclusión importante que podemos rescatar de nuestra investigación, es la necesidad de impartir buenos cursos de lógica matemática a niveles educacionales introductorios. Es decir, enseñarlo en los colegios y liceos. Estudios han destacado lo mismo, implementando sistemas de introducción a la programación más amigables para niños y jóvenes, en escuelas (Valle y Salgado, 2013).

Con respecto a la influencia del género en las notas finales, en la primera parte pudimos descartar al sexo como variable incidente en programación, como ha sido demostrado por estudios como el de Vázquez y cols. (2007). Sin embargo, en estadística si ha sido una variable influyente. No hemos podido encontrar un estudio que pueda confirmar o desmentir nuestra investigación, pero hay que tomar en cuenta, que en nuestro caso, se realizó una encuesta no probabilística que puede generar cierto sesgo. Nuestros datos decían que la mayoría de las mujeres que contestaron la encuesta no jugaban videojuegos online, por lo que podemos plantear la hipótesis de que algunas alumnas veían el título de la encuesta, y se desanimaban a contestarla. Esto no deja de ser un supuesto, pero por eso no consideraremos los resultados de la parte II como una fuente confiable de información.

## 5. Librerías y códigos

Las librerías usadas, para análisis estadístico, fueron:

xtable gallery: <https://cran.r-project.org/web/packages/xtable/vignettes/xtableGallery.pdf>

olsrr package: <https://cloud.r-project.org/web/packages/olsrr/olsrr.pdf>

survival package: <https://cran.r-project.org/web/packages/survival/survival.pdf>

Los códigos y las bases de datos se pueden encontrar en mi github:

Python, Parte I: <https://github.com/nachomim27/Selenium-Genre-Location-with-RUT>

Códigos y base de datos, Parte I y II: <https://github.com/nachomim27/Trabajo-Cuanti>

## Referencias

- Anand, V. (2007). A study of time management: The correlation between video game usage and academic performance markers. *CyberPsychology & Behavior*, 10(4), 552-559. Descargado de <https://doi.org/10.1089/cpb.2007.9991> (PMID: 17711364) doi: 10.1089/cpb.2007.9991
- Baeza Martín, M. J. (2021). *Brecha educacional se expande por pandemia: mientras en las condes no hubo alumnos afectados, en la pintana la mitad perdió el año*. Descargado 2023-01-20, de <https://www.elmostrador.cl/noticias/pais/2021/02/18/brecha-educacional-se-expande-por-pandemia-mientras-en-las-condes-no-hubo-alumnos-afectados-en-la-pintana-la-mitad-perdio-el-ano/>
- Mora-Rivera, S., Coto-Chotto, M., y Villalobos-Murillo, J. (2017, Jan.). Women's participation in the information systems career at the national university of costa rica and their performance in programming courses. *Revista Electrónica Educare*, 21(1), 1-22. Descargado de <https://www.revistas.una.ac.cr/index.php/EDUCARE/article/view/7576> doi: 10.15359/ree.21-1.12
- Trejos Buriticá, O. I. A. (2014, 09). Relaciones de aprendizaje significativo entre dos paradigmas de programación a partir de dos lenguajes de programación. *Tecnura*, 18, 91 - 102. Descargado de [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0123-921X2014000300008&nrm=iso](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-921X2014000300008&nrm=iso)
- Valle, J. E. M., y Salgado, V. C. (2013, jul.). Pensamiento lógico matemático con scratch en nivel básico. *Revista Vínculos*, 9(1), 87-95. Descargado de <https://revistas.udistrital.edu.co/index.php/vinculos/article/view/4208> doi: 10.14483/2322939X.4208
- Vázquez, R., Angulo, F., y Rodríguez, C. (2007). Las mujeres y el mundo de la computación y la informática. aportaciones de una investigación cualitativa. *Pixel-Bit. Revista de Medios y Educación*(30), 31-40.

Anexo

Parte I, probamos polinomios de varios grados usando las variables de Estadística I y Matemáticas III. Aquí algunos:

Grado 2:

|               | Estimate | Std. Error | t value | Pr(> t ) |
|---------------|----------|------------|---------|----------|
| (Intercept)   | 0.7034   | 1.8968     | 0.37    | 0.7119   |
| notas_estad   | -0.7667  | 0.7418     | -1.03   | 0.3049   |
| notasestad2   | 0.1241   | 0.0869     | 1.43    | 0.1575   |
| notas_mateIII | 1.4665   | 0.7316     | 2.00    | 0.0488   |
| notasmate2    | -0.1040  | 0.0742     | -1.40   | 0.1654   |

Multiple R-squared: 0.3154, Adjusted R-squared: 0.2769. F-statistic: 8.179 on 4 and 71 DF, p-value: 1.753e-05

Grado 3:

|               | Estimate | Std. Error | t value | Pr(> t ) |
|---------------|----------|------------|---------|----------|
| (Intercept)   | 2.3460   | 6.7371     | 0.35    | 0.7287   |
| notas_estad   | -4.2493  | 3.5893     | -1.18   | 0.2405   |
| notasestad2   | 1.0222   | 0.9044     | 1.13    | 0.2623   |
| notasestad3   | -0.0728  | 0.0727     | -1.00   | 0.3200   |
| notas_mateIII | 3.2620   | 3.9478     | 0.83    | 0.4115   |
| notasmate2    | -0.4944  | 0.9217     | -0.54   | 0.5934   |
| notasmate3    | 0.0268   | 0.0675     | 0.40    | 0.6921   |

Multiple R-squared: 0.3275, Adjusted R-squared: 0.269. F-statistic: 5.6 on 6 and 69 DF, p-value: 8.867e-05

Grado 4:

|               | Estimate | Std. Error | t value | Pr(> t ) |
|---------------|----------|------------|---------|----------|
| (Intercept)   | 54.9148  | 24.1798    | 2.27    | 0.0264   |
| notas_estad   | -1.7601  | 15.6375    | -0.11   | 0.9107   |
| notasestad2   | -0.0258  | 6.1730     | -0.00   | 0.9967   |
| notasestad3   | 0.1135   | 1.0405     | 0.11    | 0.9134   |
| notasestad4   | -0.0117  | 0.0634     | -0.18   | 0.8541   |
| notas_mateIII | -53.6369 | 20.3564    | -2.63   | 0.0104   |
| notasmate2    | 19.9836  | 7.2607     | 2.75    | 0.0076   |
| notasmate3    | -3.0674  | 1.0920     | -2.81   | 0.0065   |
| notasmate4    | 0.1679   | 0.0592     | 2.84    | 0.0060   |

Multiple R-squared: 0.4002, Adjusted R-squared: 0.3286. F-statistic: 5.588 on 8 and 67 DF, p-value: 1.994e-05

Parte II, tablas con todas las variables, primero lineales, luego exponenciales ordinales.

|                                    | Estimate | Std. Error | z value | Pr(> z ) |
|------------------------------------|----------|------------|---------|----------|
| (Intercept)                        | 6.8362   | 0.6949     | 9.84    | .2e-16   |
| sexo_var                           | 0.3033   | 0.2021     | 1.50    | 0.13346  |
| carreraExterno                     | 0.3930   | 0.2704     | 1.45    | 0.14607  |
| carreraIC MA                       | 0.4224   | 0.2904     | 1.45    | 0.14583  |
| carreraIC ME                       | 0.0736   | 0.2856     | 0.26    | 0.79663  |
| carreraICG                         | 0.2709   | 0.2721     | 1.00    | 0.31947  |
| horas_juegoEntre 15 y 20           | -0.1047  | 0.4376     | -0.24   | 0.81089  |
| horas_juegoEntre 20 y 25           | -0.4720  | 0.4158     | -1.14   | 0.25623  |
| horas_juegoEntre 5 y 10            | -0.0707  | 0.3079     | -0.23   | 0.81849  |
| horas_juegoMás de 25               | 0.5369   | 0.5424     | 0.99    | 0.32224  |
| horas_juegoMenos de 5              | -0.1765  | 0.3237     | -0.55   | 0.58554  |
| horas_juegoNo juega online         | -2.4048  | 0.6858     | -3.51   | 0.00045  |
| horas_streamEntre 15 y 20          | -1.1203  | 0.5164     | -2.17   | 0.03007  |
| horas_streamEntre 5 y 10           | -0.7635  | 0.3624     | -2.11   | 0.03514  |
| horas_streamMás de 25              | -0.6368  | 0.9090     | -0.70   | 0.48359  |
| horas_streamMenos de 5             | -1.3738  | 0.3411     | -4.03   | 5.6e-05  |
| contenido para mejorar en el juego | -0.5187  | 0.6071     | -0.85   | 0.39293  |
| contenido por diversión            | -0.8006  | 0.3983     | -2.01   | 0.04444  |
| contenido un mix de ambas          | -0.7533  | 0.4287     | -1.76   | 0.07890  |
| elo_maximo Elo alto                | -0.6342  | 0.4508     | -1.41   | 0.15946  |
| elo_maximo Elo bajo                | -0.8491  | 0.5692     | -1.49   | 0.13579  |
| elo_maximo Elo medio               | -0.8730  | 0.4294     | -2.03   | 0.04204  |
| elo_maximo Elo medio alto          | -0.3039  | 0.4329     | -0.70   | 0.48267  |
| elo_maximo Elo medio bajo          | -0.7331  | 0.5339     | -1.37   | 0.16971  |
| elo_maximo Introductorio           | -0.5534  | 0.9413     | -0.59   | 0.55657  |
| elo_maximoNunca he jugado ranked   | -0.7847  | 0.5329     | -1.47   | 0.14088  |

Gaussian distribution. Scale= 0.732, Loglik(model)= -191.1 . Chisq= 32.37 on 27 degrees of freedom, p= 0.22 Number of Newton-Raphson Iterations: 4

|                                   | Estimate | Std. Error | z value | Pr(> z ) |
|-----------------------------------|----------|------------|---------|----------|
| (Intercept)                       | 4.8606   | 0.4618     | 10.53   | 2e-16    |
| sexo_var                          | 0.3033   | 0.2021     | 1.50    | 0.133    |
| carrera externo                   | 0.3930   | 0.2704     | 1.45    | 0.146    |
| carreraIC MA                      | 0.4224   | 0.2904     | 1.45    | 0.146    |
| carreraIC ME                      | 0.0736   | 0.2856     | 0.26    | 0.797    |
| carreraICG                        | 0.2709   | 0.2721     | 1.00    | 0.319    |
| hjuego2.L                         | 5.3405   | 3.3204     | 1.61    | 0.108    |
| hjuego2.Q                         | -4.6094  | 3.2388     | -1.42   | 0.155    |
| hjuego2.C                         | 4.0652   | 2.3431     | 1.73    | 0.083    |
| hjuego2 <sup>4</sup>              | -1.7163  | 1.4078     | -1.22   | 0.223    |
| hjuego2 <sup>5</sup>              | 1.1609   | 0.6623     | 1.75    | 0.080    |
| hstream2.L                        | -2.9794  | 2.3760     | -1.25   | 0.210    |
| hstream2.Q                        | 2.1846   | 2.0055     | 1.09    | 0.276    |
| hstream2.C                        | -0.9423  | 1.0853     | -0.87   | 0.385    |
| hstream2 <sup>4</sup>             | 2.9293   | 1.4608     | 2.01    | 0.045    |
| contenidoPara mejorar en el juego | -0.5187  | 0.6071     | -0.85   | 0.393    |
| contenidoPor diversión            | -0.8006  | 0.3983     | -2.01   | 0.044    |
| contenidoUn mix de ambas          | -0.7533  | 0.4287     | -1.76   | 0.079    |
| elo2.L                            | -1.0089  | 2.4311     | -0.41   | 0.678    |
| elo2.Q                            | 0.8853   | 0.7763     | 1.14    | 0.254    |
| elo2.C                            | 1.1005   | 1.5162     | 0.73    | 0.468    |
| elo2 <sup>4</sup>                 | -1.8688  | 2.8770     | -0.65   | 0.516    |
| elo2 <sup>5</sup>                 | 2.1184   | 3.0032     | 0.71    | 0.481    |
| elo2 <sup>6</sup>                 | -1.1745  | 2.1123     | -0.56   | 0.578    |
| elo2 <sup>7</sup>                 | 0.9379   | 0.9723     | 0.96    | 0.335    |

Gaussian distribution. Scale= 0.732, Loglik(model)= -191.1 . Chisq= 32.37 on 28 degrees of freedom, p= 0.26 Number of Newton-Raphson Iterations: 4