



RAZONAMIENTO ESTADÍSTICO EN EL CONTEXTO COVID-19: UNA PROPUESTA BASADA EN GEOGEBRA

*STATISTICAL REASONING IN THE COVID-19 CONTEXT:
A PROPOSAL WITH GEOGEBRA*

Manuel González-Navarrete
magonzalez@ubiobio.cl
Universidad del Bío-Bío, Concepción, Chile

Iván Maldonado-Carrasco
ivan.maldonado06@inacapmail.cl
INACAP sede Puente Alto, Santiago, Chile

RESUMEN

En este trabajo presentamos una propuesta didáctica de análisis exploratorio de datos asociados a la evolución de la pandemia en las comunas de Chile. Las actividades son formuladas utilizando el software GeoGebra, en su versión clásica 6.0. Se seleccionaron 15 comunas de tres regiones para incluir un análisis univariado de los casos activos comunales, un análisis de regresión para identificar comunas cuyas evoluciones sean similares y análisis multivariado para comparar comunas en grupos. Finalmente, se ejemplifica un método de agrupamiento para clasificar comunas en función de sus casos activos por cada 100 mil habitantes.

PALABRAS CLAVE:

Razonamiento estadístico, Análisis exploratorio de datos, COVID-19, GeoGebra.

ABSTRACT

In this work we introduce a proposal of exploratory data analysis related to the evolution of the COVID-19 pandemic in the communes of Chile. The activities are formulated using the GeoGebra software, in its classic 6.0 version. We selected 15 communes from 3 regions to include a univariate analysis of active communal cases, a regression analysis to identify communes whose evolutions are similar, and multivariate analysis to compare communes in groups. Finally, an algorithm for community detection is used to classify communes based on their active cases per 100 thousand inhabitants.

KEYWORDS:

Statistical reasoning; Exploratory data analysis; COVID-19; GeoGebra.

1. Introducción

En el marco conceptual de la didáctica de la estadística, se destaca la jerarquía cognitiva de alfabetización estadística, razonamiento y pensamiento estadístico (Garfield, 2002). De acuerdo con Ben-Zvi y Garfield (2004), la alfabetización estadística incluye habilidades básicas de comprensión de la información estadística, destacando la capacidad de organizar datos, construir y presentar tablas y trabajar con diferentes representaciones de datos, agregando que el razonamiento estadístico se entiende como aquello que hacen las personas al razonar con ideas estadísticas y al dar sentido a la información estadística. El último nivel de pensamiento estadístico es más avanzado, puesto que implica una comprensión de por qué y cómo se realizan las investigaciones estadísticas. Esto incluye reconocer y comprender el proceso de investigación completo (Chance, 2002).

En esta misma línea, cobra relevancia el análisis exploratorio de datos (Tukey, 1977), entendido como la forma de resumir las características principales de los datos recolectados, utilizando a menudo gráficos estadísticos y otros métodos de visualización de datos. Así también se destaca el proceso de inferencia estadística informal, siendo definido como “la forma en que los estudiantes usan su conocimiento estadístico informal para crear argumentos que sustenten inferencias basados en los datos” (Zieffler et al., 2008, p. 44). La literatura en este sentido es extensa (ver, por ejemplo, a Arnold, 2008; Dvir y Ben-Zvi, 2021; Makar et al., 2011; Makar y Rubin, 2009, 2018; Pfannkuch, 2006; Rossman, 2008 y Wild et al., 2011).

De forma complementaria, la estadística ha jugado un rol esencial en el contexto de la actual pandemia de COVID-19. Diariamente se generan datos asociados a la evolución del virus en nuestro país, los cuales han sido dispuestos en una plataforma que incluye diversas variables a nivel comunal asociadas a la crisis sanitaria (Ministerio de Ciencia, 2021). El análisis de los datos disponibles es un desafío para la ciencia y una enorme oportunidad para la comunidad educativa en la enseñanza de la estadística, cobrando vital relevancia las palabras de Estrella (2017):

Nuestra actual sociedad requiere que los ciudadanos sean competentes en evaluar críticamente afirmaciones basadas en datos y en argumentar con fundamentos en la evidencia que entregan los datos. El estudio de la estadística provee a los estudiantes de herramientas, ideas y disposiciones para reaccionar inteligentemente a la ingente información del mundo que les rodea. (p. 173)

De esta forma, se busca evidenciar la importancia del estudio de datos a nivel comunal de la evolución de la pandemia, permitiendo el desarrollo del razonamiento estadístico de forma contextualizada. De hecho, recientemente se definieron lineamientos y orientaciones específicas para la enseñanza y aprendizaje en tiempos de crisis (Propuestas

Educación Mesa Social Covid-19, 2020), realizando una lectura de la priorización curricular elaborada el pasado año por la Unidad de Currículum y Evaluación del Ministerio de Educación (UCE MINEDUC). En particular para tercero medio, en el eje temático estadística y probabilidades, se sugiere una actividad basada en datos de coronavirus, argumentando que:

En el caso de Chile, se puede encontrar datos segmentados por región y comuna. Estos datos favorecen que los estudiantes puedan formular sus propias preguntas de interés y puedan analizar el comportamiento de los datos mediante los gráficos disponibles en distintas escalas. (Propuestas Educación Mesa Social Covid-19, 2020, p. 304)

En este sentido, el currículum nacional chileno plantea cuatro unidades de trabajo en los niveles de tercero y cuarto medio en su plan electivo, estos se muestran en la Tabla 1. De esta forma, la presente propuesta se relaciona a las unidades 1 y 2.

Tabla 1. Unidades y objetivos de aprendizaje (OA) del currículum nacional. Nota. MINEDUC (2019, p. 151).

Unidades	OA
¿Qué dicen los gráficos? Análisis crítico de la información	Argumentar y comunicar decisiones a partir del análisis crítico de información presente en histogramas, polígonos de frecuencia, frecuencia acumulada, diagramas de cajón y nube de puntos, incluyendo el uso de herramientas digitales.
Media muestral, dispersión y correlación	Resolver problemas que involucren los conceptos de media muestral, desviación estándar, varianza, coeficiente de variación y correlación muestral entre dos variables, tanto de forma manuscrita como haciendo uso de herramientas tecnológicas digitales.
Situaciones o fenómenos que se modelan por medio de las distribuciones binomial y normal	Modelar fenómenos o situaciones cotidianas del ámbito científico y del ámbito social, que requieran el cálculo de probabilidades y la aplicación de las distribuciones binomial y normal.
Inferencia estadística	Argumentar inferencias acerca de parámetros (media y varianza) o características de una población, a partir de datos de una muestra aleatoria, bajo el supuesto de normalidad y aplicando procedimientos con base en intervalos de confianza o pruebas de hipótesis.

Por otra parte, la utilidad de las tecnologías y específicamente del software GeoGebra en la enseñanza de la estadística ha sido expuesta en diversos artículos (por ejemplo, Chance et al., 2007; Del Pino, 2013; García y Cuadros, 2013; Guncaga et al., 2019; Mellado y Marín, 2010; Prodromou, 2014; Zenteno Ruiz et al., 2020). Se suma a esto, el hecho de que GeoGebra en todas sus versiones es un software libre de código abierto y multiplataforma que está diseñado para todo nivel educativo escolar, y a su vez para educación superior tradicional como técnico-profesional. GeoGebra reúne dinámicamente geometría, álgebra, estadística y cálculo en registros gráficos, de análisis y de organización en hojas de cálculo (<https://www.geogebra.org/>). Además, existe una comunidad en la cual los usuarios de GeoGebra pueden interactuar a través de ella.

Por dichos motivos, el presente trabajo se basa en los datos de la evolución del virus para las comunas del país, resumidos en datos de casos activos disponibles en la plataforma covid19.ubiobio.cl. Particularmente, objetivamos exponer una propuesta didáctica utilizando las herramientas de GeoGebra clásico en su versión 6.0. El propósito es contribuir a la alfabetización y razonamiento estadístico basados en el análisis exploratorio de datos y el desarrollo de inferencia informal por parte de los estudiantes. La intención es realizar un estudio comparativo entre las comunas seleccionadas, por lo que sus medidas de resumen serán confrontadas, buscando similitudes y semejanzas en la evolución de la pandemia para las distintas comunas.

Para ello, en la Sección 2 seleccionamos cinco comunas de cada una de las siguientes regiones de Antofagasta (RA), Metropolitana (RM) y Los Ríos (RLR). Las actividades en la Sección 3 serán separadas

en tres grupos: (a) análisis univariado, mediante medidas de tendencia central y dispersión, (b) análisis bivariado, mediante modelos de regresión lineal, y (c) análisis multivariado, con el objeto de comparar comunas dentro de la región o por periodos de tiempo específicos. Finalmente, basados en el análisis de regresión, proponemos una forma de agrupamiento utilizando el coeficiente de determinación entre los pares de comunas.

2. Recolección de datos comunales sobre la evolución del virus

Nuestra propuesta consiste en seleccionar un grupo de comunas para analizar la evolución de casos activos en cada una de ellas. Particularmente, utilizamos la serie de datos iniciada en abril de 2020 y terminando en el mismo mes de 2021. Los datos utilizados en esta propuesta pueden ser consultados en la plataforma de comparación comunal y regional (Universidad del Bío-Bío, 2021). La plataforma dispone de datos relacionados a casos activos y fallecidos a nivel comunal, además de las evoluciones de casos nuevos y fallecidos a nivel regional, permitiendo también descargar los datos seleccionados en el visor. Dado que buscamos comparar comunas, es importante conocer el tamaño de su población, puesto que influirá en la cantidad de casos que puedan reportarse diariamente, esto es, para una comuna más habitada se esperan más casos nuevos. En otras palabras, bajo las mismas condiciones de transmisión del virus, un grupo mayor de habitantes terminará con más personas contagiadas.

Las respectivas poblaciones de las comunas seleccionadas se resumen en la Tabla 2 y Figura 1, la cual incluye la media y mediana del número de habitantes para dichas comunas.

Tabla 2. Número de habitantes para las 15 comunas seleccionadas
Nota. Datos INE, proyección 2021 (Biblioteca del Congreso Nacional, 2021).

RA	Antofagasta	Mejillones	San Pedro de Atacama	Sierra Gorda	Tocopilla
Habitantes	425.725	14.776	10.434	1746	28.079
RM	La Pintana	Maipú	Puente Alto	Santiago	Vitacura
Habitantes	189.335	578.605	645.909	503.147	96.774
RLR	Futrono	Lanco	Los Lagos	Panguipulli	Valdivia
Habitantes	15.261	17.652	20.518	35.991	176.774

Nótese que la media está en torno a los 180 mil habitantes. Sin embargo, la mediana es cercana a los 36 mil habitantes, lo que corresponde específicamente a la población de Panguipulli. Otra observación que se desprende de esto es que la mayoría de las comunas en estudio poseen menos de 40 mil habitantes. Además, todas las comunas de la RM están por sobre la mediana de habitantes, siendo Puente Alto la que tiene la mayor población. Por otro lado, todas las

comunas de la RLR están por debajo de la media de habitantes.

En este sentido, a la hora de comparar la evolución en diferentes comunas, es importante definir un indicador ponderado por la población respectiva. De esta forma, los datos utilizados serán los casos activos por cada 100 mil habitantes.

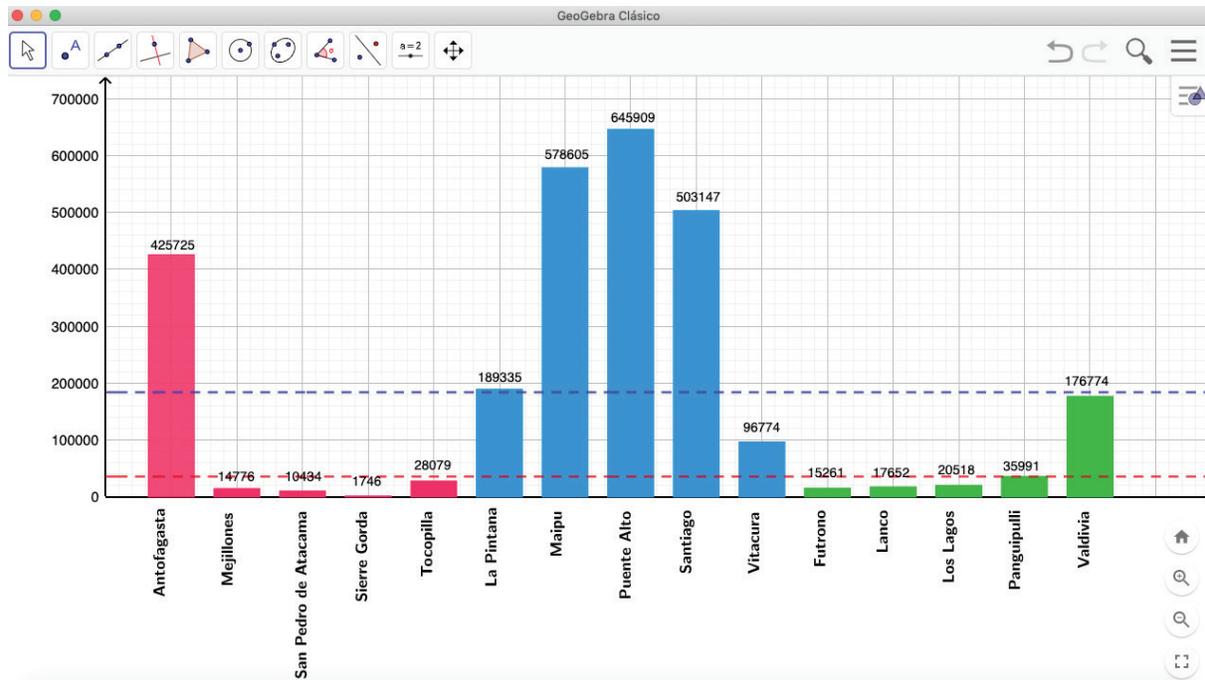


Figura 1. Gráfico de barras del número de habitantes para las 15 comunas seleccionadas

Nota. Elaboración propia. Las líneas horizontales representan la media (azul) y mediana (roja) del número de habitantes para las 15 comunas.

2.1 Visualización de datos históricos en covid19.ubiobio.cl

La plataforma utilizada es de fácil manipulación y sus diversas herramientas favorecen la visualización y adquisición de los datos, los cuales son constantemente actualizados desde la base de datos del Ministerio de Ciencia.

En la Figura 2 vemos la pantalla inicial de la plataforma, en ella se pueden seleccionar comunas escribiendo sus nombres, o bien, escoger una región para comparar todas sus comunas. Las gráficas generadas pueden ser visualizadas por periodos específicos a criterio del usuario. Al descargar en formato CSV, el usuario dispone de los datos de población, casos activos y casos activos por cada 100 mil habitantes para las comunas seleccionadas.



Figura 2. Pantalla de inicio de la plataforma de datos comunales de la Universidad del Bio-Bio.

Nota. Obtenido de covid19.ubiobio.cl.

Utilizando la plataforma, en la Figura 3 se presentan las cinco comunas de la RM seleccionadas para este estudio: La Pintana, Maipú, Puente Alto, Santiago y Vitacura. Es posible notar que estas tienen una evolución similar para el indicador seleccionado. Se observa un momento crítico (*peak*) en torno a los meses de mayo y junio de 2020, con épocas de primavera y verano con valores bajos.

Se evidencia que La Pintana ha presentado los índices más altos y actualmente (hasta el día 6 de mayo) junto a Puente Alto presentan la situación más delicada de estas cinco comunas. En cambio, Vitacura se ha mostrado con una situación más controlada.

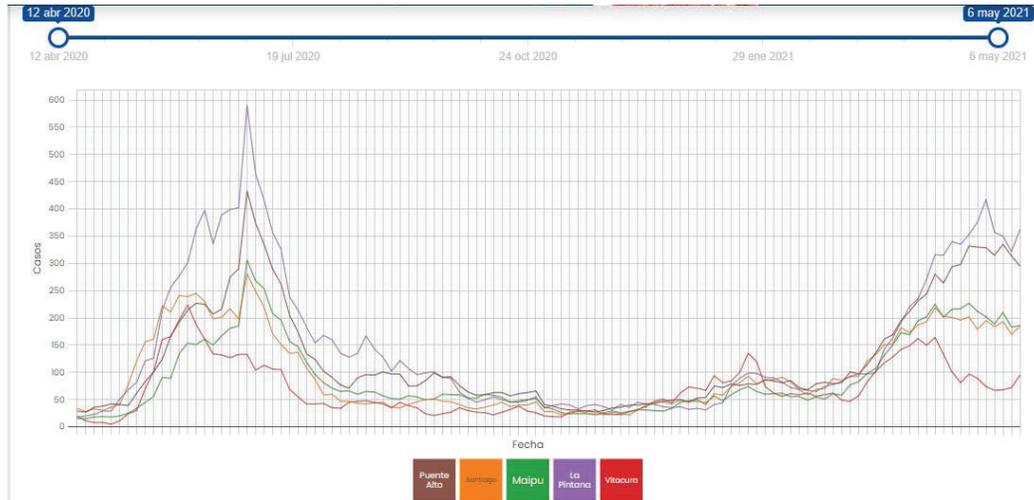


Figura 3. Evolución de casos activos por cada 100 mil habitantes para las comunas de la RM

Nota. Generado por covid19.ubiobio.cl.

En la Figura 4, al considerar las comunas de la RLR, Futrono, Lanco, Los Lagos, Panguipulli y Valdivia, se observa un momento crítico para la estación de verano 2021, antecedido por una primavera con clara tendencia al alza. Mencionamos que la comuna de Futrono presenta el valor histórico más alto de las

cinco comunas, el que fue alcanzado los primeros días del mes de marzo. Se destaca también la comuna de Los Lagos con una tendencia al alza constante durante los últimos meses, finalizando una larga cuarentena en el mes de abril de 2021, pero siendo actualmente la comuna con los peores valores.

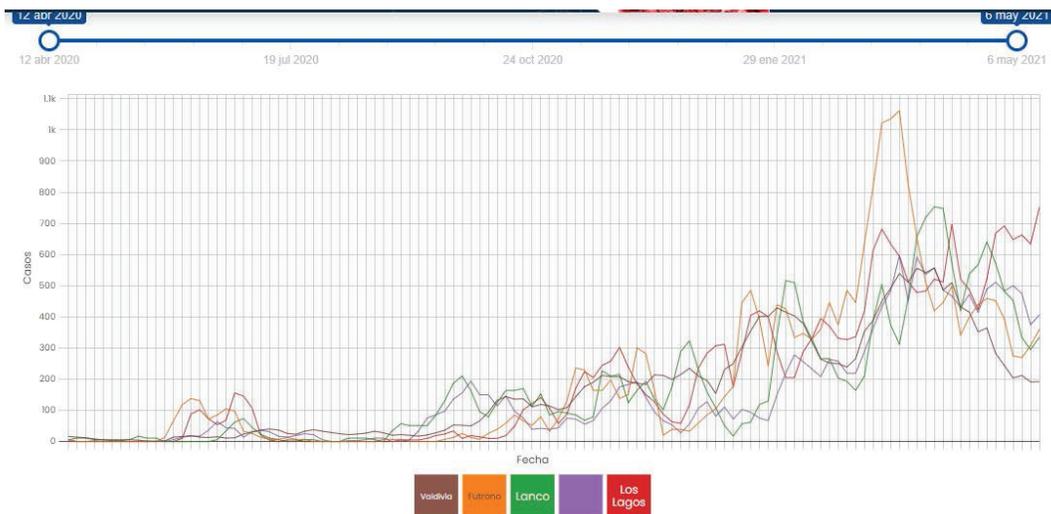


Figura 4. Evolución de casos activos por cada 100 mil habitantes para las comunas de la RLR

Nota. Generado por covid19.ubiobio.cl.

Finalmente, al observar las evoluciones para las comunas de Antofagasta, Mejillones, San Pedro de Atacama, Sierra Gorda y Tocopilla, podemos encontrar diferencias notorias. Particularmente, la comuna de Sierra Gorda presenta un comportamiento disonante con las otras comunas, esto puede deberse al tamaño de su población (ver Tabla 2). En el caso de las cuatro

restantes, tuvieron un periodo crítico en los meses de enero y febrero de 2021.

El caso de Mejillones representa el *peak* más alto de todas las comunas, en torno a los 900 casos activos por cada 100 mil habitantes a comienzos del mes de febrero de 2021.

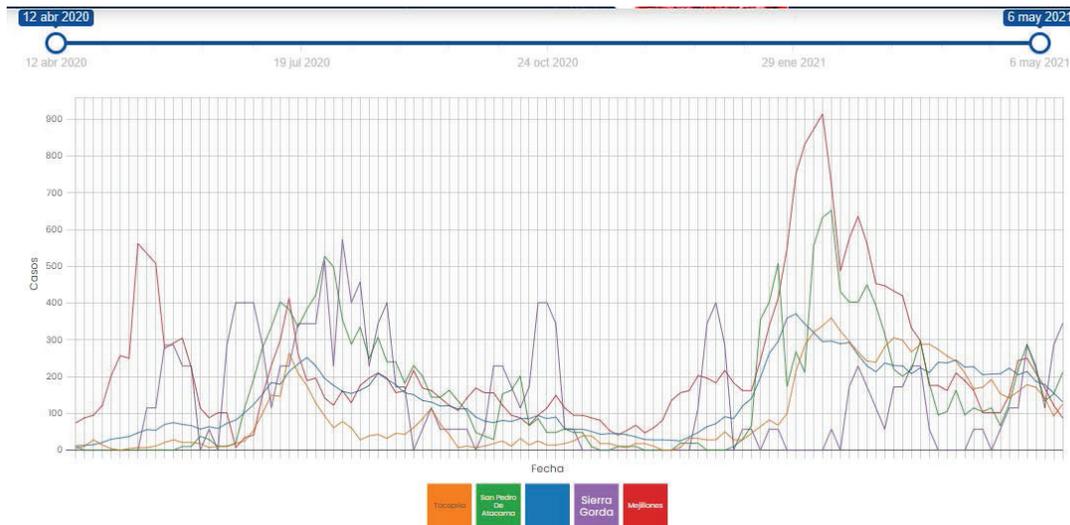


Figura 5. Evolución de casos activos por cada 100 mil habitantes para las comunas de la RA

Nota. Generado por covid19.ubiobio.cl.

3. Análisis exploratorio con GeoGebra

Una vez seleccionadas las 15 comunas, procedemos a descargar los datos en formato CSV. En este punto, se debe confeccionar una planilla simplificada para procesar en GeoGebra, dicha tarea puede ser desarrollada en Excel.

Las herramientas de GeoGebra que serán utilizadas son: (1) Análisis de una variable, (2) Análisis de regresión de dos variables y (3) Análisis multivariable (ver Figuras 21, 26 y 29, respectivamente). Los detalles de la manipulación requerida para las actividades pueden ser consultados en el Apéndice. En esta sección nos concentraremos en exponer el tipo de resultados y conclusiones a los que es posible llegar, indicando por tanto las posibilidades de desarrollar alfabetización y razonamiento estadístico.

3.1 Caracterización de variables unidimensionales

A modo de ejemplo, consideraremos las comunas que visualmente presentan *peak* más altos en sus respectivas regiones, esto es Mejillones (RA), La Pintana (RM) y Futrono (RLR). La serie de datos utilizada corresponde al periodo del 13-04-20 al 23-04-21.

En primer lugar, visualizamos el histograma, y a continuación el diagrama de cajas con su gráfico Q-Q para la distribución normal. Los detalles de cada construcción y sus propiedades pueden ser consultados en Gibbons y Chakraborti (2003), Triola (2009) y Tukey (1977).

Con el histograma de la Figura 6 podemos evidenciar que la mayor parte del tiempo la comuna de Mejillones ha tenido índices de casos activos por cada 100 mil habitantes en el intervalo de 150 a 180, esto se puede visualizar también en la tabla de frecuencias. Observando la distribución de esta comuna, vemos asimetría hacia el lado izquierdo (positiva). Desde las estadísticas se observa que la media (236.13) es mayor a la mediana (162), en esta misma línea es posible identificar las medidas de posición: cuartil 1 (108), cuartil 2 (162) y cuartil 3 (291).

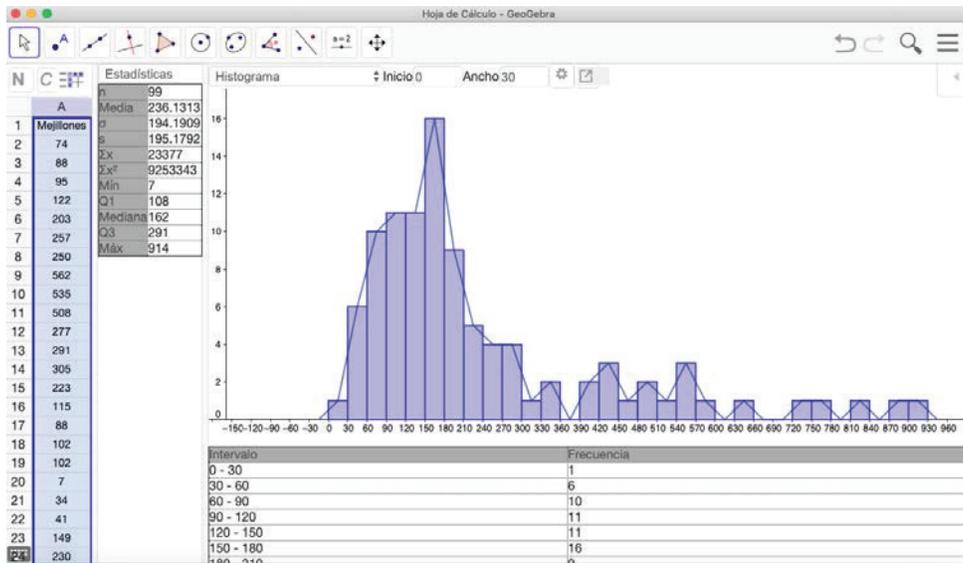


Figura 6. Histograma para los casos activos por cada 100 mil habitantes en la comuna de Mejillones
 Nota. Elaboración propia.

Las estadísticas mencionadas pueden ser visualizadas en el gráfico de cajas (*boxplot*) de la Figura 7, observando la existencia de datos atípicos (*outliers*). Bajo el *boxplot* observamos el gráfico Q-Q (cuantil-cuantil), el cual nos compara dos distribuciones: una empírica (comuna de Mejillones) versus la distribución

teórica (normal o de Gauss); en este caso los puntos están representando los datos para la comuna de Mejillones y la recta indica normalidad. Nótese que el histograma y el gráfico Q-Q señalan que la distribución empírica no aparenta ser una distribución normal.

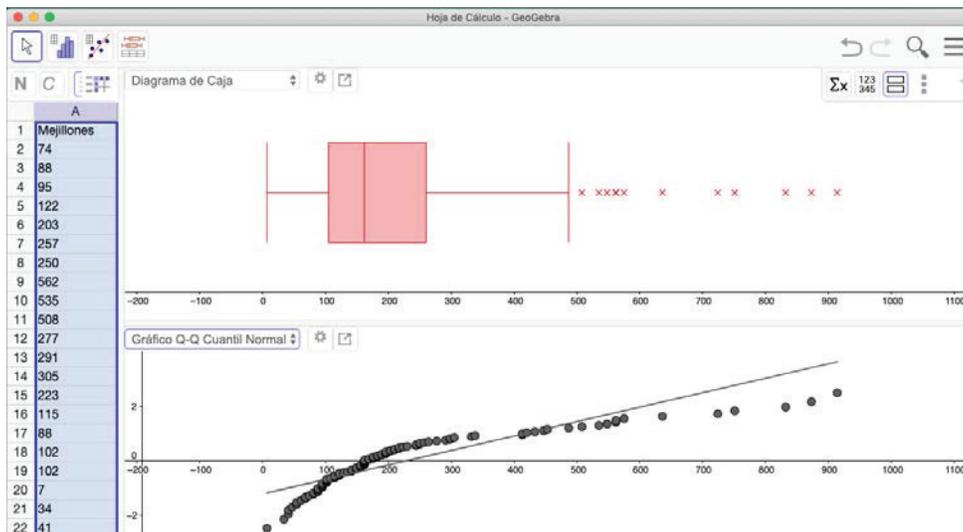


Figura 7. Diagrama de caja (*boxplot*) y gráfico Q-Q para los casos activos por cada 100 mil habitantes en la comuna de Mejillones
 Nota. Elaboración propia.

A continuación, visualizamos el histograma, diagrama de cajas y gráfico Q-Q para la comuna de La Pintana. Considerando el histograma de la Figura 8, es posible observar que la mayor parte del tiempo la comuna de La Pintana ha tenido índices de casos activos en el intervalo de 30 a 60, esta distribución presenta

cierta asimetría positiva. Al mismo tiempo, desde el resumen de estadísticas se puede distinguir que la media (142.19) es mayor a la mediana (92), es posible identificar por consiguiente las medidas de posición: cuartil 1 (45), cuartil 2 (92) y cuartil 3 (203).

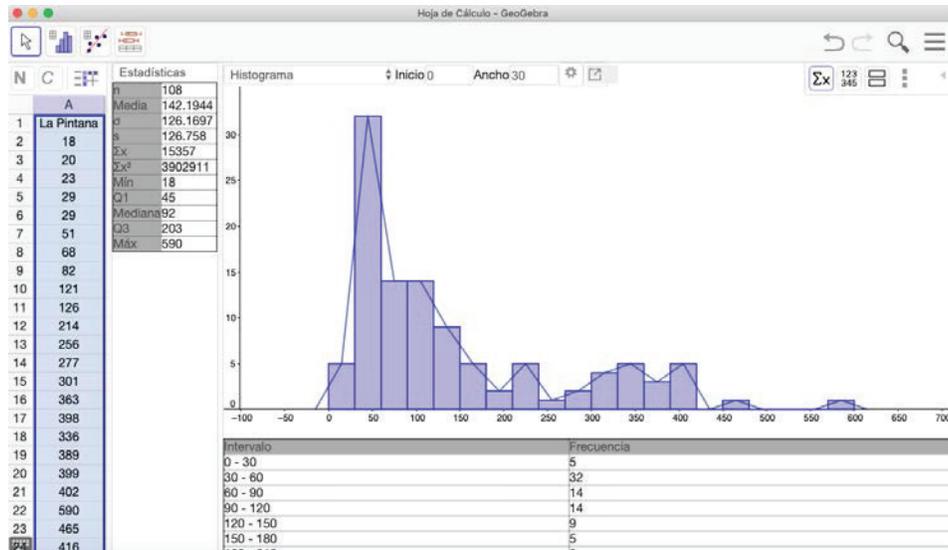


Figura 9. Diagrama de caja (boxplot) y gráfico Q-Q para los casos activos por cada 100 mil habitantes en la comuna de La Pintana
Nota. Elaboración propia.

A continuación, visualizamos el histograma, diagrama de cajas y gráfico Q-Q para Futrono. En la Figura 10 se evidencia que la mayor parte del tiempo Futrono ha pasado con bajos índices de casos activos, esto

debido a que el intervalo de 0 a 30 se presenta con mayor frecuencia. Además, la media (188.57) es mayor a la mediana (75.5) y las medidas de posición: cuartil 1 (7), cuartil 2 (75.5) y cuartil 3 (344).

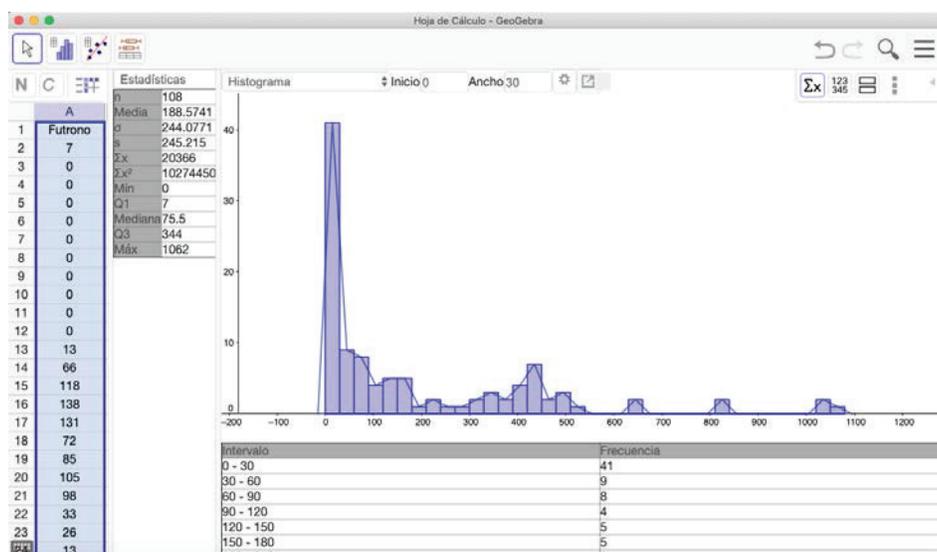


Figura 10. Histograma para los casos activos por cada 100 mil habitantes en la comuna de Futrono
Nota. Elaboración propia.

Observando el gráfico de cajas (*boxplot*) de la comuna de Futrono, visualizamos la existencia de tres datos atípicos (*outliers*). Bajo el *boxplot* observamos el

gráfico Q-Q (cuantil-cuantil), con lo que se concluye con mayor seguridad que los datos no aparentan una distribución normal.

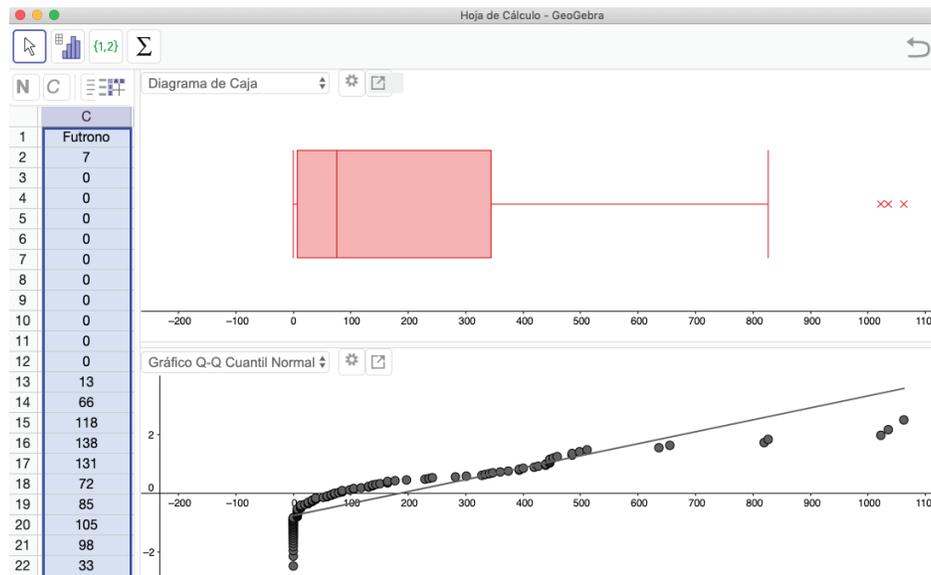


Figura 11. Diagrama de caja (*boxplot*) y gráfico Q-Q para los casos activos por cada 100 mil habitantes en la comuna de Futrono

Nota. Elaboración propia.

3.2 Análisis de regresión

Considerando la RM, realizaremos un análisis de regresión para identificar qué tan similares (o disímiles) han sido las evoluciones del indicador en estudio para las cinco comunas seleccionadas.

Comenzando con Puente Alto (eje X) versus La Pintana (eje Y), en la Figura 12 vemos el resumen entregado por GeoGebra, en el que se presenta la ecuación de la recta que describe la regresión lineal, esto es:

$$y=1.3146x-14.4293.$$

Se entiende que la recta tiene pendiente positiva, lo que indica una correlación directa entre las comunas. En el caso del intercepto, este es cercano a -14. Además, el programa nos entrega el valor de R^2 (0.9309), llamado coeficiente de determinación, el cual, mientras más cercano a 1 sea, indica mayor poder predictivo para la recta de regresión. En GeoGebra es posible obtener otros tipos de regresión (ver Figura 28), pero nos enfocamos en el caso lineal para comenzar. El coeficiente de determinación será utilizado en la Sección 3.3.1, para construir un indicador de dependencia geográfica de la evolución del virus.

En este sentido, es posible utilizar la técnica de regresión lineal para analizar tendencias en los datos.

De esta forma, basta cruzar los datos de una comuna con una secuencia de fechas o secuencia numérica consecutiva, que represente los días del periodo analizado. La idea es identificar una recta de tendencia que podría indicar un alza, estabilidad o disminución de los casos activos en el periodo analizado. Este tipo de estudios es de utilidad si se quiere analizar la época presente, por las últimas dos semanas, a modo de proyectar la evolución para los próximos días.

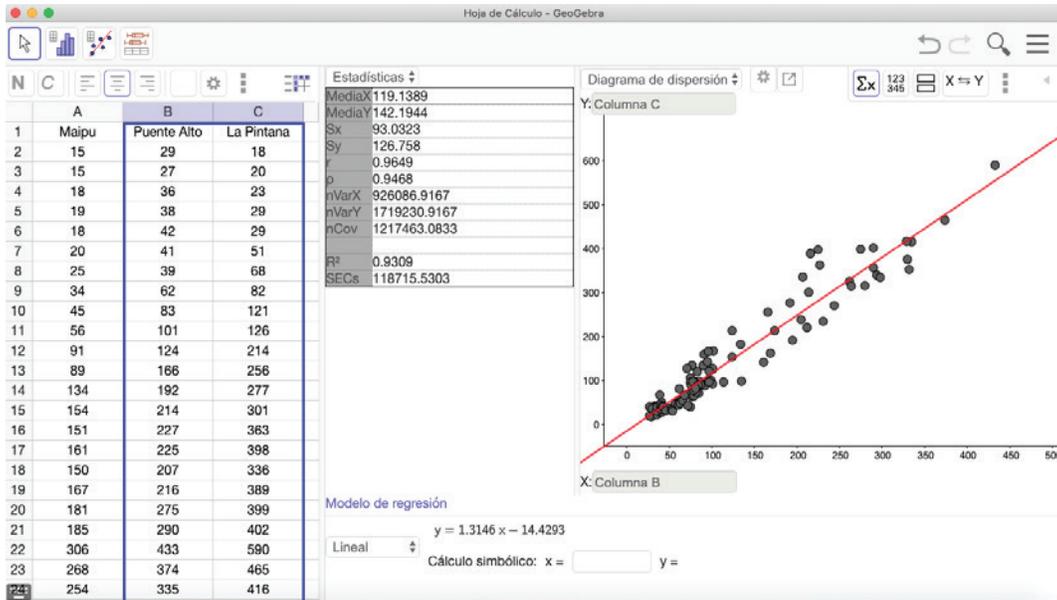


Figura 12. Análisis de regresión para los casos activos por cada 100 mil habitantes en las comunas de Pueblo Alto y La Pintana
 Nota. Elaboración propia.

Replicando este análisis con todos los pares de comunas de la RM, obtenemos un resumen de los valores de R^2 en la Tabla 3. Se observa que Vitacura es la comuna con un comportamiento diferente a

las otras cuatro comunas. Las comunas con mayor índice de dependencia son Pueblo Alto y La Pintana. Finalmente, los gráficos de dispersión y la recta regresora están incluidos en la Figura 13.

Tabla 3. Valores de R^2 obtenido en GeoGebra mediante análisis de regresión lineal
 Nota. Elaboración propia.

	La Pintana	Maipú	Pueblo Alto	Santiago	Vitacura
La Pintana	1	0.8794	0.9683	0.7834	0.5056
Maipú	0.8794	1	0.9309	0.8094	0.4954
Pueblo Alto	0.9683	0.9309	1	0.7976	0.4776
Santiago	0.7834	0.8094	0.7976	1	0.7547
Vitacura	0.5056	0.4954	0.4776	0.7547	1

En la Figura 14, nótese que la comuna de Futrono presenta una caja más grande y valores atípicos que se escapan bastante en comparación a las otras comunas. Por otro lado, todas las comunas presentan un primer cuartil cercano a cero.

De forma análoga generamos los diagramas de caja para los casos activos por cada 100 mil habitantes en las comunas de las RM y RA, los detalles están en la Figura 15, cuya identificación se encuentra en la Tabla 4.

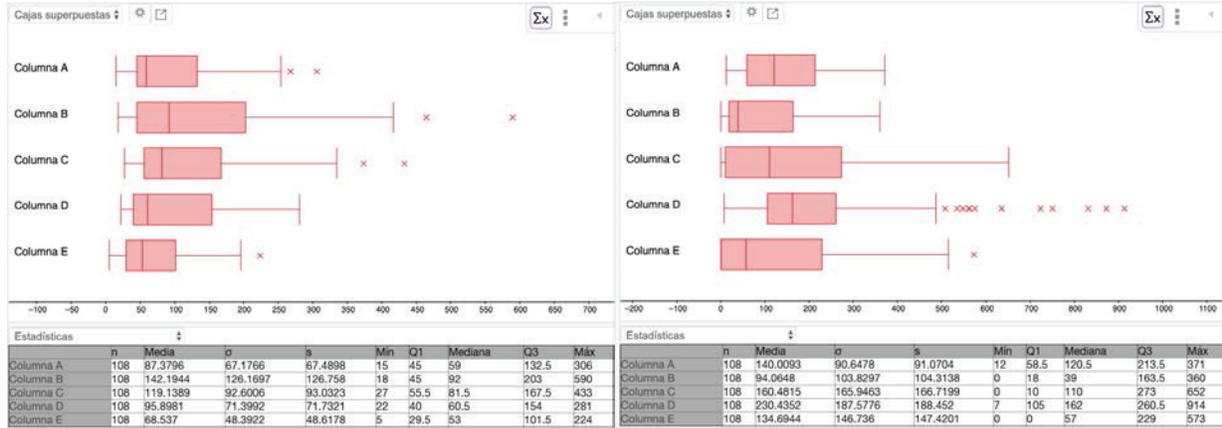


Figura 15. Boxplot para las comunas de la RM (izquierda) y la RA (derecha)

Nota. Elaboración propia.

Tabla 4. Codificación usada en la Figura 15, comunas de la RM y RA. Nota. Elaboración propia.

Columna	Comunas RM	Comunas RA
Columna A	Maipú	Antofagasta
Columna B	La Pintana	Tocopilla
Columna C	Puente Alto	San Pedro de Atacama
Columna D	Santiago	Mejillones
Columna E	Vitacura	Sierra Gorda

Siguiendo con el análisis multivariado, es interesante estudiar en comparativa las tres comunas analizadas en la Sección 3.1. Particularmente, consideraremos la estación de verano, desde el 21-12-20 al 22-03-21, correspondientes a 27 registros. Mostramos un resumen de estadísticas en la Tabla 5 y los respectivos boxplot en la Figura 16.

Al comparar media, mediana y cuartiles, es posible observar que la comuna de La Pintana durante

el periodo de verano fue menos afectada que las comunas de Mejillones y Futrono. En esta línea, es posible calcular el coeficiente de variación de estas comunas en el periodo de verano. Con los resultados de la Tabla 5, es posible observar que La Pintana presenta una mayor homogeneidad de los datos con respecto a las otras dos comunas y la comuna con mayor heterogeneidad sería Futrono.

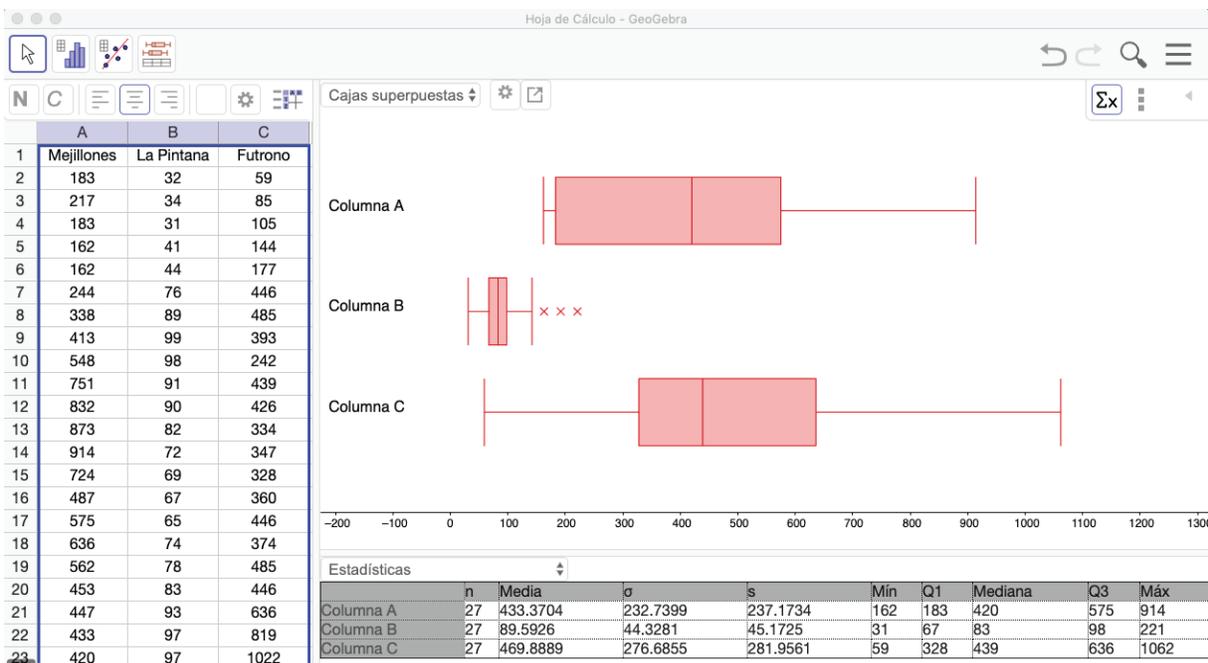


Figura 16. Boxplot para los casos activos por cada 100 mil habitantes en las comunas de Mejillones, La Pintana y Futrono, para la temporada de verano
 Nota. Elaboración propia.

Tabla 5. Coeficiente de variación en periodo de verano de las comunas de Mejillones, La Pintana y Futrono para los casos activos por cada 100 mil habitantes. Nota. Elaboración propia.

	Desviación estándar	Media	Coefficiente de variación
Mejillones	237.17	433.37	0.547
La Pintana	45.17	89.59	0.504
Futrono	281.96	469.89	0.601

3.3.1 Agrupamiento basado en el coeficiente de determinación de la regresión lineal

En esta sección se incluye un análisis de agrupamiento con respecto a todas las comunas seleccionadas. Esto será basado en el coeficiente de determinación R^2 obtenido del análisis de regresión, como fue explicado en la Sección 3.2.

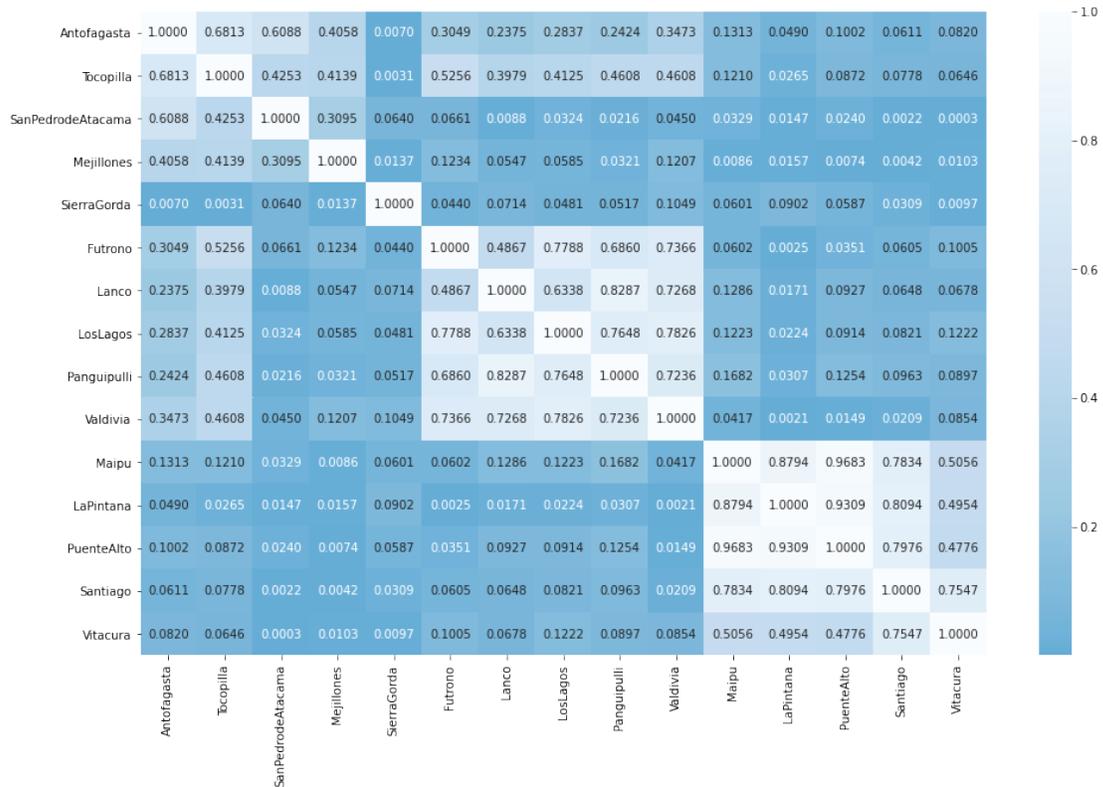


Figura 17. Resumen de los valores de R^2 asociados a los análisis de regresión
 Nota. Elaboración propia (usando Python).

Las comunas de la RM presentan indicadores de dependencia bastante altos y se podrían agrupar en un clúster, así como también se podrían agrupar las comunas de la RLR. Sin embargo, las comunas de la RA presentan menores índices de dependencia; incluso, la comuna de Sierra Gorda parece no estar significativamente correlacionada con las demás comunas de su región. Como hemos mencionado, la población de dicha comuna influye en su evolución y en su forma diferenciada de las demás.

De esta forma, un par de comunas con coeficiente más alto tendrán más probabilidad de ser identificadas en un mismo clúster. Esto ocurre con las comunas de la RM, por ejemplo. De hecho, al aplicarlo en busca de clústeres para las 15 comunas, se obtienen los cuatro clústeres identificados en la Tabla 6.

Mencionamos una posible herramienta para obtener una clasificación en clúster, el llamado algoritmo de Louvain, propuesto por Blondel et al. (2008). Para utilizar el método, se necesita una ponderación entre elementos (comunas), que podría ser considerada con el coeficiente de determinación de la regresión lineal.

Tabla 6. Clasificación de las 15 comunas mediante el algoritmo de Louvain

Nota. Elaboración propia, usando el software R.

RA	Antofagasta	Mejillones	San Pedro de Atacama	Tocopilla	Sierra Gorda
Clúster	1	1	1	1	2
RM	La Pintana	Maipú	Puente Alto	Santiago	Vitacura
Clúster	3	3	3	3	3
RLR	Futroneo	Lanco	Los Lagos	Panguipulli	Valdivia
Clúster	4	4	4	4	4

4. Comentarios finales

En el presente artículo realizamos una propuesta de análisis exploratorio de las series de evolución de casos activos para 15 comunas de Chile, entre abril de 2020 y abril de 2021. En este sentido, los análisis pueden ser realizados considerando periodos diferentes de tiempo, como el efectuado para la estación de verano en la Figura 16.

Para el caso de la RM, es importante complementar los análisis con la discusión del artículo de Mena et al. (2021), el cual caracteriza la influencia de la variable socioeconómica en la tasa de mortalidad para las comunas de dicha región.

Otras actividades pueden ser basadas en los datos regionales disponibles en la plataforma covid19.ubiobio.cl. En este caso, es posible realizar comparaciones entre sintomáticos y asintomáticos. Además, dichos datos poseen una frecuencia diaria, por lo que se pueden aplicar técnicas de suavizamiento, como las medias móviles. Sumado a esto, en el repositorio del Ministerio de Ciencia se encuentran disponibles los datos de positividad y el avance del proceso de vacunación para las regiones y comunas del país.

Complementado a lo que hemos indicado, esta propuesta es aplicable en tercero y cuarto medio de Educación Media en el contexto chileno, considerando los planes electivos, tanto por la pertinencia de las herramientas GeoGebra y Excel, como por la contingencia de los datos COVID-19. Por este mismo motivo, es totalmente factible aplicar esta propuesta en asignaturas de educación superior. En este sentido, las diversas actividades pueden ser adaptadas a los contextos en función de la disponibilidad de datos.

Como trabajo futuro, pretendemos implementar esta propuesta en el aula escolar, para comprobar los beneficios o desventajas que puedan presentarse. De esta forma, el estudio de clases se vuelve una interesante línea de investigación para ser explorada (Estrella y Estrella, 2020; Isoda y Olfos, 2020; Olfos et al., 2015).

Agradecimientos

Los autores agradecen a los revisores anónimos por sus valiosos aportes al mejoramiento de este trabajo y a Nelly Gómez (UBB) por sus comentarios y correcciones. Este trabajo fue parcialmente financiado por Fondecyt Iniciación n° 11200500.

Referencias

- Arnold, P. (2008). Developing new statistical content knowledge with secondary school mathematics teachers. En C. Batanero, G. Burrill, C. Reading y A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference* (pp. 1-6). International Commission on Mathematical Instruction and International Association for Statistical Education.
- Ben-Zvi, D., y Garfield, J. (2004). Statistical literacy, reasoning, and thinking: Goals, definitions, and challenges. En *The challenge of developing statistical literacy, reasoning and thinking* (pp. 3-15). Springer. <https://doi.org/10.1007/1-4020-2278-6>
- Biblioteca del Congreso Nacional. (2021, 20 de mayo). *Reportes comunales*. <https://www.bcn.cl/siit/reportescomunales/index.html>
- Blondel, V., Guillaume, J., Lambiotte, R., y Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, (10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Chance, B. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, 10(3), 1-14. <https://doi.org/10.1080/10691898.2002.11910677>
- Chance, B., Ben-Zvi, D., Garfield, J., y Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education Journal*, 1(1), 1-27. <https://doi.org/10.5070/T511000026>
- Del Pino, J. (2013). El uso de GeoGebra como herramienta para el aprendizaje de las medidas de dispersión. *Probabilidad Condicionada: Revista de didáctica de la estadística*, 2, 243-250.
- Dvir, M., y Ben-Zvi, D. (2021). Informal statistical models and modeling. *Mathematical Thinking and Learning*, 1-21. <https://doi.org/10.1080/10986065.2021.1925842>
- Estrella, S. (2017). Enseñar estadística para alfabetizar estadísticamente y desarrollar el razonamiento estadístico. En A. Salcedo (Ed.), *Alternativas Pedagógicas para la Educación Matemática del Siglo XXI* (pp. 173-194). Centro de Investigaciones Educativas, Escuela de Educación. Universidad Central de Venezuela. <http://saber.ucv.ve/bitstream/123456789/15712/1/Alternativas%20Pedagogicas%20para%20la%20Educa%C3%B3n%20Matem%C3%A1tica%20del%20Siglo%20XXI.pdf>
- Estrella, S., y Estrella, P. (2020). Representaciones de datos en estadística: de listas a tablas. *Revista Chilena de Educación Matemática*, 12(1), 21-34. <https://doi.org/10.46219/rechiem.v12i1.20>
- García, G., y Cuadros, P. (2013). Estrategias para mejorar la enseñanza de la Estadística con GeoGebra. En Sociedad de Educación Matemática Uruguay (Ed.), *VII Congreso Iberoamericano de Educación Matemática* (pp. 6335-6342). Autor.
- Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*, 10(3), 1-12. <https://doi.org/10.1080/10691898.2002.11910676>
- Gibbons, J. D., y Chakraborti, S. (2003). *Nonparametric Statistical Inference*. Marcel Dekker.
- Guncaga, J., Zawadowski, W., y Prodromou, T. (2019). Visualisation of Selected Mathematics Concepts with Computers – the Case of Torricelli’s Method and Statistics. *European Journal of Contemporary Education*, 8(1), 69-91. <https://doi.org/10.13187/ejced.2019.1.69>
- Isoda, M., y Olfos, R. (2020). *Teaching multiplication with Lesson Study: Japanese and Ibero-American Theories for Mathematics Education*. Springer. <https://doi.org/10.1007/978-3-030-28561-6>
- Makar, K., Bakker, A., y Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, 13, 152-173. <https://doi.org/10.1080/10986065.2011.538301>
- Makar, K., y Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8(1), 82-105.
- Makar, K., y Rubin, A. (2018). Learning about statistical inference. En D. Ben-Zvi, K. Makar y J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 261-294). Springer. https://doi.org/10.1007/978-3-319-66195-7_8
- Mellado, A., y Marín, L. A. (2010). Distribuciones estadísticas: Un ejemplo de uso de GeoGebra en enseñanza universitaria. *Épsilon*, 74, 33-42.
- Mena, G. E., Martínez, P. P., Mahmud, A. S., Marquet, P. A., Buckee, C. O., y Santillana, M. (2021). Socioeconomic status determines covid-19 incidence and related mortality in Santiago, Chile. *Science*, 372(6545). <https://doi.org/10.1126/science.abg5298>
- Ministerio de Ciencia. (2021, 20 de mayo). *Datos-COVID19* [GitHub dataset]. <https://github.com/MinCiencia/Datos-COVID19>
- Ministerio de Educación de Chile. (2019). Bases Curriculares 3° y 4° año medio. Autor. https://www.curriculumnacional.cl/614/articles-89597_recurso_10.pdf

Olfos, R., Estrella, S., y Morales, S. (2015). Clase pública de un estudio de clases de estadística: Una instancia de cambio de creencias en los profesores. *Revista Electrónica Educare*, 19(3), 1-17. <https://doi.org/10.15359/ree.19-3.21>

Pfannkuch, M. (2006). Informal inferential reasoning. En A. Rossman y B. Chance (Eds.), *Proceedings of the 7th International Conference on Teaching Statistics* (pp. 1-6). International Association for Statistics Education.

Prodromou, T. (2014). GeoGebra in Teaching and Learning Introductory Statistics. *Electronic Journal of Mathematics and Technology*, 8(5), 363-376.

Propuestas Educación Mesa Social Covid-19. (2020). *Didácticas para la proximidad: aprendiendo en tiempos de crisis*. Pontificia Universidad Católica de Chile y Universidad de Chile. https://www.pucv.cl/uuaa/site/docs/20200723/20200723181922/dida_cticas_para_la_proximidad__aprendiendo_en_tiempos_de_crisis.pdf

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5-19.

Triola, M. F. (2009). *Estadística* (10.a Ed.). Pearson Education.

Tukey, J. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Universidad del Bío-Bío. (2021, 20 de mayo). *Plataforma de comparación comunal datos covid-19*. <http://covid19.ubiobio.cl/index.html>

Wild, C. J., Pfannkuch, M., Regan, M., y Horton, N. J. (2011). Towards more accessible conceptions of statistical inference. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174, 247-295. <https://doi.org/10.1111/j.1467-985X.2010.00678.x>

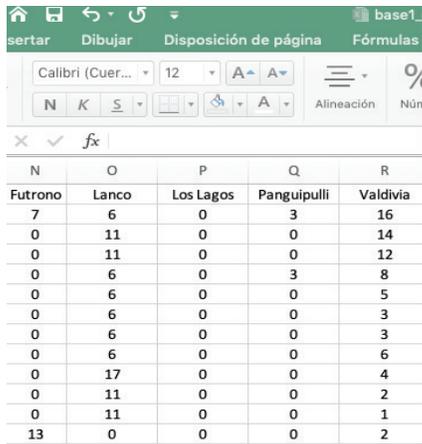
Zenteno Ruiz, F. A., Rivera Espinoza, T. A., y Pariona Cervantes, D. J. (2020). Tratamiento de las medidas de dispersión por medio del software GeoGebra. *Universidad y Sociedad*, 12(1), 244-250.

Zieffler, A., Garfield, J., Delmas, R., y Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40-58.

Apéndice: manipulación en el ambiente de trabajo de GeoGebra

En esta sección se presentan las rutinas en GeoGebra utilizadas para desarrollar el análisis exploratorio de datos de la Sección 3. Para más detalles, visitar https://youtube.com/playlist?list=PL2eb1_IpIPd-PCwipO-5ijiwOPUUi21DM.

Se destaca que la información descargada en formato CSV se consolidó en una planilla de cálculo, obteniendo una tabla como en la Figura 18.



N	O	P	Q	R
Futrono	Lanco	Los Lagos	Panguipulli	Valdivia
7	6	0	3	16
0	11	0	0	14
0	11	0	0	12
0	6	0	3	8
0	6	0	0	5
0	6	0	0	3
0	6	0	0	3
0	6	0	0	6
0	17	0	0	4
0	11	0	0	2
0	11	0	0	1
13	0	0	0	2

Figura 18. Encabezado de datos organizados en planilla de cálculo. Nota. Elaboración propia.

A continuación, se prepara el ambiente de trabajo de GeoGebra (Figura 19). Se deja habilitada sola la hoja de cálculo, dejando deshabilitada vista algebraica y gráfica (ver Figura 20).

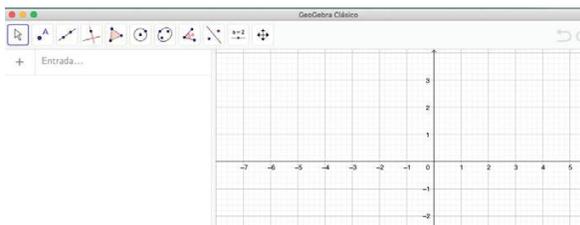


Figura 19. Vista por defecto en GeoGebra Clásico 6.0. Nota. Elaboración propia.

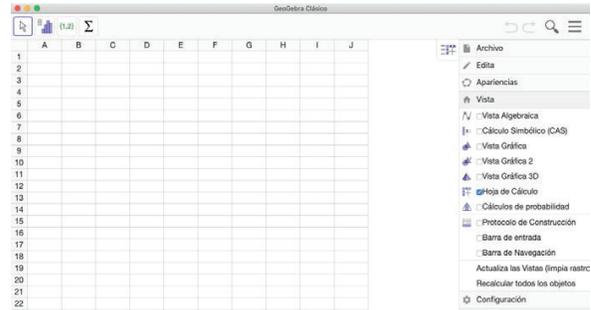
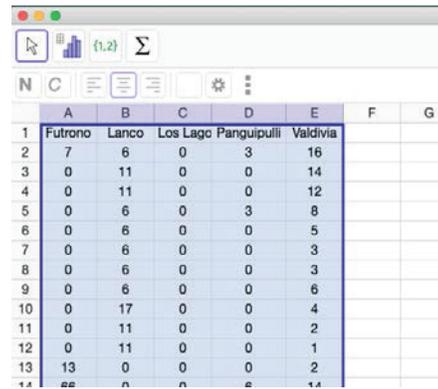


Figura 20. Vista con hoja de cálculo en GeoGebra Clásico 6.0

Nota. Elaboración propia.

Luego de configurar la vista de GeoGebra, se toma la información desde la planilla de cálculo y se inserta en la hoja de cálculo, como lo indica la Figura 21.



	A	B	C	D	E	F	G
1	Futrono	Lanco	Los Lago	Panguipulli	Valdivia		
2	7	6	0	3	16		
3	0	11	0	0	14		
4	0	11	0	0	12		
5	0	6	0	3	8		
6	0	6	0	0	5		
7	0	6	0	0	3		
8	0	6	0	0	3		
9	0	6	0	0	6		
10	0	17	0	0	4		
11	0	11	0	0	2		
12	0	11	0	0	1		
13	13	0	0	0	2		
14	13	0	0	0	2		

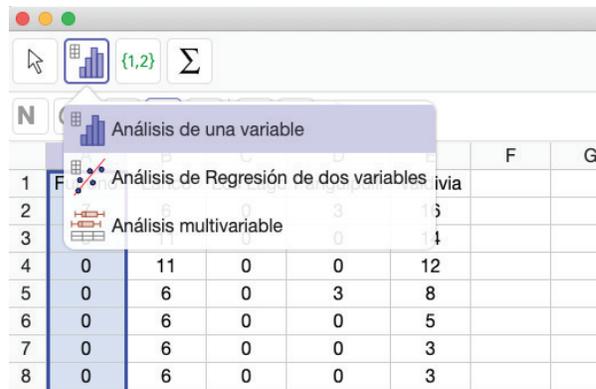


Figura 21. Hoja de cálculo comunes de RLR (arriba) y selección análisis de una variable (abajo)

Nota. Elaboración propia.

Análisis de una variable

Ejemplificamos la comuna de Futrono para realizar un análisis de una variable. Luego de seleccionar la herramienta de análisis de una variable (Figura 21), en la pantalla aparecerá un histograma asociado a la comuna de Futrono, tal como lo indica la Figura 22.

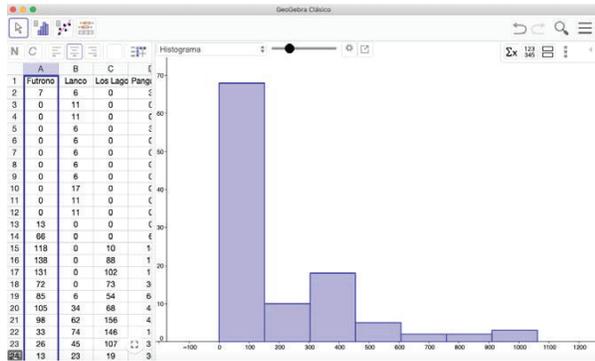


Figura 22. Histograma por defecto, comuna de Futrono
Nota. Elaboración propia.

Si se desea mostrar otro tipo de representación, se puede seleccionar el siguiente menú. Es posible hacer un diagrama de barras, de caja, de puntos, tallo y hoja y Q-Q plot. Esto siempre dependiendo del tipo de variable que estamos explorando (Figura 23).

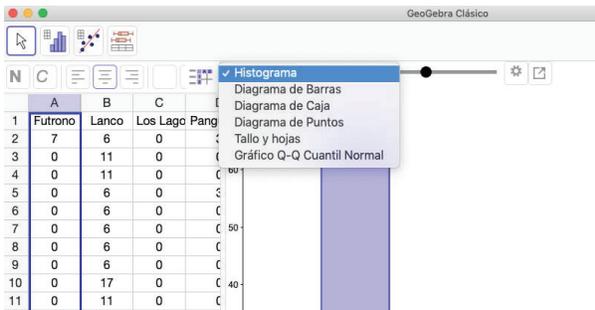


Figura 23. Opciones del tipo de representaciones que brinda GeoGebra
Nota. Elaboración propia.

Al mismo tiempo, si se presiona el icono con forma de engranaje, se puede ajustar el ancho de los intervalos, como lo indica la Figura 24, habilitando algunas opciones para las clases, el tipo de frecuencia y las opciones de habilitar histogramas y polígonos de frecuencias. También se pueden habilitar las estadísticas asociadas a la comuna de Futrono, como lo indica la Figura 25, donde se muestran otras opciones disponibles, en ellas podemos ver que se pueden realizar algunos test de hipótesis y construir intervalos de confianza.

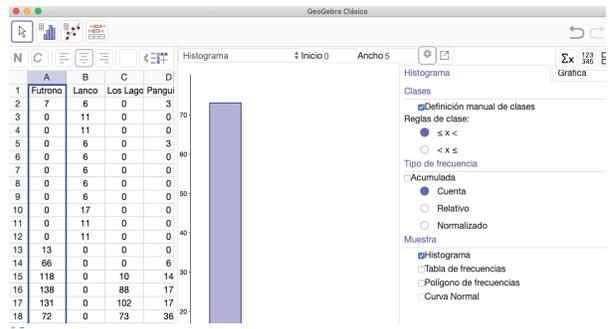


Figura 24. Icono de configuración de elementos de un histograma
Nota. Elaboración propia.

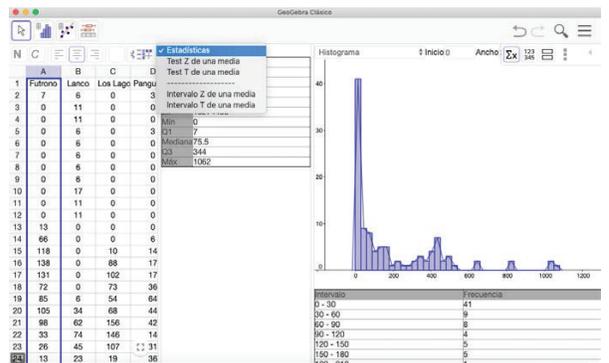


Figura 25. Otras opciones de la herramienta estadísticas
Nota. Elaboración propia.

Análisis de regresión con GeoGebra

Para la construcción de los gráficos de regresión usando GeoGebra Clásico 6.0, debemos seguir los siguientes pasos. Con la información de la planilla de cálculo de las comunas de la RM, seleccionamos dos comunas, para efectos del ejemplo Puente Alto y Vitacura, y presionamos en el menú correspondiente (Figura 26).

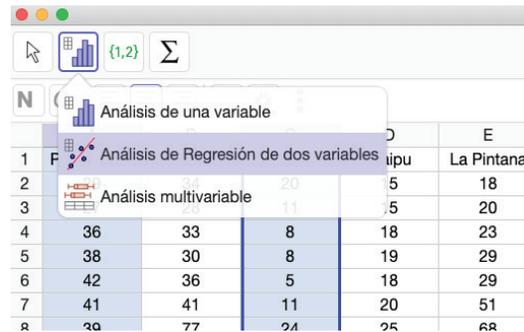


Figura 26. Selección de variables para regresión y agrupamiento
Nota. Elaboración propia.

Cuando se genera el gráfico de dispersión (Figura 27) se habilitan además algunos iconos como el de las estadísticas, visualización de puntos en caso de

querer borrar –por ejemplo, algún dato fuera de línea– y un botón que nos permite cambiar de eje la variable, en este caso las comunas de Puente Alto y Vitacura.

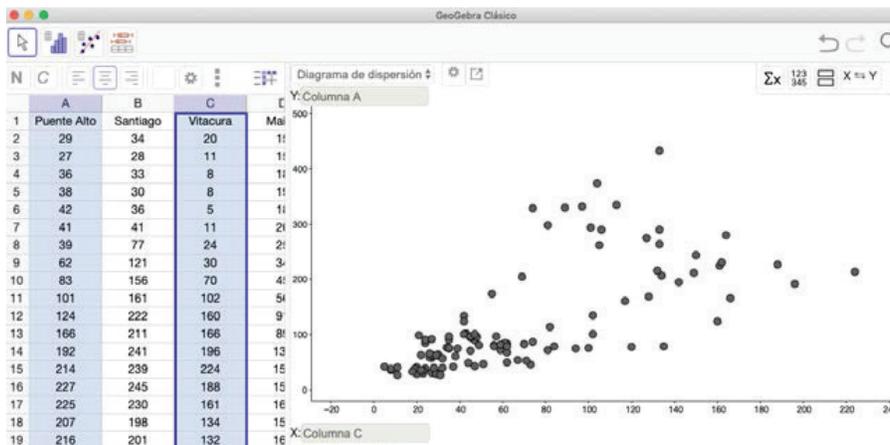


Figura 27. Regresión y agrupamiento para comunas de Puente Alto (columna A - eje Y) y Vitacura (columna C - eje X)

Nota. Elaboración propia.

En la parte inferior de la pantalla aparece un menú que nos permitirá seleccionar algún modelo de ajuste: lineal, log, polinomio (hasta grado 9), entre otros.

Para esta propuesta seleccionaremos el modelo de regresión lineal (Figura 28).

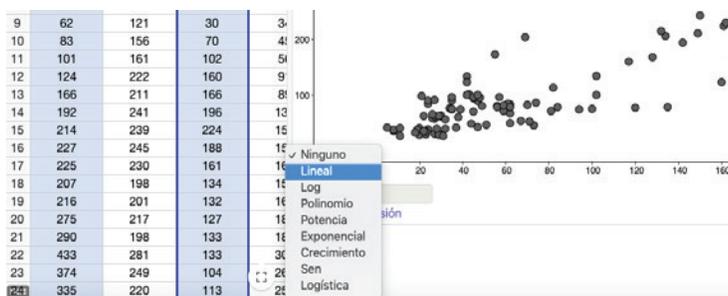


Figura 28. Selección de modelos de regresión Puente Alto (columna A - eje Y) y Vitacura (columna C - eje X)

Nota. Elaboración propia.

Cuando seleccionamos el modelo de regresión lineal, GeoGebra nos muestra de inmediato en la pantalla la línea de tendencia (color rojo) y junto con esto entrega en la parte inferior la ecuación que representa dicho modelo, el cual por lo tanto nos permite predecir asignando un valor a la variable X. Junto con esto, en la parte superior presionamos el icono de estadísticas, el cual nos genera algunos estadísticos relevantes que pueden ser parte en un análisis exploratorio, estos son medias, desviaciones estándar y varianzas de X e Y. También entrega coeficiente de correlación, covarianza, el valor de ρ (rho) y coeficiente de determinación, R^2 .

Análisis multivariado con GeoGebra

Con la información de la planilla de cálculo de la Figura 18, seleccionamos todas las comunas y pinchamos en el menú, seleccionando la herramienta análisis multivariable, como lo indica la Figura 29. A continuación, al costado derecho se generan los gráficos de cajas para cada comuna.

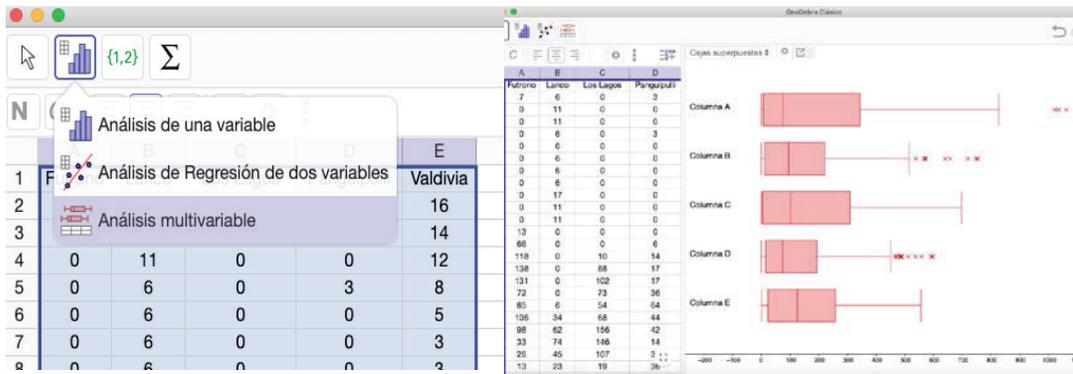


Figura 29 . Selección de comunas para un análisis multivariable (izquierda) y gráficos de caja para las comunas RLR (derecha) Nota. Elaboración propia.

Al hacer clic en el botón ubicado en el costado derecho se pueden habilitar las estadísticas asociadas a cada comuna. Además, en el menú ubicado en la parte inferior de la ventana (Figura 30), se encuentran

algunas herramientas para calcular ANOVA, test de diferencias e intervalos de confianza de medias y pareadas.

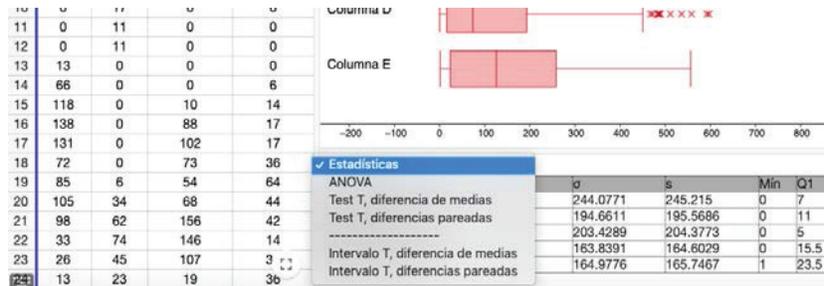


Figura 30. Otras herramientas desde el análisis multivariable Nota. Elaboración propia.