Informe sobre la variable de consumo de gobierno y una posible explicación con modelos

Introducción

El informe analizará como variable central al consumo de gobierno. Esta se explica en la capacidad del gobierno de un país para consumir respecto a una proporción de su PBI nominal en el año de corte. Aunque se pueda minimizar el consumo respecto a otras variables como el crecimiento económico, este recae en la importancia de su funcionalidad real para ver cómo otras variables recaen en este. O sea, simplemente no se puede deducir muchas características de un país al saber que el índice de consumo es alto o bajo, esto mas que solo puede predecir una política fiscal expansiva donde se amplifique el gasto. De ese modo, el consumo parecería solo una variable que pueda explicar ciertas variables en específico. Sin embargo, lo que se propone es que al momento de explicarla mediante otras variables es ver si se pueden alinear positiva o negativamente con ellas y se pueda proponer un modelo de predicción con una o más variables. Por ese motivo, se propone averiguar la funcionalidad del consumo no para ser explicada exactamente sino para averiguar si va acorde a otras variables que puedan predecirlo como modelo y abriendo más el campo a futuro a otras variables en conjunto con modelos más específicos.

Se utilizarán tres variables independientes al principio y luego una variable de control. Las independientes son el crecimiento económico, la regulación económica y libertad de expresión. Las tres variables se seleccionaron por diferentes motivos basados en su concepto. El crecimiento económico mide el crecimiento del porcentaje del PBI en la fecha de corte. Esta variable, por lo tanto, será importante en la jerarquía para concluir si dentro del modelo se puede explicar si es significante y en qué medida el crecimiento y consumo. Luego, la regulación económica mide la regulación interna general en el aspecto económico. Por lo tanto, su significancia recae si la limitación en ese aspecto influye en el consumo general del gobierno. Vale recalcar que la regulación la propone usualmente el mismo gobierno, asi que será importante en ver si la limitación afecta su manera de consumir. Esto sucede aunque el consumo se vea explicado por la regulación. Por último, la libertad de expresión no se encuentra dentro del aspecto económico. Esta variable significa en la libertad de movimientos a la demostración de expresión. Con esta variable se busca agregar al modelo un motivo social que donde la limitación de una variable en otro campo del entorno nacional pueda explicar efectivamente el consumo.

Por otro lado, se encuentra la variable de control. El dinero sólido es el valor real que ostenta el país respecto a su capacidad económica. Esta variable es útil al momento del intercambio de divisas. La inflación y los precios internos a largo plazo pueden variar si el valor del dinero sólido cambia. La variable es importante para controlar el modelo de modo de que puede agrupar las variables independientes y puede representarlas para configurar el error que independientemente puedan tener.

Entonces, habiendo explicado las variables y su importancia, se debe explicar el objetivo del análisis y qué se busca responder. La pregunta de investigación es ¿cuál es el impacto del crecimiento económico, regulación económica, libertad de expresión y finalmente el dinero sólido sobre el consumo de gobierno en el 2021? La pregunta no busca limitarse a las variables para la variable central. Sino darle un enfoque en el que pueda correlacionar todas las variables juntas y se pueda crear el modelo propuesto anteriormente. Además, dentro del análisis conglomerado se podrá especificar sobre la manera en la que se puedan etiquetar diferentes datos en la data en subagrupaciones.

Existen dos diferentes hipótesis. El consumo de gobierno se puede explicar sólo por las variables independientes antes descritas. No es necesario una variable de control. La segunda hipótesis es la necesidad de incluir la variable de control para que el modelo sí tenga validez o tenga una más alta.

Metodología y data

Como se explicó en la anterior sección, el modo de operacionalización para probar las hipótesis es crear un modelo de regresión entre las variables según la primera hipótesis. Según los resultados, también se añadirá la variable de control para ver cuál de los modelos podría ser más efectivo. En modo de cadena se añadirán nuevos procesos para fijar la efectividad de cada modelo. La regresión será gaussiana por la cantidad de datos y de la forma que están desplegadas. Esto ya que los datos no están contando. Eventualmente, por la técnica elegida, se estandarizará todas las variables involucradas en su respectivo modelo. Luego de saber cuál sería el mejor modelo. Se tiene que demostrar su factibilidad según el cumplimiento de los supuestos de la regresión de linealidad, homocedasticidad, normalidad de residuos, no multicolinealidad y valores influyentes.

Desde un análisis visual y gracias a códigos como “summary”, se puede saber que la data en el último año actualizado, 2021, provee una alta confianza para el análisis al tener solo 5 casos en la variable de consumo de gobierno en el que es 0. Su media es 5.28, cerca de la mediana, 5.37. Esta tendencia significa que el consumo de gobierno de los países se está minimizando y esto puede aludirse a la efectividad del manejo económico. Esta relación es lo que se busca deducir de los resultados del modelo. Además, se dice que se minimiza porque 10 años antes, su media se encontraba en 5.62 y la mediana 5.97. Ha disminuido.

El rango de las 5 variables continuas en la data es de 0 a 10. A mayor el índice, mayor será el concepto que explica. Por ejemplo, si el índice de libertad de expresión de un país es 9.5, se puede decir que esa variable es muy alta en el año que es analizado.

Todas las variables se encuentran en la misma data y fueron extraídas del análisis para medir la libertad humana. El título oficial es ”The human freedom Index 2023”. Fue realizado por Cato Institute. Esta institución privada recolecta todos los años los datos para que sea fiable la demostración del índice al que indaga. Por tal motivo, a pesar de que el índice se reveló en el 2023, la data solo agrupa hasta el año 2021. Este año será importante para el análisis exploratorio y la descripción de variables. En el dashboard las dos primeras secciones toman en cuenta únicamente el año 2021 para poder graficar las variables, proporciones y relaciones entre el consumo de gobierno y las demás variables.

Para extraer nuestras variables de interés se creó una nueva data. Esta selecciona el año, país, región y las cinco variables para la elaboración del modelo. Esta data sobre la que se trabajará se llama “data\_entregable”. Esta tiene 3630 datos, o sea 165 países desde el 2000. De esta data principal se extraerá la data para los modelos de las hipótesis: “data\_entregable\_2021”. Esta se utilizará como se mencionó anteriormente para los gráficos exploratorios y demás. Debido a la cantidad de datos, la data se agrupará por regiones en los gráficos descriptivos. Estas son 10 en el que el occidente de Europa representa la mayor proporción de todas las regiones. Además, para la asociación de la variable central y las otras variables en la segunda sección se realiza con la misma data. Previo a esto, en un análisis exploratorio, los 165 casos no presentan NA’s, por lo que no es necesario el uso de “complete\_cases”. La data original de los 22 años sí cuenta con gran variedad de casos perdidos. Esto posiblemente a la limitación de años anteriores para extraer los índices o que se sean efectivos.

Posteriormente, tanto para las gráficas, regresiones y la clusterización se extraerán las variables necesarias para mutar variables. Por ejemplo, para el análisis conglomerado se usará la función “row.name” con la intención de que la variable país permanezca en la data y pueda identificar cada dato y ver a qué subgrupo pertenece. Entonces, en medida de lo necesario se formará nueva data y nuevos valores principalmente con el fin de no perjudicar las datas principales y valores, pues podría afectar la regresión o algún modelo siguiente.

Hallazgos

Análisis univariado:

Desde la primera sección se busca analizar a la variable central. Como se mencionó al principio, el consumo independientemente es una variable que raramente se puede deducir de lo que significa un alto o bajo índice. En uno de los primeros gráficos se aprecia que el índice del consumo es mayor para las regiones de Asia Central y Latinoamérica. Por otro lado, una de las regiones más estables económicamente y por sus innovaciones que es Europa occidental tiene el índice más bajo en el 2021 representando una media de 2.51. Esto no se asocia al crecimiento económico en el que Oceanía toma el primer puesto pero Europa de occidente sigue siendo la región con menor índice, aunque no es bajo, 7.92.

Por otro lado, en la segunda sección hay un gráfico de boxplot para medir las diferencias entre las variables. Se puede ver que en el último año actualizado, el consumo de gobierno en África debajo del Sahara tiene una mayor variedad de índices desde el 0 lo que podría significar que es data con porcentaje de error y además dos países de la región que llegan a 10: Chad y Somalia. Tomando en cuenta información geopolítica conocida, se sabe que estos dos países están muy limitados económica, esto debido a conflictos sociales, políticos y claramente económicos. Esto puede dar indicio a que probablemente un alto índice en el consumo puede significar que el manejo económico está altamente presionado. No se puede decir lo mismo de un muy bajo índice, pues países como Noruega y Suecia tienen un índice de 0.74 y 0.85 en el 2021 respectivamente. El gráfico muestra un solo caso atípico superior. Se encuentra en Latinoamérica y el Caribe. Este es Haití y presenta 9.65 en el índice. Igualmente refuerza el indicio sobre un alto índice es producido por una gran presión sobre el manejo económico.

Entonces, tomando en cuenta el posible significado de lo que un alto índice de la variable de consumo podría inferir, se puede ver los gráficos de las asociaciones entre la variable con las variables del modelo.

Para este caso se usará el gráfico de la sección del análisis de conglomerados, exactamente el de correlación, en él se ve la relación del consumo de gobierno con las variables predictoras no son altas pero tampoco son tan bajas. Se destaca la correlación entre la variable con el crecimiento económico y dinero sólido. Además, se puede ver la relación el impacto de la variable de control sobre las otras variables mostrando una alta correlación con la regulación económica al igual que lo tiene con el consumo.

Volviendo a la segunda sección se encuentran los gráficos de dispersión de la variable de consumo y las demás variables. Los países están agrupados por los colores de las regiones que pertenecen. No se puede ver una correlación clara en los gráficos debido a que los datos están muy dispersos, en especial el gráfico del consumo y de la libertad de expresión, y no se logra encontrar un patrón claro. Sin embargo, sí se puede inferir sobre la visualización de los datos más propiamente independientes. Por ejemplo, sabiendo que la media del crecimiento económico es alta, en el gráfico de la relación con el consumo se puede ver que los puntos están cerca al tope y además manteniéndose gran parte en el medio; esto a excepción de casos donde el consumo es muy bajo o alto.

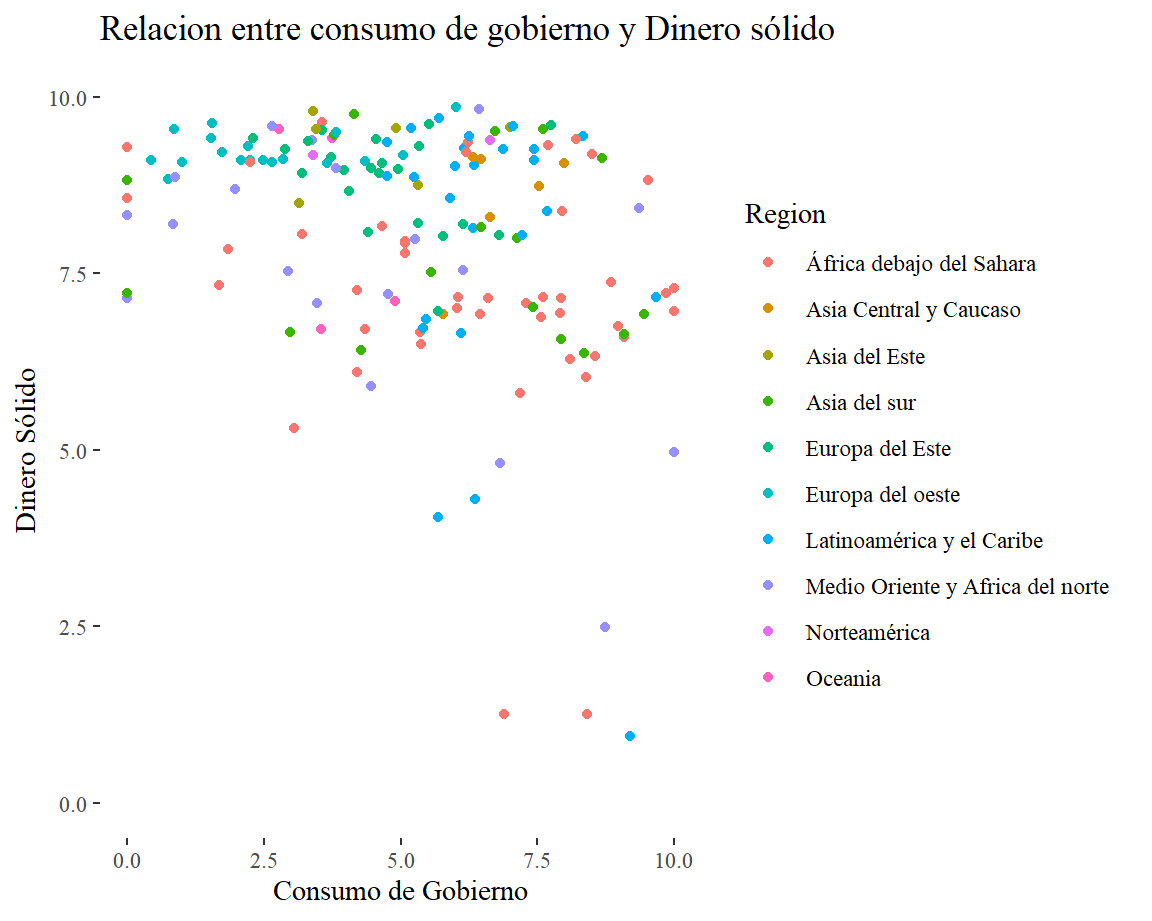
Además, se tiene que resaltar la relación de la variable de control, el dinero sólido. Anteriormente se muestra su correlación con la variable central y su alta correlación como la variable de control ante las otras variables. En el gráfico se muestra que también la media del dinero es alta con excepciones. La mayoría se encuentra en el primer y segundo cuadrante lo que también demuestra la media alrededor de 5 del consumo. Una inferencia es que los datos más altos del dinero sólido no se relacionan fácilmente con una tendencia del consumo. Sin embargo, hay una sección entre el 2.5 y 5 de la variable central que se encuentra más junto y cerca al índice 9 del dinero sólido. Gran parte de los puntos pertenecen a Europa occidental. Se recuerda el indicio de que un alto índice en el consumo podría ser indicio de mayor presión económica. Este sector se ve aliviado en esta característica.

Ilustración 1

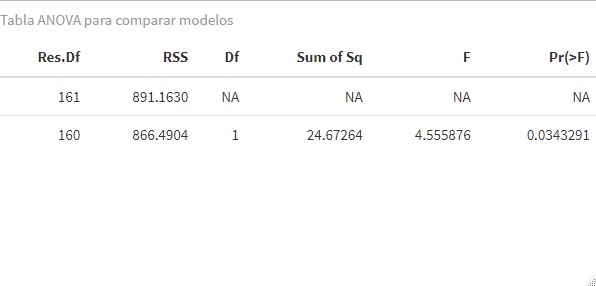
Análisis multivariado:

El siguiente paso de la verificación de la hipótesis son los modelos de regresión. El primer modelo centra a la variable central junto a las tres variables independientes: regulación económica, crecimiento económico y libertad de expresión. Este primer modelo muestra una única variable significativa: la regulación económica. Tiene un 0.01 de significancia. Sin embargo, se da a cuenta que la regresión debe tomar valores estandarizados y también añadir una variable de control. De esta forma se verá si el R2 ajustado puede ser mayor, está en 0.106.

Es necesario estandarizar para ver el verdadero efecto de las independientes sobre la central. Se ve que la variable de regulación económica tiene mayor efecto por sobre las otras aunque sean insignificantes. Es -0.264. Las otras variables también bajan su efectividad y se mantienen insignificantes. Además, el AIC baja.

Tomando en cuenta la segunda hipótesis, se añade la variable de control, el dinero sólido, su inclusión en el modelo deja ver que todas las demás variables ya no son significantes. En cambio, el dinero sólido presenta una significancia de 0.05. Cabe resaltar que el R2 ajustado sube ligeramente a 0.126, pero no demuestra una completa explicación del modelo con ese índice.

Por último en el cuarto cuadro se aprecian los dos modelos estandarizados. Como se había previsto en el tercer cuadro, solo la variable de control tiene significancia, su efecto es -0.230. En este cuadro, como los otros, también se debe apreciar que todas estas variables tienen un efecto negativo sobre el consumo. Por ejemplo, mientras más consumo de gobierno, se puede decir que hay menos dinero sólido. En el caso de Venezuela se puede ejemplificar, el índice de consumo es 9.18 y el dinero sólido es 0.94.

Por lo tanto, habiendo visto los modelos y sus estandarizaciones, se procede a hacer un cuadro de anova comparando ambos modelos. Según el Pr(>f) y el RSS se puede decir que el modelo 2 es el que se tiene que seleccionar para el análisis. Por tal motivo este es elegido que debe pasar los supuestos de la regresión.

Se toman en cuenta los 5 supuestos. En el gráfico de linealidad no se puede encontrar una alineación de la línea roja con la horizontal. Se podría decir que hay indicios que demuestran que el modelo no es explicativo. En el gráfico de homocedasticidad, se ve que la varianza de errores no es constante y no hay una dirección definida. Se puede rechazar también el supuesto de homocedasticidad. En el gráfico de normalidad de residuos, tampoco se cumple su supuesto sobre que los residuos se distribuyen de forma normal. Según la prueba de Shapiro se puede demostrar la falta del supuesto al ser menor a 0.05. Es 0.022.

Por otro lado, en el gráfico de no multicolinealidad se puede ver que si cumple el supuesto entre la correlación de las variables independiente y la de control. Según el estadístico VIF todas las variables deben ser menores a 3. Por último, en el gráfico de valores influyentes se puede ver que no hay valores influyentes. O sea que no hay ningún valor que pueda influir en un mayor error del modelo.

Según estos resultados de los modelos de regresión y los supuestos del mejor modelo seleccionado se puede decir que aparte de un R2 bajo existe una gran mayoría de los supuestos no cumplidos. Por lo tanto, la explicación del consumo en el modelo puede explicarse parcialmente pero se encuentra muy limitado al error según los supuestos. La formulación de otro modelo con variables diferentes sería apropiada para brindarle un modelo efectivo que pueda probar otros modelos como se propuso para su justificación.

Gráfico, Gráfico de barras

Descripción generada automáticamenteTeniendo en cuenta el resultado de la regresión, se procede con los pasos para la clusterización. Para este campo se toma en cuenta la correlación de la variable central con las variables independientes. Con los que tiene mayor correlación, pero no significantemente, son con las variables de regulación económica y dinero sólido con -0.34 ambos. Con la libertad de expresión solo se correlaciona en -0.26.

Se escoge la estrategia jerárquica con el algoritmo aglomerativo. El primer paso para la clusterización es saber cuántos clusters se deben usar. Sin embargo, se crea una nueva data, este solo tomará las cinco variables numéricas del modelo. Además, con la función “row.names” se selecciona a la variable “país” para que no se pierda qué dato se está incluyendo a cada cluster.

Para saber el número de clusters se usa el estadístico GAP, según el algoritmo AGNES. Este va a ser el único necesario y usado durante todo el proceso. En la función para hallar el número se usa la matriz de distancia de Gower. El resultado arroja que deben ser dos el número de clusters. Por lo tanto en la data para la clusterización se añade la columna a la que pertenece cada país según el cluster.

Luego, se puede realizar el gráfico del dendograma. Se puede ver que existe un cluster que tiene menos países en él. Después se podría inferir esta característica con todos los resultados. A continuación, se evalúa el uso del algoritmo de AGNES y ver los resultados de usar este modelo. El gráfico de las “silhouttes” muestran además valores mal clusterizados. En total son 32 países.

Por último, se extrae el promedio de las variables de los clusters. Se puede ver que el segundo cluster tiene un menor promedio y una mayor libertad de expresión y dinero sólido. Esto puede decir que los países en el segundo cluster están clasificados como países que muy probablemente no presentarían problemas como Somalia o Venezuela.

Conclusiones

Los modelos de regresión aún con variable de control muestran no ser efectivos totalmente y están sujetos a error según los supuestos. Respondiendo a la pregunta de investigación, se podría decir que los modelos planteados buscando probar las hipótesis no son efectivos tampoco para responder el consumo de gobierno. Sin embargo, las variables predictoras desde el análisis univariado y la metodología pueden hacer inferir una relación entre tales variables independientemente con el consumo. En el modelo 2, con la variable de control, se ve que el único significante es este y mantiene aún un efecto que no se puede minimizar. Entonces, el mejor modelo que contiene la variable de control aunque no sea un modelo explicativo de la variable central si puede analizar en ciertos casos pero limitándose a un margen de error, un ejemplo son los residuos hallados en los supuestos que lo pueden desacreditar. Igualmente, el indicio de presión económica podría estar probada en la relación del dinero sólido y el consumo. Esto se puede explicar mejor gracias al análisis de asociación de los gráficos de la variable central y la variable de control.

Por otro lado, en el proceso de clusterización, juntando todas las variables. Dentro de él se da a resultado a dos clusters. Al termino del proceso, siguiendo los resultados de la media de los dos grupos se puede inferir también que uno de ellos presenta una mejor sostenibilidad dentro del progreso económico y más fuerte en temas sociales como es la libertad de expresión. Es el caso del segundo cluster, esta junta menos datos y se resalta también que es el grupo en el que no se quita valores mal clusterizados.

Finalmente, se advertía de la rareza del consumo de gobierno como variable a explicar. Sin embargo, este análisis puede abrir grandes indicios a relacionar más variables que pueden tender a una asociación con la variable más directa y que en un modelo se refleje más efectividad. Seguramente, la creación de más modelos con más variables e incluso más lejanas al análisis económico de un país podrían tomar a la variable central como consumo y se puedan explicar mejor para la mayor efectividad de la ecuación de la regresión. La funcionalidad del consumo ya no es ambigua gracias a los procesos anteriores y puede dar indicio sobre la dirección del país según las variables utilizadas en el modelo.

**LINK DE DASHBOARD:**

<https://rpubs.com/Sebastian_03/Dashboard>