

Subscriptores de internet - Inferencia estadística 2020-1

Carlos Arturo Moreno Tabares - C.C 1088029924

11 Junio 2020

Contents

| | | |
|----------|--|----------|
| 1 | Introducción | 2 |
| 2 | Datos para el análisis | 2 |
| 2.1 | Pre-Procesamiento | 2 |
| 3 | Puntos a responder | 2 |
| 3.1 | Punto 1 - ¿Cual es la empresa con mas suscriptores en el año 2012? . | 2 |
| 3.2 | Punto 2 - ¿Cual es la empresa con menos suscriptores 2012? | 3 |
| 3.3 | Punto 3 - Repetir lo realizado en los puntos 1 y 2 para el año 2019 . . | 3 |
| 3.4 | Punto 4 - Seleccione una empresa y genere modelo descriptivo para el año 2019 sobre el atributo de cantidad de suscriptores | 5 |
| 3.5 | Punto 5 - Seleccione un tipo de tecnología (2,3,4 G) y segmento (empresa/personal) para el año 2019 | 7 |
| 3.5.1 | ¿Cual es el promedio de suscriptores? | 7 |
| 3.5.2 | ¿Cual es la desviación estándar de suscriptores? | 7 |
| 3.5.3 | ¿Cual es el tamaño de la muestra? | 7 |
| 3.5.4 | ¿Coincide con sus valores del modelo descriptivo? | 7 |
| 3.6 | De lo seleccionado del punto 5 Niegue o refute que el numero de suscriptores del segundo trimestre no ha cambiado respecto al promedio anual del año 2019. | 7 |

1 Introducción

Para evaluar el conocimiento adquirido en el curso de inferencia estadística 2020-1 se realizará un estudio a los datos abiertos sobre los **Suscriptores de internet móvil de cargo fijo** que entrega la web de datos abiertos de Colombia. Se resolverán varios puntos planteados donde se afronta un problema de como abarcar los datos, como analizar e interpretar y con estos medir al estudiante.

2 Datos para el análisis

Los datos a analizar se pre-visualizan mediante el siguiente enlace <https://www.datos.gov.co/Ciencia-Tecnologia-e-Innovacion/Suscriptores-de-internet-movil-de-cargo-fijo/7m3n-pxmk> estos datos contienen información de suscriptores de Internet móvil de cargo fijo reportado por los proveedores de redes y servicios que ofrezcan acceso móvil a Internet. Incluye información desde enero de 2012 hasta diciembre de 2019. Este cuenta con un total de 4409 datos.

2.1 Pre-Procesamiento

Como ocurre en la mayoría de set de datos requieren un tratamiento para poder realizarse análisis, en este caso algunas empresas no lograban tener reportes de suscriptores en algunos de los meses pero al traducirlos al R eran tomados como algo **no numérico** o **nulo** y debía darse algún valor en este caso fueron rellenados con valor de cero, además que por defecto los valores no eran tomados tipo numérico (cuando si debía de ser) para mayor detalle visualizarlo en el Notebook.

En el Notebook adjunto hay una celda que pre-procesa los datos y muestra una previa de en formato de tabla, por defecto el Kernel de R utilizado usa un encoding que no reconoce todos los caracteres mostrando caracteres erróneos en las tildes y ñ.

3 Puntos a responder

En los puntos a responder se realiza un análisis descriptivo de los datos el cual entrega información para preguntar y responder a la hipótesis a plantear, también permitirá dar unas ideas sobre el comportamiento que tengan estos datos tratando de realizar un ejercicio muy cercano al campo laboral.

3.1 Punto 1 - ¿Cual es la empresa con mas suscriptores en el año 2012?

El calculo que se realizó fue Filtrar a solo los suscriptores del 2012 y de ello agrupar sobre Meses del trimestre y trimestres dejando solo los 12 meses para el año sumando de forma directa las diferentes tecnologías y tipo de cliente de cada compañía de internet y luego agrupar por empresa para ese resultado dividirlo en el total de meses (que es 12) y así tener una tentativa del "promedio" anual de cada empresa.

A grouped_df: 4 × 3

| ID_EMPRESA | DESC_EMPRESA | CANTIDAD_SUSCRIPTORES |
|------------|---|-----------------------|
| <int> | <fct> | <dbl> |
| 800153993 | COMUNICACION CELULAR S A COMCEL S A | 1087714.50 |
| 830122566 | COLOMBIA TELECOMUNICACIONES S.A. ESP | 1023686.50 |
| 830114921 | COLOMBIA MOVIL S.A. E.S.P. | 589853.90 |
| 899999115 | EMPRESA DE TELECOMUNICACIONES DE BOGOTÁ S.A. ESP. | 43588.75 |

Figure 1: Promedio de suscriptores por empresa para el 2012 - Descendente

Por cada empresa $prom = \sum_{i=1}^n \frac{x_i}{n}$

De acuerdo al resultado de la figura 1 muestra que la empresa con mas suscriptores para el 2012 es **COMUNICACION CELULAR S A COMCEL S A**

3.2 Punto 2 - ¿Cual es la empresa con menos suscriptores 2012?

Obteniendo los datos de la misma forma que en el punto 1 solo que ordenando de forma ascendente se obtiene la siguiente tabla

A grouped_df: 4 × 3

| ID_EMPRESA | DESC_EMPRESA | CANTIDAD_SUSCRIPTORES |
|------------|---|-----------------------|
| <int> | <fct> | <dbl> |
| 899999115 | EMPRESA DE TELECOMUNICACIONES DE BOGOTÁ S.A. ESP. | 174355 |
| 830114921 | COLOMBIA MOVIL S.A. E.S.P. | 1966180 |
| 830122566 | COLOMBIA TELECOMUNICACIONES S.A. ESP | 4094746 |
| 800153993 | COMUNICACION CELULAR S A COMCEL S A | 4350858 |

Figure 2: Promedio de suscriptores por empresa para el 2012 - Ascendente

De acuerdo al resultado de la figura 2 muestra que la empresa con menos suscriptores para el 2012 es **EMPRESA DE TELECOMUNICACIONES DE BOGOTÁ S.A. ESP.**

3.3 Punto 3 - Repetir lo realizado en los puntos 1 y 2 para el año 2019

Para obtener este dato se hace el filtro para suscriptores de el 2019, se agrupan por mes del trimestre y trimestre sumando directamente y luego son agrupados por empresa y a ese agrupamiento dividido sobre el total de meses que es 12.

Por cada empresa $prom = \sum_{i=1}^n \frac{x_i}{n}$

De acuerdo a los resultados en la figura 4 muestra que para el 2019 la empresa con mas suscriptores en el año es **COMUNICACION CELULAR S A COMCEL S A**, la

A grouped_df: 7 × 3

| ID_EMPRESA | DESC_EMPRESA | CANTIDAD_SUSCRIPTORES |
|------------|---|-----------------------|
| <int> | <fct> | <dbl> |
| 800153993 | COMUNICACION CELULAR S A COMCEL S A | 6388195.5 |
| 830122566 | COLOMBIA TELECOMUNICACIONES S.A. ESP | 3661153.3 |
| 830114921 | COLOMBIA MOVIL S.A. E.S.P. | 1435672.6 |
| 830016046 | AVANTEL S.A.S. | 546536.6 |
| 899999115 | EMPRESA DE TELECOMUNICACIONES DE BOGOTÁ S.A. ESP. | 157804.0 |
| 900420122 | VIRGIN MOBILE COLOMBIA S.A.S. | 0.0 |
| 901193332 | NESH MOVIL S A S | 0.0 |

Figure 3: Promedio de subscriptores por empresa para el 2019 - Descendente

A grouped_df: 7 × 3

| ID_EMPRESA | DESC_EMPRESA | CANTIDAD_SUSCRIPTORES |
|------------|---|-----------------------|
| <int> | <fct> | <dbl> |
| 900420122 | VIRGIN MOBILE COLOMBIA S.A.S. | 0.0 |
| 901193332 | NESH MOVIL S A S | 0.0 |
| 899999115 | EMPRESA DE TELECOMUNICACIONES DE BOGOTÁ S.A. ESP. | 157804.0 |
| 830016046 | AVANTEL S.A.S. | 546536.6 |
| 830114921 | COLOMBIA MOVIL S.A. E.S.P. | 1435672.6 |
| 830122566 | COLOMBIA TELECOMUNICACIONES S.A. ESP | 3661153.3 |
| 800153993 | COMUNICACION CELULAR S A COMCEL S A | 6388195.5 |

Figure 4: Promedio de subscriptores por empresa para el 2019 - Ascendente

cual es la misma que para el 2012. con la figura 3 se identifica que la empresa con menos subscriptores y valor diferente a cero es **EMPRESA DE TELECOMUNICACIONES DE BOGOTÁ S.A. ESP.**, sin embargo, en estos mismos datos se visualiza que hay dos empresas que están registradas pero no tienen subscriptores son **VIRGIN MOBILE COLOMBIA S.A.S.** y **NESH MOVIL S A S**

3.4 Punto 4 - Seleccione una empresa y genere modelo descriptivo para el año 2019 sobre el atributo de cantidad de subscriptores

La empresa que se selecciona para el modelo descriptivo es la que mas subscriptores tuvo que es **COMUNICACION CELULAR S A COMCEL S A** a este se le hace el tratamiento de datos correspondiente para analizar Tecnologías, Trimestres, Terminal y tipo de subscriber obteniendo los siguientes datos.

A grouped_df: 12 × 3

| TRIMESTRE | MES_DEL_TRIMESTRE | CANTIDAD_SUSCRIPTORES |
|-----------|-------------------|-----------------------|
| <dbl> | <dbl> | <dbl> |
| 1 | 1 | 0.07988242 |
| 1 | 2 | 0.08003932 |
| 1 | 3 | 0.08077785 |
| 2 | 1 | 0.08121253 |
| 2 | 2 | 0.08228706 |
| 2 | 3 | 0.08278176 |
| 3 | 1 | 0.08377152 |
| 3 | 2 | 0.08452272 |
| 3 | 3 | 0.08532270 |
| 4 | 1 | 0.08606773 |
| 4 | 2 | 0.08634702 |
| 4 | 3 | 0.08698737 |

Figure 5: Porcentaje de subscriptores en el año 2019 de la empresa comcel

Los datos obtenidos en el modelo descriptivo se observa que de-acuerdo a la figura 5 se ve que a medida que va pasando el año aumenta la cantidad de subscriptores. Con

A tibble: 3 × 2

| DESC_TECNOLOGIA | CANTIDAD_SUSCRIPTORES |
|-----------------|-----------------------|
| <fct> | <dbl> |
| 2G | 78892.06 |
| 3G | 273751.19 |
| 4G | 1244405.62 |

Figure 6: Porcentaje de suscriptores en el año 2019 de la empresa comcel por tecnología

A tibble: 2 × 2

| DESC_SEGMENTO | CANTIDAD_SUSCRIPTORES |
|---------------|-----------------------|
| <fct> | <dbl> |
| Empresas | 0.1612705 |
| Personas | 0.8387295 |

Figure 7: Porcentaje de suscriptores en el año 2019 de la empresa comcel por segmento

los datos de la figura 6 se observa que la tecnología que mas suscriptores tiene es la 4, de hecho son aproximadamente 4 veces mas que las tecnologías 2G y 3G juntas. En la figura 7 se aprecia que la mayoría de suscriptores para comcel son Personas.

Otros datos que se sacaron en este modelo es el **promedio** a razón de meses (con todos los datos juntos) es 532349,625, el número de meses es 12 el cual seria el **tamaño de muestra** y la **desviación estándar** bajo esos 12 meses con todo junto es 16191,0511345868

3.5 Punto 5 - Seleccione un tipo de tecnología (2,3,4 G) y segmento (empresa/personal) para el año 2019

Para este punto se seleccionó la **tecnología 4G** y **segmento personas** por que son los que mayor cantidad presenta, con estos datos se han de responder una serie de preguntas, en cada una se aplicó el proceso respecto de filtros y tratamiento de datos para llegar a la respuesta.

3.5.1 ¿Cual es el promedio de suscriptores?

El promedio de suscriptores para 4G personas es 1815717,875

3.5.2 ¿Cual es la desviación estándar de suscriptores?

La desviación estandar de suscriptores a 4G personas es 113229,914849093

3.5.3 ¿Cual es el tamaño de la muestra?

El tamaño de la muestra son los 12 meses

3.5.4 ¿Coincide con sus valores del modelo descriptivo?

Los análisis del punto 5 son tan solo con una parte especifica de los datos, si vemos bien con todos los datos la desviación estándar corresponde al 3% del promedio relacionado mientras que el de este segmento especifico corresponde al 6,2% de del promedio relacionado es decir que al doble respecto al modelo respectivo. Si miramos para otras tecnologías se cumpliría que ese promedio ha de ser menor para el año 2019. Es decir que si hay ciertas coincidencias en el modelo descriptivo.

3.6 De lo seleccionado del punto 5 Niegue o refute que el numero de suscriptores del segundo trimestre no ha cambiado respecto al promedio anual del año 2019.

La hipotesis que se plantea es la siguiente:

- $Hipotesisnula(H_o) \rightarrow promedio_Trimestre_2 = promedio_anual$
- $Hipotesisalternativa(H_a) \rightarrow promedio_Trimestre_2 \neq promedio_anual$

Donde la condición para rechazar a H_o es $t_{0.025} > z > t_{0.975}$ es decir que el Z se encuentre en el rango de 0.025 y 0.975 debido a que se quiere dar un 95% de confianza. De los datos se filtra 2019, tecnología 4G, empresa comcel y solo personas donde se saca promedio, media, número de muestras, desviación estandar y se tiene que:

- Promedio del trimestre 2 $\rightarrow prom_trim_22165935.83333333$
- Tamaño de la muestra $\rightarrow n = 12$
- Media $\bar{x} = 2219190,54166667$
- Desviación estándar $\sigma = 131604,66217848$
- Valor de $t_{0.025}$ (Cola izquierda) -2.17881282966723
- Valor de $t_{0.975}$ (Cola derecha) 2.17881282966723

Calculo de la hipotesis

$$\bullet \frac{\bar{x} - prom_trim_2}{\sigma / \sqrt{n}} = \frac{2219190,54166667 - 22165935.83333333}{131604,66217848 / 12} = 1.40177192887742$$

El valor de 1.40177192887742 se encuentra dentro del rango de la hipótesis nula (H_o) por ende se acepta la hipótesis nula (H_o) confirmando que si ha cambiado la cantidad de suscriptores en el trimestre 2 respecto al promedio anual del 2019.