

Detección de datos estadísticamente atípicos en compras.

Elaborado por Alejandro Herrera López (alejandroherrera@gmail.com)

Este documento está cubierto en su prosa por la licencia



, siendo un documento de cultura libre.

En su código fuente está cubierto por GPLv3



Contexto del presente trabajo:

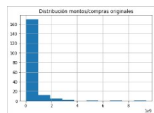
Es un trabajo completa y exclusivamente a título personal, se debe tener el cuidado en el momento de su lectura, que un dato estadísticamente atípico no implica por sí mismo, una actuación irregular o extraña de la Administración Pública. El trabajo se realiza a un alto nivel, sin ingresar a información del expediente de cada contratación y para la realización del mismo se utiliza como fuente el Sistema Integrado de la Actividad Contractual, de la Contraloría General de la República, el cual a su vez se alimenta de información ingresada por la misma.

Datos analizados:

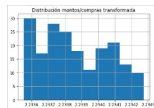
Para el mismo se utilizó como dataset las contrataciones directas registradas por la **CCSS, por un monto por encima de los 50 millones de colones**. A las variables extraídas y disponibles en el SIAC, se le incorporó un cálculo de los días entre las fechas de invitación a apertura de cada procedimiento, y los días entre las fechas de apertura y adjudicación. A continuación se hace un detalle de las variables disponibles:

```
out[ ]:
```

```
INSTITUCION LICITANTE      object
ANNIO                     int64
NUMERO_PROCEDIMIENTO      object
OBJETO_CONTRACTUAL        object
FECHA_APERTURA           datetime64[ns]
FECHA_INVITACION          datetime64[ns]
DIAS_INVITACION_APER      int64
FECHA_ADJUDICACION        datetime64[ns]
DIAS_APERTURA_ADJU       int64
FECHA_COMUNICO            datetime64[ns]
LIMITE_LEGAL              object
MODALIDAD_CONTRATACION    object
PRESUPUESTO_ESTIMADO      float64
EXCEPCION_CONTRATACION_DIRECTA object
LINEA                    int64
ESTADO_LINEA              object
BIEN_SERVICIO             object
CEDULA_ADJUDICADO         int64
ADJUDICADO                object
CANTIDAD_REQUERIDA        int64
CANTIDAD_ADJUDICADA       int64
INESTIMABLE               object
MONTO_ADJUDICADO          float64
SUBPARTIDA_GASTO          object
dtype: object
```

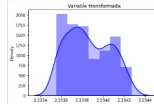


Se evidencia que la distribución del ingreso es muy asimétrica a la derecha, por lo tanto se aplica la transformación de Box-Cox. Se dispone el histograma con los datos transformados.



```
Out[ ]:
```

```
Text(0.5, 1.0, 'Variable transformada')
```



Aplicación de la regla de Tukey para ubicar outliers

```
[2.23355339 2.23370148 2.23383383 2.23403521 2.23427244]
```

```
Out[ ]:
```

```
no_outlier    192
Name: outlier_tukey, dtype: int64
```

Atipicidad por algoritmo Local Outlier Factor

Se realiza el análisis con las variables numéricas disponibles, a saber:

[DIAS_INVITACION_APER, DIAS_APERTURA_ADJU, CANTIDAD_ADJUDICADA, MONTO_ADJUDICADO]

Para detectar atipicidades se aplica el algoritmo de Local Outlier Factor, implementado en la librería de sklearn. Más información del algoritmo en https://en.wikipedia.org/wiki/Local_outlier_factor

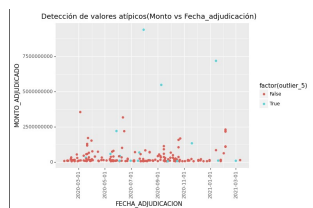
Se escalan los datos por las diferencias de magnitud entre datos de días y montos.

Se estima los puntajes basados en la distancia con 5 vecinos y un valor de densidad mayor a 2 para clasificar un procedimiento como atípico, además se estima un 10% de valores atípicos

```
Out[ ]:
```

```
19
```

Con el algoritmo LOF se ubican 19 procedimientos con un score mayor a 2 (mayor atipicidad)



```
Out[ ]:
```

```
<ggplot: (8748381097385)>
```

Un gráfico interactivo y más detallado lo puede visualizar en el siguiente enlace. Puede ver el detalle dándole click a cada punto, valores atípico en el extremo derecho. [Aquí](#)

Los procedimientos de contratación que conforme a este algoritmo se consideran atípicos, con un valor de densidad superior a 2 son los siguientes:

```
Out[ ]:
```

INSTITUCION LICITANTE	ANIO	NUMERO_PROCEDIMIENTO	OBJETO_CONTRACTUAL	FECHA_APERTURA	FECHA_INVITACION_APER	FECHA_ADJUDICACION	DIAS_INVITACION_APER	DIAS_APERTURA_ADJU	FECHA_COMUNICO	LIMITE_LEGAL	MODALIDAD_CONTRATACION	PRESUPUESTO_ESTIMADO	EXCEPCION_CONTRATACION_DIRECTA	LINEA	ESTADO_LINEA	BIEN_SERVICIO	CEDULA_ADJUDICADO	ADJUDICADO	CANTIDAD_REQUERIDA	CANTIDAD_ADJUDICADA	INESTIMABLE	MONTO_ADJUDICADO	SUBPARTIDA_GASTO
-----------------------	------	----------------------	--------------------	----------------	-----------------------	--------------------	----------------------	--------------------	----------------	--------------	------------------------	----------------------	--------------------------------	-------	--------------	---------------	-------------------	------------	--------------------	---------------------	-------------	------------------	------------------

8/5/2021 19:15

PROCEDIMIENTOS ATÍPICOS														PROCEDIMIENTOS ATÍPICOS													
INSTRUMENTOS CONTRACTUALES		FECHA DE EMISIÓN DEL INSTRUMENTO		FECHA DE EJECUCIÓN DEL INSTRUMENTO		FECHA DE VENCIMIENTO DEL INSTRUMENTO		FECHA DE EMISIÓN DEL INSTRUMENTO		FECHA DE EJECUCIÓN DEL INSTRUMENTO		FECHA DE VENCIMIENTO DEL INSTRUMENTO		INSTRUMENTOS CONTRACTUALES		FECHA DE EMISIÓN DEL INSTRUMENTO		FECHA DE EJECUCIÓN DEL INSTRUMENTO		FECHA DE VENCIMIENTO DEL INSTRUMENTO		FECHA DE EMISIÓN DEL INSTRUMENTO		FECHA DE EJECUCIÓN DEL INSTRUMENTO		FECHA DE VENCIMIENTO DEL INSTRUMENTO	
CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
132		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
138		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
144		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
144		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
162		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
164		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
168		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
177		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
178		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
179		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
180		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
181		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
188		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
189		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	

Un gráfico interactivo y más detallado lo puede visualizar en el siguiente enlace. Puede ver el detalle dándole click a cada punto, valores atipico en el extremo derecho. Aquí

Procedimientos que resultaron atípicos en ambos algoritmos de LOF y de Isolation Forest

La siguiente lista de procedimientos son de aquellos que resultaron atípicos usando ambos algoritmos.

PROCEDIMIENTOS ATÍPICOS														PROCEDIMIENTOS ATÍPICOS													
INSTRUMENTOS CONTRACTUALES		FECHA DE EMISIÓN DEL INSTRUMENTO		FECHA DE EJECUCIÓN DEL INSTRUMENTO		FECHA DE VENCIMIENTO DEL INSTRUMENTO		FECHA DE EMISIÓN DEL INSTRUMENTO		FECHA DE EJECUCIÓN DEL INSTRUMENTO		FECHA DE VENCIMIENTO DEL INSTRUMENTO		INSTRUMENTOS CONTRACTUALES		FECHA DE EMISIÓN DEL INSTRUMENTO		FECHA DE EJECUCIÓN DEL INSTRUMENTO		FECHA DE VENCIMIENTO DEL INSTRUMENTO		FECHA DE EMISIÓN DEL INSTRUMENTO		FECHA DE EJECUCIÓN DEL INSTRUMENTO		FECHA DE VENCIMIENTO DEL INSTRUMENTO	
CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
38		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
40		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
138		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
162		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
164		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
168		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
180		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
189		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	
190		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)		2020		2020		2020		2020		2020		2020	

Detalle en la adquisición de las pruebas PCR

Haciendo un detalle del costo de adquisición de las pruebas PCR tenemos los siguientes costos unitarios.

PROCEDIMIENTOS ATÍPICOS														PROCEDIMIENTOS ATÍPICOS													
INSTRUMENTOS CONTRACTUALES		FECHA DE EMISIÓN DEL INSTRUMENTO		FECHA DE EJECUCIÓN DEL INSTRUMENTO		FECHA DE VENCIMIENTO DEL INSTRUMENTO		FECHA DE EMISIÓN DEL INSTRUMENTO		FECHA DE EJECUCIÓN DEL INSTRUMENTO		FECHA DE VENCIMIENTO DEL INSTRUMENTO		INSTRUMENTOS CONTRACTUALES													
ORDEN	FECHA DE EMISIÓN DEL INSTRUMENTO	FECHA DE EJECUCIÓN DEL INSTRUMENTO	FECHA DE VENCIMIENTO DEL INSTRUMENTO	FECHA DE EMISIÓN DEL INSTRUMENTO	FECHA DE EJECUCIÓN DEL INSTRUMENTO	FECHA DE VENCIMIENTO DEL INSTRUMENTO	FECHA DE EMISIÓN DEL INSTRUMENTO	FECHA DE EJECUCIÓN DEL INSTRUMENTO	FECHA DE VENCIMIENTO DEL INSTRUMENTO	FECHA DE EMISIÓN DEL INSTRUMENTO	FECHA DE EJECUCIÓN DEL INSTRUMENTO	FECHA DE VENCIMIENTO DEL INSTRUMENTO	FECHA DE EMISIÓN DEL INSTRUMENTO	FECHA DE EJECUCIÓN DEL INSTRUMENTO	FECHA DE VENCIMIENTO DEL INSTRUMENTO												
1	CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)	4000042147-04-2020CD-000225-5101	2020	2020CD-000225-5101	SISTEMA DE PRUEBAS PCR	2020-09-25	2020-09-24	2020-09-30	2020-10-01	CONTRATACION ADMINISTRATIVA (EXCLUYE OBRA PUBL...)	ORDINARIA	296113950	CONTRATACION DIRECTA AUTORIZADA POR LA CGR POR...	1	ADJUDICADA EN FIRME	SISTEMA DE PRUEBAS PCR	310118223	TECNO DIAGNOSTICA SOCIEDAD ANONIMA	5000								
2	CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)	4000042147-04-2020CD-000238-5101	2020	2020CD-000238-5101	REACTIVO PARA DIAGNOSTICO MOLECULAR (PCR) DE C...	2020-10-07	2020-10-06	2020-10-23	2020-10-23	CONTRATACION ADMINISTRATIVA (EXCLUYE OBRA PUBL...)	ORDINARIA	1664749200	CONTRATACION DIRECTA AUTORIZADA	1	ADJUDICADA EN FIRME	REACTIVO PARA DIAGNOSTICO MOLECULAR (PCR) DE C...	3102098890	BIOCIENTIFICA INTERNACIONAL S D R L LIMITADA	22500								
3	CAJA COSTARRICENSE DE SEGURO SOCIAL (CCSS)	4000042147-04-2020CD-000243-5101	2020	2020CD-000243-5101	REACTIVO PARA DIAGNOSTICO	2020-10-07	2020-10-06	2020-10-23	2020-10-23	CONTRATACION ADMINISTRATIVA (EXCLUYE OBRA PUBL...)	ORDINARIA	1664749200	CONTRATACION DIRECTA AUTORIZADA	1	ADJUDICADA EN FIRME	REACTIVO PARA DIAGNOSTICO	3102098890	BIOCIENTIFICA INTERNACIONAL S D R L	11500								

Edad	Valores unitarios pruebas PCR
0.0	4.0
0.2	0.0
0.4	0.0
0.6	0.0
0.8	0.0
1.0	0.0
1.2	1.0
1.4	0.0

Para acceder a los datos usados en el presente análisis puede descargarlos de [aquí](#)