

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO

Econometría Aplicada I

TAREA 1

MARCO ANTONIO RAMOS JUÁREZ

142244

Índice

Nota	2
I	3
(a)	3
(b)	6
(c)	6
(d)	7
II	7
(a)	7
(b)	8
(c)	9
(d)	10
(e)	10
(f)	11
III	13
(a)	14
(b)	14
(c)	15
(d)	17
(e)	17
IV	18

Nota

El *script* de esta tarea se realizó en R Markdown y está diseñado para que se ejecute sin ninguna modificación (ni siquiera de los ficheros donde están las bases de datos pues se descargan directamente de un repositorio online). Asimismo, todos los cálculos (como varianzas, promedios, intervalos de confianza y estadísticos) se hicieron de manera “manual” y se explica el proceso en el script. Finalmente, cabe destacar que en cuanto al *bootstrap* y demás procesos aleatorios, cada vez que se ejecute el documento se van a generar diferentes datos, por lo que las cifras y gráficas serán diferentes en caso de que se ejecute de nuevo.

I

(a)

Utilizando la base BaseCOVIDm se calculo la tasa de positividad (POS) de los países de la muestra. En la siguiente cuadro se muestran la tasas.

Cuadro 1. Tasa de positividad

País	Confirmados	Pruebas	Tasa de positividad (POS)
Mexico	408449	809561	0.5045315
Ecuador	83193	176208	0.4721295
Panama	63269	139808	0.4525421
Qatar	110153	326596	0.3372760
Chile	351575	1059198	0.3319257
Argentina	178996	560862	0.3191445
Bangladesh	232194	761911	0.3047521
Costa Rica	16800	71428	0.2352019
Nigeria	42208	182749	0.2309616
Pakistan	276288	1293895	0.2135320
Iran	298909	1611999	0.1854275
South Africa	471123	2830635	0.1664372
India	1581963	11888829	0.1330630
Indonesia	104432	841027	0.1241720
El Salvador	15841	155235	0.1020453
Spain	282641	2912633	0.0970397
Ghana	35142	382752	0.0918140
Nepal	19273	232685	0.0828287
Belarus	67518	850992	0.0793403
Turkey	228924	3125944	0.0732336
Kazakhstan	87664	1221686	0.0717566
Ethiopia	15810	221958	0.0712297
Switzerland	34802	521036	0.0667938
Israel	68299	1047818	0.0651821
Philippines	85486	1327782	0.0643826
Romania	48235	773367	0.0623701
Ireland	25942	418510	0.0619866
Italy	246776	3998585	0.0617158

País	Confirmados	Pruebas	Tasa de positividad (POS)
Paraguay	4866	79905	0.0608973
Netherlands	53895	962965	0.0559678
Belgium	67335	1211031	0.0556014
Bahrain	40311	808276	0.0498728
United Kingdom	303063	6108983	0.0496094
Russia	827509	18472476	0.0447969
Japan	33382	765346	0.0436169
Iceland	1861	46310	0.0401857
Germany	208546	5364347	0.0388763
Canada	117357	3863930	0.0303724
Tunisia	1488	59889	0.0248460
Slovenia	2115	86008	0.0245907
Norway	9172	414789	0.0221124
Uruguay	1237	65309	0.0189407
Morocco	22213	1188801	0.0186852
Denmark	13868	980586	0.0141426
Slovakia	2245	172556	0.0130103
Greece	4336	342673	0.0126535
Malta	720	75872	0.0094897
Latvia	1224	130878	0.0093522
Australia	16298	2701600	0.0060327
New Zealand	1560	308258	0.0050607
Vietnam	459	261004	0.0017586
Honduras	40944	NA	NA
Jordan	1187	NA	NA
Jamaica	856	NA	NA
Andorra	918	NA	NA
Kosovo	7846	NA	NA
Peru	400683	NA	NA
Dominica	18	NA	NA
Burkina Faso	1105	NA	NA
North Macedonia	10503	NA	NA
Botswana	804	NA	NA
Sudan	11496	NA	NA
Guyana	398	NA	NA

País	Confirmados	Pruebas	Tasa de positividad (POS)
San Marino	699	NA	NA
Namibia	1986	NA	NA
Cameroon	17255	NA	NA
Algeria	29229	NA	NA
China	87680	NA	NA
Eritrea	265	NA	NA
Sierra Leone	1803	NA	NA
Mongolia	291	NA	NA
Saint Vincent and the Grenadines	52	NA	NA
Trinidad and Tobago	156	NA	NA
Azerbaijan	31221	NA	NA
Albania	5105	NA	NA
Guatemala	47605	NA	NA
Somalia	3212	NA	NA
Eswatini	2551	NA	NA
Uzbekistan	22585	NA	NA
Equatorial Guinea	3071	NA	NA
Armenia	37937	NA	NA
Iraq	118300	NA	NA
Nicaragua	3672	NA	NA
Brazil	2552265	NA	NA
Uganda	1140	NA	NA
Afghanistan	36471	NA	NA
Antigua and Barbuda	91	NA	NA
Gambia	326	NA	NA
Liechtenstein	88	NA	NA
Haiti	7378	NA	NA
Tajikistan	7320	NA	NA
Liberia	1179	NA	NA
Bosnia and Herzegovina	11127	NA	NA
Niger	1132	NA	NA
Lesotho	576	NA	NA
Burundi	387	NA	NA
Montenegro	3016	NA	NA
Saint Lucia	24	NA	NA

$H_A: \mu - .024 \neq 0$

El estadístico t que resulta es de 2.6022561 con un valor ρ de 0.010668. De esta manera, los valores t y ρ indican que podemos rechazar la hipótesis nula con más del 99 % de confianza.

(d)

El intervalo de confianza relevante para nuestra prueba de hipótesis ya no incluiría el valor de .024 a partir de un nivel de confianza α de 0.989332. Donde 0.989332 es igual a $1 - 0.010668$ es decir a $1 - \rho$.

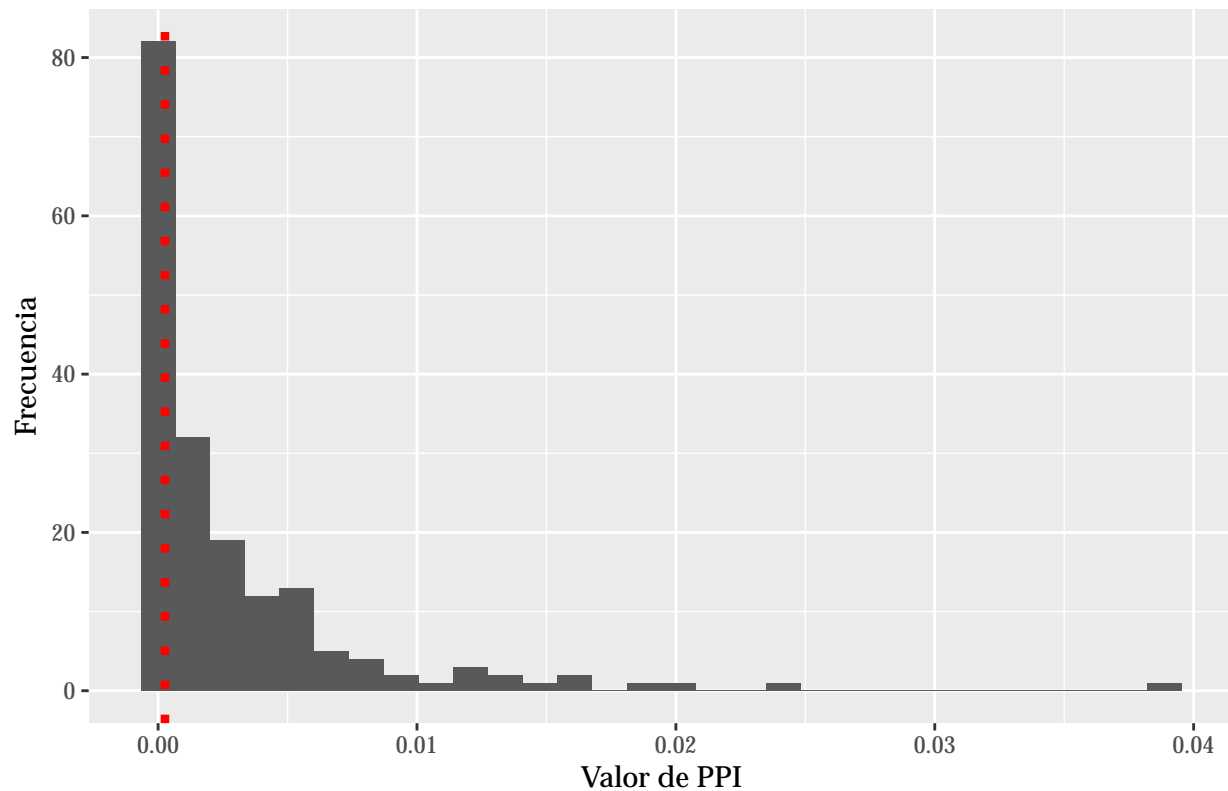
II

Ahora estamos interesados en conocer el porcentaje de la población total de cada país que ha contraído el virus (PPI), por lo que calculamos esa variable con el cociente $\frac{Confirmed}{Population}$.

(a)

Para poder analizar mejor la información, creamos un histograma del valor de PPI de la base BaseCOVIDp. Asimismo, **señalamos con una línea roja en el histograma al valor del primer cuartil (0.0002654)**. La información se muestra en el gráfico siguiente.

Grafica 1: Frecuencia de la variable Población Infectada en BaseCOVIDp

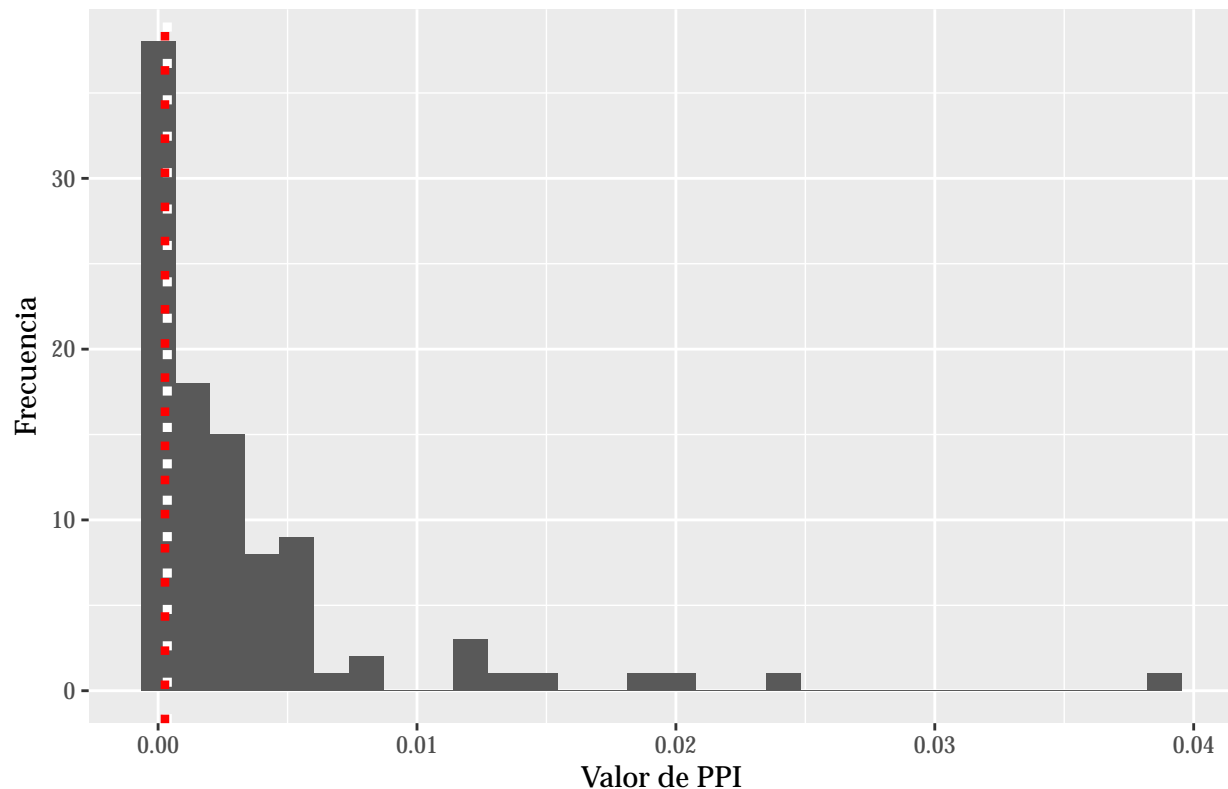


A partir de la gráfica podemos observar que el valor de PPI alrededor de los países dentro de BaseCOVIDp está muy concentrado hacia los valores pequeños.

(b)

De manera similar, realizamos el mismo ejercicio para la base de datos BaseCOVIDm, que recordemos se refiere a una muestra de 100 países. Sin embargo, ahora **usaremos una línea de color blanco para indicar el primer cuartil de BaseCOVIDm mientras que seguiremos indicando con una línea roja al primer cuartil de la BaseCOVIDp** (la misma del ejercicio anterior) con el fin de compararlas.

Grafica 2: Frecuencia de la variable Población Infectada en BaseCOVIDm



Notemos que aunque ambos cuartiles están muy cerca, el valor del primer cuartil de BaseCOVIDm es ligeramente mayor al de BaseCOVIDp por 0.0000898 unidades.

(c)

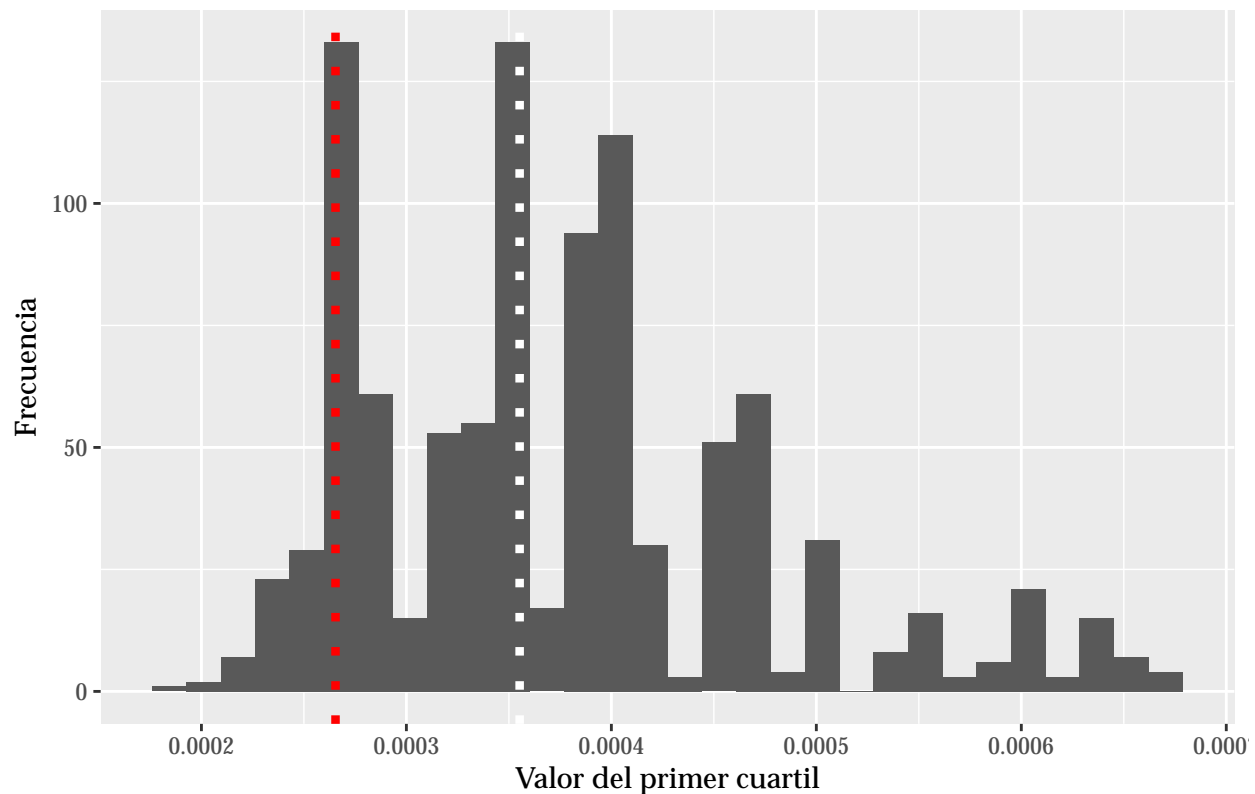
El segundo histograma está incluido en el primero, y debido a que es de un $n=100$ (más de la mitad del n poblacional) se espera ex ante que sea un histograma muy similar. A pesar de ello, el simple hecho de que BaseCOVIDm tenga menos datos hace que la dispersión de los valores sea mayor y que el pico más alto del histograma sea en un punto más bajo, debido a una menor frecuencia posible.

Como se mencionó en el inciso anterior, cabe notar que el primer cuartil tiene un valor diferente y que el PPI muestral sea más grande que el del supuesto poblacional. Esta situación indica que la muestra cuenta con una proporción mayor de países con un PPI mediano o alto. Finalmente, es interesante que el cuarto cuartil es el mismo para ambos histogramas lo que indica que comparten el país del PPI máximo.

(d)

En esta parte procederemos a utilizar el método *Bootstrap* que consiste en generar J submuestras de tamaño L . En este caso se realizaron 1000 iteraciones de tamaño 100. En cada iteración se calculo el valor del primer cuartil. En la siguiente gráfica se muestra un histograma con los valores de los 1000 primeros cuartiles calculados. **Notese que se dejaron las líneas anteriores (blanca para el primer cuartil muestral y roja para el poblacional).**

Grafica 3: Bootstrap para el primer cuartil

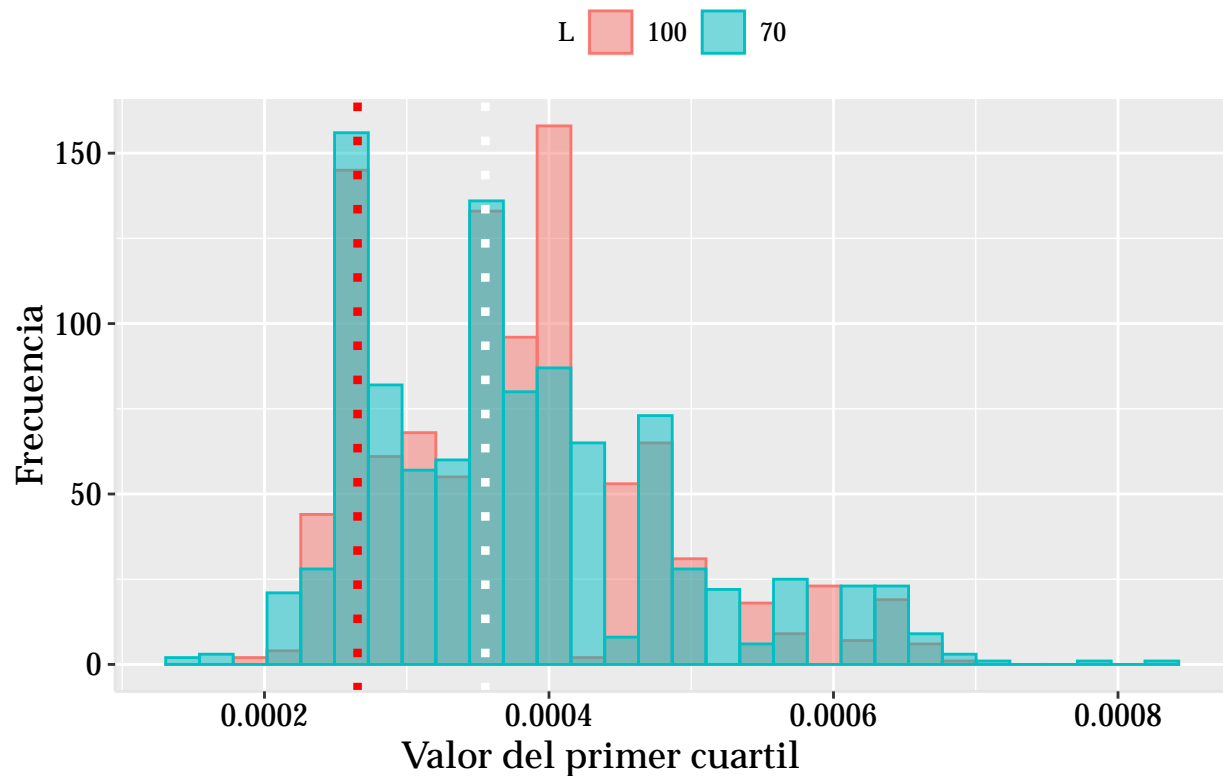


Lo primero que notamos es que la distribución está sesgada hacia la derecha. Asimismo, en esta gráfica es más fácil notar la diferencia entre los cuartiles de BaseCOVIDm y BaseCOVIDp.

(e)

Ahora realizaremos el mismo ejercicio solo que en lugar de un L de tamaño 100 usaremos uno de 70. De esta manera, en la siguiente gráfica podemos comparar ambos ejercicios. Asimismo, dejamos las líneas blanca muestral y roja poblacional para poder hacer un mejor análisis visual.

Grafica 4: Comparación entre Bootraps

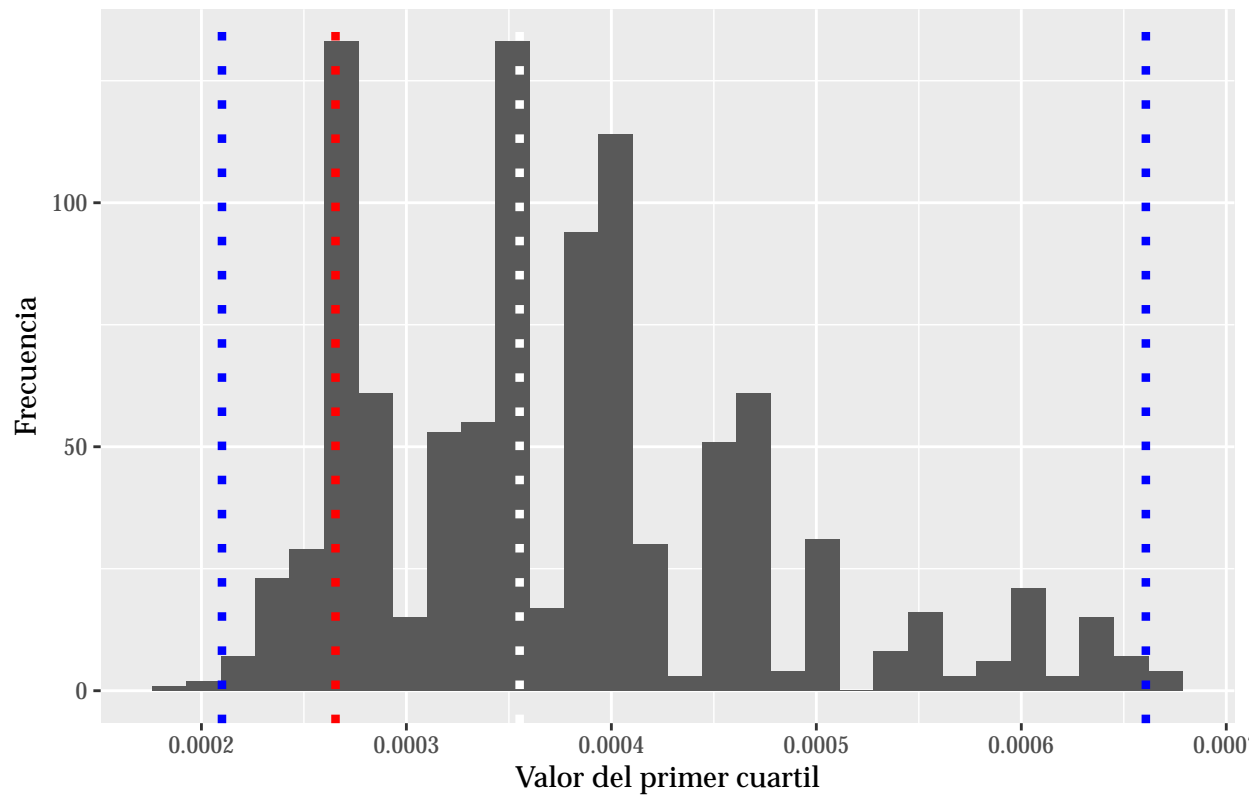


Sin duda siguen la misma distribución solo que el histograma de $L=70$ tiene mayor dispersión mientras que el de $L=100$ se concentra más en la línea blanca del primer cuartil de BaseCOVIDm.

(f)

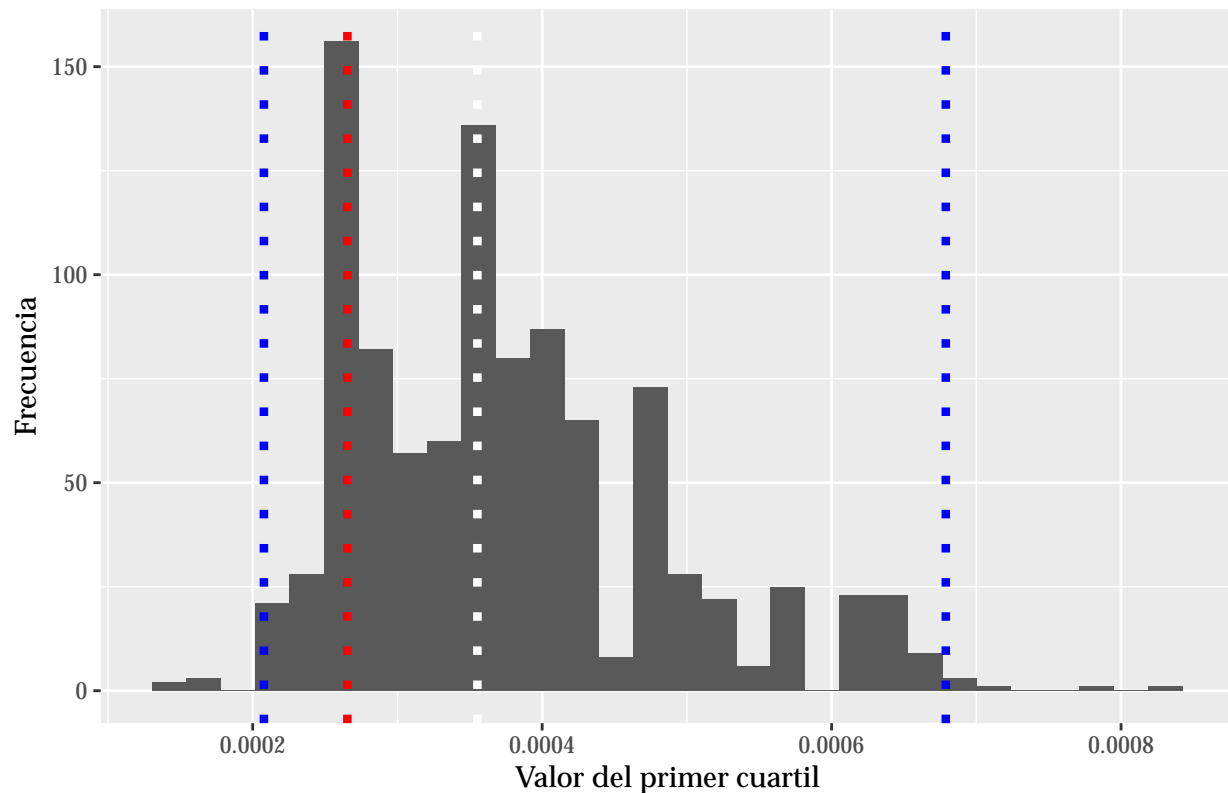
Con base en el ejercicio anterior podemos consturir un intervalo de confianza. Lo fácil de usar este método esque no tienes que asumir supuestos sobre la distribución porqué ya se simulo una. En este sentio, el intervalo de confianza al 99% solo corresponde al valor 5 y al valor 995 de los 1000 primeros cuartiles medios simulados. Asimismo, dependiendo la muestra que usemos, los intervalos de confianza estarán más o menos acotados.

A continuación se muestra el ejercicio para el caso de la $L=100$, se grafican con azul el intervalo de confianza, y como en los ejercicios previos, con blanco el primer cuartil muestral y con rojo el poblacional.

Grafica 5: Intervalo de confianza al 99% proveniente de un Bootstrap con $L=100$ 

En este caso, el intervalo de confianza es $0.00021 \leq q1 \leq 0.0006608$. A continuación se muestra el ejercicio para el caso de la $L=70$.

Grafica 6: Intervalo de confianza al 99% proveniente de un Bootstrap con L=70



En este caso, el intervalo de confianza es $0.0002077 \leq q1 \leq 0.0006789$. Mientras que la longitud del primer intervalo es de 0.0004507 la del segundo es de 0.0004712. Por lo que parece que es mejor usar el primero.

En conclusión, el intervalo de confianza proveniente del *Bootstrap* con $L=100$ es mejor debido a que es un intervalo más corto. Esto se debe a que su histograma presentaba menor varianza y estaba más concentrado en el primer cuartil de la muestra.

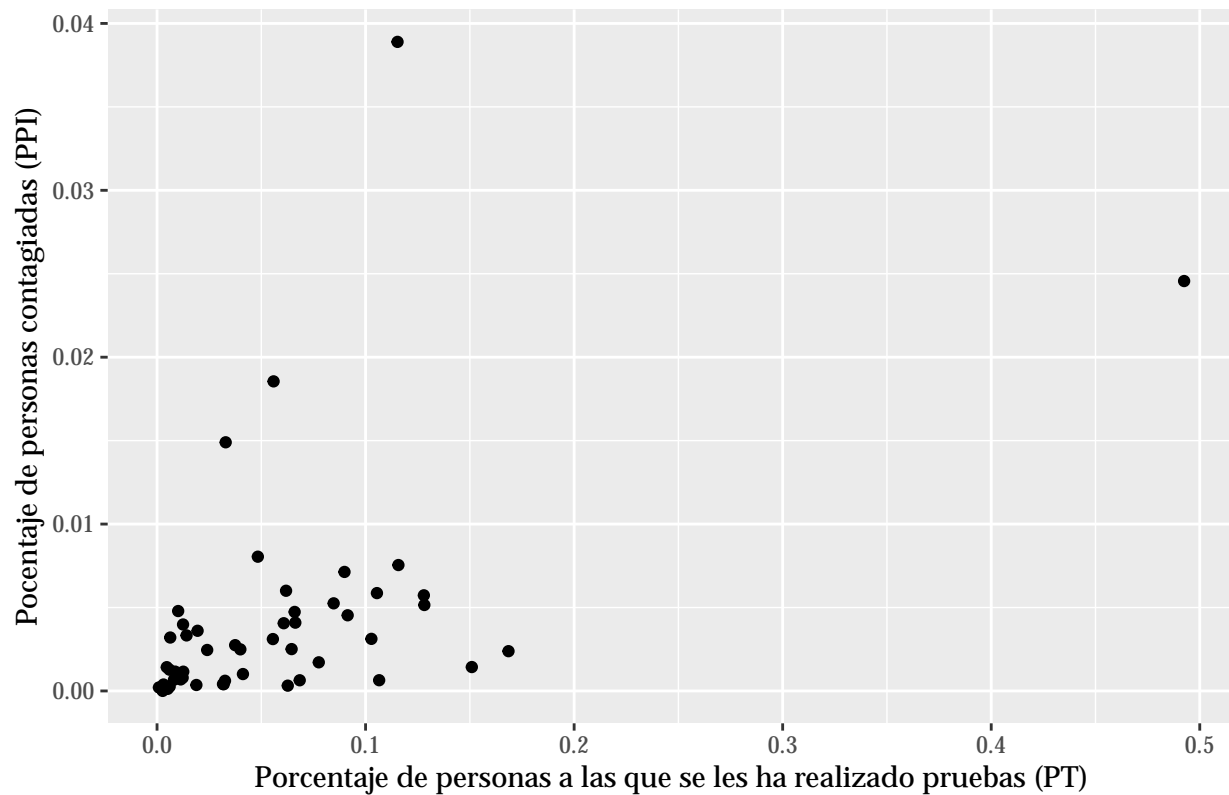
III

A lo largo de los meses que ha durado la pandemia han surgido diversas preguntas. Una de ellas se enfoca en la relación que existe entre el porcentaje de personas a las que se les han realizado pruebas ($PT = \frac{Tests}{Population}$) y el porcentaje de personas contagiadas ($PPI = \frac{Confirmed}{Population}$). Por lo que el primer paso de este ejercicio es crear dichas variables en nuestras bases de datos.

(a)

A continuación se presenta un diagrama de dispersión para visualizar la relación entre PT y PPI con respecto a la base BaseCOVIDm.

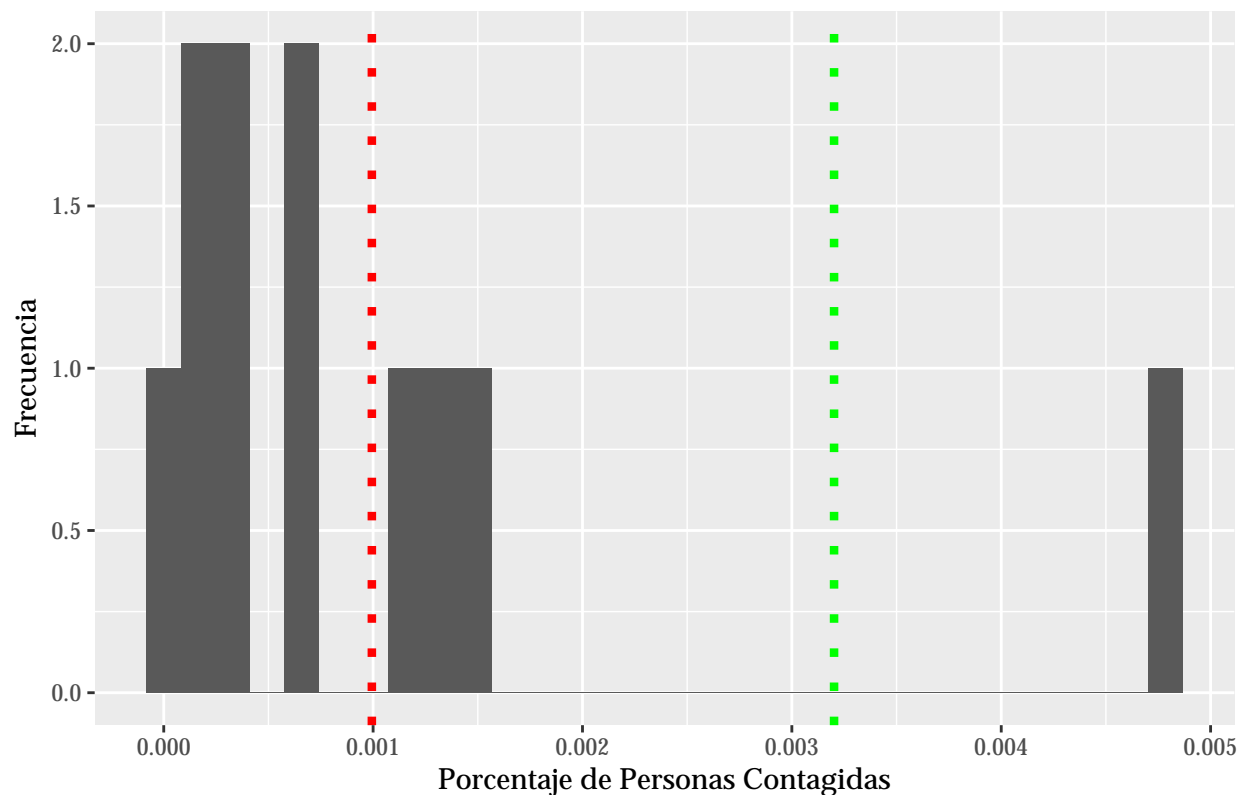
Grafica 7: Diagrama de dispersión para PT y PPI



(b)

En esta sección propondremos una vecindad de radio .005 para el valor PT de México para analizar su valor PT con respecto a sus vecinos. A continuación se realiza dicha gráfica, asimismo, dentro de ella se indica con una línea de color verde el valor PT de México (0.0063457) y con una línea roja el valor promedio PT de la vecindad, sin contar a México(0.000994).

Grafica 8: Histograma para PPI de la vecindad de Mexico

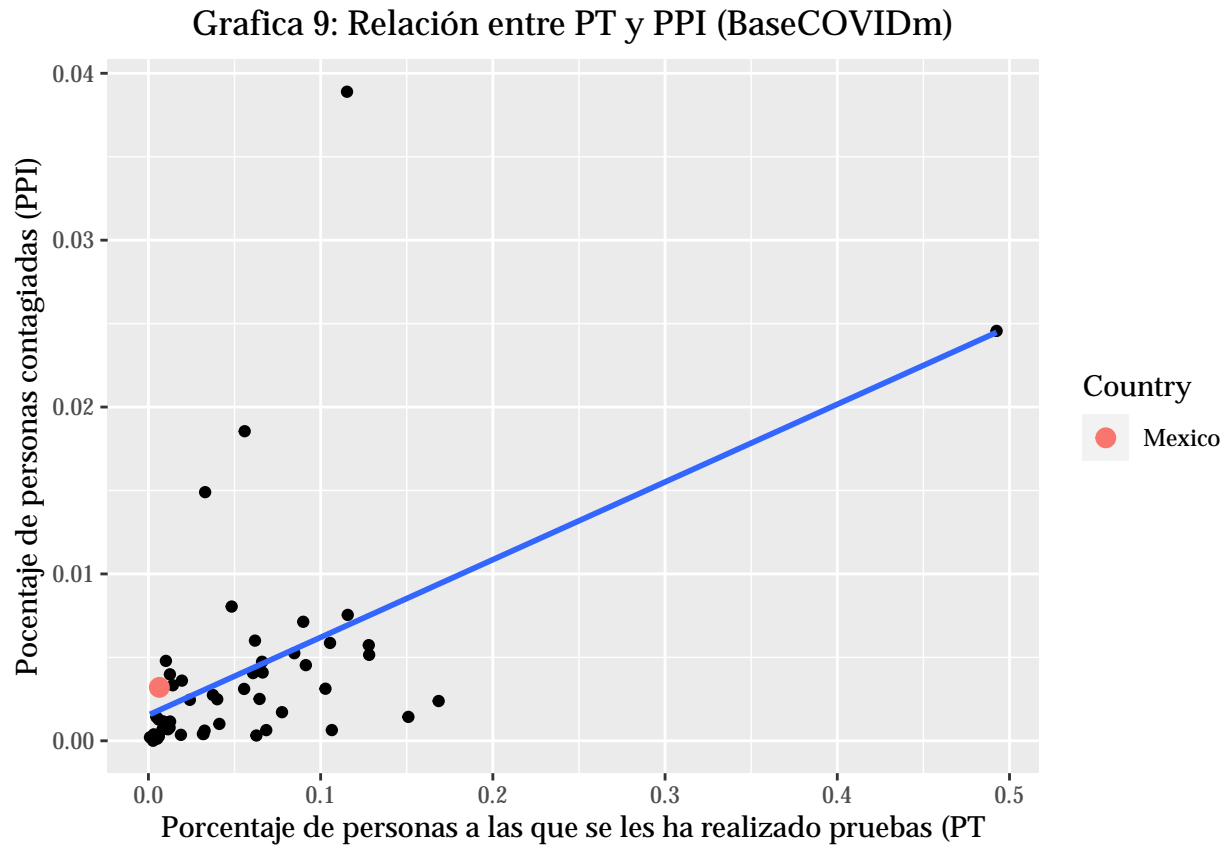


A partir de la gráfica podemos concluir que el PPI de México se encuentra por arriba de la media de PPI de la vecindad. En este sentido, parece que México tiene un PPI muy alto como para tener un PT tan bajo.

(c)

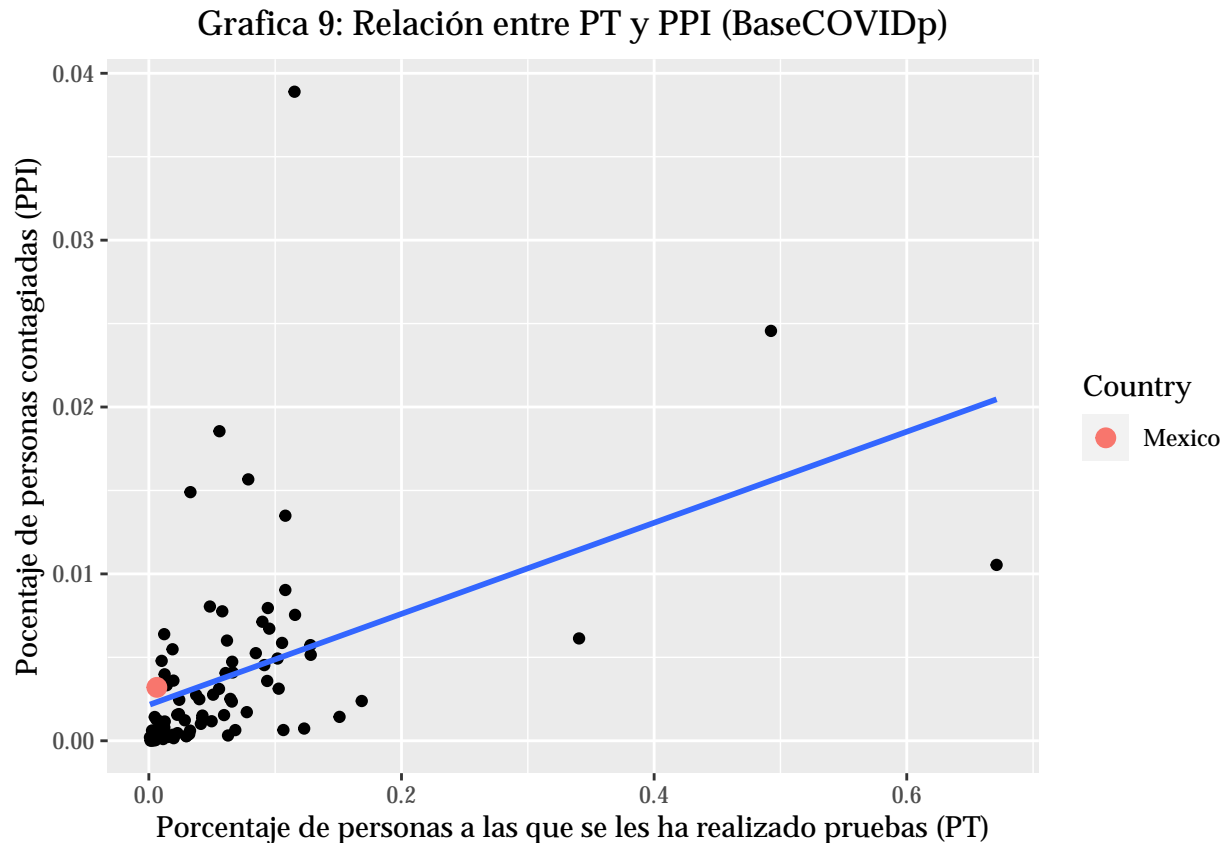
De manera alternativa, procederemos a estimar una regresión con mínimos cuadrados ordinarios de la forma $PPI_i = B_0 + B_1 PT_i + U_i$.

En la siguiente gráfica de dispersión mostramos nuevamente la relación entre PT y PPI sin embargo ahora mostraremos también la regresión estimada encima y enfatizaremos al punto que representa a México.



En la gráfica podemos observar que el punto de México se encuentra por arriba de la regresión. Esto lo podemos interpretar como que México tiene un PPI por arriba del PP esperado de acuerdo a su nivel de PT.

De manera similar podemos hacer el ejercicio para la BaseCOVIDp.



(d)

Entre las similitudes están que ambas estrategias se basan en comparar la situación de un país con respecto a la relación entre dos variables. Entre las diferencias están que el primer enfoque es más atómico pues baja la lupa hasta el país de interés y su vecindad y a partir de ahí se compara el valor del país con el valor promedio de la vecindad; por el contrario, el segundo enfoque es más general pues la regresión aprovecha la información de todos los países para generar una línea que abarca todo el plano.

(e)

La diferencia radica en que tendríamos un n mayor por lo que habría mayor certeza en los resultados. La BaseCOVIDp es de casi la totalidad de la población (considerando que existen 195 países actualmente). En este sentido, tendríamos más información para hacer una regresión y más información para comparar el valor de una variable de un país con respecto a la media de su vecindad.

IV

Los países que no aparecen en la BaseCOVIDp van a estar subrepresentados en los resultados que obtengamos a partir de la base muestral. Probablemente sea porque no cuentan con la infraestructura institucional suficiente para recolectar y reportar datos o incluso como para entablar relaciones internacionales multilaterales o mantener sus obligaciones ante las instituciones encargadas de la publicación de datos, por ejemplo la OMS. En ese sentido nuestros parámetros poblacionales siguen siendo desconocidos y aquellos que podamos calcular con base en BaseCOVIDp serían muestrales (aunque muy valiosos pues la n es casi el poblacional). En este sentido, nuestros estimadores serían muy optimistas pues estarían ligeramente sesgados en favor de los países “cumplidores” (es decir, aquellos que tienen la **capacidad** e interés en generar datos y reportarlos). Asimismo, esto quiere decir que nuestros estimadores tienen una varianza mayor a la que pensábamos pues nuestra muestra poblacional (BaseCOVIDm) en realidad es una muestra de una muestra.