



UNIVERSIDAD NACIONAL DE COLOMBIA

ESPECIALIZACIÓN EN CIENCIAS ESTADÍSTICA

DEPARTAMENTO DE ESTADÍSTICA

FACULTAD DE CIENCIAS

— MÉTODOS MULTIVARIADOS APLICADOS —

*Inferencia Estadística Sobre Parámetros de Una
Población p - Variada
tarea IV*

Integrantes:

Hugo Hernán Rodríguez Mesa C.C. 1022952094

Medellín, Colombia

Medellin, Marzo 25 de 2024

Contents

1	Introducción	1
1.1	Aclaración sobre la Temporalidad de los Datos	1
2	Carga de Datos	1
3	Formación de Subconjuntos de Datos	2
4	Pruebas de hipótesis del vector de medias por decenios	2
4.1	Resultados para el primer decenio (2010-2019):	2
4.2	Resultados para el segundo decenio (2000-2009)	3
5	Pruebas de hipótesis de igual de vectores de medias	4
6	Conclusiones	5
	Bibliografía	5

1 Introducción

En este trabajo, analizaremos el indicador “ingreso nacional por adulto” utilizando datos de la [World Inequality Database](#) para varios países del Caribe. Nos enfocaremos en los siguientes países: Bahamas, Barbados, Cuba, Haití, Jamaica, Puerto Rico, República Dominicana, Santa Lucía, San Vicente y las Granadinas, y Trinidad y Tobago.

1.1 Aclaración sobre la Temporalidad de los Datos

Es importante tener en cuenta que los datos utilizados en este análisis corresponden a los años anteriores a la pandemia de COVID-19. Dado que la pandemia ha tenido un impacto significativo en la economía global y en la distribución de ingresos, es posible que las tendencias observadas en estos datos no reflejen la situación actual.

Durante los años para los cuales se disponen datos (1999-2019), no se consideran los efectos económicos directos de la pandemia. Sin embargo, es importante reconocer que pueden haber ocurrido cambios importantes en la distribución de ingresos y en la economía en general desde entonces.

Se recomienda interpretar estos resultados en el contexto de la temporalidad de los datos y considerar el posible impacto de la pandemia en las tendencias de ingresos y desigualdad económica en los países analizados.

2 Carga de Datos

Comenzamos cargando los datos y realizando las transformaciones necesarias para su análisis. A continuación una muestra de la data final.

Table 1: Head Datos paises caribe 01

Percentile	Year	Mundo	Barbados	Bahamas	Cuba	Haití	Jamaica
p0p100	1999	12396	15447	43897	11041	5019	12173
p0p100	2000	12801	16016	44185	11642	4922	12116
p0p100	2001	12843	15378	43956	11985	4804	12013
p0p100	2002	12967	15359	43446	12070	4760	11884
p0p100	2003	13180	15563	41931	12339	4737	12129
p0p100	2004	13700	15498	41514	12907	4617	12156

Table 2: Head Datos Paises caribe 02

Percentile	Year	Mundo	Puerto Rico	República Dominicana	Santa Lucía	Trinidad y Tobago
p0p100	1999	12396	21411		12343	15174
p0p100	2000	12801	21040		12600	14819
p0p100	2001	12843	21205		12607	13824
p0p100	2002	12967	21406		12857	13770
p0p100	2003	13180	22368		12087	13791
p0p100	2004	13700	23265		11936	14245

3 Formación de Subconjuntos de Datos

Dado que tenemos datos disponibles desde 1999 hasta 2019, formaremos dos subconjuntos correspondientes a los últimos dos decenios:

Table 3: Head Datos Decada 01

Percentile	Year	Mundo	Bahamas	Barbados	Cuba	Haití	Jamaica	Puerto Rico	República Dominicana	Santa Lucía	San Vicente y las Gr
p0p100	2010	15245	37299	16086	17574	4506	10968	19938		15968	14788
p0p100	2011	15662	36867	16422	18101	4597	10994	19976		16111	15770
p0p100	2012	15921	37384	15067	18571	4568	10981	20523		16190	15213
p0p100	2013	16217	35148	15217	19025	4600	10771	20522		16444	14407
p0p100	2014	16596	34317	13029	19083	4592	10665	20363		17204	13662
p0p100	2015	16908	34134	15114	19946	4591	10410	20379		18199	12809
p0p100	2016	17193	33095	14174	19917	4588	10251	20188		19040	13547
p0p100	2017	17605	33978	14168	20242	4596	10368	19628		19436	14008
p0p100	2018	17941	33121	13954	20712	4532	10287	19455		20491	14211
p0p100	2019	18108	33291	13714	20662	4338	10317	19197		21103	13983

Table 4: Head Datos Decada 02

Percentile	Year	Mundo	Bahamas	Barbados	Cuba	Haití	Jamaica	Puerto Rico	República Dominicana	Santa Lucía	San Vicente y las Gr
p0p100	2000	12801	44185	16016	11642	4922	12116	21040		12600	14819
p0p100	2001	12843	43956	15378	11985	4804	12013	21205		12607	13824
p0p100	2002	12967	43446	15359	12070	4760	11884	21406		12857	13770
p0p100	2003	13180	41931	15563	12339	4737	12129	22368		12087	13791
p0p100	2004	13700	41514	15498	12907	4617	12156	23265		11936	14245
p0p100	2005	14079	42134	16073	14323	4598	12015	22890		13189	13792
p0p100	2006	14620	41964	16240	16088	4638	12253	22310		14189	14922

p0p100	2007	15092	41882	17116	17010	4695	12200	21846	14917	14720
p0p100	2008	15237	40690	17556	17453	4738	12051	20997	15298	15172
p0p100	2009	14753	37778	16218	17260	4742	11103	20307	15107	14814

4 Pruebas de hipótesis del vector de medias por decenios

Al realizar pruebas de normalidad multivariada, estamos evaluando si los datos siguen una distribución multivariada normal, lo cual es un supuesto necesario para el análisis de varianza. Estas pruebas nos permiten verificar la validez de este supuesto antes de realizar un análisis.

4.1 Resultados para el primer decenio (2010-2019):

Table 5: Normalidad Multivariada

multivariateNormality.Test	multivariateNormality.H	multivariateNormality.p.value	multivariateNormality.MVN
Royston	6.093884	0.0742902	YES

Table 6: Normalidad Univariada

univariateNormality.Test	univariateNormality.Variable	univariateNormality.Statistic	univariateNormality.p.value	univariateNormality.Normality
Shapiro-Wilk	Barbados	0.9594	0.7789	YES
Shapiro-Wilk	Bahamas	0.8482	0.0552	YES
Shapiro-Wilk	Cuba	0.9403	0.5566	YES
Shapiro-Wilk	Haití	0.6554	0.0003	NO
Shapiro-Wilk	Jamaica	0.8486	0.0559	YES
Shapiro-Wilk	Puerto Rico	0.9172	0.3341	YES
Shapiro-Wilk	República Dominicana	0.8961	0.1983	YES
Shapiro-Wilk	Santa Lucía	0.9842	0.9836	YES
Shapiro-Wilk	Trinidad y Tobago	0.9541	0.7174	YES

La prueba de Royston no rechaza la hipótesis de normalidad multivariada ($p\text{-value} = 0.0743 > 0.05$), lo que sugiere que los datos podrían seguir una distribución multivariada normal.

Para las pruebas de normalidad univariada de Shapiro-Wilk, la mayoría de las variables no rechazan la hipótesis de normalidad a un nivel de significancia del 5%, ya que sus $p\text{-values}$ son mayores que 0.05. Sin embargo, la variable “Haití” tiene un $p\text{-value}$ muy bajo (0.0003), lo que sugiere que no sigue una distribución normal.

4.2 Resultados para el segundo decenio (2000-2009)

Table 7: Normalidad Multivariada

multivariateNormality.Test	multivariateNormality.H	multivariateNormality.p.value	multivariateNormality.MVN
Royston	14.79117	0.0103569	NO

Table 8: Normalidad Univariada

univariateNormality.Test	univariateNormality.Variable	univariateNormality.Statistic	univariateNormality.p.value	univariateNormality.Normality
Shapiro-Wilk	Barbados	0.8752	0.1149	YES

Shapiro-Wilk	Bahamas	0.8860	0.1528	YES
Shapiro-Wilk	Cuba	0.8423	0.0470	NO
Shapiro-Wilk	Haití	0.9387	0.5388	YES
Shapiro-Wilk	Jamaica	0.6719	0.0004	NO
Shapiro-Wilk	Puerto Rico	0.9668	0.8596	YES
Shapiro-Wilk	República Dominicana	0.8854	0.1505	YES
Shapiro-Wilk	Santa Lucía	0.8411	0.0454	NO
Shapiro-Wilk	Trinidad y Tobago	0.9267	0.4160	YES

La prueba de Royston rechaza la hipótesis de normalidad multivariada ($p\text{-value} = 0.0104 < 0.05$), lo que sugiere que los datos no siguen una distribución multivariada normal.

Para las pruebas de normalidad univariada de Shapiro-Wilk, las variables “Cuba”, “Jamaica” y “Santa Lucía” tienen $p\text{-values}$ inferiores a 0.05, lo que indica que no siguen una distribución normal. El resto de las variables no rechazan la hipótesis de normalidad a un nivel de significancia del 5%.

5 Pruebas de hipótesis de igual de vectores de medias

Primero calculamos la matriz de varianzas-covarianzas para cada decenio usando la función COV.

Table 9: Matriz de covarianza decada 2009 2019

	Mundo	Bahamas	Barbados	Cuba	Haití	Jamaica	Puerto Rico	República Dominicana
Mundo	961957.38	-1540977.71	-792494.22	1045248.47	-31062.7556	-282163.689	-284553.156	1829816.16
Bahamas	-1540977.71	2959014.49	1417847.78	-1711288.36	27837.0889	512305.467	333567.822	-2864369.49
Barbados	-792494.22	1417847.78	1148889.39	-828218.83	19693.5556	227675.000	131462.389	-1372582.56
Cuba	1045248.47	-1711288.36	-828218.83	1170622.23	-26439.7111	-314607.956	-259410.856	1949255.02
Haití	-31062.76	27837.09	19693.56	-26439.71	6581.7333	6040.044	26426.867	-81363.87
Jamaica	-282163.69	512305.47	227675.00	-314607.96	6040.0444	95450.622	69573.467	-543493.91
Puerto Rico	-284553.16	333567.82	131462.39	-259410.86	26426.8667	69573.467	215394.322	-665798.71
República Dominicana	1829816.16	-2864369.49	-1372582.56	1949255.02	-81363.8667	-543493.911	-665798.711	3671302.71
Santa Lucía	-490113.20	1127269.98	552665.44	-596833.38	968.5111	201729.489	-5397.356	-863657.09
San Vicente y las Granadinas	843861.29	-1304591.51	-583989.06	904921.03	-33285.4667	-253116.089	-304532.011	1689530.96
Trinidad y Tobago	-304818.91	-322472.31	-236717.33	-116940.96	73061.1556	-30291.600	610572.911	-1001593.58

Table 10: Matriz de covarianza decada 1999 2009

	Mundo	Bahamas	Barbados	Cuba	Haití	Jamaica	Puerto Rico	República Dominicana
Mundo	914780.40	-1239147.4	600862.844	2266594.51	-45442.133	-47623.333	-65582.87	1085894.0
Bahamas	-1239147.44	3386617.1	-541205.667	-3305861.67	65349.000	412433.444	368143.33	-1463364.3
Barbados	600862.84	-541205.7	546067.789	1486861.23	-6583.078	12266.889	-184175.09	805502.6
Cuba	2266594.51	-3305861.7	1486861.233	5825227.57	-89569.411	-227668.000	-529464.31	2926137.0
Haití	-45442.13	65349.0	-6583.078	-89569.41	9237.656	-4441.667	-64653.60	-14515.3
Jamaica	-47623.33	412433.4	12266.889	-227668.00	-4441.667	108762.444	189060.11	-151790.6
Puerto Rico	-65582.87	368143.3	-184175.089	-529464.31	-64653.600	189060.111	874054.27	-599825.1
República Dominicana	1085893.96	-1463364.3	805502.567	2926137.01	-14515.300	-151790.556	-599825.09	1650401.1
Santa Lucía	360326.47	-434314.9	317285.744	942435.41	7078.900	-14959.222	-213897.73	522005.0
San Vicente y las Granadinas	977411.89	-1455774.2	594569.833	2441990.17	-52390.167	-66015.778	-53725.00	1146224.4
Trinidad y Tobago	3639619.11	-5285277.2	2235696.889	8842980.22	-221751.222	-163678.000	160463.33	3992587.8

Luego verificamos si ambas matrices son positivas definidas

Table 11: Matriz determinante decada 2009 2019

x

-4.144082e+21

Table 12: Matriz determinante 1999 2009

x
-5.230975e+24

Realizamos la prueba de Hotelling's T^2 ver (1931) usando la función `hotelling.test` ver (2023). Si alguna de las matrices no es positiva definida, indicamos que no podemos realizar la prueba de Hotelling's T^2 .

Hotelling's two sample T2-test

```
data: data_decade1[, -c(1, 2)] and data_decade2[, -c(1, 2)]
T.2 = 16.749, df1 = 11, df2 = 8, p-value = 0.0002446
alternative hypothesis: true location difference is not equal to c(0,0,0,0,0,0,0,0,0,0,0)
```

6 Conclusiones

Basándonos en los resultados de la prueba de Hotelling's T^2

- **Estadística de Prueba:** El valor de la estadística de prueba T^2 es de 16.749.
- **Grados de Libertad:** Los grados de libertad para esta prueba son $df1 = 11$ y $df2 = 8$.
- **Valor p:** El valor p obtenido es de 0.0002446.
- **Hipótesis Nula y Alternativa:** La hipótesis nula es que no hay diferencia en los vectores de medias entre los dos decenios. La hipótesis alternativa es que hay una diferencia en los vectores de medias entre los dos decenios.

Dado que el valor p es menor que cualquier nivel de significancia común (como 0.05), rechazamos la hipótesis nula. Esto sugiere que hay suficiente evidencia para concluir que hay diferencias significativas en los vectores de medias entre los dos decenios para las sub-regiones de la región asignada.

En resumen, los resultados de la prueba indican que hay cambios significativos en los ingresos nacionales por adulto entre los dos decenios analizados. Esto puede ser de interés para comprender mejor las tendencias económicas y sociales en la región durante ese período.

Bibliografía

- Hotelling, H. (1931). The generalization of student's ratio. *Annals of Mathematical Statistics*, **2**, 360–378.
- Luna, A. & Tran, L.N. (2023). *R documentation: Hotelling test*.