

**El turismo y la herramienta estadística
como elemento analítico**

Tourism and statistics as an analytical tool

Alfredo Ascanio¹

RESUMEN: Este artículo forma parte de un conjunto de propuestas relativas a la utilización de la estadística cuando se analizan los datos turísticos. La estadística descriptiva integrada a la estadística evaluativa permite realizar mejores interpretaciones de los hallazgos encontrados bien sea cuando se realizan encuestas o bien cuando se correlacionan informaciones.

PALABRAS CLAVES: Estadística. Turismo. Análisis. Datos.

ABSTRACT: This article is part of a set of proposals for the use of statistics when analyzing tourism data. Descriptive statistics with evaluative statistics integrated allows better interpretation of the findings either when surveys are conducted or when data are correlated.

KEY WORDS: Statistics. Tourism. Analysis. Data.

¹ Profesor titular jubilado del Dpto. de Ciencias Económicas y Administrativas de la Universidad Simón Bolívar, Venezuela. Economista con maestría y doctorado en Ciencia Política. Por más de 12 años fue funcionario internacional en cargos importantes en el Banco Interamericano de Desarrollo, en la Organización de Estados Americanos y en el Organismo Mundial de Turismo. Dirección postal: CCS: 90372, PO BOX 025323, MIAMI, FL, USA, 33102-5323.

INTRODUCCIÓN

Siempre se ha señalado que la estadística como herramienta analítica es fundamental para sacar mejor partido de los hechos observados. La estadística nos intenta decir cosas para comprobar hipótesis. Del conocimiento que podemos tener al relacionar datos cuantitativos bien sea con una recta de regresión para conocer la tendencia de un fenómeno, o bien las relaciones en una tabla de contingencia, no nos queda otra posibilidad para disminuir la incertidumbre al utilizar las muestras y comprobar o no nuestra hipótesis, que recurrir a los tests de significación, ante la imposibilidad manifiesta de analizar la población o el universo.

Recordemos que la evolución y la correlación de datos no nos ponen de manifiesto las causas que le dieron origen bien sea económica o psicológica (PIN, 1962, p. 159).

La tendencia del turismo receptivo

Es importante analizar las cuentas de llegadas de visitantes internacionales con varios modelos a los fines de obtener un valor promedio que es muy útil para posteriormente desde esa base hacer una adecuada valoración.

En seguida veremos que si la proyección lineal se logra aplicando logaritmos o bien recurriendo a una proyección no logarítmica, siempre el promedio obtenido puede ser muy parecido con un método o con otro.

Si la proyección se hace por ejemplo con una serie de 20 años, que es lo recomendable, y utilizando los logaritmos, obtendremos una tasa de incremento interanual muy útil con fines estadísticos (ASCANIO, 1996).

Los valores de la tendencia serán diferentes pues ello depende del número de años seleccionados y de sus valores. Cuando utilizamos la regresión entre los años aplicando logaritmos, veremos, como aparece en el cuadro no. 1 siguiente, que los datos de la tendencia pueden ser hasta un 50% más elevados o incluso apenas un 13 % ó 7% en más o en menos, aunque siempre la tendencia es que aparezcan valores mayores.

Tabla 1 - Tendencia lineal del turismo en Venezuela durante 20 años

AÑO	y	Log y	x	x ²	x Log y	Tendencia	Log de Tendencia
1	117	2,0682	-9	81	--18,6138	178	2,2504
2	144	2,1584	-8	64	--17,2672	188	2,2742
3	171	2,2329	-7	49	--15,6303	198	2,2967
4	285	2,4548	-6	36	--14,7288	209	2,3201
5	425	2,6284	-5	25	--13,1420	221	2,3444
6	436	2,6395	-4	16	--10,5580	233	2,3674
7	535	2,7284	-3	9	-8,1852	246	2,3909
8	653	2,8149	-2	4	-5,6298	259	2,4133
9	256	2,4082	-1	1	-2,4082	274	2,4378
10	270	2,4314	0	0	0	289	2,4609
11	200	2,3010	1	1	2,3010	305	2,4843
12	213	2,3284	2	4	4,6568	322	2,5079
13	200	2,3010	3	9	6,9030	340	2,5315
14	233	2,3674	4	16	9,4696	359	2,5551
15	268	2,4281	5	25	12,1405	379	2,5786
16	310	2,4914	6	36	14,9484	399	2,6009
17	337	2,5276	7	49	17,6932	422	2,6253
18	372	2,5705	8	64	20,5640	445	2,6484
19	415	2,6180	9	81	23,5620	470	2,6721
20	525	2,7202	10	100	27,2020	496	2,6955
		49,2187		670	15,7394		

Fuente: datos oficiales y su tendencia

$$49,2187 / 20 = 2,460935$$

$$15,7394 / 670 = 0,02349$$

$$\text{Log } y = 2,460935 + 0,02349$$

$$y = (289) (1,0555) \text{ al cuadrado}$$

$$y = (289) (1,11408)$$

$$y = 322$$

Este valor promedio crece más o menos al 6% al año, según las 20 cifras analizadas. Los valores de la tendencia se acercan bastante bien a los datos originales.

Ahora bien, la función que mejor representó la evolución del turismo receptivo en Venezuela desde 1971 a 1991 y con un modelo de regresión múltiple fue la siguiente (ASCANIO, 1996, p.143):

$$NT = - 342.417,5 + 2.316,89 CR (Bs./\$) + 3.921,43 PIB + 0,7998 NT_{-1} + Ut$$

$$R \text{ al cuadrado} = 93,1; DW = 2,39; h = 1,11 \text{ y } F = 76,4$$

En donde:

NT = número de turistas internacionales llegados a Venezuela en el año "t"

NT menos 1 = número de turistas internacionales llegados al país en el año anterior.

PIB = Índice del PIB real de los países industrializados en el año t

CR (Bs./\\$) = tipo de cambio real Bs./\\$ en el año "t".

Ut = término de perturbación en el año "t"

La regresión lineal no logarítmica desde el año uno con 117 mil visitantes hasta el año 16 con 310 mil visitantes es como sigue:

$$y = 1.03109 x^3 + -30.3049 x^2 + 253.299x + -216.712$$

Con los siguientes valores de tendencia: 7,31; 176,9; 298,3; 377,6; 421,0; 434,8; 425,1; 398,8; 359,9; 316,9; 275,1; 240,7; 219,9; 219,0; 244,1; 301,3.

Con un valor promedio de 295 mil visitantes por año.

La ecuación lineal desde el año 5 con 425 mil visitantes hasta el año 20 con 525 mil visitantes es como sigue:

$$y = 0.431511 x^3 + -6.06132 x^2 + -11.2538x + 516.563$$

Con los siguientes valores de tendencia: 499,7; 473,3; 439,9; 402,2; 362,7; 324,0; 288,8; 259,5; 233,7; 254,3; 293,9; 355,1; 440,3; 552,3.

Con un valor promedio de 370 mil visitantes al año.

El promedio anual de las tres estimaciones es de **329 mil visitantes al año.**

Se puede observar en el cuadro No. 2 siguiente que la relación entre valores elevados, medios y bajos, según el número de años utilizados arroja una asociación positiva, pues sus valores medios son muy parecidos.

Cuadro 1 - Relación entre promedio de visitantes y años de proyección

Promedio Visitantes	Número años elevado	Número años bajo
Elevado	-	370
Media	322	-
Baja	-	295
TOTAL	322	665

El coeficiente Q es como sigue:

$$Q = 665 \times 322 - 370 - 322 / 322 \times 370 + 370 \times 322$$

$$Q = 214.130 - 48 / 222.824$$

$$Q = 0,96$$

El resultado indica una asociación importante entre las dos variables y la probabilidad de acertar para predecir una variable conociendo la otra.

En este caso del turismo receptivo hacia Venezuela lo podemos establecer con mayor rigurosidad en base al promedio anual de visitantes de las tres estimaciones que fue de más o menos 329 visitantes al año, con un crecimiento interanual del 6%.

Esto lo que nos indica es que e posible que para el año **2010** Venezuela sólo podrá captar **entre 550 a 600 turistas internacionales** y no más que eso, siempre que se siga ignorando una política agresiva de marketing y se deje sólo en libertad lo que pueda pasar con el turismo, sin hacer nada para promoverlo.

Ahora bien si la recta de regresión se realiza con el método de los mínimos cuadrados, como aparece en cuadro No.3 siguiente, los datos reales y los estimados presentan una diferencia que puede ser mínima o mayor, dependiendo del comportamiento de la nube de puntos que se observaría en el cuadrante.

Cuadro 2 - Datos para estimar la recta de regresión lineal entre x e y

X	Y	Y prima	y prima por x	xyprima
1	117	- 153	-153 por 1	- 153
2	144	- 126	- 126 por 2	- 252
3	171	- 99	- 99 por 3	- 297
4	285	+ 15	+ 15 por 4	+ 60

5	425	+ 155	+ 155 por 5	+ 775
6	436	+ 166	+ 166 por 6	+ 996
7	535	+ 265	+ 265 por 7	+ 1.855
8	653	+ 383	+ 383 por 8	+ 3.064
9	256	- 14	- 14 por 9	- 126
10	270	0	0	0
11	200	- 70	- 70 por 11	- 770
12	213	- 57	- 57 por 12	- 684
13	200	- 70	- 70 por 13	- 910
14	233	- 37	- 37 por 14	- 518
15	268	- 2	- 2 por 15	- 30
16	310	+ 40	+ 40 por 16	+ 640
17	337	+ 67	+ 67 por 17	+ 1.139
18	372	+ 102	+ 102 por 18	+ 1.836
19	415	+ 145	+ 145 por 19	+ 2.755
20	525	+ 255	+ 255 por 20	+ 5.100
		+ 965		+14.480

Con los datos del cuadro 3 haremos los siguientes cálculos prácticos:

$$n = 20$$

$$A = 21 / 2 = 10,5$$

$$B = 20 (20 \text{ al cuadrado menos } 1) / 12 = 20 \times 399 / 12 = 665$$

$$A \text{ por sumatoria de } y \text{ prima} = 10,5 \times 965 = 10.132,5$$

$$\text{Sumatoria de } x \text{ por } y \text{ prima menos } 10.132,5 = 14.480 - 10.132,5 = 4.347,5$$

$$b = 4.347,5 / 665 = 6,54$$

$$x \text{ media} = A = 10,5$$

$$y \text{ media} = 270 + 965 / 20 = 270 + 48,25 = 318,25$$

$$a = 318,25 + (6,54 \times 10,5) = 386,92$$

Entonces la ecuación de la recta de regresión lineal es:

$$y = 386,92 + 6,54 (x)$$

Con esta recta de regresión para el año 20 el número de visitantes internacionales estimados sería igual a: $y = 386,92 + 6,54 (20) = 517,72$

El resultado nos indica que la tendencia es 7.000 visitantes menos que el dato real, o sea: 1,33% menor. Esta diferencia se debe al comportamiento de la nube de punto y debido al ajuste de la curva de regresión con el método de los mínimos cuadrados.

El gasto medio del turista: otro dato relevante

Cuando realizamos una encuesta para saber el gasto en millones de dólares de 5 regiones turísticas durante el IV trimestre del año 2008, encontramos estas respuestas:

Cuadro 3 - Gasto en el IV trimestre del 2.008-Turismo receptivo

Regiones	Gasto en millones de US\$	Número de visitantes	Gasto medio por visitante en US\$
Europa	66,5	42.624	1.560
América del Sur	32,8	22.512	1.457
América del Norte	49,8	38.271	1.301
Centro América y el Caribe	18,7	12.907	1.449
Otras regiones	8,2	3.302	2.483
TOTAL	176,0	119.616	1.471

Estos de datos de una muestra tomada al azar nos indican que hemos encontrado 42.624 visitantes europeos con un gasto medio por visitante igual a US\$ 1.560 y 38.271 visitantes de América del Norte con un gasto medio por visitante igual a US\$ 1.301.

La pregunta básica en este momento es: ¿Qué probabilidad tenemos de equivocarnos al afirmar que, en la población, los turistas Europeos son más numerosos y tienen un gasto medio mayor que los visitantes procedentes de América del Norte?

Una primera aproximación para conocer esa probabilidad es utilizar un ábaco ya diseñado y *ad-hoc*. Lo primero que haremos es obtener las diferencias en más o en menos de esas dos regiones que deseamos comparara:

$$43 + 38 = 81$$

$$43 - 38 = 5$$

Ahora interceptamos estos valores en el ábaco para conseguir el tanto por ciento de la probabilidad de equivocarnos, considerando la significación de visitantes procedentes de

Europa más los de América del Norte; y los visitantes de Europa menos América del Norte; es decir buscamos la intersección vertical de 81 con la horizontal de 5 y nos arroja **50% a 60% de posibilidades de equivocarnos**.

En efecto, tenemos casi una posibilidad por cada dos casos de equivocarnos al afirmar que los visitantes europeos son más numerosos que los visitantes de América del Norte en el universo considerado. Como el riesgo es alto no podemos afirmar lo anterior. Es decir la diferencia no es significativa pues es más bien aleatoria.

Si ahora hacemos el mismo test para varias frecuencias, en este caso para los visitantes Europeos en número de 43, América del Norte con 38 y América del Sur con 23, con un total de 104.

Cuadro 4 - Test de significación para tres frecuencias de visitantes receptivos

	F	KF	KF - N	KF - N al cuadrado	X al cuadrado
1	43	129	+25	+ 625	
2	38	114	+10	+ 100	
3	23	69	- 35	+ 1.225	
	104	312	cero		1.225/312= 3,93

La prueba Chi cuadrado anterior nos arroja el valor de **3,93** y **para 2 grados de libertad** ($3 - 1 = 2$); pero en la tabla de Chi cuadrado para 5% de probabilidad y 2 grados de libertad el valor es igual a **5,991** (digamos 6) y en nuestro cálculo hemos obtenido 3,93 (o sea un dato inferior a 6), lo cual nos indica que tenemos un poco menos del 20% de probabilidad de equivocarnos al afirmar que, en la población, los resultados de las tres categorías no son iguales.

En la prueba de dos frecuencias (Visitantes Europeos y Visitantes de América del Norte) bastaba con conocer el dato límite de 4 (3,84) para un grado de libertad, pero cuando nos enfrentamos a más frecuencias es necesario consultar la tabla de Chi Cuadrado para poder hacer el test de significación.

En la tabla de Chi cuadrado para 1, 2 y 3 grados de libertad las probabilidades con el 5% aceptable son: 3,8 (4), 5,99 (6) y 7,82 (8) y existe una regla práctica, en el caso de no tener la tabla de los valores tolerables, que consiste en aplicar la siguiente fórmula: $3,5 + 1,5 \times GL$.

Así pues, si el grado de libertad es digamos 10, entonces con la fórmula anterior: $1,5 \text{ por } 10 = 15$ y luego $3,5 + 15 = 18,5$, que es más o menos la misma intercepción entre 10 grados

de libertad y 5% de probabilidad. Esta regla sencilla nos arroja un valor aproximado (en la tabla el valor es 18,307) y en la medida en que los grados de libertad sea mayores de 10, entonces los valores estimados hay que tomarlos con más cautela.

Veamos ahora los datos para los visitantes Europeos, América del Sur, América del Norte, así como Centra América y Caribe con los valores de visitantes igual a: 43, 23,38 y 13 y un total de 117 (en miles de turistas).

Según esos datos, los visitantes europeos representan el 37% del total. Ahora utilizando otro nomograma o ábaco que relaciona el porcentaje del total (P) con N, o sea en este caso: 117 y con el indicador de tolerancia T, ello nos arroja una tolerancia igual a 9%, o sea más elevado que la tolerancia aceptada del 5%.

Si quisiéramos una tolerancia máxima del 2%, cuál debería ser el porcentaje de los visitantes europeos en relación al total?

Tomemos el porcentaje del 50% y la tolerancia del 2% y entonces podemos decir que se requieren más o menos 2.400.000 visitantes europeos, lo cual parecería imposible en el contexto actual, pero para una tolerancia del 3% se necesitarían 1.100.000 visitantes europeos, que todavía parece inalcanzable y finalmente con una tolerancia del 4% deberíamos captar al menos 600.000 europeos; y con el límite ideal del 5% se necesitaría realizar una promoción para captar al menos 300.000 europeos en un trimestre; y con una tolerancia del 8%, el ábaco nos indica que se deberían captar en un trimestre 300.000 visitantes totales de los cuales 150.000 deberían ser europeos.

Lo anterior supone un esfuerzo promocional para obtener una cuota de mercado anual de al menos 1.200.000 turistas de los cuales Europa participaría con 600.000 turistas anuales. Una meta posible de lograr siempre que se negocie con un Tour Operador un programa de marketing agresivo.

Cuando no se tiene a la mano el nomograma o ábaco, entonces la tolerancia se puede estimar con la siguiente fórmula:

$$T = t \text{ raíz de } P \times Q / N$$

Si la muestra esta formada por 300 sujetos y se estima que de esos sujetos el 40% es una parte del total, quedando el 60% para la otra parte, y con un umbral del 5% (1,96), entonces:

$$T = 1,96 \times \text{raíz de } 40 \times 60 / 300 = 1,96 \times \text{raíz de } 8 = 1,96 \times 2,28 = 5,547\%$$

Este es un buen resultado, pues si se utiliza el ábaco la tolerancia es igual a 5,5%.

Muchas veces cuando queremos hacer una encuesta, deseamos calcular el número de sujetos a interrogar para obtener una tolerancia (T) elegida de antemano, como puede ser digamos el 4% y con un umbral de significación del 5% (1,96), así:

$$N = (50 \times 1,96 / 4) \text{ elevado a la dos}$$

$$N = (98 / 4) \text{ elevado a la dos}$$

$$N = (24,5) \text{ elevado a la dos}$$

$$N = 600$$

Las tablas de contingencia y su significación

Una tabla de contingencia es una técnica utilizada en la estadística comercial con la idea de comparar varias eventualidades o confrontar los datos obtenidos unos con los otros, datos que por lo general surgen de las encuestas. Si las diferencias son significativas se puede concluir que existe una determinada relación entre los datos e igualmente nos permite evaluar el riesgo de no equivocarnos.

Cuadro 5 - Tabla de contingencia 2 x 2

	Impactados	No Impactados	Totales
Visita sin impacto publicitario	34 (9,9%)	72 (21,1%)	106 (31%)
Visita con impacto publicitario	108 (31,6%)	128 (37,4)	236 (69%)
Totales	142 (41,5%)	200 (58,5%)	342 (100%)

Esta tabla realmente se organiza de este modo:

Cuadro 6 - datos resumidos

	Impactados	No Impactados
Visitaron el sitio sin conocer la publicidad	34	72
	24%	36%
Visitaron el sitio al conocer la publicidad	108	128
	76%	64%

TOTALES	142	200
	100%	100%

La tabla de contingencia anterior resume una encuesta sobre la atención a programas publicitarios turísticos. Se han interrogado a visitantes que han estado sometidas a ver esas promociones y de aquellas que no la habían conocido. De la encuesta a 342 turistas, en el cuadro o tabla de contingencia 2 x 2 aparecen las respuestas de los conocedores de la publicidad e impactados por ella (en la columna Impactados) y los visitantes que no conocieron la publicidad (No impactados).

Los que turistas que han visitado el sitio publicitado e impactados por la promoción figuran en segunda la fila y los demás en la primera fila (lo visitaron sin influencia del impacto publicitario).

Según la tabla resumen No.7, el sitio turístico han sido visitados por el 76% (108) de los turistas que conocieron la promoción y por el 64% de los que no estuvieron impactados por la publicidad (128). O sea un total de 236 visitantes (el 69% de 342). La pregunta clave es: ¿es significativa esta diferencia?

Teóricamente, si la publicidad y la promoción no guardan ninguna relación con la visita a un sitio turístico esto quiere decir que existe en la población o universo la misma proporción de visitantes en el lugar entre los turistas impactados por la publicidad y aquellos que no estuvieron impactados. Siendo esta proporción de la población total como vimos de **0,69 (69%)**.

Siendo la proporción de impactados igual a 0,415 (41,5%), y la proporción de visitantes igual a 0,69 (69%), entonces el porcentaje teórico de visitantes impactos por la publicidad será igual a 41,5% del 69%, o sea: $0,415 \times 0,69 = 0,286$ (28,6%) y el número teórico de visitantes de un total $N = 342$, será pues igual a $0,286 \times 342 = 97,81$, mientras que en el cuadro 6 es igual a 108. Esa diferencia se debe a los errores del muestreo? Si la probabilidad es pequeña la diferencia es significativa y sin la probabilidad NO es pequeña, es posible que exista un error de muestreo.

Si hacemos un aprueba de Chi Cuadrado con el siguiente grado de libertad: 2 columnas menos 1 por 2 filas menos 1, o sea: grado de libertad 1 y en base a la tabla siguiente:

34	72	106
108	128	236
142	200	342

$$108 \times 72 = 7.776$$

$$128 \times 34 = 4.352$$

Suma 3.424

$$3.424 \text{ al cuadrado} = 11.700.000$$

$$142 \times 200 = 28.400$$

$$236 \times 106 = 25.000$$

$$\text{Chi cuadrado} = 11.700.000 \times 342 / 28.400 \times 25.000$$

$$\text{Chi cuadrado} = 4.000.000.000 / 710.000.000 = \mathbf{5,64}$$

El resultado es significativo pues es mayor que el indicador tolerable del 3,841 para 1 grado de libertad y 5% de probabilidad. El valor 5,64 es algo menos del 2% de probabilidad de equivocarnos.

Todavía nos hallamos lejos de los verdaderos estudios comerciales del turismo como sector terciario de la economía, pero es necesario trabajar los datos con las herramientas estadísticas de que disponemos.

Quizá lo más complejo sea realizar una buena evaluación de los hechos. La palabra evaluación es muy elástica porque con ella podemos abarcar muchas clases de juicios. En realidad lo que esa palabra quiere decir es con ella podemos tener una noción de los méritos de un asunto e incluso aumentar la racionalidad de las decisiones (WEISS, 1978, p. 13-14).

La publicidad, la promoción turística y la venta de espacios lúdicos es un fenómeno sociopsicológico. Podemos ver cómo evoluciona la llegada de turistas, pero no sabemos las causas, que pueden ser económicas o psicológicas. Incluso las causas psicológicas son las más poderosas.

Entonces, los test de significación se adaptan a los estudios psicológicos y en especial los *tests* de contingencia.

Un espacio turístico ensalzado como un lugar único y de elevado precio, se vende rápidamente y a lo mejor no pasa de moda. Los móviles de compra son, en gran parte, inconscientes; el turista no puede, por lo tanto, decirnos sus motivos. Si le preguntamos, corremos el riesgo que responda cualquier cosa. Sólo si provocamos un comportamiento en el interrogado, podemos detectar algunos motivos útiles para ser sometidos a la estadística evaluativa.

Consideraciones finales

Los métodos estadísticos pueden ser deductivos (descriptivos) o inductivos (probabilístico). Los métodos deductivos permiten aplicar técnicas básicas a datos reales y así poder saber la frecuencia con que aparecen los datos, estimar los valores promedios, las variaciones temporales, a los fines de interpretar los resultados. Con los métodos inductivos nos acercamos mucho más al conocimiento de lo que ocurre con los datos obtenidos de muestras para luego hacer inferencias hacia el universo o la población, es decir probar las hipótesis.

La necesidad de comprender mejor las causas y el comportamiento de los turistas cuando se desplazan de su lugar de residencia habitual, es un área en la cual pueden cooperar los especialistas en estadísticas.

La estadística evaluativa o inductiva se ha convertido en una importante fuente de conocimientos para tomar decisiones, pues con esta herramienta podemos conocer la eficiencia o deficiencia de los esfuerzos para promover el turismo.

REFERÊNCIAS

ASCANIO, Alfredo, et. al. *La demanda turística internacional en Venezuela y el comportamiento del visitante temporal*. Estudios y Perspectivas en turismo, vol. 5, 1996, p.127-146.

GARCIA PEREZ, Andrés. *Elementos del método estadístico*. México: Imprenta Universitaria (UNAM), 1960.

IBARRA MARTINEZ, Oscar Mario. *Estadística para la administración turística*. México: Editorial Trillas, 2006.

PIN, Henri. *Primeras aplicaciones de estadística comercial*. Barcelona: Editorial Sagiario, 1964.

WEISS, Carol. *Investigación evaluativa*. México: Editorial Trillas, 1978.

Artigo recebido em 12 de setembro de 2009

Aprovado para publicação em 01 de outubro de 2009