Vamos a contar taxis

Cada ayuntamiento español tiene adjudicadas N licencias de taxi (donde N depende del ayuntamiento)

Las licencias están numeradas del 1 al N

Cada taxi lleva su número de licencia en las puertas y en el luminoso

Vamos a contar taxis

Cada ayuntamiento español tiene adjudicadas N licencias de taxi (donde N depende del ayuntamiento)

Las licencias están numeradas del 1 al N

Cada taxi lleva su número de licencia en las puertas y en el luminoso

UNIVERSIDAD DE CANTABRIA

STATISTICS2013.ORG AÑO INTERNACIONAL DE LA ESTADÍSTICA

Vamos a contar taxis

Cada ayuntamiento español tiene adjudicadas N licencias de taxi (donde N depende del ayuntamiento)

Las licencias están numeradas del 1 al N

Cada taxi lleva su número de licencia en las puertas y en el luminoso

Vamos a contar taxis

Cada ayuntamiento español tiene adjudicadas N licencias de taxi (donde N depende del ayuntamiento)

Las licencias están numeradas del 1 al N

Cada taxi lleva su número de licencia en las puertas y en el luminoso

Nos plantamos en una esquina cualquiera de una ciudad cualquiera

Vamos apuntando los números de licencia de los taxis que pasan

Al final, tenemos los números de n taxis,

¿Podremos averiguar el número total de taxis de esta ciudad?

Estimando N

Vamos a llamar $X_{(1)} \leq \ldots \leq X_{(n)}$ a los números que tenemos apuntados

Estimando N

Vamos a llamar $X_{(1)} \leq \ldots \leq X_{(n)}$ a los números que tenemos apuntados

Es obvio que $N \geq X_{(n)} = \max(X_1, \ldots, X_n)$, pero ¿Cuánto mayor?

Estimando N

Vamos a llamar $X_{(1)} \leq \ldots \leq X_{(n)}$ a los números que tenemos apuntados

Es obvio que $N \geq X_{(n)} = \max(X_1, \ldots, X_n)$, pero ¿Cuánto mayor?

Idea: La diferencia $N - X_{(n)}$ debería comportarse como $X_{(1)} - 1$

Estimando N

Vamos a llamar $X_{(1)} \leq \ldots \leq X_{(n)}$ a los números que tenemos apuntados

Es obvio que $N \geq X_{(n)} = \max(X_1, \ldots, X_n)$, pero ¿Cuánto mayor?

Idea: La diferencia $N - X_{(n)}$ debería comportarse como $X_{(1)} - 1$

$\Rightarrow$ podemos estimar $\hat{N} = X_{(n)} + (X_{(1)} - 1)$

Estimando N

Vamos a llamar $X_{(1)} \leq \ldots \leq X_{(n)}$ a los números que tenemos apuntados

Es obvio que $N \geq X_{(n)} = \max(X_1, \ldots, X_n)$, pero ¿Cuánto mayor?

Idea: La diferencia $N - X_{(n)}$ debería comportarse como $X_{(1)} - 1$

$\Rightarrow$ podemos estimar $\hat{N} = X_{(n)} + (X_{(1)} - 1)$

Ejemplo. Si $n = 2$ y los datos son 1 y 73 $\Rightarrow \hat{N} = 73$

Estimando N

Vamos a llamar $X_{(1)} \leq \ldots \leq X_{(n)}$ a los números que tenemos apuntados

Es obvio que $N \geq X_{(n)} = \max(X_1, \ldots, X_n)$, pero ¿Cuánto mayor?

Idea: La diferencia $N - X_{(n)}$ debería comportarse como $X_{(1)} - 1$

$$\Rightarrow \quad \text{podemos estimar } \hat{N} = X_{(n)} + (X_{(1)} - 1)$$

Ejemplo. Si $n = 2$ y los datos son 1 y 73 $\Rightarrow \hat{N} = 73$

Este estimador es malo porque sólo tiene en cuenta dos datos

Estimando N

Vamos a llamar $X_{(1)} \leq \ldots \leq X_{(n)}$ a los números que tenemos apuntados

Es obvio que $N \geq X_{(n)} = \max(X_1, \ldots, X_n)$, pero ¿Cuánto mayor?

Idea: La diferencia $N - X_{(n)}$ debería comportarse como $X_{(1)} - 1$

$$\Rightarrow \quad \text{podemos estimar } \hat{N} = X_{(n)} + (X_{(1)} - 1)$$

Pero ¡tenemos más datos para mirar!

Estimando N

Vamos a llamar $X_{(1)} \leq \ldots \leq X_{(n)}$ a los números que tenemos apuntados

Es obvio que $N \geq X_{(n)} = \max(X_1, \ldots, X_n)$, pero ¿Cuánto mayor?

Idea: La diferencia $N - X_{(n)}$ debería comportarse como $X_{(1)} - 1$

$$\Rightarrow \quad \text{podemos estimar } \hat{N} = X_{(n)} + (X_{(1)} - 1)$$

La diferencia $N - X_{(n)}$ también debería comportarse como

$$\begin{array}{c} X_{(2)} - X_{(1)} \\ X_{(3)} - X_{(2)} \\ \ldots \end{array}$$

Estimando N

Vamos a llamar $X_{(1)} \leq \ldots \leq X_{(n)}$ a los números que tenemos apuntados

Es obvio que $N \geq X_{(n)} = \max(X_1, \ldots, X_n)$, pero ¿Cuánto mayor?

Idea: La diferencia $N - X_{(n)}$ debería comportarse como $X_{(1)} - 1$

$\Rightarrow$ podemos estimar $\hat{N} = X_{(n)} + (X_{(1)} - 1)$

La diferencia $N - X_{(n)}$ también debería comportarse como $X_{(2)} - X_{(1)}$, $X_{(3)} - X_{(2)}$, $\ldots$

Podemos aceptar que

$$N - X_{(n)} \simeq \frac{1}{n}\left((X_{(1)} - 1) + (X_{(2)} - X_{(1)}) + \ldots + (X_{(n)} - X_{(n-1)})\right)$$

Estimando N

Vamos a llamar $X_{(1)} \leq \ldots \leq X_{(n)}$ a los números que tenemos apuntados

Es obvio que $N \geq X_{(n)} = \max(X_1, \ldots, X_n)$, pero ¿Cuánto mayor?

Idea: La diferencia $N - X_{(n)}$ debería comportarse como $X_{(1)} - 1$

$$\Rightarrow \quad \text{podemos estimar } \hat{N} = X_{(n)} + (X_{(1)} - 1)$$

La diferencia $N - X_{(n)}$ también debería comportarse como $X_{(2)} - X_{(1)}$, $X_{(3)} - X_{(2)}$, $\ldots$

Podemos aceptar que

$$\begin{aligned} N - X_{(n)} &\simeq \frac{1}{n}\left((X_{(1)} - 1) + (X_{(2)} - X_{(1)}) + \ldots + (X_{(n)} - X_{(n-1)})\right) \\ &= \frac{1}{n}\left(X_{(n)} - 1\right) \end{aligned}$$

Estimando N

Vamos a llamar $X_{(1)} \leq \ldots \leq X_{(n)}$ a los números que tenemos apuntados

Es obvio que $N \geq X_{(n)} = \max(X_1, \ldots, X_n)$, pero ¿Cuánto mayor?

Idea: La diferencia $N - X_{(n)}$ debería comportarse como $X_{(1)} - 1$

$\Rightarrow$ podemos estimar $\hat{N} = X_{(n)} + (X_{(1)} - 1)$

La diferencia $N - X_{(n)}$ también debería comportarse como

$$\begin{array}{c} X_{(2)} - X_{(1)} \\ X_{(3)} - X_{(2)} \\ \ldots \end{array}$$

Podemos aceptar que

$$\begin{aligned} N - X_{(n)} &\simeq \frac{1}{n}\left((X_{(1)} - 1) + (X_{(2)} - X_{(1)}) + \ldots + (X_{(n)} - X_{(n-1)})\right) \\ &= \frac{1}{n}\left(X_{(n)} - 1\right) \quad \Rightarrow \quad \hat{N} = \frac{n+1}{n} X_{(n)} - \frac{1}{n} \end{aligned}$$

Estimando N

Vamos a llamar $X_{(1)} \leq \ldots \leq X_{(n)}$ a los números que tenemos apuntados

Es obvio que $N \geq X_{(n)} = \max(X_1, \ldots, X_n)$, pero ¿Cuánto mayor?

Idea: La diferencia $N - X_{(n)}$ debería comportarse como $X_{(1)} - 1$

$$\Rightarrow \quad \text{podemos estimar } \hat{N} = X_{(n)} + (X_{(1)} - 1)$$

La diferencia $N - X_{(n)}$ también debería comportarse como $X_{(2)} - X_{(1)}$, $X_{(3)} - X_{(2)}$, $\ldots$

Podemos aceptar que

$$\begin{aligned} N - X_{(n)} &\simeq \frac{1}{n}\left((X_{(1)} - 1) + (X_{(2)} - X_{(1)}) + \ldots + (X_{(n)} - X_{(n-1)})\right) \\ &= \frac{1}{n}\left(X_{(n)} - 1\right) \quad \Rightarrow \quad \hat{N} = \frac{n+1}{n} X_{(n)} - \frac{1}{n} \end{aligned}$$

es posible dar una expresión para la certidumbre del resultado

Estimando N

Ejemplo En Santander hay $N = 218$ licencias

Me senté en una terraza céntrica (Paseo de Pereda)
y apunté taxis hasta 40 números

Estimaciones:

n	Máximo	Estimador	Intervalos

Estimando N

Ejemplo En Santander hay $N = 218$ licencias

Me senté en una terraza céntrica (Paseo de Pereda) y apunté taxis hasta 40 números

62 55 **140** 130 93

Estimaciones:

n	Máximo	Estimador	Intervalos
5	140	168	(140,255)

Estimando N

Ejemplo En Santander hay $N = 218$ licencias

Me senté en una terraza céntrica (Paseo de Pereda)
y apunté taxis hasta 40 números

62 55 **140** 130 93 86 3 116 7 102

Estimaciones:

n	Máximo	Estimador	Intervalos
5	140	168	(140,255)
10	140	154	(140,189)

Estimando N

Ejemplo En Santander hay $N = 218$ licencias

Me senté en una terraza céntrica (Paseo de Pereda)
y apunté taxis hasta 40 números

62	55	**140**	130	93	86	3	116	7	102	174	149	140	127
190	85	**200**	123	51	149								

Estimaciones:

n	Máximo	Estimador	Intervalos
5	140	168	(140,255)
10	140	154	(140,189)
20	200	210	(200,232)

Estimando N

Ejemplo En Santander hay $N = 218$ licencias

Me senté en una terraza céntrica (Paseo de Pereda)
y apunté taxis hasta 40 números

62	55	**140**	130	93	86	3	116	7	102	174	149	140	127
190	85	**200**	123	51	149	16	171	67	62	139	144	**209**	112
193	66	178	129	181	107	57	55	51	174	184	28		

Estimaciones:

n	Máximo	Estimador	Intervalos
5	140	168	(140,255)
10	140	154	(140,189)
20	200	210	(200,232)
40	209	214	(209,225)

¿Cuál es el interés de adivinar un máximo?

¿Cuál es el interés de adivinar un máximo?

Vamos a la Segunda Guerra Mundial

Uno está (muy) interesado en la capacidad de producción del enemigo

Las piezas del equipo llevan información sobre

1. El productor y la planta en que se fabricaron
2. Fecha de fabricación
3. Número de serie
4. Otra información:
    4.1 Identificación del molde u horno usado,
    4.2 marca comercial,

    ...

No todos las piezas traen todos estos datos
y, además, la información estaba codificada,
pero no fue excesivamente difícil romper los códigos

Neumáticos

Se disponía de 2.000 neumáticos alemanes
procedentes del derribo de aviones y tomados en el norte de África

Se concluyó que cada fabricante los numeraba correlativamente

Y, también, que todos traían el mes y año de fabricación

Por lo tanto, estamos en el problema de los taxis:

Si estimamos el número máximo hasta fin de enero
el número máximo hasta fin de febrero

Su diferencia da la producción de febrero

Neumáticos

Se disponía de 2.000 neumáticos alemanes
procedentes del derribo de aviones y tomados en el norte de África

Se concluyó que cada fabricante los numeraba correlativamente

Y, también, que todos traían el mes y año de fabricación

Por lo tanto, estamos en el problema de los taxis:

Si estimamos el número máximo hasta fin de enero
el número máximo hasta fin de febrero

Su diferencia da la producción de febrero

Utilidad: La producción de neumáticos de avión informa de la de aviones

Se puede comparar la producción antes y después de un bombardeo para estimar sus efectos

...

Otros productos

Se aplicó la misma metodología a

1. Tanques
2. Motores de vehículos
3. Artillería
4. Artillería autopropulsada
5. Cohetes

. . .

Resultados

Producto	Período	Este método	Espionaje	Producción real
Neumáticos	1º trim. 43	526.500	3.000.000	558.300
Camiones	año 1.942	97.300	200.000	79.827
Tanques	Junio 1941	244	1.550	271
	Agosto 1942	327	1.550	342
Bombas V1	hasta Jun. 1944	13.000	-	12.000
	Jun-Jul 1944	18.000	-	$<$ 6.000
Cohetes V2	Set-Nov 44	1.730	-	1.500
	Nov. 44-Feb. 45	1.800	-	1.800

Resultados

Producto	Período	Este método	Espionaje	Producción real
Neumáticos	1º trim. 43	526.500	3.000.000	558.300
Camiones	año 1.942	97.300	200.000	79.827
Tanques	Junio 1941	244	1.550	271
	Agosto 1942	327	1.550	342
Bombas V1	hasta Jun. 1944	13.000	-	12.000
	Jun-Jul 1944	18.000	-	$<$ 6.000
Cohetes V2	Set-Nov 44	1.730	-	1.500
	Nov. 44-Feb. 45	1.800	-	1.800

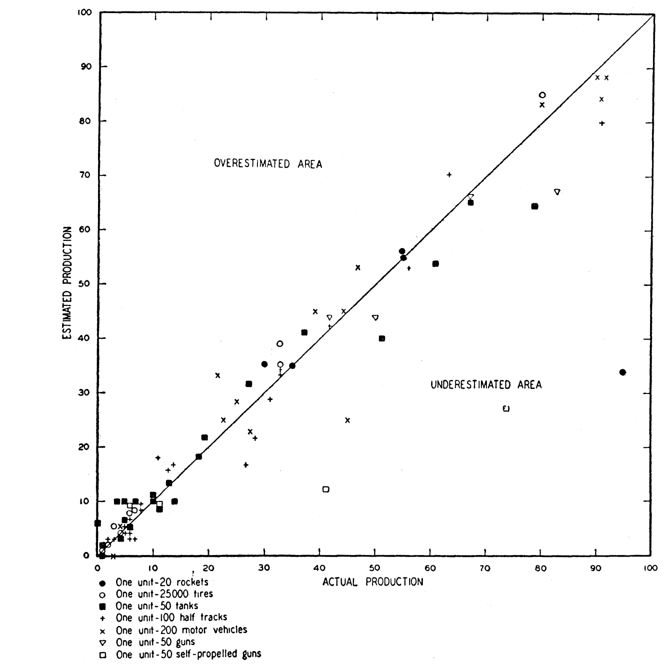
Bombas V1 — Probablemente, 18.000 era el número esperado.
Pero apareció un cuello de botella en
la fabricación de carcasas (no llevaban numeración)

Resultados

Producto	Período	Este método	Espionaje	Producción real
Neumáticos	1º trim. 43	526.500	3.000.000	558.300
Camiones	año 1.942	97.300	200.000	79.827
Tanques	Junio 1941	244	1.550	271
	Agosto 1942	327	1.550	342
Bombas V1	hasta Jun. 1944	13.000	-	12.000
	Jun-Jul 1944	18.000	-	$<$ 6.000
Cohetes V2	Set-Nov 44	1.730	-	1.500
	Nov. 44-Feb. 45	1.800	-	1.800

Avanzada la guerra, prisioneros informaron de falta de munición
Esta metodología concluyó que no había problemas de fabricación
Al finalizar la guerra se comprobó que el problema era de comunicaciones

Resultados



OVERESTIMATED AREA
UNDERESTIMATED AREA
ESTIMATED PRODUCTION
ACTUAL PRODUCTION
0 10 20 30 40 50 60 70 80 90 100
One unit-20 rockets
One unit-25000 tires
One unit-50 tanks
One unit-100 half tracks
One unit-200 motor vehicles
One unit-50 guns
One unit-50 self-propelled guns

Anécdota

Antes de 1.991, el coronel Trevor Dupuy (USA) visitó la fábrica de tanques Merkava en Israel.

Al finalizar la visita, preguntó las cifras de producción

Y le dijeron que era información clasificada

El coronel Dupuy dijo (su correspondencia personal)

"La respuesta fue divertida porque yo estaba viendo los números de serie de los tanques"

!!! M U C H A S
G R A C I A S !!!

e-mail: cuestaj@unican.es
http://personales.unican.es/cuestaj

Año internacional: http://personales.unican.es/navarrop/

http://www.youtube.com/user/CanalSEIO