CONTADE CON NÓS

IES DE CACHEIRAS
2ºESO

Alumnado:

Anxo Amigo González
Martiño Amigo González
Juan Ferreirós Estévez
Brais Suárez Varela

Titoras:

Trinidad Pazos Celis
Ana Mª Rodríguez Sánchez

Índice

Introdución.....3

Investigación 3

Obxectivos.....3

Deseño inicial do experimento e recollida de datos.....4

Transcurso e transformacións do experimento. Traballo de campo 7

Análise estatística e interpretación dos resultados obtidos.....8

Conclusións 18

Posibles melloras 18

Introdución

Somos catro rapaces de 2º da ESO do IES de Cacheiras: Anxo, Brais, Juan e Martiño.

Decidimos presentarnos de novo ao concurso "Incubadora de sondaxes" porque algúns de nós xa participamos o ano pasado, co proxecto *Emociónate connosco,* e gustaranos moito a experiencia. Pero este ano imos con reforzo: uníusenos un amigo que oíra falar do noso traballo.

A idea da nosa investigación xorde da nosa curiosidade por saber de que xeito se podería contar a xente que acode a unha manifestación, xa que sempre os datos que se trasladan na prensa son moi dispares:

A policía e as distintas asociacións de manifestantes nunca están de acordo, como por exemplo nestes casos:

- Exemplo 1: Marcha en Madrid contra os recortes en educación (25.000 asistentes segundo a Delegación do Goberno e máis de 100.000 segundo os organizadores)
- Exemplo 2: Manifestación contra Altri (100.000 persoas segundo a organización e 40.000 para a policía local)

Será que non temos un método para contar os asistentes a unha concentración?

Investigación

Como se calculan os asistentes a unha manifestación?

Ao buscar información, vemos que hai diferentes métodos. Por exemplo:

- Herbert Jacobs, un profesor de xornalismo da Universidade de California, Berkeley, é o responsable da modernización do cálculo das multitudes na década dos '60. Desde a fiestra da súa oficina, Jacobs podía ver os estudantes reunidos nunha praza para protestar contra a guerra do Vietnam. A praza estaba organizada en cadrados, así que Jacobs contou a cantidade de estudantes nalgúns cadrados para obter un promedio de alumnos. Despois, multiplicou esa cifra pola cantidade de cadrados totais. Jacobs tamén creou unha regra básica de densidade que di que unha "multitude fluída" conta cunha persoa por cada 10 pés cadrados e unha "multitude densa " ten unha persoa por cada 4,5 pés cadrados. Unha multitude moi densa, chamado un "*mosh-pit*" por Watson e Yip, tería unha persoa por cada 2,5 pés cadrados.

(Fonte: noticia de ABC)

- A Policía Nacional realiza o cálculo dividindo a superficie pola que transcorre a manifestación entre o número de persoas que caben por metro cadrado e que segundo as súas estimacións pode ser entre dous e catro dependendo de se a manifestación está parada, é concentración, ou é marcha porque en movemento o número de persoas sería menor.

(Fonte: noticia de COPE)

Obxectivos

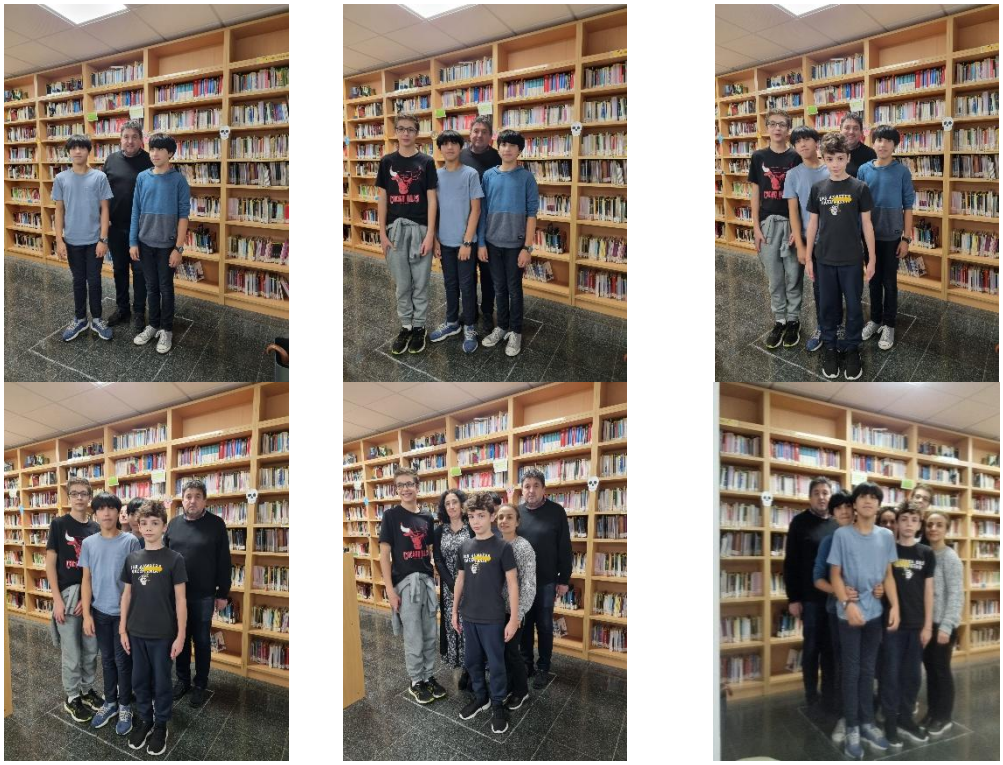
Obxectivo xeral do traballo: Crear o noso método para estimar o número de persoas asistentes a unha concentración contando os asistentes en pequenas rexións aleatorias.

Obxectivos secundarios:

1. Adquirir máis experiencia no uso dos programas informáticos de uso común (excel, word, etc).
2. Mellorar as nosas habilidades estatísticas e matemáticas.
3. Aprender a redactar e comunicar os resultados da nosa investigación.

Deseño inicial do experimento e recollida de datos

a) O primeiro que pensamos foi que hai moita diferenza entre concentracións multitudinarias (de moita xente) e aquelas menos concorridas. Así que fixemos o noso propio estudo: **Cantas persoas poden caber nun metro cadrado?**
Para calcular isto só tivemos que pintar un metro cadrado no chan da biblioteca e meternos nel.



A nosa conclusión, e pensando nunha manifestación masiva, foi que o máximo número de persoas por metro cadrado é 8.

b) Como podiamos facer unha **simulación de concentración** de persoas para realizar os nosos experimentos?
Ao principio pensamos en crear nós algunha maqueta, pero usando o que tiñamos á man, ocorréusenos utilizar taboleiros de xadrez que xa estaban divididos en seccións e podían aforrarnos moito traballo. Para simular as persoas, compramos pezas de tamaño proporcional (fixámonos que entrasen como moito 8 pezas en cada cuadrícula) e quedounos algo así:

Con isto, poderiamos simular calquera concentración dun máximo de 512 (8x64) pezas por taboleiro.

Para facer as simulacións bastaría con coller un número aleatorio de pezas entre 0 e 512 e tiralas enriba do taboleiro obtendo a nosa maqueta. Ademais, como o taboleiro xa está dividido en seccións, podiamos utilizalas como rexións independentes.

c) Tiñamos claro que o noso método consistiría en contar as pezas de certas cuadrículas ao chou e con iso estimar o total. Pero **cantos cadrados contamos?** Está claro que cantos máis contemos mellor debería ser a estimación, pero contar moitos cadrados ía ser pouco operativo nun caso real. Decidimos facelo de varias formas para comparar os resultados obtidos e ao final poder decidir cal delas sería a máis apropiada:

- ✓ CASO1: contar o 50% do territorio, é dicir, 32 cadrados.
- ✓ CASO2: contar o 25% do territorio, é dicir, 16 cadrados.
- ✓ CASO3: contar o 12,5% do territorio, é dicir, 8 cadrados.
- ✓ CASO4: contar o 6,25% do territorio, é dicir, 4 cadrados.
- ✓ CASO5: contar o 3,125% do territorio, é dicir, 2 cadrados.

d) Por suposto, para estimar o número total de persoas é moi importante estimar primeiro a densidade, é dicir, se hai moita ou pouca xente concentrada. Decidimos facer a seguinte distinción (dividimos en tres seccións a cantidade máxima de poboación):

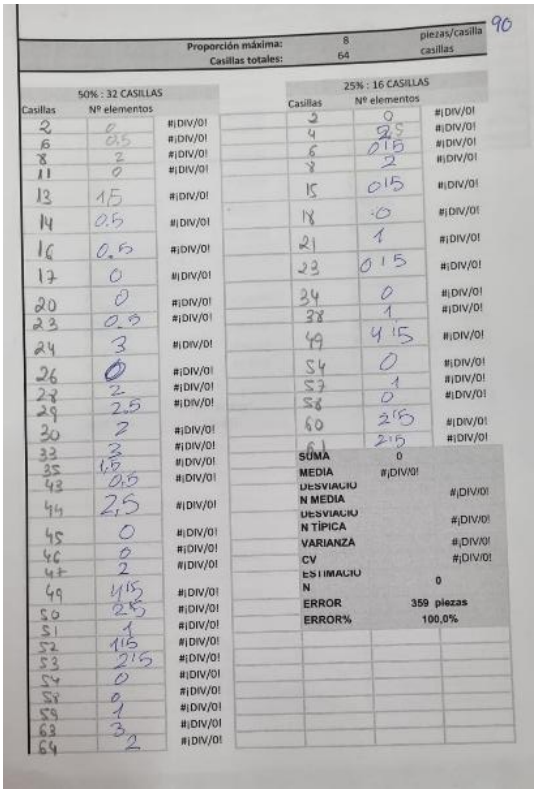
- ✓ Densidade alta: entre 341 e 512 pezas, é dicir, entre 5,3 e 8 persoas por metro cadrado.
- ✓ Densidade media: entre 171 e 341 pezas, é dicir, entre 2,7 e 5,3 persoas por metro cadrado.
- ✓ Densidade baixa: entre 0 e 171 pezas, é dicir, entre 0 e 2,7 persoas por metro cadrado (aínda que a isto tivemos que facerlle un apaño máis adiante)

Aquí temos un exemplo de taboleiro de cada caso:

e) Así xa tiñamos o **experimento inicial** deseñado (dicimos inicial porque ao longo do proceso foi sufrindo modificacións). Para cada unha das densidades faríamos o seguinte proceso:

- ✓ Cunha calculadora de números aleatorios elixiriamos un número de pezas ao chou (sempre dentro dos límites desa densidade)
- ✓ Tiramos ese número de pezas enriba do taboleiro de xadrez (no que previamente numeramos os cadrados do 1 ao 64)
- ✓ Outra vez coa calculadora de números aleatorios elixiriamos 32 cadrados ao azar para cubrir os datos do caso1, outros 16 cadrados para cubrir os datos do caso2, outros 4 para o caso3 e dous cadrados para o caso 4.

Todos estes datos empezamos cubríndoos nun papel, para pasalos despois a unha folla excel.



90

Proporción máxima:	8	piezas/casilla
Casillas totales:	64	casillas

50% : 32 CASILLAS		
Casillas	Nº elementos	
2	0	#¡DIV/0!
6	0.5	#¡DIV/0!
8	2	#¡DIV/0!
11	0	#¡DIV/0!
13	1.5	#¡DIV/0!
14	0.5	#¡DIV/0!
16	0.5	#¡DIV/0!
17	0	#¡DIV/0!
20	0	#¡DIV/0!
23	0.5	#¡DIV/0!
24	3	#¡DIV/0!
26	0	#¡DIV/0!
28	2	#¡DIV/0!
29	2.5	#¡DIV/0!
30	2	#¡DIV/0!
33	3	#¡DIV/0!
35	1.5	#¡DIV/0!
43	0.5	#¡DIV/0!
44	2.5	#¡DIV/0!
45	0	#¡DIV/0!
46	0	#¡DIV/0!
47	2	#¡DIV/0!
49	4.5	#¡DIV/0!
50	2.5	#¡DIV/0!
51	1	#¡DIV/0!
52	1.5	#¡DIV/0!
53	2.5	#¡DIV/0!
54	0	#¡DIV/0!
58	0	#¡DIV/0!
59	1	#¡DIV/0!
63	3	#¡DIV/0!
64	2	#¡DIV/0!

25% : 16 CASILLAS		
Casillas	Nº elementos	
2	0	#¡DIV/0!
4	2.5	#¡DIV/0!
6	0.5	#¡DIV/0!
8	2	#¡DIV/0!
15	0.5	#¡DIV/0!
18	0	#¡DIV/0!
21	1	#¡DIV/0!
23	0.5	#¡DIV/0!
34	0	#¡DIV/0!
38	1	#¡DIV/0!
49	4.5	#¡DIV/0!
54	0	#¡DIV/0!
57	1	#¡DIV/0!
58	0	#¡DIV/0!
60	2.5	#¡DIV/0!
61	2.5	#¡DIV/0!

SUMA	0	
MEDIA	#¡DIV/0!	
DESVIACIÓN MEDIA		#¡DIV/0!
DESVIACIÓN TÍPICA		#¡DIV/0!
VARIANZA		#¡DIV/0!
CV		#¡DIV/0!
ESTIMACIÓN	0	
ERROR	359 piezas	
ERROR%	100,0%	

f) **Cantas veces faríamos o experimento?** Outra vez temos que falar da nosa idea inicial, porque despois isto cambiaría totalmente. Pensamos que poderíamos facer 50 veces o experimento en cada unha das densidades. Cando comezamos co traballo era o mes de outubro de 2024 e críamos que teríamos tempo de sobra para facer todo isto

g) **Como estimamos o total de persoas?** Decidimos empezar partindo da idea de Herbert Jacobs e facelo da maneira máis sinxela. En cada un dos casos: calculamos o número medio de persoas por cadrado e multiplicamos ese promedio polo número total de cadrados. Como tiñamos claro que esa estimación non sería acertada na maioría dos casos decidimos construír intervalos utilizando o erro cometido (parecía fácil cando o pensamos ao principio pero non sabiamos onde nos estabamos metendo)

h) **Como estimamos o erro cometido?** Tamén era moi fácil, tiñamos o dato real, o dato estimado e a resta era o erro cometido.

i) **A recompilación destes datos** iámola facendo en tres documentos excel creados para tal fin (un documento para cada densidade). Estes documentos fixémolos coas fórmulas correspondentes, de maneira que ao proporcionarlle as fichas en cada un dos cadrados, nos calculara todo o que podíamos necesitar: a suma das fichas contadas, a media de fichas por cadrado segundo a estimación, o erro cometido en valor absoluto e en porcentaxe e outros datos que nos podían facer falla despois como a desviación típica, a varianza e o coeficiente de variación de Pearson. Para isto dedicamos un par de sesións do taller de estatística, xa que non sabiamos como se podían introducir estas fórmulas usando o excel e aínda que en clase xa déramos a parte de estatística fíxonos falla que as nosas profesoras nos recordaran que era e para que servía cada unha destas medidas de centralización e dispersión.

Así comezou inicialmente o noso experimento alá polo mes de novembro de 2024.

Transcurso e transformacións do experimento. Traballo de campo

Estivemos facendo isto hasta mediados de decembro, e foi cando empezamos a darnos conta das dificultades respecto ao deseño inicial: só levabamos feito 32 tiradas en total (14 en densidade baixa, 11 en densidade media e 7 en densidade alta). A pesar de que as nosas profesoras tamén estaban facendo tiradas con nós e axudándonos era imposible que a este ritmo fixéramos as 50 tiradas por densidade e elaboráramos as conclusións a tempo de presentar este traballo á incubadora. A isto engadimos que cando nos tocaba unha tirada de densidade alta era moi difícil contar as fichas nos cadrados e tamén notamos que cando tirabamos as fichas encima do taboleiro era moi probable que non se repartían ao chou senón que tendían a concentrarse nalgunha zona.

Como podíamos solucionalo? Ou cambiábamos a forma de realizar o experimento ou baixábamos as nosas expectativas co número de tiradas (o cal faría que os resultados perderan moita fiabilidade).

Decidímonos por botar man da Intelixencia artificial. As nosas profesoras avisáronnos de que elas non estaban moi postas neste campo, polo que sería algo que teríamos que aprender entre todos.

Para comezar intentamos crear imaxes de concentracións de xente nun espazo de dimensións coñecidas. Necesitabamos imaxes vistas dende arriba para poder contar as seccións elixidas. Isto non funcionou, non sabemos se foi culpa nosa ou isto é algo que polo de agora e coas nosas posibilidades non se pode crear con IA.

A seguinte idea foi crear unha versión dos nosos taboleiros mediante IA. Para isto usamos (despois de moitos intentos que fracasaron por múltiples razóns) o seguinte promt en CHATGPT onde os números a e b cambiaban segundo a densidade dos taboleiros que queríamos crear:

> Por favor, genera un archivo excel con 100 hojas creadas de forma similar
>
> • En cada hoja:
>
> -La primera columna tendrá 64 celdas con un número aleatorio entre a y b (llamado p) en la primera celda y en el resto copiará ese mismo número.
>
> -La segunda columna será una lista de números naturales en orden creciente del 1 al 64.
>
> -La tercera columna contendrá 64 números aleatorios menores o iguales que 8 (puede usarse el cero) que sumen exactamente p (el número aleatorio generado en la primera columna). Es muy importante que en esta columna los números sean aleatorios y que no superen el 8.

Neste punto descubrimos algunhas cousas de CHATGPT, seguramente debido tamén a nosa inexperiencia:

- A CHATGPT gústalle máis o castelán que o galego.
- Nunca di que non o sabe facer, polo que hai que revisar cada cousa que fai e asegurarse que efectivamente é o que ti querías.
- Se non pagas a versión premium tes un número limitado de intentos cando xeras un contido, polo que o proceso de proba e erro dilatouse no tempo xa que enseguida se nos acababan estes intentos.
- O promt ten que ter ordes moi precisas en non deixar nada ao azar, e incluso nese caso non sempre consegues o que esperas.

A pesar de todo, conseguimos ter cen taboleiros creados con IA en cada unha das densidades.

O seguinte paso foi elixir os números aleatorios que nos indicaran cales eran os cadrados que tiñamos que contar en cada un dos taboleiros. Este foi o promt que finalmente usamos:

> Por favor, crea un documento excel, que se llamará números aleatorios, con 5 hojas con las siguientes características:
>
> - En la primera hoja debe haber 300 columnas, en cada columna habrá 32 números aleatorios que no se repitan mayores o iguales que 1 y menores o iguales que 64. Aparecerán ordenados de menor a mayor
> - En la segunda hoja debe haber 300 columnas, en cada columna habrá 16 números aleatorios que no se repitan mayores o iguales que 1 y menores o iguales que 64. Aparecerán ordenados de menor a mayor
> - En la tercera hoja debe haber 300 columnas, en cada columna habrá 8 números aleatorios que no se repitan mayores o iguales que 1 y menores o iguales que 64. Aparecerán ordenados de menor a mayor
> - En la cuarta hoja debe haber 300 columnas, en cada columna habrá 4 números aleatorios que no se repitan mayores o iguales que 1 y menores o iguales que 64. Aparecerán ordenados de menor a mayor
> - En la quinta hoja debe haber 300 columnas, en cada columna habrá 2 números aleatorios que no se repitan mayores o iguales que 1 y menores o iguales que 64. Aparecerán ordenados de menor a mayor

Con isto puxémonos a traballar de novo, botando man destes excel creamos a base de datos suficiente con máis de cen experimentos realizados en cada unha das densidades. A pesar de estar usando IA, trasladar estes datos ao noso excel levounos moito tempo e chegamos ao mes de febreiro con este traballo de campo.

Análise estatística e interpretación dos resultados obtidos

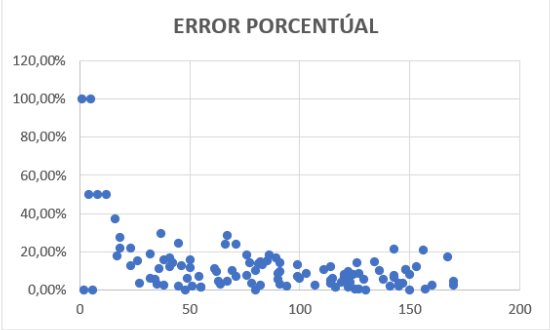
a) Estudo das estimacións e dos erros cometidos.

Estudamos os resultados de cada densidade por separado. Aquí explicaremos polo miúdo o estudo realizado en densidade baixa e presentaremos as conclusións dos realizados en densidade media e en densidade alta (xa que son estudos similares cada un cos seus datos).

Ademais en cada unha das densidades consideramos os cinco casos segundo a porcentaxe do territorio estudado:

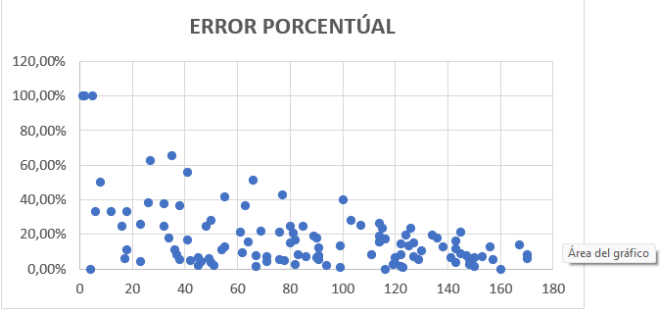
No caso de densidade baixa, o primeiro que observamos e que na primeira parte da nube de puntos os erros son moi grandes polo que parece lóxico pensar que este método non é valido cando hai poucas persoas (no resto das densidades non ocorría isto). Nas seguintes táboas podemos ver a comparativa entre ter en conta o total de datos (cor verde), eliminar os datos de 20 persoas ou menos (cor azul) e eliminar os datos de corenta persoas ou menos (cor violeta):

CANDO ANALIZAMOS O 50% DO TERRITORIO



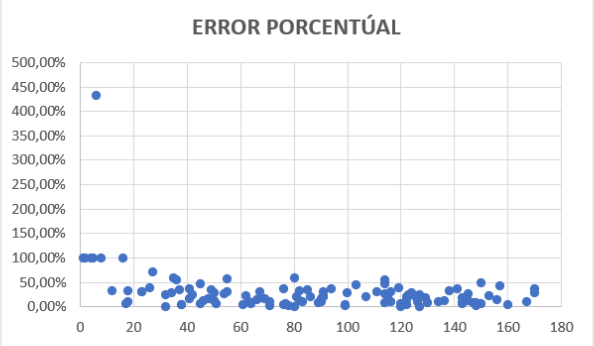
	ERROR PORCENTÚAL MEDIO	DESVIACIÓN TÍPICA	CVP	ERRO MÍNIMO	ERRO MÁXIMO	Número de datos
Valor real entre 0 y 171 0 ≤ d < 2,7	12,12%	0,153952467	1,26996	0	100,00%	114
Valor real entre 20 y 171 0,3 ≤ d < 2,7	9,00%	0,067972556	0,75538752	0	29,73%	103
Valor real entre 40 y 171 0,6 ≤ d < 2,7	8,57%	0,064786523	0,75588185	0	28,36%	91

CANDO ANALIZAMOS O 25% DO TERRITORIO



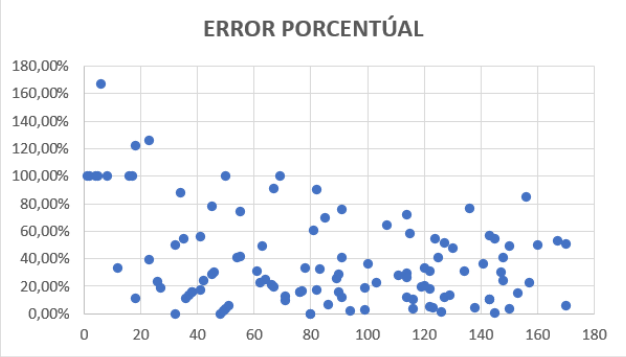
	ERROR PORCENTÚAL MEDIO	DESVIACIÓN TÍPICA	CVP	ERRO MÍNIMO	ERRO MÁXIMO	Número de datos
Valor real entre 0 y 171 0 ≤ d < 2,7	18,05%	0,19277357	1,06814915	0	100,00%	114
Valor real entre 20 y 171 0,3 ≤ d < 2,7	15,20%	0,13545759	0,8912739	0	65,71%	103
Valor real entre 40 y 171 0,6 ≤ d < 2,7	13,48%	0,11353712	0,84247672	0	56,10%	91

CANDO ANALIZAMOS O 12,5% DO TERRITORIO



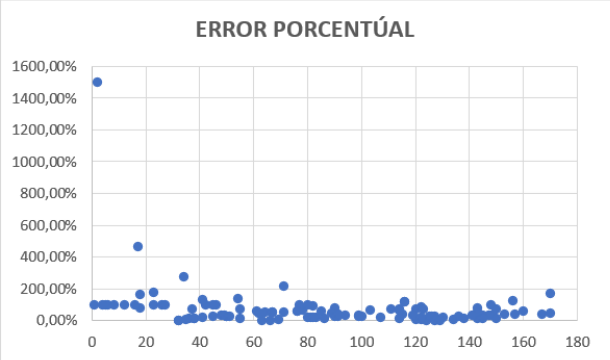
	ERROR PORCENTÚAL MEDIO	DESVIACIÓN TÍPICA	CVP	ERRO MÍNIMO	ERRO MÁXIMO	Número de datos
Valor real entre 0 y 171 0 ≤ d < 2,7	29,25%	0,446526088	1,526532455	0	433,33%	114
Valor real entre 20 y 171 0,3 ≤ d < 2,7	21,53%	0,15637557	0,726304576	0	70,37%	103
Valor real entre 40 y 171 0,6 ≤ d < 2,7	20,14%	0,14159787	0,703240205	0	60,00%	91

CANDO ANALIZAMOS O 6,25% DO TERRITORIO



	ERROR PORCENTÚAL MEDIO	DESVIACIÓN TÍPICA	CVP	ERROR MÍNIMO	ERROR MÁXIMO	Número de datos
Valor real entre 0 y 171 0 ≤ d < 2,7	38,44%	0,33707034	0,87679644	0	166,67%	114
Valor real entre 20 y 171 0,3 ≤ d < 2,7	32,52%	0,269368815	0,82840298	0	126,09%	103
Valor real entre 40 y 171 0,6 ≤ d < 2,7	31,80%	0,255080081	0,80217285	0	100,00%	91

CANDO ANALIZAMOS O 3,125% DO TERRITORIO



	ERROR PORCENTÚAL MEDIO	DESVIACIÓN TÍPICA	CVP	ERROR MÍNIMO	ERROR MÁXIMO	Número de datos
Valor real entre 0 y 171 0 ≤ d < 2,7	73,00%	1,479745543	2,02714251	0	1500,00%	114
Valor real entre 20 y 171 0,3 ≤ d < 2,7	52,55%	0,472329182	0,89884941	0	276,47%	103
Valor real entre 40 y 171 0,6 ≤ d < 2,7	49,82%	0,397052536	0,79699552	0,00775194	215,49%	91

O estudo en densidade baixa supuxo todo un reto, xa que os erros eran altísimos cando se estudaba pouco territorio e non tiña sentido cando o número de persoas era moi baixo. En densidade media e densidade alta o estudo foi moito máis sinxelo.

As conclusións xerais deste estudo foron:

		Erro porcentual mínimo	Erro porcentual máximo	Erro porcentual medio
DENSIDADE BAIXA (ENTRE 0,6 E 2,7 PERSOAS POR m²)	50% do territorio	0%	28,36%	8,57%
	25% do territorio	0%	56,10%	13,48%
	12,5% do territorio	0%	60%	20,14%
	6,25% do territorio	0%	100%	31,8%
	3,125% do territorio	0%	215,49%	49,82%
DENSIDADE MEDIA (ENTRE 2,7 E 5,3 PERSOAS POR m²)	50% do territorio	0%	31,03%	7,54%
	25% do territorio	0%	42,25%	10,98%
	12,5% do territorio	0%	54,62%	16,36%
	6,25% do territorio	0,65%	91,96%	27,33%
	3,125% do territorio	0%	146,15%	37,77%
DENSIDADE ALTA (ENTRE 5,3 E 8 PERSOAS POR m²)	50% do territorio	0%	15,45%	2,58%
	25% do territorio	0%	14,75%	3,73%
	12,5% do territorio	0%	29,14%	6,51%
	6,25% do territorio	0,2%	39,3%	10,24%
	3,125% do territorio	0,20%	62,35%	12,75%

Como é lóxico, o erro diminúe a medida que aumenta o número de persoas a estimar e aumenta a medida que diminúe o terreo estudado. No caso de densidade baixa, para poder facer unha boa estimación necesitaremos coller, como mínimo un 12,5% do territorio e asegurar un mínimo de 40 persoas.

b) Agora, a partir destas estimacións e usando os erros cometidos propuxémonos crear intervalos que servirían como estimación.

A nosa primeira idea foi moi sinxela: colleríamos o erro medio das nosas tiradas (que ao usar a IA eran máis de cen) e sumando e restando ese erro á estimación obtida tiñamos o intervalo de estimación. Así nos quedou:

✓ Cando a densidade é baixa:

TERRITORIO ESTUDADO	INTERVALO DE ESTIMACIÓN
50%	$(valor\ estimado - 8,57\%, valor\ estimado + 8,57\%)$
25%	$(valor\ estimado - 13,48\%, valor\ estimado + 13,48\%)$
12,5%	$(valor\ estimado - 20,14\%, valor\ estimado + 20,14\%)$
6,25%	$(valor\ estimado - 31,8\%, valor\ estimado + 31,8\%)$
3,125%	$(valor\ estimadoo - 49,82\%, valor\ estimado + 49,82\%)$

✓ Cando a densidade é media:

TERRITORIO ESTUDADO	INTERVALO ESTIMADO
50%	$(valor\ estimado - 7,45\%, valor\ estimado + 7,54\%)$
25%	$(valor\ estimado - 10,98\%, valor\ estimado + 10,98\%)$
12,5%	$(valor\ estimado - 16,36\%, valor\ estimado + 16,36\%)$
6,25%	$(valor\ estimado - 27,33\%, valor\ estimado + 27,33\%)$
3,125%	$(valor\ estimadoo - 37,77\%, valor\ estimado + 37,77\%)$

✓ Cando a densidade é alta:

TERRITORIO ESTUDADO	INTERVALO ESTIMADO
50%	$(valor\ estimado - 2,58\%, valor\ estimado + 2,58\%)$
25%	$(valor\ estimado - 3,73\%, valor\ estimado + 3,73\%)$
12,5%	$(valor\ estimado - 6,51\%, valor\ estimado + 6,51\%)$
6,25%	$(valor\ estimado - 10,24\%, valor\ estimado + 10,24\%)$
3,125%	$(valor\ estimado - 12,75\%, valor\ estimado + 12,75\%)$

Con isto xa tiñamos o noso **primeiro método de estimación**:

Contar o territorio que decidamos segundo o erro que esteamos dispostos a cometer, estimar o número de persoas e, segundo a densidade dese número de persoas, sumar e restar esas porcentaxes para ter un intervalo de estimación.

Por suposto, quixemos poñer a proba o noso método. Pensamos en facer unha proba na realidade intentando xuntar nun espazo a moita xente e estimar este número co noso método, pero démonos conta que só sería unha proba e no caso de que saíra mal non poderíamos saber se fallaba nese caso ou en xeral. Como non era viable (por razóns de loxística) facer moitas probas reais, decidimos volver usar a IA.

Como xa estabamos no mes de marzo e non nos daba tempo a repetir todo o proceso de xerar outros taboleiros, números aleatorios e pasar os datos ao noso excel, pensamos que poderíamos intentar encargar á IA, que, ademais de xerar os taboleiros e os números aleatorios, tamén nos dera as fórmulas necesarias para que a partir deses datos apareceran cubertas as celas no excel. (A verdade é que se isto se nos houbera ocorrido antes, teríamos aforrado moito traballo pero, como nos din as profesoras, tamén é importante o traballo que fixemos para poder valorar todo o que nos aforran as novas tecnoloxías sen as cales non poderíamos ter feito o noso traballo).

Este foi o resultado de poñer a proba o método con cen novos taboleiros:

- ✓ Densidade de poboación é baixa:

TERRITORIO ESTUDADO	INTERVALO ESTIMADO	TAMAÑO MEDIO	PORCENTAXE DE ACIERTO
50%	$(valor\ estimado - 8,57\%, valor\ estimado + 8,57\%)$	18	43%
25%	$(valor\ estimado - 13,48\%, valor\ estimado + 13,48\%)$	27	20%
12,5%	$(valor\ estimado - 20,14\%, valor\ estimado + 20,14\%)$	41	27%
6,25%	$(valor\ estimado - 31,8\%, valor\ estimado + 31,8\%)$	64	33%
3,125%	$(valor\ estimadoo - 49,82\%, valor\ estimado + 49,82\%)$	91	31%

- ✓ Cando a densidade é media:

TERRITORIO ESTUDADO	INTERVALO ESTIMADO	TAMAÑO MEDIO	PORCENTAXE DE ACIERTO
50%	$(valor\ estimado - 7,54\%, valor\ estimado + 7,54\%)$	40	56%
25%	$(valor\ estimado - 10,98\%, valor\ estimado + 10,98\%)$	58	53%
12,5%	$(valor\ estimado - 16,36\%, valor\ estimado + 16,36\%)$	85	62%
6,25%	$(valor\ estimado - 27,33\%, valor\ estimado + 27,33\%)$	142	68%
3,125%	$(valor\ estimado - 37,77\%, valor\ estimado + 37,77\%)$	238	70%

- ✓ Cando a densidade é alta:

TERRITORIO ESTUDADO	INTERVALO ESTIMADO	TAMAÑO MEDIO	PORCENTAXE DE ACIERTO
50%	$(valor\ estimado - 2,58\%, valor\ estimado + 2,58\%)$	22	36%
25%	$(valor\ estimado - 3,73\%, valor\ estimado + 3,73\%)$	32	43%
12,5%	$(valor\ estimado - 6,51\%, valor\ estimado + 6,51\%)$	54	46%
6,25%	$(valor\ estimado - 10,24\%, valor\ estimado + 10,24\%)$	87	47%
3,125%	$(valor\ estimado - 12,75\%, valor\ estimado + 12,75\%)$	109	41%

Que chasco! A porcentaxe de éxito é moi baixa, incluso cando estudamos unha gran parte do territorio. Ademais, chama a atención que no caso de densidade media a porcentaxe de acerto é bastante grande pero os intervalos son moi amplos; no caso de densidade alta redúcense estes tamaños considerablemente, pero baixa a porcentaxe de acerto.

Este método non parece adecuado para estimar a poboación nunha concentración de xente.

c) Tras este fracaso, tivemos que buscar outra idea. Esta vez coa axuda das nosas profesoras pensamos en usar os errores máximos. A idea é a seguinte: fixamos a porcentaxe de casos que queremos acertar (nivel de confianza), e collemos o erro máximo cometido nesa porcentaxe de casos. Este erro será o que usemos para crear os intervalos de estimación. Está claro que os intervalos serán máis grandes que no primeiro intento, pero esperemos aumentar considerablemente a porcentaxe de acertos.
Por suposto, agora que xa tiñamos dominado a creación de taboleiros con IA, despois de crear estes novos intervalos puxémolos a proba con cen novos taboleiros:

- ✓ Comezamos por construír os intervalos de densidade baixa. Na seguinte táboa construímos todos intervalos pero observaremos que nos casos 4 e 5 non teñen moito sentido (como xa viñemos observando), xa que os intervalos son moi grandes e a porcentaxe de acerto non compensa. Ademais a diferenza entre a porcentaxe de erro e a confianza é considerable. Supoñemos que isto pasa porque unha mostra de cen elementos non aporta toda a fiabilidade que buscamos. Pero non está tan mal.

TERRITORIO ESTUDADO	CONFIANZA	MAX ERROR	INTERVALO ESTIMADO	TAMAÑO MEDIO	PORCENTAJE DE ACIERTO	ERROR MEDIO
50%	99% casos	24,44%	$(valor\ estim. - 24,44\%, valor\ estim. + 24,44\%)$	54	84%	9
	95% casos	21,15%	$(valor\ estim - 21,15\%, valor\ estim. + 21,15\%)$	42	78%	9
	90% casos	17,07%	$(valor\ estim. - 17,07\%, valor\ estim. + 17,07\%)$	34	69%	9
	85% casos	14,63%	$(valor\ estim. - 14,63\%, valor\ estim. + 14,63\%)$	30	64%	9
	80% casos	14,29%	$(valor\ estim. - 14,29\%, valor\ estim. + 14,29\%)$	29	61%	9
	75% casos	12,42%	$(valor\ estim. - 12,42\%, valor\ estim. + 12,42\%)$	25	56%	9
	70% casos	11,48%	$(valor\ estim - 11,48\%, valor\ estim. + 11,48\%)$	23	53%	9
25%	99% casos	51,52%	$(valor\ estim. - 51,52\%, valor\ estim. + 51,52\%)$	104	74%	24
	95% casos	36,51%	$(valor\ estim - 36,51\%, valor\ estim. + 36,51\%)$	74	64%	21
	90% casos	25,23%	$(valor\ estim. - 25,23\%, valor\ estim. + 25,23\%)$	50	46%	21
	85% casos	23,48%	$(valor\ estim. - 23,48\%, valor\ estim + 23,48\%)$	46	36%	20
	80% casos	21,05%	$(valor\ estim. - 21,05\%, valor\ estim. + 21,05\%)$	42	35%	21
	75% casos	19,10%	$(valor\ estim. - 19,10\%, valor\ estim + 19,10\%)$	38	31%	22
	70% casos	17,07%	$(valor\ estim. - 17,07\%, valor\ estim. + 17,07\%)$	34	28%	23
12,5%	99% casos	56,36%	$(valor\ estim - 56,36\%, valor\ estim. + 56,36\%)$	116	67%	42
	95% casos	46,67%	$(valor\ estim. - 46,67\%, valor\ estim. + 46,67\%)$	96	63%	40
	90% casos	37,59%	$(valor\ estim - 37,59\%, valor\ estim. + 37,59\%)$	77	52%	36
	85% casos	34,69%	$(valor\ estim - 34,69\%, valor\ estim. + 34,69\%)$	60	49%	39
	80% casos	31,53%	$(valor\ estim. - 31,53\%, valor\ estim. + 31,53\%)$	65	48%	37
	75% casos	29,41%	$(valor\ estim. - 29,41\%, valor\ estim. + 29,41\%)$	60	44%	36
	70% casos	26,90%	$(valor\ estim. - 26,9\%, valor\ estim + 26,9\%)$	55	40%	36
6,25%	99% casos	100%	$(valor\ estim. - 100\%, valor\ estim. + 100\%)$	200	73%	57
	95% casos	77,78%	$(valor\ estim. - 77,78\%, valor\ estim. + 77,78\%)$	155	70%	53
	90% casos	71,93%	$(valor\ estim. - 71,93\%, valor\ estim. + 71,93\%)$	144	66%	48
	85% casos	56,64%	$(valor\ estim - 56,64\%, valor\ estim. + 56,64\%)$	113	61%	45
	80% casos	53,29%	$(valor\ estim - 53,29\%, valor\ estim. + 53,29\%)$	106	54%	40
	75% casos	49,21%	$(valor\ estim. - 49,21\%, valor\ estim. + 49,21\%)$	98	51%	39
	70% casos	40,74%	$(valor\ estim. - 40,74\%, valor\ estim. + 40,74\%)$	81	43%	39
3,125%	99% casos	172,94%	$(valor\ estim. - 172,9\%, valor\ estim. + 172,9\%)$	251	55%	69
	95% casos	120,69%	$(valor\ estim. - 120,7\%, valor\ estim. + 120,7\%)$	213	80%	23
	90% casos	100%	$(valor\ estim. - 100\%, valor\ estim. + 100\%)$	195	78%	42
	85% casos	83,61%	$(valor\ estim. - 83,61\%, valor\ estim. + 83,61\%)$	154	49%	68
	80% casos	73,98%	$(valor\ estim - 73,98\%, valor\ estim. + 73,98\%)$	136	45%	64
	75% casos	68,93%	$(valor\ estim - 68,93\%, valor\ estim. + 68,93\%)$	127	45%	66
	70% casos	62,35%	$(valor\ estim - 62,35\%, valor\ estim. + 62,35\%)$	114	42%	64

- ✓ Vexamos como son estas porcentaxes en densidade media: nos casos 4 e 5 seguen sen ter moito sentido, xa que os intervalos son moi grandes aínda que agora a porcentaxe de acerto é maior.

TERRITORIO ESTUDADO	CONFIANZA	MAX ERROR	INTERVALO ESTIMADO	TAMAÑO MEDIO	PORCENTAJE DE ACIERTO	ERROR MEDIO
50%	99% casos	28,64%	$(valor\ estim-28,64\%, valor\ estim+28,64\%)$	151	99%	9
	95% casos	17,65%	$(valor\ estim-17,65\%, valor\ estim+17,65\%)$	93	97%	10
	90% casos	16,55%	$(valor\ estim-16,55\%, valor\ estim.+16,55\%)$	87	96%	9
	85% casos	15,35%	$(valor\ estim-15,35\%, valor\ estim+15,35\%)$	81	93%	7
	80% casos	11,95%	$(valor\ estim-11,95\%, valor\ estim.+11,95\%)$	63	84%	8
	75% casos	9,92%	$(valor\ estim-9,92\%, valor\ estim+9,92\%)$	52	75%	9
	70% casos	9,09%	$(valor\ estim-9,09\%, valor\ estim+9,09\%)$	48	71%	10
25%	99% casos	38,93%	$(valor\ estim-38,93\%, valor\ estim+38,93\%)$	207	99%	6
	95% casos	27,42%	$(valor\ estim-27,42\%, valor\ estim+27,42\%)$	146	94%	13
	90% casos	23,62	$(valor\ estim-23,62\%, valor\ estim+23,62\%)$	125	89%	13
	85% casos	19,64%	$(valor\ estim-19,64\%, valor\ estim+19,64\%)$	104	82%	15
	80% casos	17,48%	$(valor\ estim-17,48\%, valor\ estim+17,48\%)$	93	80%	18
	75% casos	15,25%	$(valor\ estim.-15,25\%, valor\ estim+15,25\%)$	81	69%	16
	70% casos	13,21%	$(valor\ estim-13,21\%, valor\ estim.+13,21\%)$	70	62%	18
12,5%	99% casos	49,65%	$(valor\ estim-49,65\%, valor\ estim+49,65\%)$	258	90%	31
	95% casos	40,82%	$(valor\ estim-40,82\%, valor\ estim+40,82\%)$	212	88%	37
	90% casos	33,85%	$(valor\ estim-33,85\%, valor\ estim+33,85\%)$	176	86%	41
	85% casos	30,71%	$(valor\ estim-30,71\%, valor\ estim+30,71\%)$	133	63%	35
	80% casos	25,79%	$(valor\ estim.-25,79\%, valor\ estim.+25,79\%)$	134	82%	45
	75% casos	23,02%	$(valor\ estim-23,02\%, valor\ estim+23,02\%)$	120	80%	46
	70% casos	21,08%	$(valor\ estim-21,08\%, valor\ estim+21,08\%)$	110	74%	38
6,25%	99% casos	84,62%	$(valor\ estim-84,62\%, valor\ estim+84,62\%)$	437	93%	46
	95% casos	62,03%	$(valor\ estim-62,03\%, valor\ estim+62,03\%)$	320	92%	66
	90% casos	54,40%	$(valor\ estim.-54,40\%, valor\ estim.+54,40\%)$	281	91%	66
	85% casos	44,72%	$(valor\ estim-44,72\%, valor\ estim+44,72\%)$	231	85%	49
	80% casos	43,72%	$(valor\ estim-43,72\%, valor\ estim+43,72\%)$	226	85%	51
	75% casos	38,93%	$(valor\ estim-38,93\%, valor\ estim+38,93\%)$	201	82%	48
	70% casos	36,28%	$(valor\ estim-36,28\%, valor\ estim+36,28\%)$	187	80%	48
3,125%	99% casos	100,00%	$(valor\ estim-100\%, valor\ estim+100\%)$	540	84%	58
	95% casos	85,84%	$(valor\ estim.-85,84\%, valor\ estim.+85,84\%)$	460	95%	23
	90% casos	79,15%	$(valor\ estim-79,15\%, valor\ estim.+79,15\%)$	420	95%	36
	85% casos	69,01%	$(valor\ estim.-69,01\%, valor\ estim+69,01\%)$	373	81%	74
	80% casos	64,04%	$(valor\ estim-64,04\%, valor\ estim+64,04\%)$	346	81%	79
	75% casos	51,76%	$(valor\ estim-51,76\%, valor\ estim.+51,76\%)$	280	80%	87
	70% casos	48,84%	$(valor\ estim-48,84\%, valor\ estim+48,84\%)$	264	77%	78

- ✓ E, por último, vexamos como son en densidade alta: volvemos as mesmas conclusións, os casos 4 e 5 non compensan.

TERRITORIO ESTUDADO	CONFIANZA	MAX ERROR	INTERVALO ESTIMADO	TAMAÑO MEDIO	PORCENTAJE DE ACIERTO	ERROR MEDIO
50%	99% casos	12,5%	$(valor\ estim-12,5\%, valor\ estim+12,5\%)$	106	98%	11
	95% casos	7,6%	$(valor\ estim.-7,6\%, valor\ estim.+7,6\%)$	64	87%	12
	90% casos	5,44%	$(valor\ estim-5,44\%, valor\ estim+5,44\%)$	46	73%	12
	85% casos	4,88%	$(valor\ estim.-4,88\%, valor\ estim+4,88\%)$	41	69%	13
	80% casos	4,31%	$(valor\ estim-4,31\%, valor\ estim+4,31\%)$	36	60%	12
	75% casos	3,9%	$(valor\ estim-3,9\%, valor\ estim+3,9\%)$	33	58%	13
	70% casos	3,07%	$(valor\ estim.-3,07\%, valor\ estim.+3,07\%)$	38	47%	26
25%	99% casos	13,54%	$(valor\ estim.-13,54\%, valor\ estim+13,54\%)$	116	83%	17
	95% casos	11,48%	$(valor\ estim-11,48\%, valor\ estim+11,48\%)$	98	80%	22
	90% casos	7,62%	$(valor\ estim-7,62\%, valor\ estim+7,62\%)$	65	67%	25
	85% casos	6,48%	$(valor\ estim-6,48\%, valor\ estim+6,48\%)$	55	64%	27
	80% casos	6,08%	$(valor\ estim-6,08\%, valor\ estim+6,08\%)$	52	57%	27
	75% casos	5,41%	$(valor\ estim-5,41\%, valor\ estim+5,41\%)$	46	51%	24
	70% casos	4,43%	$(valor\ estim.-4,43\%, valor\ estim.+4,43\%)$	38	47%	26
12,5%	99% casos	28,69%	$(valor\ estim-28,69\%, valor\ estim+28,69\%)$	239	94%	41
	95% casos	17,09%	$(valor\ estim-17,09\%, valor\ estim+17,09\%)$	143	79%	33
	90% casos	14,85%	$(valor\ estim-14,85\%, valor\ estim+14,85\%)$	124	71%	31
	85% casos	12,5%	$(valor\ estim-12,5\%, valor\ estim+12,5\%)$	88	48%	34
	80% casos	11,11%	$(valor\ estim-11,11\%, valor\ estim+11,11\%)$	93	62%	35
	75% casos	9,5%	$(valor\ estim-9,5\%, valor\ estim+9,5\%)$	79	57%	37
	70% casos	8,11%	$(valor\ estim-8,11\%, valor\ estim+8,11\%)$	68	53%	39
6,25%	99% casos	35,67%	$(valor\ estim-35,67\%, valor\ estim+35,67\%)$	303	91%	49
	95% casos	25,66%	$(valor\ estim-25,66\%, valor\ estim+25,66\%)$	218	84%	46
	90% casos	24,88%	$(valor\ estim.-24,88\%, valor\ estim+24,88\%)$	211	84%	49
	85% casos	20,43%	$(valor\ estim-20,43\%, valor\ estim.+20,43\%)$	173	82%	56
	80% casos	17,39%	$(valor\ estim-17,39\%, valor\ estim+17,39\%)$	147	72%	44
	75% casos	16,36%	$(valor\ estim-16,36\%, valor\ estim+16,36\%)$	139	66%	40
	70% casos	13,98%	$(valor\ estim-13,98\%, valor\ estim.+13,98\%)$	119	59%	41
3,125%	99% casos	43,82%	$(valor\ estim-43,82\%, valor\ estim+43,82\%)$	374	55%	69
	95% casos	38,59%	$(valor\ estim-38,59\%, valor\ estim+38,59\%)$	330	89%	50
	90% casos	26,11%	$(valor\ estim-26,11\%, valor\ estim+26,11\%)$	221	86%	102
	85% casos	22,49%	$(valor\ estim-22,49\%, valor\ estim.+22,49\%)$	192	68%	74
	80% casos	20%	$(valor\ estim-20\%, valor\ estim+20\%)$	171	64%	74
	75% casos	17,48%	$(valor\ estim-17,48\%, valor\ estim+17,48\%)$	149	53%	64
	70% casos	16,53%	$(valor\ estim-16,53\%, valor\ estim.+16,53\%)$	141	51%	65

En xeral, podemos dicir que este método si parece ser acertado sempre que poidamos observar, como mínimo, un dez por cento do territorio e na concentración baixa haxa un mínimo de 40 persoas.

Ata o de agora sempre traballamos nas distintas densidades por separado. Por último, e de novo con 100 taboleiros creados e contados pola IA fixemos unha proba xeral estimando tamén dende o inicio a densidade da poboación a contar. Estes foron os resultados:

TERRITORIO ESTUDADO	CONFIANZA	MAX ERROR	TAMAÑO MEDIO	PORCENTAJE DE ACIERTO	ERROR MEDIO
50%	99% casos	28,64%	105	95%	10
	95% casos	17,65%	68	93%	11
	90% casos	16,55%	56	85%	9
	85% casos	15,35%	51	80%	9
	80% casos	11,95%	43	78%	9
	75% casos	9,92%	37	73%	10
	70% casos	9,09%	32	64%	9
25%	99% casos	38,93%	142	99%	21
	95% casos	27,42%	80	93%	9
	90% casos	23,62	69	90%	8
	85% casos	19,64%	62	90%	10
	80% casos	17,48%	55	86%	9
	75% casos	15,25%	81	69%	16
	70% casos	13,21%	47	81%	9
12,5%	99% casos	49,65%	215	99%	19
	95% casos	40,82%	154	99%	22
	90% casos	33,85%	104	97%	13
	85% casos	30,71%	114	99%	26
	80% casos	25,79%	94	95%	12
	75% casos	23,02%	88	96%	8
	70% casos	21,08%	66	86%	8
25%	99% casos	84,62%	323	98%	5
	95% casos	[illegible]	237	98%	12
	90% casos	[illegible]	221	98%	14
	85% [illegible]	[illegible]	178	98%	19
	[illegible]	[illegible]	164	98%	20
	[illegible]	38,93%	154	98%	21
	[illegible] casos	36,28%	133	98%	24
3,125%	[illegible]	100,00%	398	100%	5
	[illegible]	85,84%	341	100%	12
	[illegible]	[illegible]	280	99%	14
	[illegible]	69,01%	241	99%	19
	80% casos	64,04%	164	98%	20
	75% casos	51,76%	154	98%	21
	70% casos	48,84%	133	98%	24

Esta proba resultou todo un éxito, as porcentaxes de acerto son moi altas e o tamaño do intervalo parece razoable (sobre todo se a densidade de xente é alta e non exiximos demasiada confianza)

Así que tras este gran resultado e tendo en conta que xa estabamos no mes de marzo, consideramos dúas alternativas: facer unha proba real (algo que se nos complicaba moito por temas de loxística e tempo) ou construír un estimador de xente segundo estas características estudadas ao longo do traballo.
Fixémolo nun excel introducindo as fórmulas que calcularan os intervalos segundo a densidade da concentración, o territorio observado e a confianza requirida na estimación. Unha vez ocultas esas fórmulas aquí temos o resultado:

ESTIMADOR DE XENTE NUNHA CONCENTRACIÓN

Conclusións

- ✓ Está claro que si que é difícil estimar o número total de xente cando coñeces só certas partes.
- ✓ Non chega con facer unha estimación proporcional sumándolle e restándolle o error medio cometido nesa estimación. Este método acerta moi pouco.
- ✓ Hai que recorrer aos errores máximos para crear uns intervalos adecuados. E por suposto:
    - Cantas máis posibilidades queiras ter de acertar (confianza) maior será o intervalo de estimación.
    - Canto máis terreo poidas estudar, mellor será esta estimación e máis pequeno será o intervalo.
    - Para que este método funcione, a densidade da concentración debe ser de máis de 0,6 persoas por metro cadrado e debes poder observar máis dun 10% do territorio.

Posibles melloras

Poderíamos mellorar o seguinte:

- ✓ Se tiveramos utilizado a IA dende o principio poderíamos utilizar unha mostra moito máis grande e seguro que os intervalos de estimación serían moito mellores. (A cambio teríamos perdido unha parte moi divertida do traballo).
- ✓ Todos os taboleiros que utilizamos para poñer a proba o método poderíamos usalos tamén para mellorar os intervalos (cando maior é a mostra, máis fiables son os resultados), pero non tivemos tempo para facelo. Ademais unha vez feita a mellora tamén habería que poñela a proba (esta mellora sería un proceso infinito).
- ✓ Gustaríanos, e así o tiñamos pensado ao inicio do proxecto, facer unha proba real (concentrando aos nosos compañeiros no patio), pero por cuestión de tempo resultounos imposible.
- ✓ Gustaríanos ter forma de comprobar se usando o noso estimador nos casos das noticias que nomeamos ao principio, evitabamos esa gran diferenza que se producía.

Por suposto, temos moitas cousas que mellorar no noso método, pero para ser alumnos de 2º ESO opinamos que non está nada mal.