

ANEXO II
CONCURSO ESCOLAR DE TRABAJOS ESTADÍSTICOS
(Curso escolar 2009-2010)
PORTADA

TÍTULO:

COMPROBACIÓN DE LA LEY DE LOS GRANDES NÚMEROS

FRECUENCIA DE LAS CINCO VOCALES EN LAS LENGUAS ROMANCES

Número de alumnos: CINCO

Nivel que cursan los alumnos:

1º de E.S.O.

2º de E.S.O.

3º de E.S.O. (CUATRO ALUMNOS)

4º de E.S.O. (UNA ALUMNA)

1º de Bachillerato. Modalidad:

2º de Bachillerato. Modalidad:

Ciclo Formativo (grado medio).

Ciclo Formativo (grado superior).

Rama y especialidad:

COMPROBACIÓN DE LA LEY DE LOS GRANDES NÚMEROS

FRECUENCIA DE LAS CINCO VOCALES
EN LAS LENGUAS ROMANCES

INDICE

1		
	Notas introductorias	3
2		
	Cómo surgió la idea	5
3		
	Qué vamos a analizar en el trabajo	6
4		
	Comprobación de la Ley de los Grandes Números	7
5		
	Frecuencia comparada de las cinco vocales en el latín y las cinco lenguas romances más habladas	10
6		
	Frecuencias comparadas de las cinco vocales en las lenguas de la península ibérica	12
7		
	Conclusión	14

1 NOTAS INTRODUCTORIAS

Al llegar al IES la publicación del Concurso, conseguimos formar un grupo de trabajo con alumnos de 3º de ESO. Al poco tiempo un alumno se descolgó y entró otra de 4º curso dónde no se había podido formar grupo.

En el desarrollo de la Estadística de 3º y 4º curso una de las cosas que les parece más misteriosa es la Ley de los Grandes Números. Esto, unido a lo adecuado para el nivel de estos alumnos, nos llevó a decantarnos por comprobar dicha ley.

Tras leer las bases publicadas en el folleto de la convocatoria, hemos procurado ajustarnos a ellas, tanto en formato como en contenido.

Contenido Estadístico de tema libre: distribución de las cinco vocales en las lenguas romances. Comenzamos analizando el castellano y continuamos abarcando las cinco romances más importantes, así como el resto de lenguas de la península, incluido el euskera.

Método de trabajo: Hemos abordado la recogida de datos y comprobación de la Ley de los Grandes Números y la comparación de las distribuciones de las vocales en las distintas lenguas romances. Para hacer el conteo y comprobar la ley de los grandes números hemos usado los primeros 25 capítulos del Quijote, que contienen más de 80.000 vocales. Para el segundo apartado, por un lado, hemos tomado el latín como texto comparativo, y hemos analizado las cinco lenguas romances más habladas, y por el otro, hemos usado el castellano como modelo y lo hemos comparado con las lenguas romances peninsulares, además del euskera.

En las bases del folleto se dice que se puede usar datos extraídos del ICANE, aunque no se expresa como requisito obligatorio. Hemos revisado la página www.icanes.es para tomar modelo sobre el uso de tablas, gráficos, y la estructura de los trabajos, aunque hemos aportado nuestro aire personal. Aparte de esto, nos ha venido muy bien el conocimiento de la página para la obtención de datos en los trabajos de clase, tanto para matemáticas como para las asignaturas de sociales.

Al principio diseñamos un modo de conteo manual pero después de una semana de trabajo nos dimos cuenta que con ayuda del contador de Word podíamos hacerlo de forma automática, pudiendo ampliar el tamaño de la muestra indefinidamente. Los alumnos así han aprendido a:

- Buscar textos en Internet.
- Pasarlos a Word.
- Usar el contador de Word

Los alumnos sabían usar Excel y así hicieron un diseño de gráficos y de ellos elegimos los más adecuados.

Comprobación de la Ley de los Grandes Números

Dónde más he tenido que orientarles ha sido en la búsqueda de textos en las lenguas distintas del español por sus pocos conocimientos de literatura universal, ya que son de 3º y 4º de Eso.

El trabajo ha sido gratificante, aunque a veces hemos tenido dificultades para las reuniones, porque los recreos dan poco de sí y por lo que ha sido necesario reunirnos alguna tarde.

2 CÓMO SURGIÓ LA IDEA

Cuando la profesora nos invitó a participar en este concurso, tuvimos que pensar sobre qué tema trabajar.

Alguien conocía el cuento de Allan Poe “El escarabajo de Oro” y de ahí surgió la idea de comprobar si era verdad lo que en este cuento se decía: que un idioma puede identificarse por la frecuencia de sus letras.

Nos pareció que podíamos simplificarlo y estudiar sólo las frecuencias de sus vocales y así empezamos a contar vocales de diez hojas del Quijote. Vimos que al llevar contadas cerca de 8.000 vocales aún había variaciones.

Entonces, pensamos que quizás podríamos utilizar el ordenador para facilitar el trabajo, y de este modo, hemos llegado a contar las vocales de los 25 primeros capítulos del Quijote.

3 QUÉ VAMOS A ANALIZAR EN EL TRABAJO

- a. **Comprobación de la Ley de los Grandes Números, en la frecuencia relativa de las cinco vocales en la lengua Española.** El texto lo sacamos de la edición del Instituto Cervantes del Quijote, lo copiamos en Word y con su contador contamos las cinco vocales con y sin acento (también están contadas las ues mudas del “qu” y del “gu”). El trabajo está detallado en la carpeta “castellano”. Adjuntamos el trabajo manual previo que hicimos para que se vea el método que usamos, ahí se ve la gran dificultad por lo laborioso y los muchos errores de cálculo que cometimos y que están sin corregir porque abandonamos el modo manual. Esto está en unas fotocopias identificadas como “conteo inicial manual”.

- b. Como habíamos encontrado un modo fácil de realizar el trabajo decidimos hacer **un estudio comparativo de la frecuencia de las cinco vocales en las cinco lenguas romances más habladas, entre ellas y con el Latín.** Los textos elegidos para ello son: En Latín, un texto de Cicerón. En Portugués, un capítulo de “Os lusiadas” de Camoes. En Francés una escena de “Le malade imaginaire” de Molière. En Italiano un capítulo de “I Promesa sposi” de Manzoni. En rumano un capítulo de “Făt-Frumos din lacrima” de Mihai Emines. Esto está detallado en la carpeta “textos de referencia”

- c. **Comparación de las frecuencias de las vocales en las lenguas romances de la península y también comparación con la lengua no romance el euskera.** Los textos elegidos han sido en Catalán una parte de “El Canigó” de Jacinto Verdaguer y en Gallego “Poesías” de Rosalía y su prólogo. En portugués y castellano las usadas en el apartado anterior. En euskera hemos usado los escritos más antiguos que hemos encontrado que han sido unas Poesías del siglo XVI

NOTA: Hemos usado en todos los casos obras clásicas que estuvieran libres en Internet.

4 COMPROBACIÓN DE LA LEY DE LOS GRANDES NÚMEROS

Después de contar las vocales de los 25 primeros capítulos del Quijote, construimos esta tabla, síntesis de la que hay en el fichero Excel “ley comprobada”.

	a%	e%	i%	o %	u%	Total vocales contadas
cap I	27,74	28,85	12,09	20,86	10,46	3854
cap II	28,54	28,69	11,68	20,54	10,56	6535
cap III	28,87	29,14	11,15	20,38	10,46	11323
cap IV	28,39	29,08	11,10	20,88	10,55	13739
cap V	28,31	29,17	11,16	20,87	10,48	16960
cap VI	28,02	29,26	11,27	21,06	10,40	19200
cap VII	27,66	29,48	11,14	21,20	10,53	22914
cap VIII	27,45	29,51	11,18	21,35	10,51	25520
cap IX	27,57	29,44	11,20	21,25	10,53	26668
cap X	27,37	29,59	11,13	21,37	10,54	30489
cap XI	27,22	29,65	11,19	21,38	10,57	34816
cap XII	27,18	29,55	11,22	21,43	10,62	39386
cap XIII	27,15	29,65	11,13	21,52	10,55	43216
cap XIV	27,08	29,66	11,19	21,49	10,58	45648
capXV	27,08	29,69	11,19	21,48	10,56	48605
capXVI	27,20	29,62	11,17	21,39	10,62	51460
cap XVII	27,14	29,66	11,13	21,47	10,60	54954
capXVIII	27,15	29,80	10,91	21,53	10,61	59159
cap XIX	27,08	29,76	10,97	21,58	10,61	62569
cap XX	27,01	29,79	10,97	21,58	10,64	66558
cap XXI	27,03	29,80	10,96	21,57	10,64	69763
cap XXII	26,97	29,85	10,97	21,56	10,65	73331
cap XXIII	26,97	29,87	10,99	21,53	10,64	77821
cap XXIV	26,93	29,95	11,04	21,42	10,65	81304
cap XXV	26,86	30,00	11,08	21,38	10,67	86084

En esta tabla se muestran los porcentajes de cada vocal conforme se van acumulando los capítulos.

Como se puede comprobar, esta tabla permite comprobar la Ley de los Grandes Números.

LEY DE LOS GRANDES NÚMEROS:

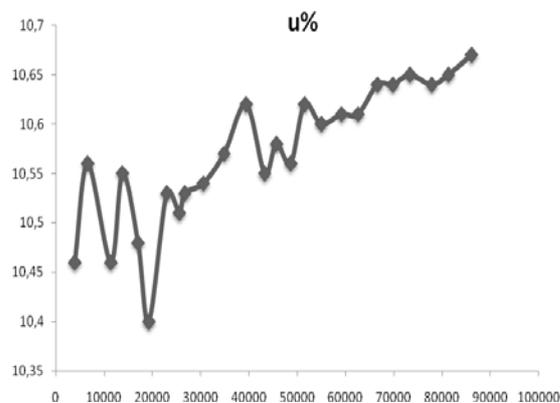
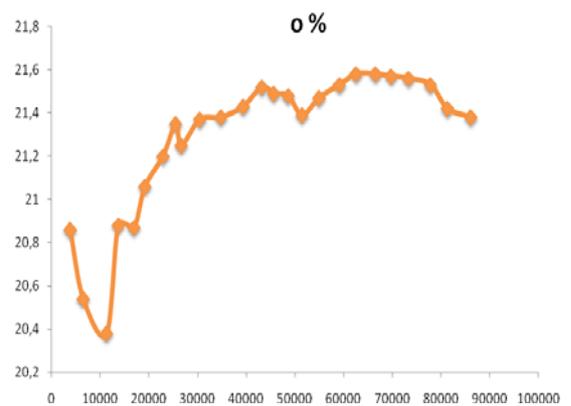
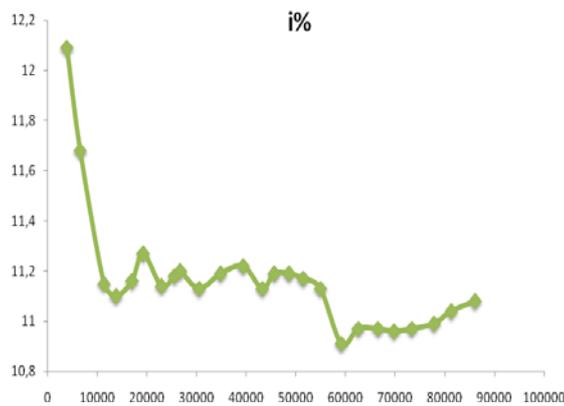
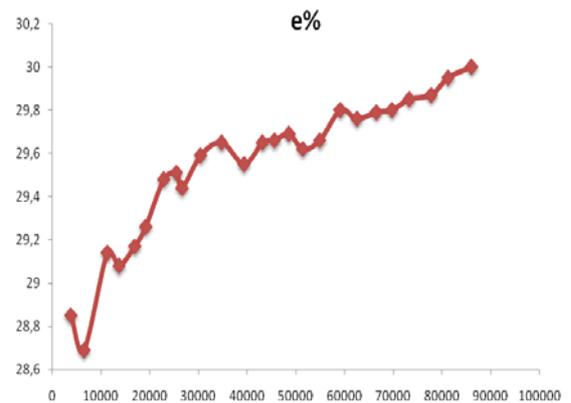
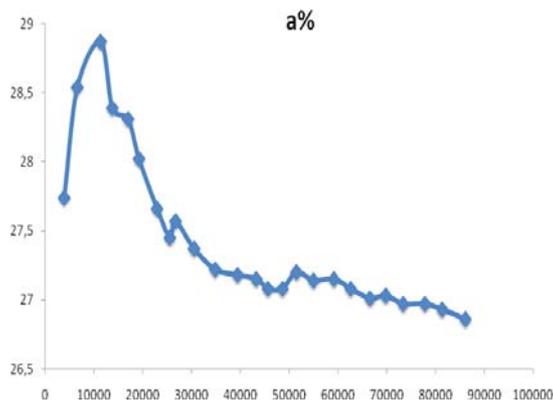
“En un experimento aleatorio, la frecuencia relativa de un suceso se aproxima cada vez más a su probabilidad teórica a medida que aumenta el número de experiencias que se realizan”.

Comprobación de la Ley de los Grandes Números

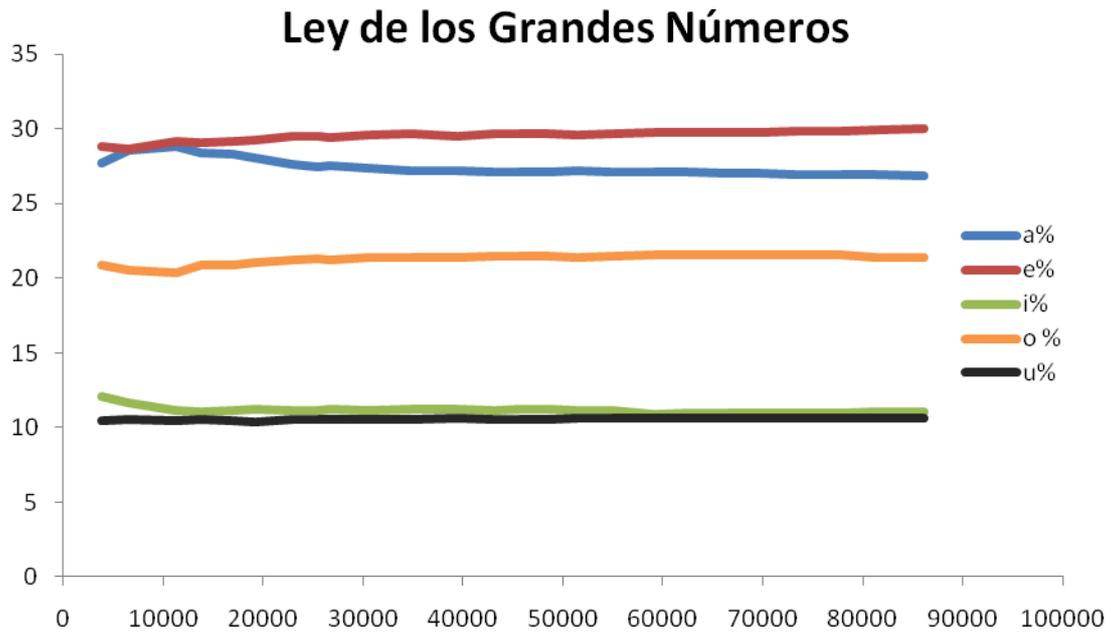
La probabilidad teórica a que se refiere podría enunciarse así:

Según los datos obtenidos del conteo realizado, en español, la probabilidad de elegir una vocal en un texto al azar es:

P(a)	P(e)	P(i)	P(o)	P(u)
0,269	0,300	0,111	0,214	0,106



Nota: Obsérvese detalladamente la escala de los ejes verticales de cada gráfica ya que son diferentes



Estas gráficas muestran de forma clara que se cumple la Ley de los Grandes Números al contar 80.000 vocales de un texto español.

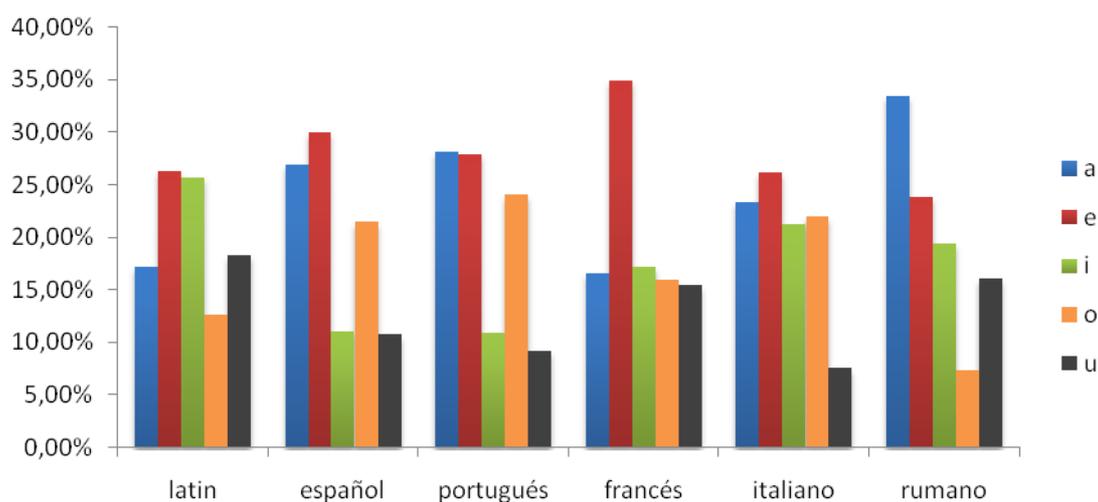
En estas gráficas como se ve, el eje vertical mide las frecuencias cuando se llevan contadas tantas vocales en total como marca el eje horizontal.

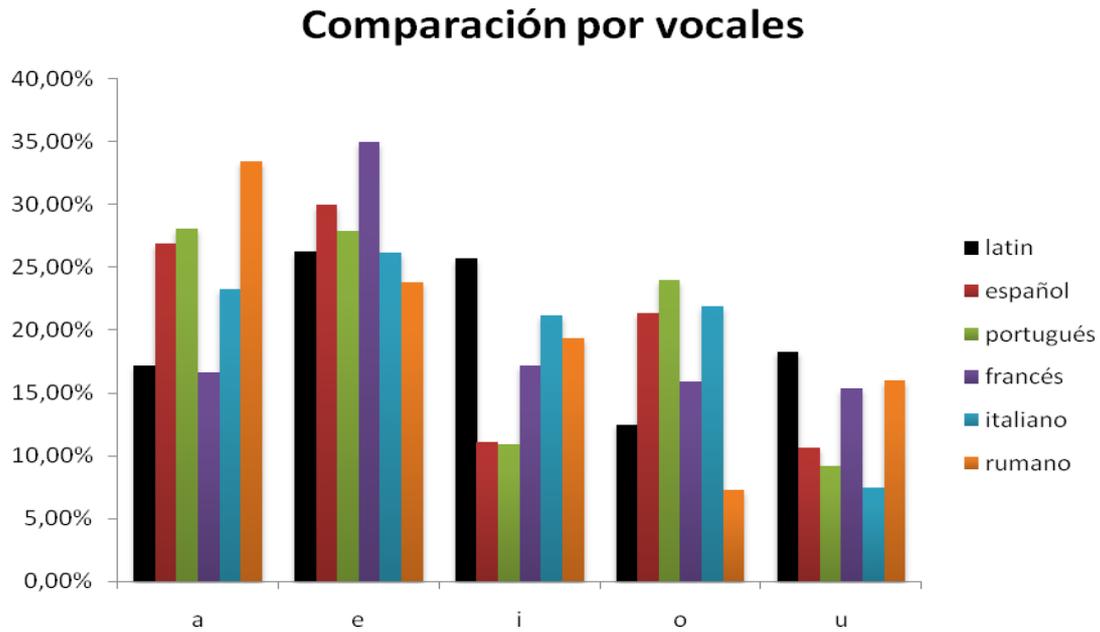
5 FRECUENCIA COMPARADA DE LAS CINCO VOCALES EN EL LATÍN Y LAS CINCO LENGUAS ROMANCES MÁS HABLADAS

A continuación, presentamos las tablas que son síntesis de la tabla en Excel de todos los datos que se encuentran en el fichero " todos los idiomas + gráficos". Hay que decir que en algunas lenguas como el francés hemos contado **los símbolos gráficos**, que no coinciden con los fonéticos, agrupando las vocales con sus distintos acentos (como a y e hay cinco sonidos diferentes).

	a	e	i	o	u
latin	17,19%	26,32%	25,70%	12,51%	18,28%
español	26,86%	30,00%	11,08%	21,38%	10,67%
portugués	28,08%	27,91%	10,89%	23,97%	9,15%
francés	16,60%	34,94%	17,21%	15,88%	15,37%
italiano	23,30%	26,16%	21,19%	21,89%	7,47%
rumano	33,41%	23,82%	19,40%	7,32%	16,05%

Distribución de las vocales en las lenguas romances





Según se puede ver en las gráficas, las lenguas más parecidas al latín en las frecuencias de la vocal i son el italiano y el rumano. Esta alta frecuencia de la i en italiano se debe a la formación del plural, al igual que en el rumano, donde el plural se forma con la e y la i.

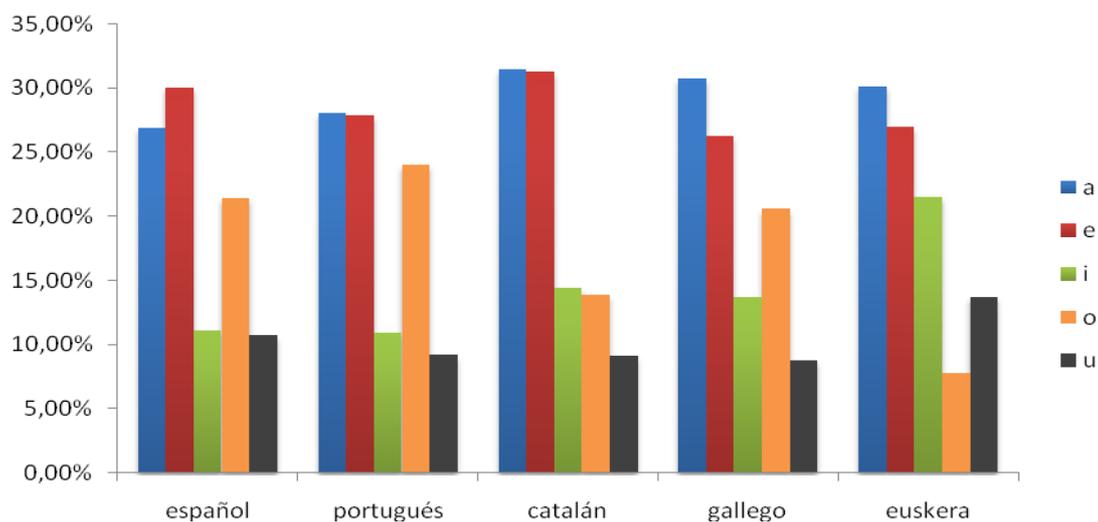
Creemos que los lingüistas podrían dar más explicaciones de estos hechos.

6 FRECUENCIAS COMPARADAS DE LAS CINCO VOCALES EN LAS LENGUAS DE LA PENÍNSULA IBÉRICA

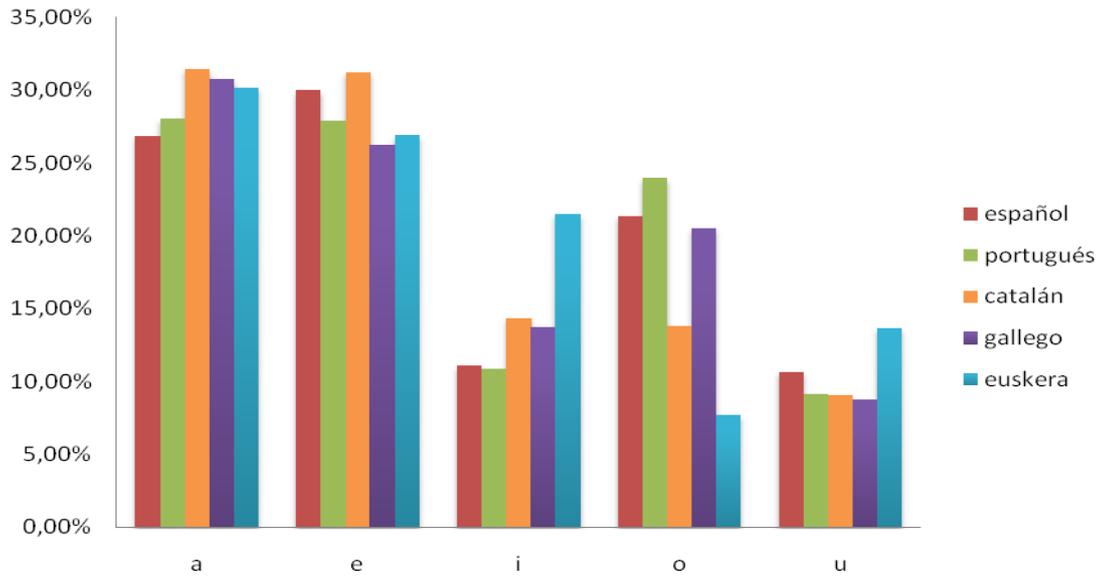
En este apartado, presentamos las tablas que son síntesis de la tabla en Excel de todos los datos que se encuentran en el fichero " todos los idiomas + gráficos".

	a	e	l	o	u
español	26,86%	30,00%	11,08%	21,38%	10,67%
portugués	28,08%	27,91%	10,89%	23,97%	9,15%
catalán	31,49%	31,26%	14,38%	13,79%	9,07%
gallego	30,76%	26,24%	13,72%	20,52%	8,75%
euskera	30,14%	26,93%	21,52%	7,72%	13,69%

Distribución de las vocales en las lenguas peninsulares



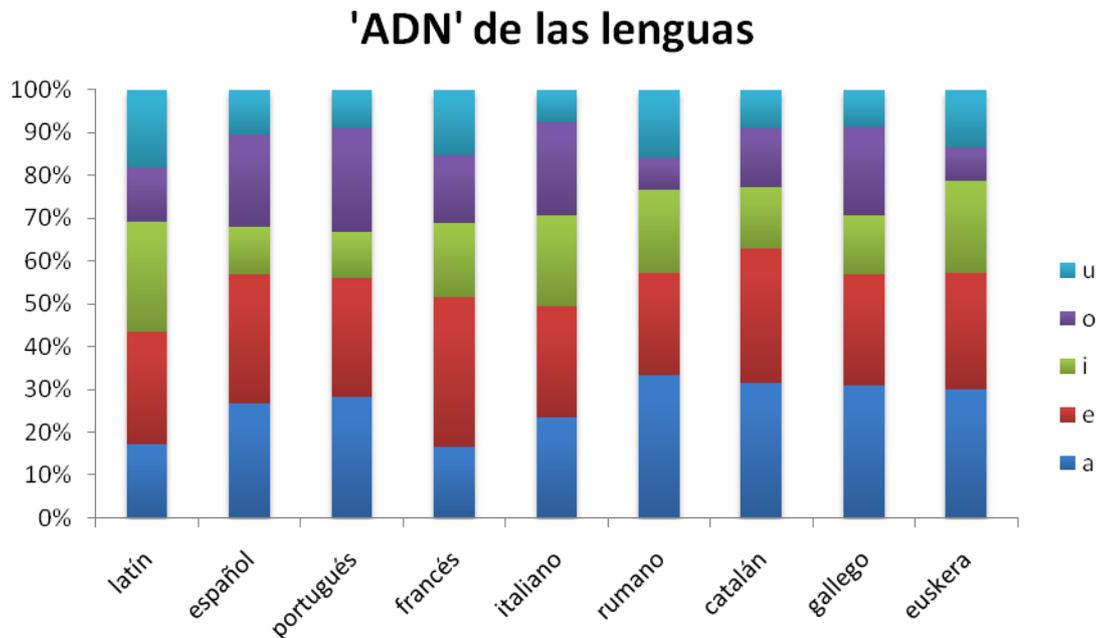
Comparación de las vocales



Hay que hacer notar el gran parecido de las cuatro lenguas romances de la península Ibérica, lo cual se comprende por el tipo de lengua que aquí había cuando llegó el latín y por el tipo de lenguas que trajeron los bárbaros (visigodos, suevos, alanos...) que fue bastante común.

Se puede observar que el euskera es la lengua que más se diferencia de las lenguas romances en la frecuencia de las vocales i, o y u.

7 CONCLUSIÓN



Esta tabla pone de manifiesto que es cierto lo que se dice en el cuento 'El escarabajo de Oro' de Edgar Allan Poe, cada lengua puede ser identificada por la frecuencia relativa de sus vocales.

Con este trabajo hemos podido comprobar además que las letras y las ciencias no están reñidas. Hemos utilizado las lenguas para comprobar una ley matemática y del mismo modo, hemos usado las matemáticas para estudiar algunos aspectos de las lenguas.