

INFORME SOBRE ANÁLISIS DE DATOS COVID-19.

TAREA 26. ESTADÍSTICA DESCRIPTIVA.

ÍNDICE

1. INTRODUCCIÓN	3
2. EVOLUCIÓN POR PROVINCIAS	3
3. EVOLUCIÓN POR MUNICIPIOS	5
4. COMPARATIVA ENTRE CENTROS DE SALUD	7
5. COMPARATIVA ENTRE GÉNEROS Y EDADES	10
6. CONCLUSIONES	16

1. INTRODUCCIÓN

Para poder analizar los datos sobre la evolución del covid 19, se han cogido muestras de datos referentes a los casos en la comunidad autónoma vasca, cedidos de forma gratuita por el gobierno.

Los datos se pueden obtener de la siguiente dirección:

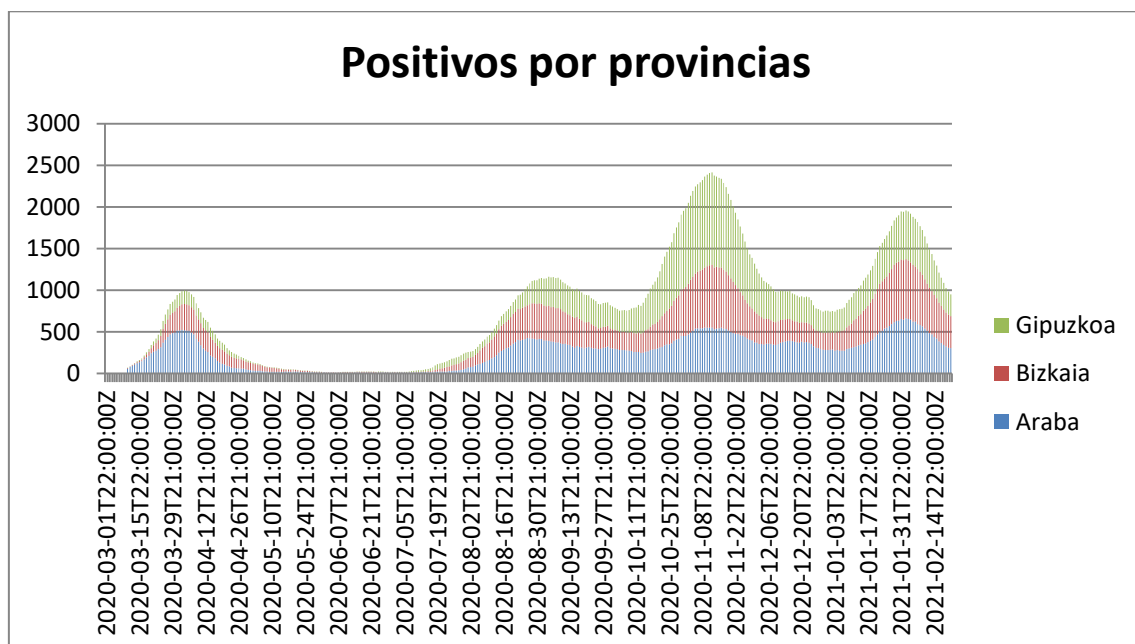
<https://datos.gob.es/catalogo/a16003011-evolucion-del-coronavirus-covid-19-en-euskadi1>

Con esos datos y los conocimientos obtenidos en estadística descriptiva, se plantea desarrollar un análisis estadístico sobre el covid 19, y su evolución, con el fin de completar la tarea 26 y entender mejor como analizar datos de forma correcta.

2. EVOLUCIÓN POR PROVINCIAS

Para comenzar con el análisis descriptivo, se presentan los datos referentes infectados por provincias y positivos por tesis PCR para relacionar ambos entre sí.

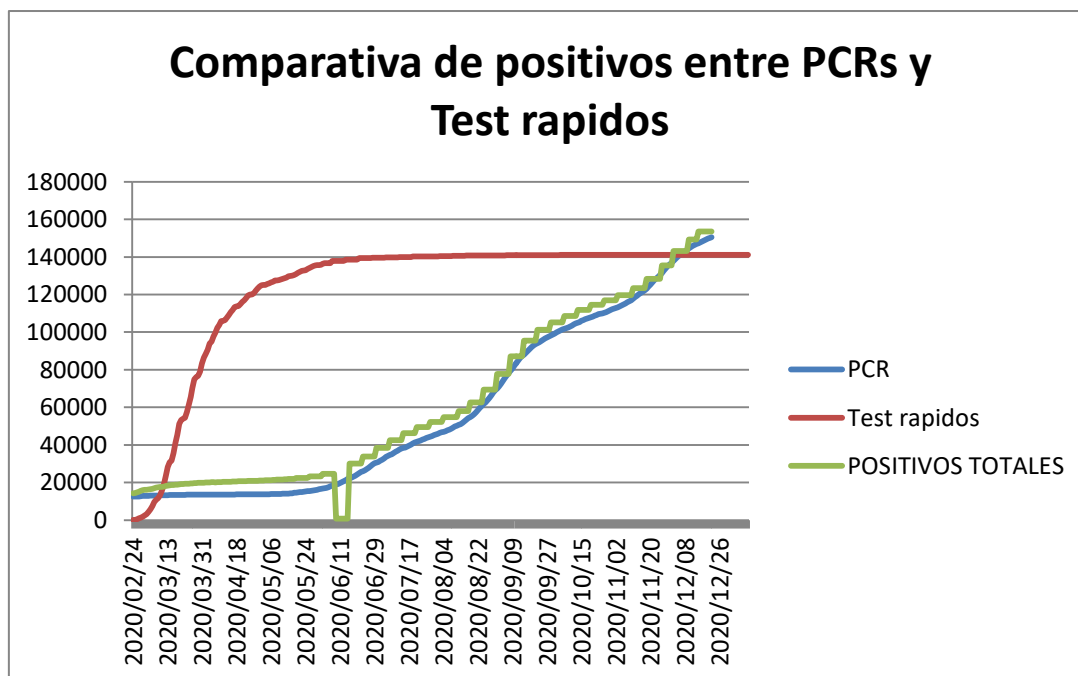
Como dato inicial se relacionan la cantidad de infectados por cada provincia desde el comienzo de la pandemia hasta la actualidad.



Gráfica 1. Evolución del virus por provincias en el País Vasco.

Se puede observar que las 3 provincias mantienen su nivel de infección en prácticamente el mismo recorrido, siendo la provincia más afectada Guipúzcoa. El recorrido que generan todas juntas refleja la evolución de la pandemia desde que empezó a partir del tercer mes del año, marcando el primer pico de infección en Marzo. A partir de ahí, y con el primer confinamiento, los casos se fueron reduciendo con el paso de los meses hasta la llegada de verano, desde la cual se generan 3 olas de infección que engloban a todo el país vasco. La primera en Agosto, la segunda y más alta en Noviembre y la última hasta la fecha en Febrero de 2021.

Otro dato importante a la hora de analizar las 3 provincias es ver la cantidad de PCR que se han hecho desde el comienzo de todo, ya que es necesario saber en todo momento el número de infectados detectado con la cantidad de pruebas realizadas.



Gráfica 2. Relación entre test rápidos y PCRs.

Se ve que al comienzo de la pandemia, el uso de tests rápidos fue crucial en la detección del virus, de modo que hasta que se regularizó el uso de tests PCR, estos fueron los predominantes. Se aprecia que la cantidad de PCR usados cada vez son más, ya que estos son los que confirman con seguridad la presencia del virus en los pacientes, llegando a superar a la cantidad de tests rápidos hechos hasta el momento. Hay que tener en cuenta que hay mucha gente que no se ha hecho ningún tipo de test y aun así son positivos, ya sea por ser asintomático o por otras causas, lo que podría modificar los resultados de esta gráfica. Pero como punto de referencia es suficiente dato a tener en cuenta.

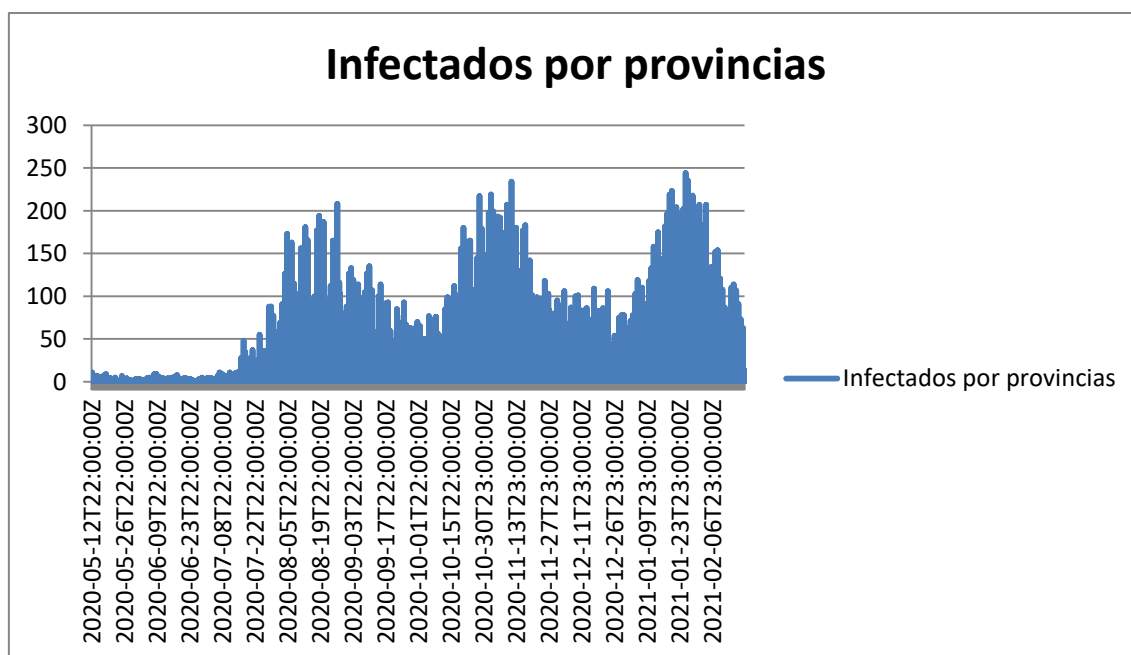
Ahora que se sabe cómo varía el uso de los PCR y el efecto que tiene el virus en las 3 provincias, se va a proceder a analizar los datos de una forma más concreta, aislando datos sobre municipios y centros de salud, además de comparar infecciones y evolución del virus entre diferentes rangos de edad y géneros.

3. EVOLUCIÓN POR MUNICIPIOS

Como ya se ha mencionado antes, se va a tratar de aislar poco a poco la cantidad total de datos, de forma que pasando de una visión general del virus por provincias, se pasa a aislar la muestra en municipios, y posteriormente en grupos de personas.

Dado que los municipios son los que componen a las provincias como tal, las gráficas que representan el avance del virus por municipios son muy similares al recorrido que genera la primera gráfica (gráfica 1), de forma que todos ellos componen un primer pico de infección al comienzo de la pandemia, y posteriormente 3 picos más de infección, los cuales representan a las 3 olas de contagio que hemos sufrido.

En la siguiente gráfica se representan los datos obtenidos con mediciones diarias en todos los municipios del país vasco desde después del primer pico de infección hasta la actualidad.



Gráfica 3. Evolución de infectados por municipios.

A partir de Julio de 2020, el número de infectados crece considerablemente, alcanzando un pico de en torno a 200 infectados en algunas zonas. Esto puede ser muy crítico dependiendo del tamaño del municipio en el que se encuentren los casos.

A partir de agosto, se observa que la tendencia de infectados comienza a bajar hasta el mes de octubre, donde comienza a haber nuevos infectados, replicando el efecto anterior de la gráfica y superando el máximo por primera vez. Finalmente se observa que a partir de diciembre de 2020 el proceso se vuelve a repetir por tercera vez, volviendo a superar el máximo establecido anteriormente hasta la fecha de hoy.

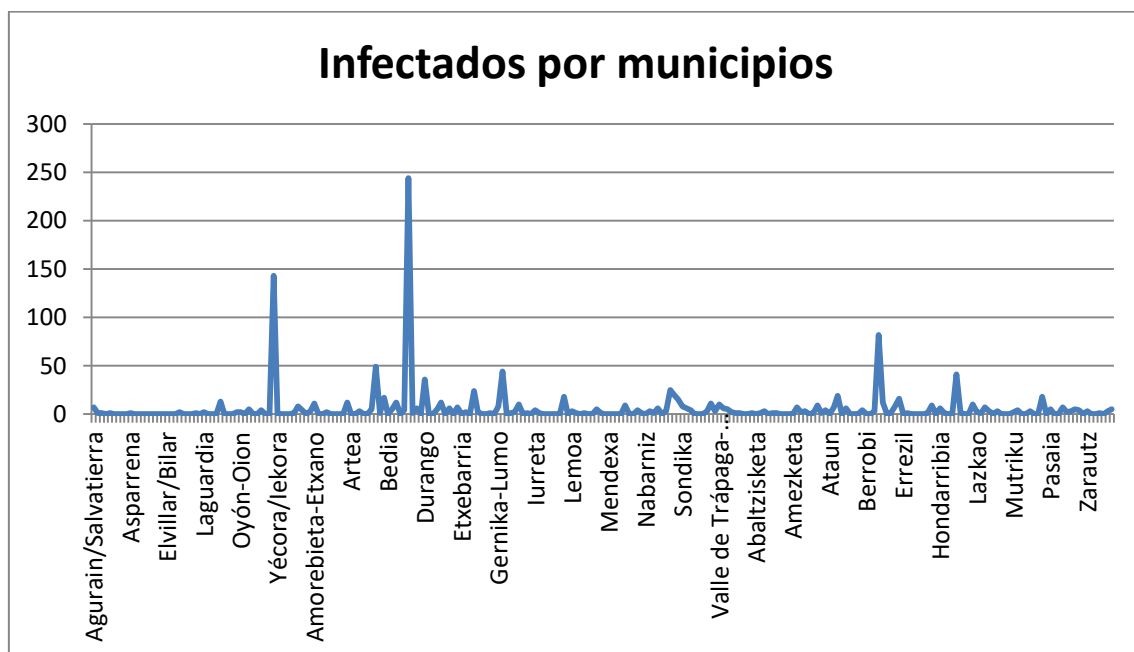
De todos estos datos se va a aislar una muestra que abarca el día con valores máximos de contagio por municipios (26-01-2021), donde se observan los municipios más infectados.

Dado la cantidad de nombres de municipio que hay para representar la gráfica, y teniendo en cuenta que no hay espacio en esta hoja para graficar todos los nombres en una única imagen, se mencionan aquí los municipios con los 3 picos más altos:

Máximo 1: Bilbao - 244 infectados

Máximo 2: Vitoria - 143 infectados

Máximo 3: Donosti - 83 infectados



Gráfica 4. Muestra aislada del gráfico anterior en la zona más representativa.

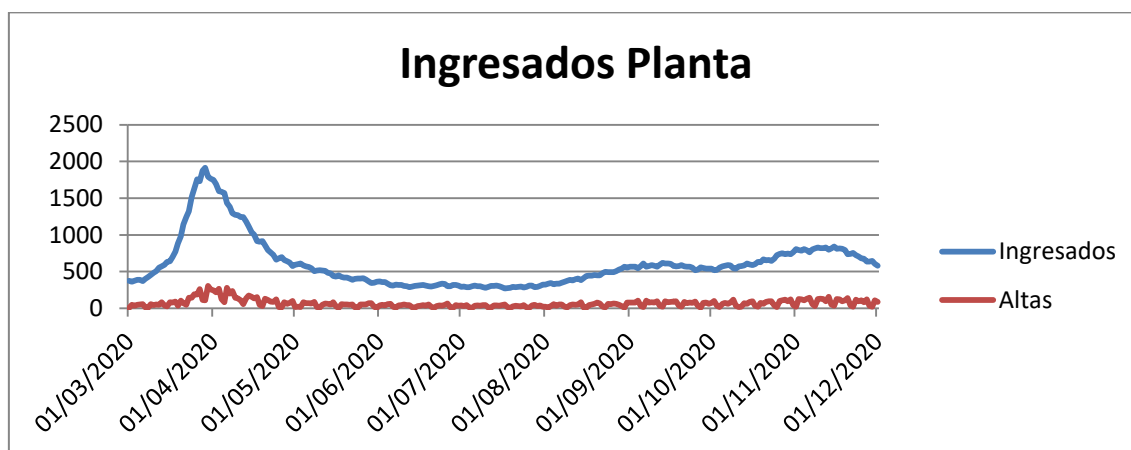
La conclusión principal es que la mayor cantidad de contagios se dan en las 3 capitales del país vasco. Comparando estos datos con los de la primera gráfica, se puede ver como la tendencia de infectados en Guipúzcoa está decreciendo, pasando a primer puesto de infectados Bizkaia. (Todos los datos se han sacado de los CSV y JSON de la web mencionada en la introducción).

4. COMPARATIVA ENTRE CENTROS DE SALUD

En este apartado se muestran los resultados del análisis de cada centro de salud, y una visión general de todos los centros juntos desde el 1 de Marzo hasta el 1 de Diciembre del mismo año (2020). Primero se explican los datos de ingresos y altas POR PLANTA y posteriormente los de UCI.

Los centros sanitarios que abarcan este análisis son: Araba (centros sanitarios de todo el territorio excepto el psiquiátrico), Cruces, Donosti (todo el territorio), Basurto, Galdakao, Zumarraga, Bidasoa, Mendaro, Alto de Deba, San Eloy, Urduliz, Eibar, Leza, Sta Marina, Gorliz, Bermeo, Zaldibar, Zamudio y Psiquiátrico de Araba.

Como se puede observar en la siguiente gráfica, el pico máximo de ingresados **por planta** se encuentra en Abril de 2020, con casi 2000 casos entre todos los centros sanitarios mencionados anteriormente. Parece que a partir del mes siguiente, la tendencia a recibir ingresados se ve reducida hasta Agosto donde parece que vuelve a subir ligeramente. Por último, en Noviembre vuelve a crecer, teniendo el máximo en diciembre. Aún así, la cantidad de gente ingresada se ve reducida considerablemente en los dos últimos picos, teniendo en cuenta que son los momentos en los que se supera al pico inicial de infectados en la gráfica 1. Esto podría significar que una vez superado el colapso inicial de ingresados en centros de salud en Marzo, la cantidad de ingresados se ve reducida por las restricciones decretadas por el gobierno, en cuanto a que no todo el que este infectado debe ser hospitalizado, ya que la gente que puede superar el virus sin complicaciones, lo puede superar en su propio domicilio.



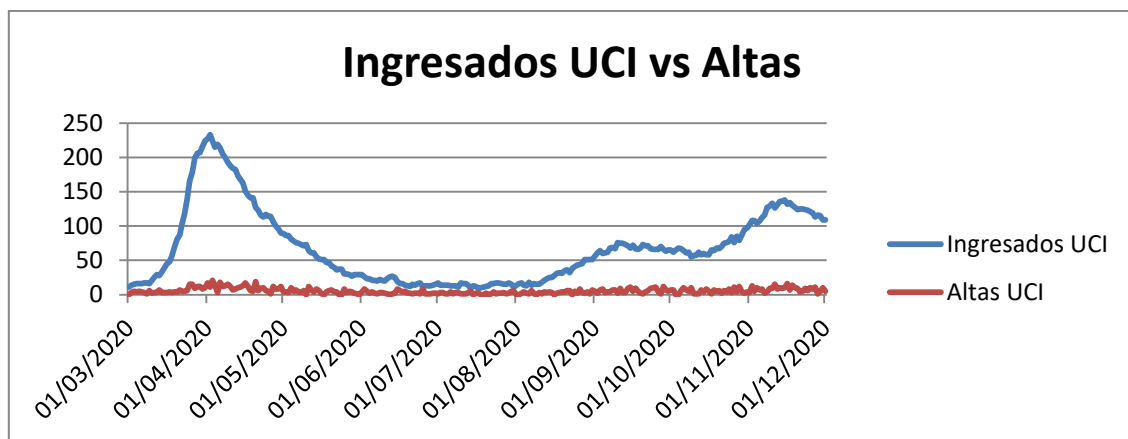
Gráfica 5. Total de infectados en centros de salud por planta hospitalaria VS altas recibidas en el territorio Vasco.

La línea de seguimiento de las altas conseguidas (línea roja en el grafico 5), es representada con picos que se mantienen constantes con cierta frecuencia. Esto se debe al

tiempo mínimo que necesita un paciente para superar los síntomas, de forma que los que consiguen pasar el virus y no tienen complicaciones, reciben el alta al de X días de ser ingresados. La tendencia de la gráfica muestra como el máximo de altas se da en la franja horaria del máximo de infectados, y que a medida que hay menos ingresos también hay menos altas, resultado que se repite en cada franja máxima de ingresos.

La conclusión en cuanto a ingresos y altas en los centros de salud (en planta) es que la mayor parte de la gente que es ingresada en planta, se queda ingresada. En los casos que el paciente se queda en planta, se espera un tiempo para ver si consiguen recibir el alta, o deben ser metidos en la UCI (suponiendo que empeore su estado).

A continuación se muestran los datos referentes a ingresos y altas de UCI.

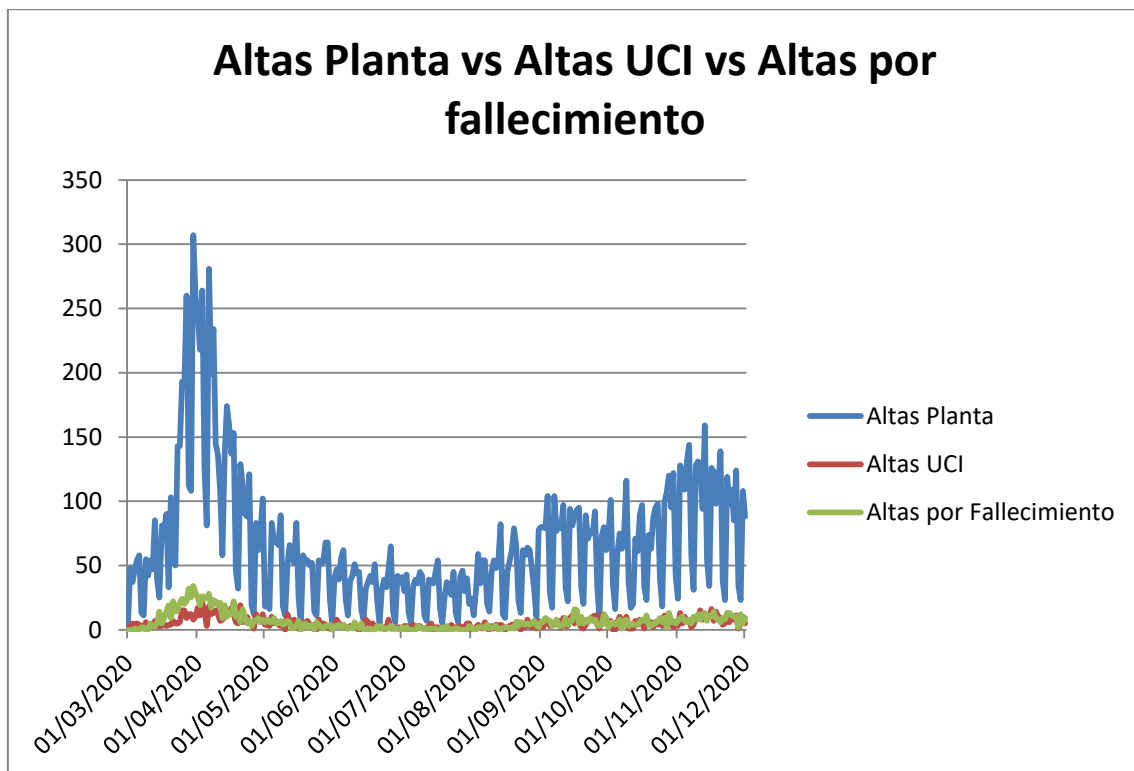


Gráfica 6. Total de infectados en centros de salud por UCI VS altas recibidas en el territorio Vasco.

A simple vista se ve como el gráfico sigue el mismo recorrido que el de ingresos, con la diferencia que entre los casi 2000 ingresos que se encuentran en Abril, solamente 225 son ingresos en UCI. Otro dato característico es que a partir de Agosto, los ingresos en UCI generan una tendencia superior a la del gráfico de ingresos, es decir, a partir de agosto se ingresa más gente en la UCI en proporción a los ingresos de esas fechas.

Además, la mayor parte de los ingresos en UCI no reciben el alta, al contrario que los de planta, que si que tienen una probabilidad mayor de conseguirla.

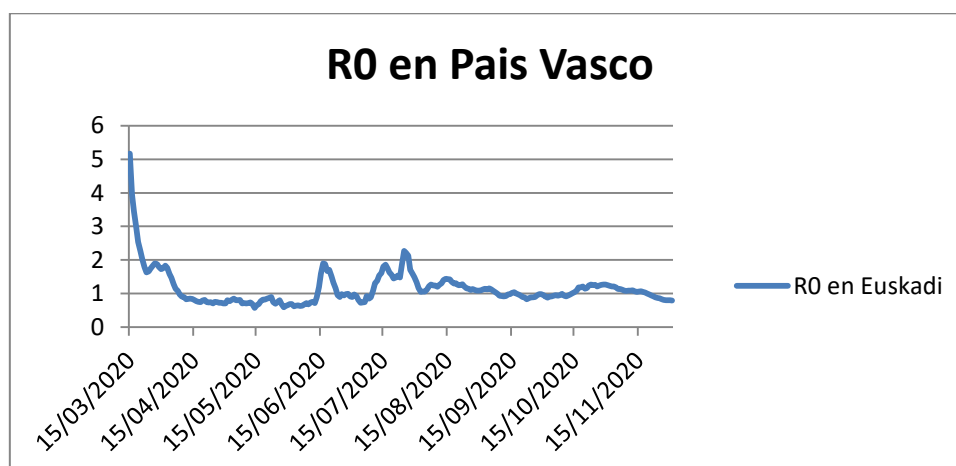
Para concluir con este punto, se presentan los datos que comparan la cantidad de altas obtenidas en Planta y UCI vs la cantidad de altas por fallecimiento.



Gráfica 7. Total de Altas en Planta y UCI vs Altas por fallecimiento.

Se observa que del total de altas de ingresados, la mayor parte se consigue en plantas de hospital, sin pasar por la UCI. Además, la cantidad de altas por fallecimiento supera a la cantidad de altas en UCI, de forma que en las fechas de Abril es cuando se consigue sacar a más gente del hospital, aunque también es cuando más gente fallece debido a la gran cantidad de personas ingresadas.

Teniendo en cuenta ese dato y el recuento de gente que ha recibido altas hospitalarias y por fallecimiento, el resultado del R_0 en todo el territorio sería el siguiente:



Gráfica 8. Factor R_0 en el territorio vasco a lo largo de la pandemia.

5. COMPARATIVA ENTRE GÉNEROS Y EDADES

Uno de los análisis que se plantean es observar si el covid19 tiene más influencia en un rango de edad o en otro, además de saber si afecta más a un género o a otro. Para ello, se recogen todos los datos de hombres y mujeres infectadas en un rango de casi 1 año de pandemia y se filtran de forma que se sacan datos como la letalidad, la tasa por cada 100.000 habitantes, el porcentaje de infectados por cada genero...

Para ello se presentan las siguientes tablas que recogen el resumen de todos esos datos mencionados.

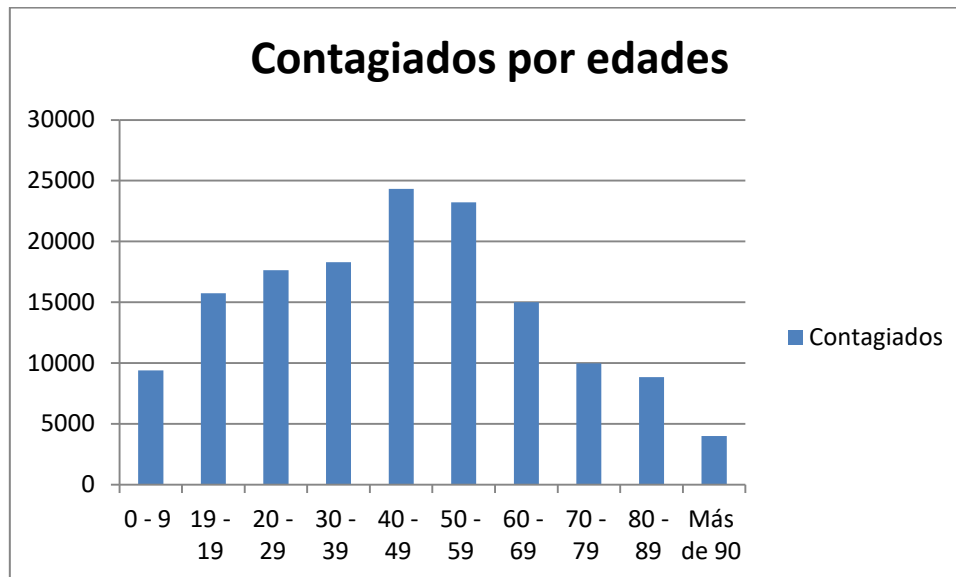
Edad	Positivos	Mujeres: Positivos	Hombres: Positivos	Población	Mujeres: Población	Hombres: Población	Tasa por 100.000 habitantes	Mujeres: Tasa por 100.000 habitantes	Hombres: Tasa por 100.000 habitantes	% sobre el total de los casos
0 - 9	9399	4646	4753	190667	93085	97582	4929,53684	4991,13713	4870,77535	0,06295674
19 - 19	15737	7631	8105	210862	101692	109170	7463,17497	7504,03178	7424,20079	0,10541017
20 - 29	17626	9225	8397	199866	98373	101493	8818,90867	9377,57312	8273,47699	0,11806314
30 - 39	18286	10033	8247	258272	128980	129292	7080,13257	7778,72538	6378,58491	0,12248397
40 - 49	24339	12924	11413	358698	177352	181346	6785,37377	7287,20285	6293,4942	0,16302841
50 - 59	23217	12396	10817	336769	170989	165780	6894,0431	7249,58916	6524,91254	0,15551299
60 - 69	15009	7444	7556	277319	145053	132266	5412,17875	5131,9173	5712,73041	0,10053385
70 - 79	9984	5027	4957	210914	115791	95123	4733,68292	4341,44277	5211,14767	0,06687521
80 - 89	8834	5466	3367	124844	77854	46990	7076,03089	7020,83387	7165,35433	0,05917223
Más de 90	4001	3050	951	31500	23326	8174	12701,5873	13075,538	11634,4507	0,02679965
No consta	2861	1507	1345	-	-	-	-	-	-	0,01916366

Tabla 1. Resumen de datos filtrados de infectados por rangos de edades. Parte 1.

Fallecidos	Mujeres: Fallecidos	Hombres: Fallecidos	Letalidad	Mujeres: Letalidad	Hombres: Letalidad		MEDIA	14643,2
1	0	1	0,00010639	0	0,00024534		MEDIANA	15373
0	0	0	0	0	0		DESV TIPICA	6222,77176
3	1	2	0,0001702	0,0001173	0,00026031		VARIANZA	43025431,5
11	4	7	0,00060155	0,0004353	0,00092838		ASIMETRIA	-0,01413491
31	10	21	0,00127368	0,00085361	0,00204658		CURTOSIS	-0,83867742
73	30	43	0,00314425	0,00263435	0,00438462		RANGO	8575,75
255	75	180	0,01698981	0,01100191	0,02611345		Q1	9545,25
645	207	438	0,06460337	0,04498044	0,09611587		Q2 (MEDIANA)	15373
1528	737	791	0,17296808	0,14307901	0,25175048		Q3	18121
1051	727	324	0,26268433	0,25034435	0,36404494		MINIMO	4001
							MAXIMO	24339

Tabla 2. Resumen de datos filtrados de infectados por rangos de edades. Parte 2.

Graficando los valores medios de contagiados por cada rango de edad, de 10 en 10 años, queda la siguiente representación:



Gráfica 9. Variación de contagios por rangos de edad.

Se puede decir que el valor medio de contagiados entre los diferentes rangos de edad es de 14643,2 personas, aunque lo importante en este caso es saber que grupos son los más afectados. Al parecer las personas con edades comprendidas entre 40 y 60 son las que más cantidad de contagios obtienen, pero si se quiere conseguir un dato más concreto, se debe calcular la cantidad de infectados por cada 100.000 habitantes, lo que nos daría un resultado más concreto ya que redistribuye los datos de manera que se reflejan los 2 grupos que más afección tienen por cada zona localizada, lo que ajusta los datos para el resultado que necesitamos. Este proceso se hace justo después de la siguiente gráfica (gráfica 10), ya que ahora se procede a explicar los valores de media, mediana, rango intercuartílico... de los datos de la gráfica 9.

Como bien se ha dicho, el valor medio de contagiados es de 14643,2 del total de datos recogidos por edades. La mitad de contagiados se encuentran por debajo de 15373, por lo que la media de contagios está por debajo del punto central de contagiados. Con esto, hay que tener en cuenta que la desviación típica es de 6222,77 casos, lo que viene a marcar el margen de error que hay en cada media de infección.

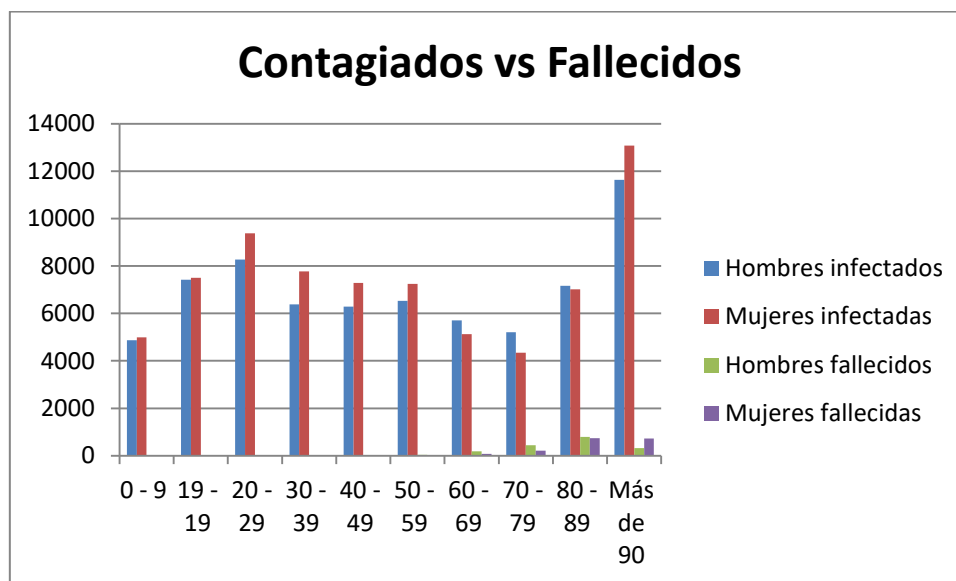
El grado de asimetría en este caso es de -0,01413 lo que significa que la deformación de la distribución del gráfico anterior tiende hacia la izquierda, es decir, el pico está más hacia la izquierda, consiguiendo una relajación del gráfico a medida que avanzamos hacia la derecha. En este caso el valor es tan cercano a cero, que apenas se nota esa deformación.

En cuanto a la curtosis, también es negativa en este caso, con un valor de -0,83. Esto quiere decir que la gráfica es platicúrtica (coeficiente negativo = gráfico aplastado) ya que la curtosis

mide el grado de apuntamiento/achatamiento que tiene la gráfica. En este caso también es un valor muy cercano a cero, de forma que apenas varía respecto al valor de una gráfica mesocúrtica.

Por último, se menciona el rango intercuartílico que es el que marca la diferencia que hay entre el grupo más afectado y el que menos. En este caso el valor es de 8575,75 infectados.

Todos estos cálculos deberían hacerse dentro de cada rango de edades, de forma que los resultados obtenidos sean más concisos.



Gráfica 10. Variación de contagios por rangos de edad.

En este caso se aprecia que los valores máximos se obtienen en personas de edad avanzada, y que el segundo pico más alto corresponde a jóvenes de entre 20 y 30 años.

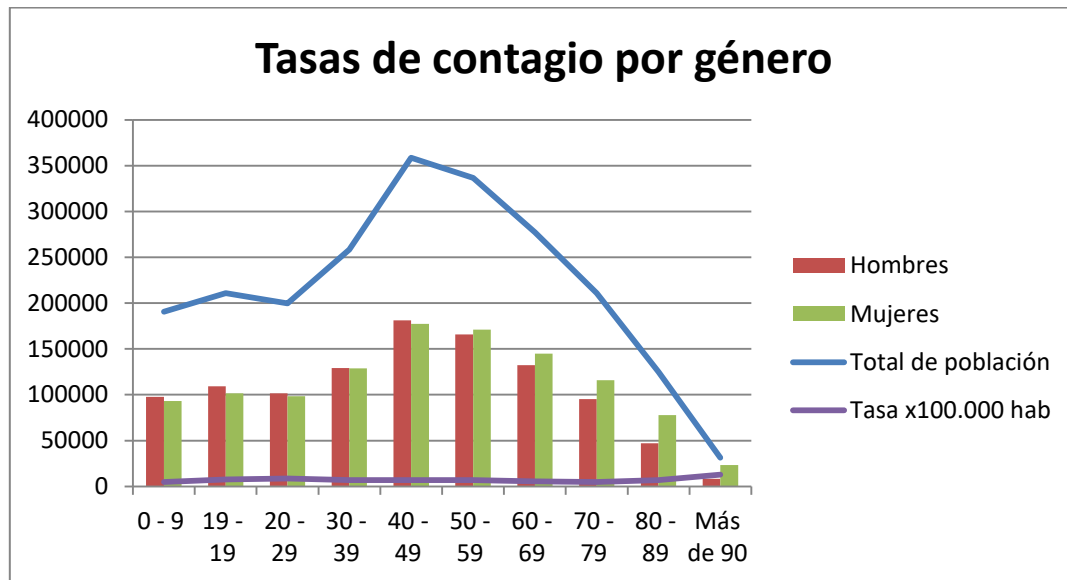
Esto quiere decir que la mayor cantidad de contagios se concentra en esos dos rangos de edad cuando el total de contagios se acota por cada 100.000 habitantes. Este dato es importante ya que de la otra forma se estaría observando un resultado de un grupo más generalizado, haciendo así, que este resultado se adapte mejor a la realidad de los habitantes en grupos más reducidos.

Se aprovecha también a representar que porcentaje de la población corresponde a infecciones en mujeres y hombres, para poder comprobar así si el virus tiene más eficacia en un género o en otro, por lo que viendo la gráfica 10, el número de mujeres contagiadas es ligeramente superior al de hombres, de forma que si se pudiera comparar este dato con resultados más globales, quizá podría deducirse si el virus afecta más a un género que a otro. Por lo pronto, en nuestra tierra parece que el género femenino es más afectado por el covid-19, aunque, en las gráficas que vienen a continuación, se observa que la tasa de contagio más alta corresponde a los hombres en ciertos rangos de edad.

Por último, comparando la cantidad máxima de contagios con los fallecimientos en esos rangos de edad, se aprecia que la mortalidad comienza a partir de los 60 años, manteniéndose constante entre 80 y 100 años. Por supuesto, sigue habiendo mortalidad

en el resto de rangos de edad, solo que la cantidad de víctimas mortales es muchísimo menor que en edades superiores.

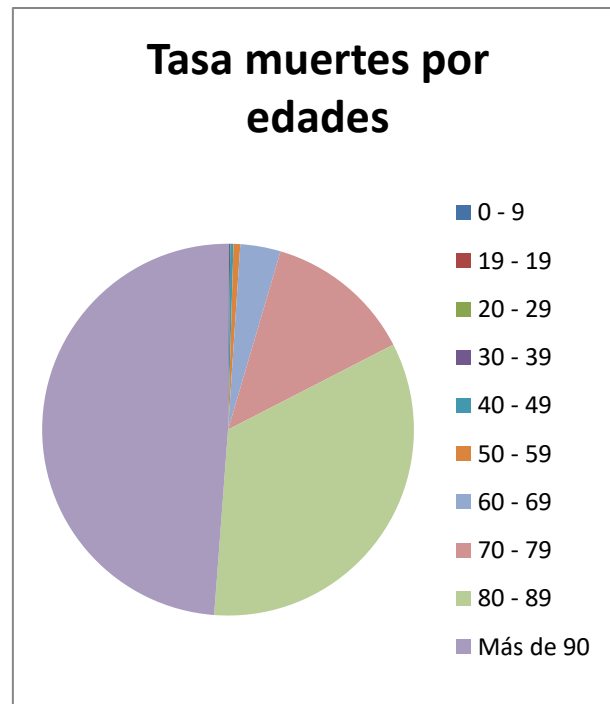
Una vez sabido en que grupos de edades se concentran la mayor cantidad de contagiados, es conveniente saber si uno de los dos géneros tiene mayor tasa de contagio que el otro.



Gráfica 11. Tasas de contagio por géneros.

Como se ha mencionado anteriormente, aunque las mujeres son las más afectadas por el virus en cuanto a cantidad de contagios, son los hombres los que más tasa de contagio tienen respecto a la población total de contagiados. El máximo de este dato se encuentra en franjas de edad entre 40-49 años. A partir de esa edad, la probabilidad de contagiarse más fácilmente pasa a ser parte del género femenino.

Por último, recopilando todos los datos recogidos de muertes en esos rangos de edad, el resultado es el siguiente:



Gráfica 12. Tasas de muerte por edades

Claramente se observa lo mencionado anteriormente. A partir de los 60 años de edad, la probabilidad de morir por covid crece considerablemente, siendo más crítico para las personas comprendidas entre esas edades.

6. CONCLUSIONES

Como conclusión general se plantea que aunque unas provincias estén más afectadas que otras, ese dato puede cambiar a la inversa en un margen de tiempo razonable, ya sea por mejores o peores condiciones de contingencia del virus, o por las restricciones que tenga cada territorio, de forma que la provincia que más datos de contagio tiene en un comienzo, puede ser la que menos tenga al final del proceso. También hay que tener en cuenta la cantidad de tesis que se hace a la gente y la efectividad de los mismos, de forma que no haya falseo de datos por errores de ese tipo.

Una vez obtenidos los datos generales, se opta por ir acotando datos de forma que se obtenga un resultado más localizado, consiguiendo así resultados por zonas más pequeñas que las provincias, tales como centros de salud y grupos de edades, y por las personas que las componen. En este caso, se ve que la relación entre gente hospitalizada y altas tanto en planta como en UCI sigue el mismo patrón, siendo claramente más crítico el hecho de acabar en la UCI por complicación de los síntomas. En estos casos, la probabilidad más alta de salir victorioso de esa situación, corresponde a los pacientes ingresados en planta, ya que gran parte de los que acaban en UCI acaban falleciendo.

En cuanto a los grupos de personas por rangos de edad y género, vuelcan datos de cómo los grupos más afectados son las personas de edades muy avanzadas (90 años en

adelante) y los jóvenes de en torno a 25 años. Ambos casos concuerdan por una parte en que las personas de 25 años tienden a estar más tiempo en sociedad (no siempre por supuesto, para todo hay excepciones), por lo que la cantidad de contagios es mayor, además de la seriedad con la que se tome esta situación, respetando el uso de medidas de protección y demás, y por otra parte las personas de avanzada edad, que son más susceptibles a los síntomas del virus.

Para finalizar, comentar que la cantidad de hombres contagiados por el virus es superior a partir de los 60 años de edad y que el efecto de contagiar el virus con mayor facilidad es mayor en el género femenino a partir de los 50 años de edad. Como bien he dicho anteriormente, haría falta contrastar este dato con fuentes más amplias de información, pero suponiendo que este dato fuera correcto, se podría concluir que hay más contagios masculinos a partir de 60 años de edad con menos tasa de contagio, es decir, los hombres a partir de 60 necesitan tener más cantidad de gente alrededor para acabar estando contagiados ya que la tasa de contagio es ligeramente menor que la de las mujeres, y que en mujeres a partir de 50 años, la tasa de contagio es mayor por lo que no hace falta que estén rodeadas de tanta gente para acabar estando infectada también.

Por último, mencionar que la tasa de mortalidad crece exponencialmente a partir de los 60 años de edad, lo que supone un riesgo mayor para ese grupo de personas.

Con toda esta información, doy por concluido el informe sobre el avance del covid mediante estadística descriptiva. Muchas gracias