

Instrucciones sobre cómo presentar la estadística en un trabajo científico

Francisco Javier Barón López

baron@uma.es

<http://www.bioestadistica.uma.es/baron>

1 Introducción

Estamos acostumbrados a ver cómo la práctica totalidad de los artículos y comunicaciones a congresos requieren del uso de algún tipo de estadísticas. Siendo muy extenso el número de las técnicas estadísticas existentes, el objetivo de esta obra no consistirá en exponer muchas en detalle, sino en destacar algunas de ellas, por lo extendido de su uso, y explicar qué pasos deben seguirse para su utilización en la redacción de un trabajo científico.

2 Estadística descriptiva

La primera tarea consiste siempre en describir los datos que tenemos, divididos en tantos grupos como esté formado el estudio. Por ejemplo si estamos presentando los resultados de un estudio realizados sobre individuos que han seguido uno de tres posibles tratamientos dermatológicos, la presentación de los resultados descriptivos debe hacerse para cada grupo por separado.

Por tanto, en primer lugar daremos unas ideas sobre la manera de presentar la estadística descriptiva de los datos en un grupo, para pasar posteriormente a cómo presentar las comparaciones de los resultados obtenidos en varios grupos.

2.1 Descriptiva de una variable

Normalmente en cada grupo de pacientes se suelen observar, no una sino varias variables que pueden ser cualitativas, como sexo, tabaquismo o “presencia de psoriasis”; o numéricas como edad, peso o *Psoriasis Area and Severity Index* (PASI).

2.1.1 Datos cualitativos

Decir que el 60% de los pacientes se curaron no es suficiente. No debemos conformarnos con exponer simplemente los porcentajes. Hemos de dar información suficiente como para saber qué número exacto de individuos presentó cada modalidad

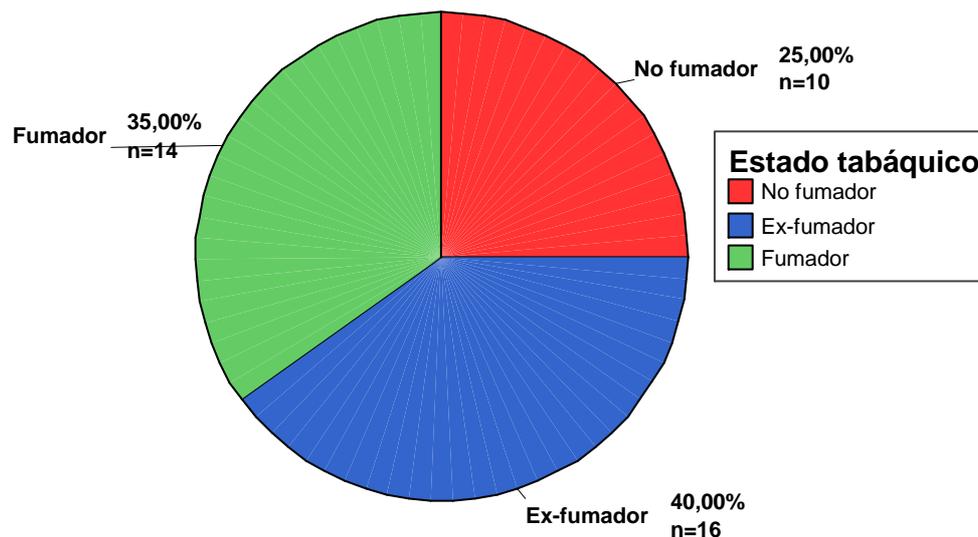
de la variable. Un ejemplo de cómo presentar correctamente este tipo de datos sigue a continuación, acompañado de la tabla de donde han sido obtenidos.

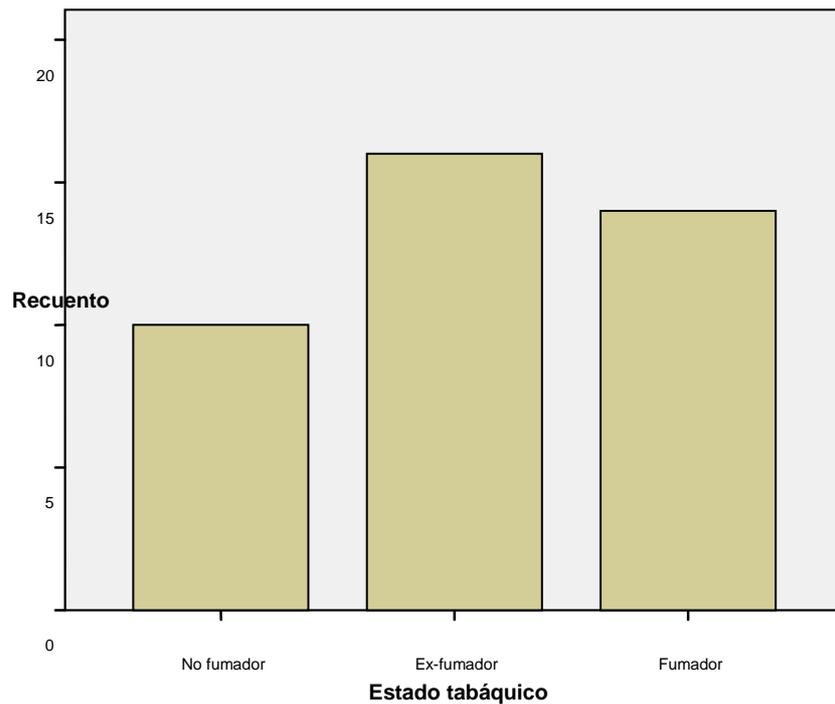
Se estudiaron 50 pacientes con de liquen plano asociado a prurito intenso de más de un mes. Se les trató con enoxaparina y se evaluó el resultado tras la primera semana de tratamiento. A 30 pacientes (60%) le desapareció el prurito, 15 (30%) experimentaron mejoría, y los 5 restantes (10%) no experimentaron ningún cambio.

Prurito tras una semana de uso de enoxaparina

		Frecuencia	Porcentaje
Válidos	Desaparece	30	60,0
	Mejora	15	30,0
	Sin cambios	5	10,0
	Total	50	100,0

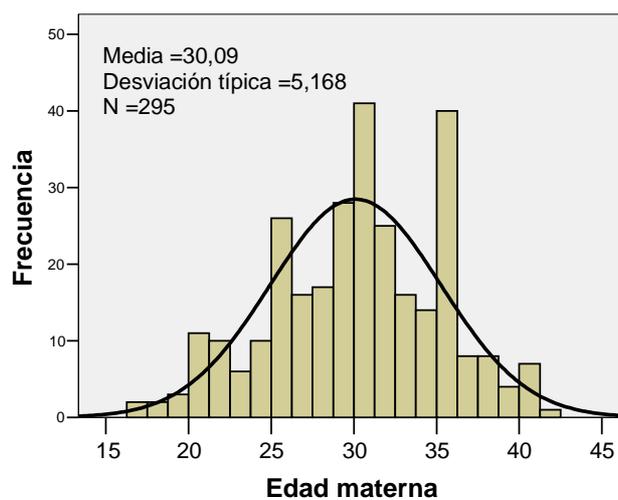
El espacio que ocupa una tabla de frecuencias es, como se ve, bastante grande, así que si hay problemas de espacio en la redacción del trabajo, no se suele exponer (a menos que sea necesario por claridad). Lo mismo ocurriría con los gráficos equivalentes, que para variables cualitativas suelen ser diagramas de sectores o de barras (a elegir libremente) como los siguientes:





2.1.2 Datos Numéricos.

Los datos numéricos no suelen presentarse como los cualitativos, a menos que la variable numérica presente muy pocas modalidades. Lo habitual es presentar una serie de medidas que los resuman, denominadas estadísticos, y cuya elección depende de la distribución que presenten los datos.



En múltiples ocasiones los datos presentan cierta distribución acampanada como la de la figura adjunta, denominada distribución normal. En estos casos con sólo dos medidas, como son la media y la desviación típica, tenemos resumida prácticamente toda la

información contenida en las observaciones. Así, la forma habitual de presentar los resultados que se observan en la gráfica mencionada sería:

Las mujeres que habían tenido hijos formaban un grupo de 295 pacientes, cuyas edades estaban distribuidas de forma aproximadamente normal, con una media de 30,09 años y desviación típica de 5,17 años.

Como se ve es una forma de redactar que ocupa mucho menos espacio que el gráfico. Por otro lado, es difícil decidir en ocasiones si los datos se alejan mucho o poco de una distribución normal, y en este caso, es mejor dar alguna información adicional como el mínimo y el máximo. En este caso los resultados podríamos exponerlos del siguiente modo:

El grupo formado por las mujeres que habían tenido hijos estaba formado por 295 pacientes con edades comprendidas entre los 16 y los 43 años, siendo la media de 30,09 años y la desviación típica de 5,2 años.

Hay más estadísticos que podemos añadir en estas situaciones donde existe cierto alejamiento de la normalidad, como son la asimetría y la curtosis (o apuntamiento). Estos son estadísticos adimensionales (sin unidades). Su utilización indiscriminada haría que un párrafo como el anterior, que de forma natural ya es farragosa, se haga aún peor. De estos estadísticos debemos saber lo siguiente:

- **Asimetría:** Valores negativos (positivos) indican la presencia de una cola a la izquierda (derecha) de la distribución.
- **Curtosis o apuntamiento:** Valores negativos indican datos que presentan un menor grado de apuntamiento que la distribución normal. Es lo que suele ocurrir cuando las observaciones se distribuyen con una cierta tendencia a la uniformidad, desde el valor menor hasta el mayor. Valores positivos indican apuntamiento.

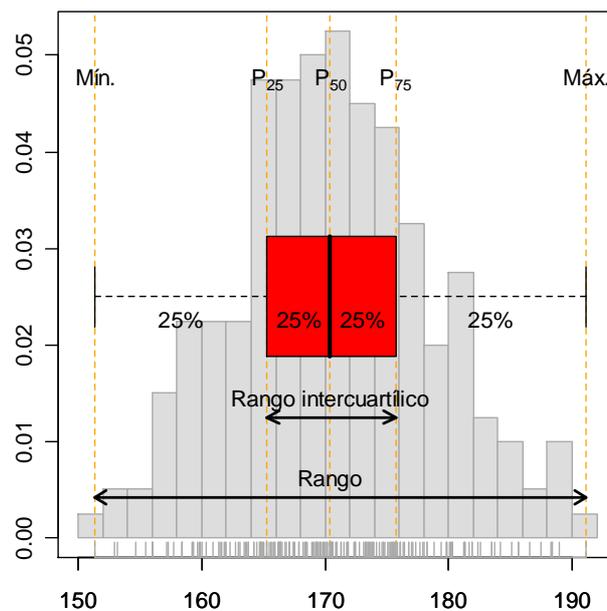
Hay situaciones de claro alejamiento de la normalidad, por ejemplo cuando se observan distribuciones que presentan largas colas u observaciones anómalas¹ (outliers). En ese caso, resumir los datos mediante medias y desviaciones típicas no es suficiente, y debemos considerar medidas más resistentes a estas influencias.

¹ Observaciones demasiado grandes o pequeñas con respecto al resto de valores.

Es frecuente encontrarnos con que haciendo alguna transformación a los datos, recuperemos la normalidad. Un ejemplo de esta situación es cuando el histograma presenta una larga cola a la derecha. A veces una transformación logarítmica de la variable o la toma de una raíz cuadrada en variables que representen conteos, pueden devolvemos a una situación próxima a la normalidad.

En otros casos debemos saber que presentar sólo la media y la desviación de las variables no es suficiente, y debemos indicar la situación, acompañándola de otras medidas:

- La **mediana**, que es aquel valor que deja la mitad de los datos por debajo de él. Se puede presentar junto a la media para mostrar la diferencia entre ellas, pero más frecuentemente se hace sustituyéndola.
- La **media recortada**, utilizada en datos preferentemente simétricos, con muchas observaciones anómalas. Se obtiene eliminando un determinado porcentaje de los datos menores y mayores.



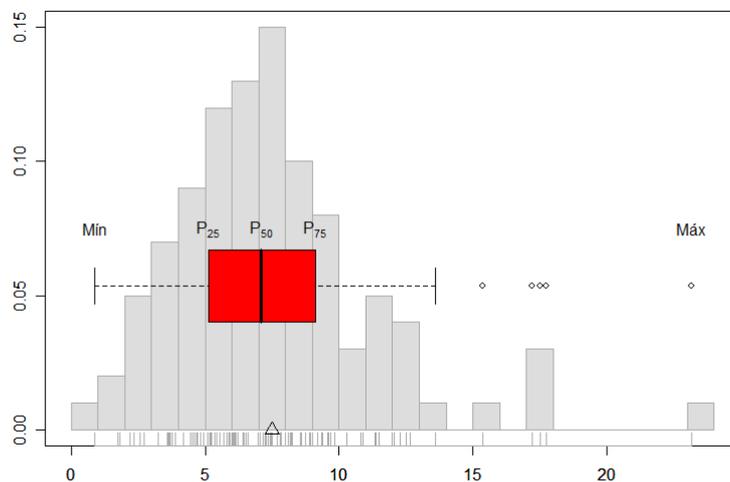
En cuanto a las medidas de dispersión podemos sustituir la desviación típica por el **rango intercuartílico (IQR)**, que es la diferencia entre el tercer cuartil y el primero. El **primer cuartil** (percentil 25, P_{25}) deja al 25% de los datos por debajo de él y el **tercer cuartil** (Percentil 75, P_{75}) deja al 75%. Por tanto sabemos que en esa zona central que tiene una longitud de un IQR, se encuentra el 50% central de las observaciones.

Cuando la muestra no parece tener una distribución normal, para obtener una idea aproximada de la distribución de los datos, se acostumbra a mostrar **un resumen en cinco números**, que son el valor mínimo, el primer cuartil, la mediana, el tercer cuartil,

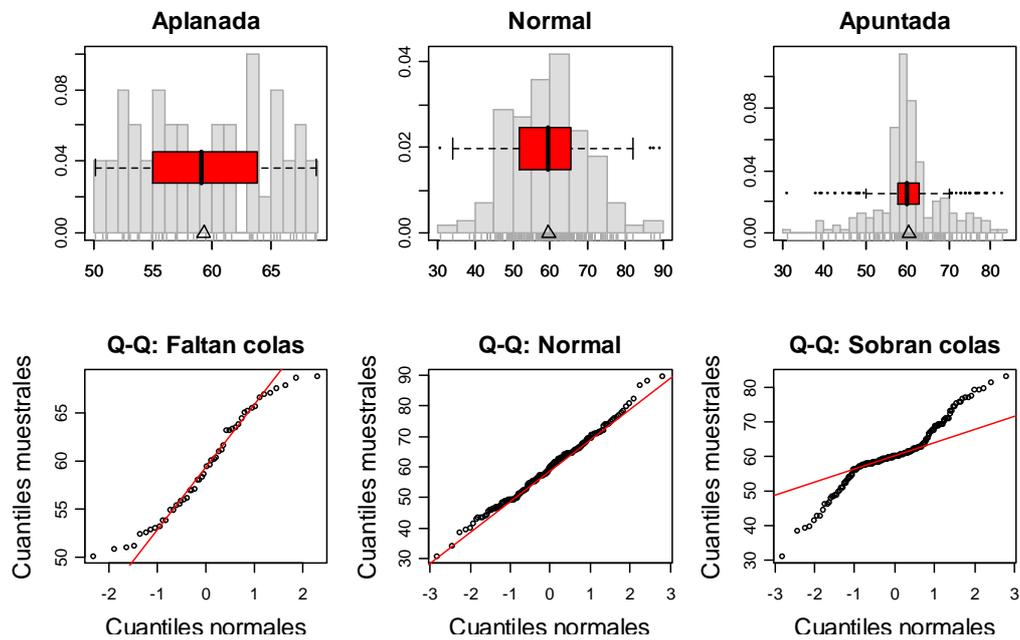
y el valor máximo. Estas cantidades vienen reflejadas en el **diagrama de cajas**. Este permite ver rápidamente si los datos son simétricos o si incluyen observaciones anómalas. Su composición se basa en una **caja** cuyos extremos son el primer y tercer cuartil (aproximadamente), con una marca interior para la mediana, y dos **bigotes**, cuya misión es delimitar hasta donde podemos considerar los datos de las colas como no anómalos. Cualquier valor que quede fuera de los bigotes es marcado como anómalo.

Cuando la muestra no parece tener una distribución normal, no debemos resumir la descripción de la muestra en media y desviación típica. La mediana (percentil 50) da una mejor idea del centro. Para dar una idea de cómo se distribuyen los datos a su alrededor, se complementa con otras cantidades: mínimo, primer cuartil, tercer cuartil y máximo, lo que forma un resumen en cinco números.

Estas cantidades vienen reflejadas en el diagrama de cajas de Tukey. Éste permite ver rápidamente si los datos son simétricos o si incluyen observaciones anómalas. Su composición se basa en una caja cuyos extremos son el primer y tercer cuartil (percentiles 25 y 75 respectivamente), con una marca interior para la mediana, y dos bigotes, cuya misión es delimitar hasta donde podemos considerar los datos de las colas como no anómalos. Cualquier valor que quede fuera de los bigotes es marcado como anómalo.



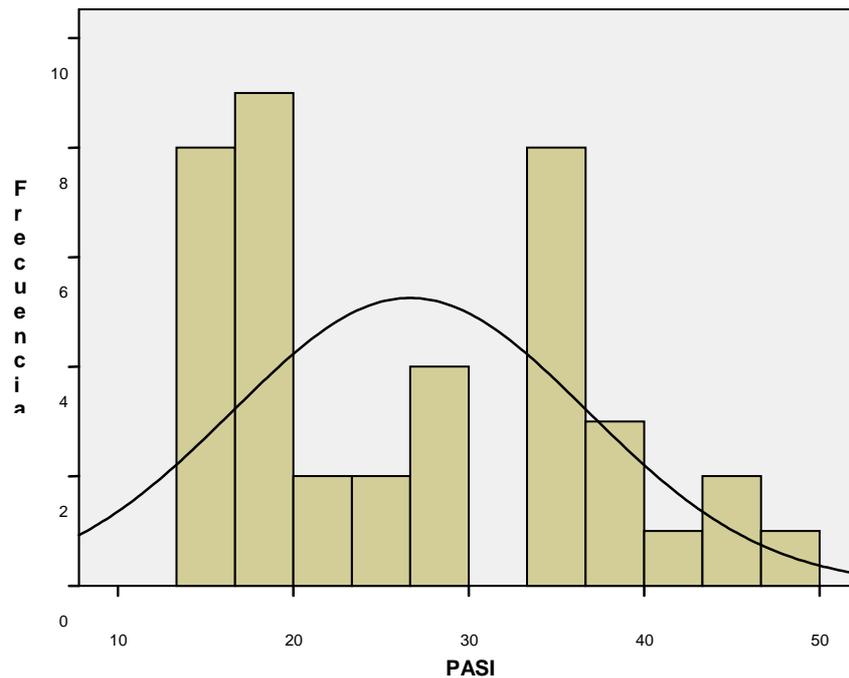
Un tipo de representación gráfica, que es muy útil para estudiar si unos datos sigue una distribución normal (o cualquier otra), es la de los gráficos cuantil-cuantil (Q-Q). En ellos se compara la distribución de los datos observados con respecto a una distribución de referencia. Si los puntos aparecen más o menos alineados, es que ambas distribuciones son similares. Las desviaciones con respecto a la distribución de referencia se aprecian como desviaciones con respecto a una línea recta en el gráfico Q-Q.



Ejemplo: En un estudio sobre psoriasis de 40 individuos, se obtuvieron los siguientes resultados al medir el indicador “*Psoriasis Area and Severity Index*” (PASI).

Estadísticos

pasi		
N	Válidos	40
	Perdidos	0
Media		26,88
Mediana		24,00
Desv. típ.		10,773
Asimetría		,635
Error típ. de asimetría		,374
Curtosis		-,789
Error típ. de curtosis		,733
Mínimo		15
Máximo		48
Percentiles	25	18,00
	50	24,00
	75	34,00



Los datos presentan cierta asimetría positiva que la alejan de la normalidad. Cualquiera de las siguientes alternativas podría ser suficiente para presentar los resultados.

La variable PASI presentaba valores comprendidos entre 15 y 48, con cierta asimetría positiva (0,635) y aplanamiento (curtosis=-0,79). El valor medio de PASI fue de 26,9 (desviación típica=10,8).

Otra posibilidad sería la siguiente.

Segunda opción: *La variable PASI presentaba valores comprendidos entre 15 y 48, con cierta asimetría positiva (0,635). Los cuartiles tomaron los valores: 18; 24 y 34.*

Una última forma más simple de presentar la información (sobre todo cuando queremos comparar diferentes grupos) sería:

La variable PASI presentaba una valor mediano de 24, con un rango intercuartílico de 16.

2.2 Estadística descriptiva bivalente

Si resumir la información de una variable es de por sí interesante, en investigación lo es mucho más el poner de manifiesto la posible relación entre dos de ellas. Para ello realizamos estudios donde intervienen ambas variables simultáneamente. Según sean los tipos de cada una de ellas usaremos técnicas diferentes.

2.2.1 Categórica-categórica

Cuando ambas variables son categóricas (o discretas con pocas modalidades), se suele presentar las observaciones en una tabla de contingencia. Esta es una tabla de doble entrada donde se presentan la distribución de frecuencias conjunta de las dos variables. Con ellas podemos estudiar si existe asociación entre ambas. Cuando no hay asociación, la distribución de porcentajes de una de las variables es similar para cada valor de la otra.

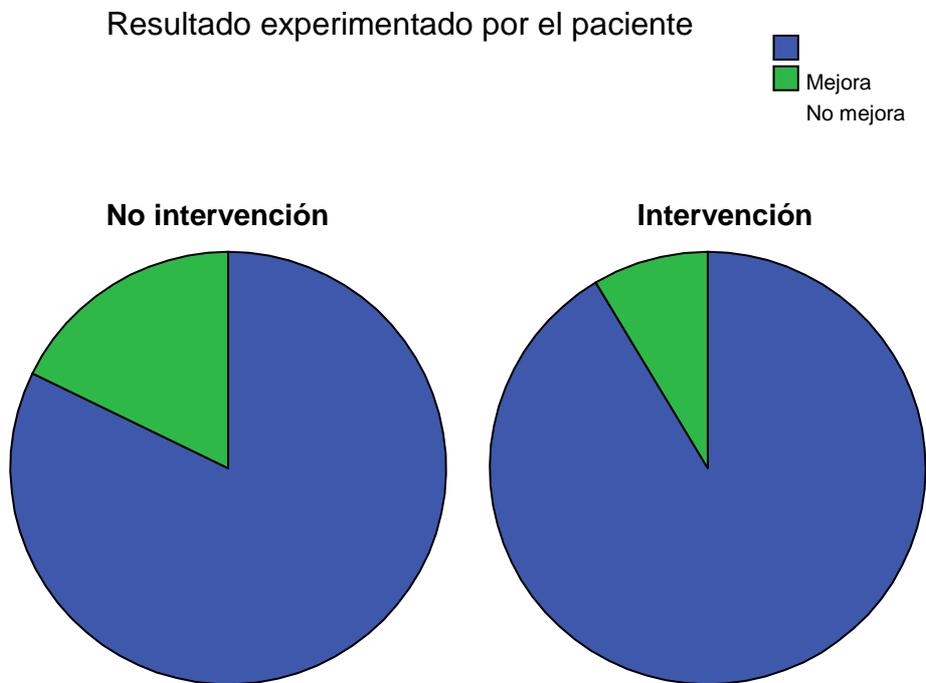
Adicionalmente, se suele presentar un estadístico (chi-cuadrado) que nos proporciona una idea del grado de asociación. Si el estadístico no toma un valor “*muy grande*”, se considera que no hay asociación entre ambas. Pero, ¿cómo saber cuando un valor del estadístico chi-cuadrado es grande? Para ello se acompaña de un valor, denominado significación, del que hablaremos más adelante. En general es costumbre admitir que los valores de un estadístico no se pueden considerar “*extremos*”, a menos que su significación sea inferior a 0,05.

Ejemplo: En un estudio sobre la efectividad de un fármaco se deseaba evaluar la mejoría de los pacientes. Para ello se formó dos grupos de pacientes, donde por azar, se les decidía aplicar bien el tratamiento, o bien un placebo (grupo de no intervención)

La tabla de contingencia adjunta muestra los resultados. El valor del estadístico chi-cuadrado fue de 4,28, con una significación de 0,028.

Tabla de contingencia Grupo * Resultado experimentado por el paciente

			Resultado experimentado por el paciente		Total
			Mejora	No mejora	
Grupo	No intervención	Recuento	111	24	135
		% de Grupo	82,2%	17,8%	100,0%
	Intervención	Recuento	117	11	128
		% de Grupo	91,4%	8,6%	100,0%
Total		Recuento	228	35	263
		% de Grupo	86,7%	13,3%	100,0%



La forma de exponer los resultados de forma puramente descriptiva sería la siguiente.

El grupo de intervención estaba formado por 128 pacientes frente a 135 del grupo control. Se observó en el grupo de intervención una mejoría en el 91,4% de los casos (117 pacientes), notablemente superior al 82,2% (111 pacientes) conseguido en el grupo de control.

En la práctica, es muy extraño que no se ofrezca, al mismo tiempo, una medida sobre lo diferente que son ambos porcentajes. Por ello, aún a riesgo de adelantar conceptos que se exponen más adelante, indicamos que lo habitual sería expresar el resultado de un modo más completo de la siguiente forma:

El grupo de intervención estaba formado por 128 pacientes y 135 el grupo de control. Se observó en el grupo de intervención una mejoría en el 91,4% de los casos (117 pacientes), notablemente superior al 82,2% (111

pacientes) conseguido en el grupo de control. La diferencia es estadísticamente significativa ($\chi^2 = 4,81$; $p=0,028$).

2.2.2 Categórica-Numérica

Supongamos que tenemos datos numéricos para varias categorías. Un ejemplo consistiría en un experimento donde hacemos mediciones numéricas en dos grupos: tratamiento y control. Podemos describir los resultados del experimento con sólo dos variables: Una variable categórica que representa el grupo de tratamiento, y otra que representa el resultado numérico

En estos casos, lo que se realiza es un estudio descriptivo de la variable numérica en cada una de las muestras (como se hizo anteriormente) y comparamos los resultados.

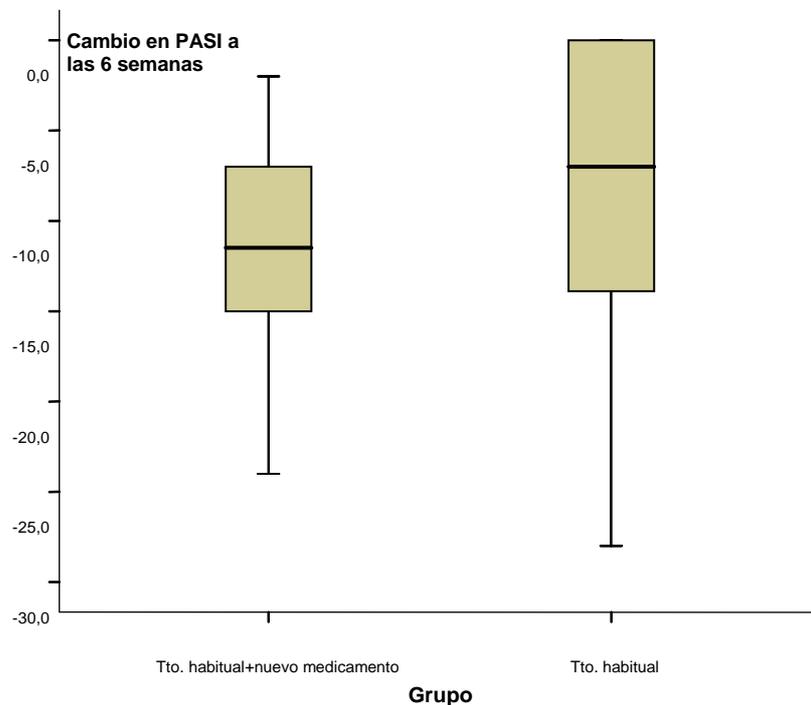
Si queremos presentar la comparación gráficamente, la opción más sencilla consiste en enfrentar los diagramas de caja generados para cada categoría de la variable categórica.

Ejemplo: Se quiere comparar la efectividad de un tratamiento en el alivio de la psoriasis cuando se aplica a pacientes que se encuentran con un ataque agudo. Para ello se observa el índice PASI de 40 pacientes en la primera consulta, tras el inicio de la fase aguda y se les asigna aleatoriamente a uno de dos grupos de tratamiento, cada uno formado por el mismo número de individuos. En uno de ellos (controles) se aplica el fármaco habitual, mientras que en el segundo se aplica, además del fármaco habitual, un medicamento experimental. Tras 6 semanas de tratamiento se mide de nuevo el valor de PASI. El efecto que ha tenido el tratamiento se observa en una variable que es calculada como la diferencia entre los valores medidos en ambos momentos.

Con gran frecuencia, estas variables, que son calculadas como las diferencias o cambios observados en cada paciente, tienen tendencia a presentar distribución normal. Una justificación, viene de la eliminación de ciertas fuentes de variabilidad que son perjudiciales para la normalidad: Cuando sólo se mide a las variables en un instante de tiempo, debido a que hay muchos factores que influyen enormemente en la variable que se mide y además estos se presentan de forma muy dispar en cada individuo, podemos encontrarnos con distribuciones muy asimétricas, y por tanto alejadas de la normalidad. La suma de estos efectos a nivel de individuo suele quedar corregida al utilizarse a cada individuo en tiempos diferentes para medir la diferencia entre dos momentos (efecto del tratamiento). Esta es una situación ideal para presentar los resultados de ambos grupos como media y desviación típica. Los resultados de la comparación podrían formularse como sigue:

El grupo experimental experimentó una mejoría media en PASI de 12 puntos (desviación típica=10) frente a los 8 puntos (desviación típica=11) del grupo de control.

En estas comparativas debe destacarse adicionalmente, si la diferencia observada es interesante desde el punto de vista clínico. Otra cuestión que trataremos más adelante, es si la diferencia observada es además significativa desde un punto de vista estadístico.



2.2.3 Numérica-Numérica.

Cuando hablamos de comparar dos variables numéricas, pensamos en establecer la posible relación entre ellas. La vía más directa para estudiar la posible asociación, consiste en inspeccionar visualmente un **diagrama de dispersión** (nube de puntos). Con ella se pretende buscar patrones en los datos.

Uno de los patrones más básicos para reconocer, es el de las tendencias lineales, que es lo que ocurre cuando los puntos del diagrama tienden a no alejarse demasiado de una línea recta. Otro tipo de patrones pueden ser reconocibles, como sería la aproximación a una curva cuadrática.

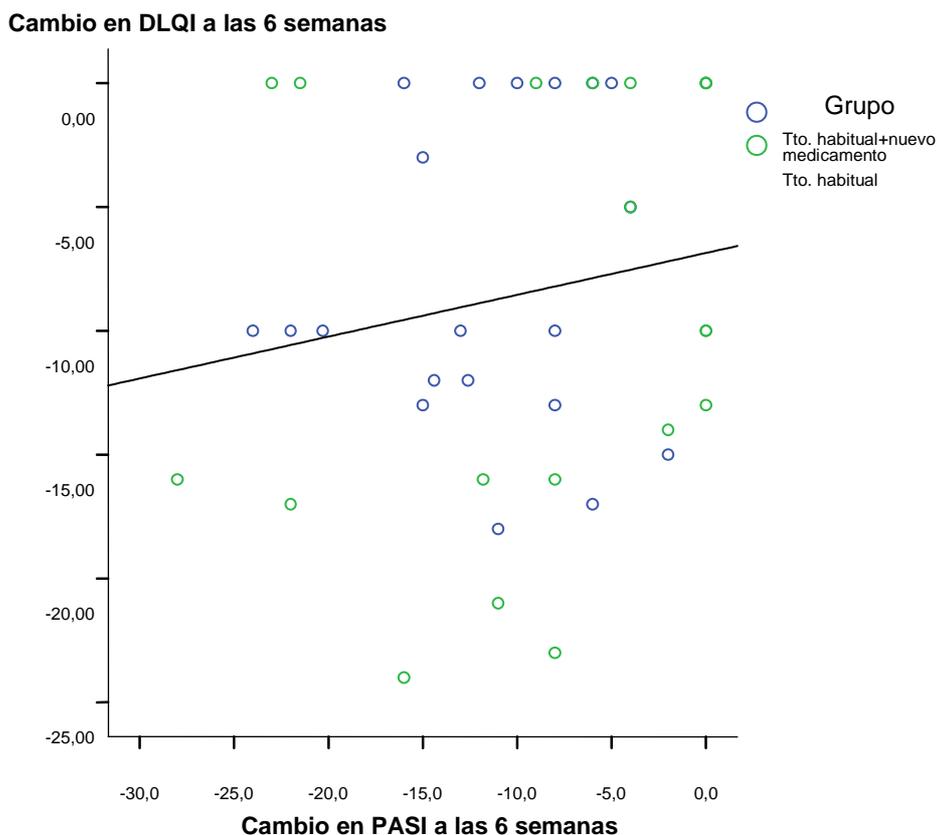
Si reconocemos una tendencia como las mencionadas, es una indicación de que puede valer la pena explorar con más profundidad dichas relaciones con modelos que incluyan tal vez más variables. Si es el caso, puede interesarnos proseguir con un análisis de regresión múltiple.

Desde un punto de vista puramente descriptivo, el estadístico más interesante (en la mayoría de los casos) donde sólo se explora la posible asociación lineal entre dos variables numéricas, es el coeficiente de correlación lineal de Pearson. Este nos indica con signo positivo si hay una relación directa entre ambas variables. Análogamente, valores negativos, indican relación inversa. Los valores cercanos a cero son muestras de

ausencia de relación lineal, mientras que los cercanos a 1 ó -1 indican una buena relación lineal.

Hay otra forma de exponer la relación existente entre dos variables, y que se extiende más allá del caso de la simple relación lineal entre dos variables. Se puede utilizar para describir cualquier otro tipo de relación (cuadrática, cúbica, etc.) En estos casos se utiliza un coeficiente que se suele escribir como R^2 , y al que se le denomina como **porcentaje de variabilidad explicado** o **bondad de ajuste**. Toma valores entre 0 y 1, pero también lo solemos encontrar expresado en porcentaje. Además sirve para explicar relaciones no sólo entre dos, sino entre múltiples variables. En el caso del modelo lineal, R^2 no es más que el coeficiente de correlación lineal al cuadrado.

Ejemplo: Continuando con el ejemplo anterior, se evaluó durante las 6 semanas de tratamiento, no sólo el cambio producido en la variable Psoriasis Area and Severity Index (PASI), sino también “Dermatology Life Quality Index”, DLQI. Los resultados se observan en la gráfica adjunta, siendo el coeficiente de correlación de $r=0,17$, lo que supone un valor de $R^2=0,029$



Esto se corresponde con una ligerísima asociación directa entre los cambios producidos en ambas variables en ese periodo, pero que desde el punto de vista práctico es irrelevante. Podríamos expresarlo del siguiente modo:

El coeficiente de correlación de Pearson tomó un valor $r=0,17$ ($p=0,34$); lo que nos indica que no existe una asociación lineal entre los cambios producidos en PASI y DLQI

En el párrafo anterior hemos introducido un p-valor que contrasta la hipótesis de no asociación entre las variables. Sobre p-valores hablaremos más adelante.

De forma equivalente podemos utilizar R^2 , para expresar lo mismo en un lenguaje ligeramente diferente:

El porcentaje de variabilidad explicado en la variación de DLQI por la variabilidad de PASI fue de tan sólo $R^2=2,9\%$. Por tanto el efecto de los tratamientos se expresaba en cada índice de modo independiente.

3 Estimación y contrastes de hipótesis

Con frecuencia, queremos hacer algo más que describir los resultados con unos pacientes. Buscamos poder extender los resultados a alguna población más general de la que se supone que los individuos de la muestra son una representación. A veces, dadas las limitaciones que nos encontramos, es difícil imaginar cuál es esta población sobre la que se podrían inferir conclusiones. Como nuestro objetivo no es más que ayudar a presentar los resultados del análisis estadístico, de ahora en adelante supondremos que esta es una cuestión resuelta por el investigador.

3.1 p-valores

Para presentar los resultados disponemos de varias posibilidades, de la que la más común es la presentación de p-valores (significación estadística). **Un p-valor es una cantidad que mide la evidencia en contra de la hipótesis nula.** Esta última afirma que no hay ningún patrón interesante en los datos, como por ejemplo la no existencia de diferencia entre grupos de tratamiento. La forma habitual con la que nos lo encontramos suele ser:

Los resultados anteriores muestran que hay diferencias significativas entre los grupos de tratamiento ($p=0,003$)

A la hora de suministrar los p-valores significativos, debemos utilizar al menos 3 cifras decimales de precisión. Cuando p es muy pequeño es posible indicarlo de forma más breve escribiendo por ejemplo $p<0,001$.

Poco a poco se ha ido extendiendo la costumbre de aceptar que, p-valores inferiores a 0,05 indiquen evidencia estadísticamente significativa a favor de diferencias entre tratamientos, pero vamos a observar las siguientes cuestiones sobre su uso:

- Especialmente en estudios con grandes muestras, los p-valores suelen ser extremadamente pequeños, sin que por ello las diferencias numéricas que observe el médico sean en la práctica (clínicamente) significativas. Es posible que no queramos considerar entonces las diferencias como relevantes.
- Cuando los p valores son grandes ($p>0,05$), seguramente no consideremos las diferencias como estadísticamente significativas, pero eso no quiere decir que no lo sean desde un punto de vista clínico. Esto es especialmente cierto cuando las muestras en estudio son pequeñas.
- No debemos considerar que un p-valor grande es una prueba de que los tratamientos en comparación son equivalentes. Para esto hay pruebas específicas, denominadas de equivalencia de tratamiento. La no evidencia de la ausencia de efecto no tiene nada que ver con la evidencia de la ausencia.

3.2 Intervalos de confianza y errores típicos

Mejor que ofrecer p-valores, es ofrecer intervalos de confianza, sobre todo si los tamaños muestrales son muy grandes, pues en ese caso, es sencillo obtener p-valores muy pequeños (estadísticamente significativos), sin que por ello tengan significación clínica. El suministro de intervalos de confianza permite juzgar si este es el caso. Cuando esto es muy engorroso o no es posible, se pueden sustituir por errores típicos o algunos estadísticos de interés que permitan al lector obtener la información que le interese para juzgar lo interesante (desde el punto de vista práctico) de las diferencias observadas.

3.3 Tablas de contingencia

Si el resultado del estudio que estamos haciendo es dicotómico, se espera que los resultados obtenidos en la muestra los ofrezcamos como porcentaje, y por tanto, las cantidades a estimar en la población deben corresponderse a un intervalo de confianza, alrededor de dicho porcentaje. Sin embargo, normalmente estos porcentajes no son calculados en un sólo grupo de pacientes, lo que no permite estimarlos fácilmente.

Cuando la tabla de contingencia no es de dimensiones 2x2, no hay mucho más que hacer que mostrarla y referir la significación del estadístico chi-cuadrado. Hemos de tener cuidado con las tablas grandes, pues es frecuente que una buena cantidad de

casillas tenga frecuencias esperadas inferior a 5, lo que invalida las aproximaciones que suelen estar programadas, y los programas de análisis suelen especificarlo. En estos casos debemos utilizar las pruebas exactas, que normalmente pueden usarse para tablas de dimensiones no muy grandes, dependiendo del software y del ordenador que usemos. Si la situación no es evitable, debemos considerar redefinir las variables.

En el caso de tablas 2x2 hay más información que deberíamos ofrecer además de la significación, y a eso está dedicado el resto de la sección.

En el caso de que se trate de comparar los porcentajes en sólo dos grupos, lo más conveniente es ofrecer junto a los porcentajes de cada grupo, el valor del estadístico z para la diferencia de proporciones, que no es más que la raíz del estadístico chi-cuadrado en este caso. Así, retomando un ejemplo anterior, y donde queremos remarcar que lo que nos interesa realmente es la diferencia entre dos proporciones, podríamos presentar los resultados del siguiente modo:

Tabla de contingencia Grupo * Resultado experimentado por el paciente

			Resultado experimentado por el paciente		Total
			Mejora	No mejora	
Grupo	No intervención	Recuento	111	24	135
		% de Grupo	82,2%	17,8%	100,0%
	Intervención	Recuento	117	11	128
		% de Grupo	91,4%	8,6%	100,0%
Total		Recuento	228	35	263
		% de Grupo	86,7%	13,3%	100,0%

El grupo de intervención estaba formado por 128 pacientes frente a 135 del grupo control. Se observó en el grupo de intervención una mejoría en el 91,4% de los casos (117 pacientes), significativamente superior al 82,2% (111 pacientes) del grupo de control ($z=2,2$; $p=0,028$).

Otro tipo de pruebas con variables cualitativas que también conduce a una tabla de doble entrada de dimensiones 2x2, es aquel en el que se observa la misma variable dicotómica (síntomas presentes/ausentes) en un sólo grupo de individuos, pero se hace en **dos ocasiones diferentes** para saber si los pacientes han evolucionado con el tiempo. En este caso se utiliza la prueba de McNemar. Un ejemplo de uso sería el siguiente, donde se estudia si un grupo de individuos muy sensibles a la luz solar tiene tendencia a cambiar en 2 años consecutivos.

Tabla de contingencia Irritaciones verano 2006 * Irritaciones verano 2007

Recuento		Irritaciones verano 2007		Total
		No	Sí	
Irritaciones verano 2006	No	40	16	56
	Sí	4	80	84
Total		44	96	140

Se estudiaron 140 pacientes durante dos años consecutivos. De los 56 pacientes que no presentaron irritaciones el primer verano, 16 (28,6%) si lo presentaron el segundo, mientras que sólo 4 de los 44 (9,1%) que no presentaron problemas el segundo verano, lo presentaron el primero. Esto muestra un empeoramiento en la condición del primer al segundo año, siendo la diferencia estadísticamente significativa (Mc Nemar $p=0,012$).

Otra forma de enunciarlo, más clara, pero que no refleja tan directamente lo que se contrasta en la prueba de Mc Nemar, que son los cambios de status en diferentes momentos, sería:

Se estudiaron 140 pacientes durante dos años consecutivos. El 60% presentaba irritaciones el primer año, porcentaje que aumentó al 68,6% el segundo año. La diferencia es estadísticamente significativa (McNemar $p=0,012$).

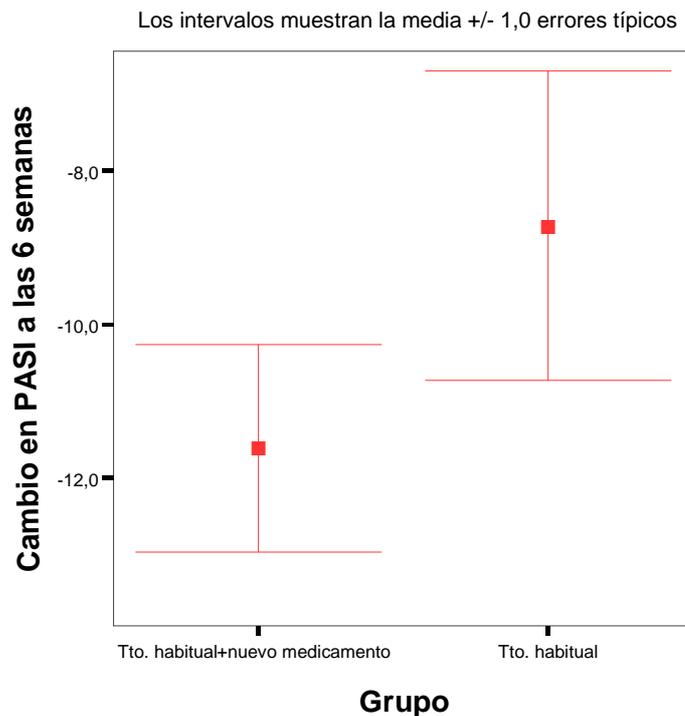
3.4 Comparando medias en varios grupos

En el caso de que estudiemos variables de distribución aproximadamente normal, lo más adecuado es que las comparaciones entre grupos lo hagamos basándonos en estimaciones de la media dentro de cada grupo. En realidad lo podemos extender a distribuciones que presenten cierta asimetría, siempre que estas no sean muy grandes, y vayan acompañadas de un tamaño muestral que lo compense.

Cuando sea posible se espera un intervalo de confianza asociado a las medias, aunque esto no siempre es interesante, sobre todo cuando se está comparando los resultados de un tratamiento en múltiples grupos de pacientes. En cualquier caso, siempre que se trate de hacer inferencia, y tratándose de medias, hay que ofrecer una cantidad con cada una denominada **error típico**. Este nos da una idea de la precisión de la estimación en cada grupo de tratamiento. Cuando las medias en los grupos queden reflejadas en una gráfica a efecto de comparaciones, hemos de intentar acompañarlas de las respectivas barras de error, aclarando si las estamos eligiendo del tamaño de un error típico, cuando se intenta reflejar este directamente, o bien, dos errores típicos, si intentamos reflejar cierta noción

de intervalo de confianza al 95%. Aunque no hay que engañarse en este último caso, pues estos intervalos no son reales.

Retomemos de nuevo un ejemplo anterior, donde se observaba la variable cambio en PASI tras 6 semanas de tratamiento para dos grupos de pacientes. Si nuestra intención es comunicar diferencias de efecto entre grupos, la forma de hacerlo gráficamente podría ser la siguiente



Para comunicar lo mismo en una tabla lo haríamos del siguiente modo:.

Grupo	Tamaño del grupo	Media \pm E. Típico
Tratamiento habitual (controles)	20	-8,7 \pm 2,0
Tratamiento habitual +medicamento (experimental)	20	-11,6 \pm 1,3

Una vez que exponemos las estimaciones de la media en cada grupo, trataremos de aplicar contrastes que tratan sobre si los efectos medidos en cada grupo de tratamiento son similares.

3.4.1 Comparando dos medias: t-student

Cuando el número de grupos a comparar es dos (como en el último ejemplo) se utiliza la prueba t-student. De nuevo hay que recordar que estamos utilizando esta prueba porque

admitimos un cierto grado de normalidad en los datos. Debemos hacer mención (normalmente en la sección *Material y Métodos*), sobre cómo comprobaremos esta posibilidad (normalmente usando métodos gráficos como los Q-Q, o bien usando directamente contrastes de normalidad, como el del Kolmogorov-Smirnov), y también mencionar qué hubiésemos hecho en caso de no ser cierta, lo que consiste normalmente en aplicar una prueba no paramétrica (como la de Mann-Whitney o la de Wilcoxon). Se podría redactar el resultado como sigue:

En el grupo de control se experimentó una reducción media de PASI de 8,7 (ET=2), y en el grupo experimental de 11,6(ET=1,3). La diferencia de efecto medio fue 2,9 (ET=2,42) a favor del grupo experimental, no siendo estadísticamente significativa ($t=1,19;p=0,238$).

Este último es un ejemplo de un resultado no significativo. Es frecuente que se expliquen con más detalles los resultados significativos, y el resto, se agrupen bajo una expresión más general y se expongan todos en la misma frase como sigue:

No se observó diferencias significativas para el resto de variables y tratamientos ($p>0,2$).

El valor de p mostrado, es uno que cubre a todos los p de esos resultados no significativos. Esta frase resumen, no se utiliza para aquellos resultados que estén en el límite de la significación (por ejemplo $p=0,071$), sino que se hace a partir de algún valor de significación muy grande (como el mostrado).

3.4.2 Comparando tres o más medias: ANOVA

Cuando hay más de dos grupos de tratamiento, tenemos la posibilidad de utilizar la prueba ANOVA. Esta no sólo requiere la normalidad, sino que no haya diferencias de variabilidad muy grandes en cada grupo. Como en el caso anterior debemos mencionar en la sección *Material y Métodos* cómo verificaremos la validez de las premisas (que en el caso de la igualdad de variabilidad puede ser la prueba de Levene), y qué alternativa usaremos en caso de que no se cumpla, que normalmente consiste en aplicar la prueba no paramétrica de Kruskal-Wallis.

En este caso, el p -valor muestra la evidencia que hay en contra de que las medias de todos los grupos de tratamiento son iguales. Si se rechaza esta hipótesis ($p<0,05$ como es habitual), pasaremos a contrastar qué grupos de tratamiento son los responsables de este rechazo. Estos son los análisis *post-hoc* de los resultados de ANOVA. Hay múltiples contrastes a utilizar, dependiendo de las características de las muestras: *Diferencia Honestamente Significativa de Tukey, Scheffé, Dunnet...*

3.4.3 Contrastes no paramétricos

Los contrastes no paramétricos suelen ser menos potentes que los anteriores, es decir, tienen mayor riesgo de no mostrar resultados significativos, cuando en realidad sí hay diferencia entre los grupos de tratamiento. A cambio tienen la ventaja de que los requerimientos que se necesitan se cumplen casi siempre en los estudios médicos.

Los utilizamos para comparar valores de variables numéricas en varios grupos, cuando la aproximación mediante el uso de medias no ha sido posible. En este caso los resultados de cada grupo de tratamiento no los expresaremos en formas de media±error típico, sino ofreciendo la mediana y rango intercuartílico de cada grupo, además del valor obtenido en el contraste de significación. Un ejemplo de cómo presentar los resultados sigue a continuación:

<u>Grupo</u>	<u>Tamaño del grupo</u>	<u>Mediana±IQR</u>
Tratamiento habitual (controles)	20	-7±12
Tratamiento habitual +medicamento (experimental)	20	-11,5±8,5

No se encontró que el grupo experimental tuviese resultados significativamente diferentes de los controles (Wilcoxon $p=0,103$)

4 Guía para elegir el contraste adecuado para variables numéricas

