acadèmica

Estadística aplicada al ámbito sanitario

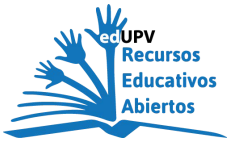
Raffaele Vitale | Marta Hermenegildo Caudevilla
Alberto J. Ferrer Riquelme

Editorial
Universitat Politècnica
de València

Raffaele Vitale
Marta Hermenegildo Caudevilla
Alberto J. Ferrer Riquelme

Estadística aplicada al ámbito sanitario

edUPV
Universitat Politècnica de València



http://tiny.cc/edUPV_rea

Colección *Académica*

Para referenciar esta publicación utilice la siguiente cita:
Vitale, Raffaele; Hermenegildo Caudevilla, Marta; Ferrer Riquelme, Alberto J. (2022). *Estadística aplicada al ámbito sanitario*. Valencia: edUPV

Autoria
© Raffaele Vitale
Marta Hermenegildo Caudevilla
Alberto J. Ferrer Riquelme

Editado por edUPV, 2022
www.lalibreria.upv.es / Ref.: 6368_01_01_01

ISBN: 978-84-1396-024-1
DOI: https://doi.org/10.4995/REA.2022.636801

Si el lector detecta algún error en el libro o bien quiere contactar con el autor, puede enviar un correo a edicion@editorial.upv.es

Autores

RAFFAELE VITALE

Doctor en Estadística y Optimización por la Universitat Politècnica de València. Está especializado en el análisis de datos de naturaleza química, biológica y biomédica. Profesor titular de la Universidad de Lille, en Francia, donde combina su experiencia en estadística con su sólida formación en ciencias de la vida, desarrollando tareas docentes e investigadoras. Su trabajo se centra principalmente en el diseño y aplicación de metodologías y algoritmos para el procesamiento de imágenes hiperespectrales y de microscopía óptica. Colabora activamente con entidades e instituciones públicas y privadas en Italia, Países Bajos, Brasil, Noruega, Polonia y España.

MARTA HERMENEGILDO CAUDEVILLA

Doctora en Farmacia por la Universitat de València y farmacéutica especialista en Farmacia Hospitalaria. Trabaja en el Servicio de Farmacia del Hospital Universitario Dr. Peset de Valencia donde coordina la Unidad de Atención Farmacéutica a Pacientes Externos. Ha sido jefa de servicio de Gestión y Coordinación de la Investigación Sanitaria de la Conselleria de Sanidad. Ha publicado numerosos artículos en revistas de ámbito nacional e internacional, así como capítulos de libros y monografías relacionadas con la farmacia hospitalaria y con la gestión de la investigación sanitaria.

ALBERTO J. FERRER RIQUELME

Catedrático de universidad del Departamento de Estadística e Investigación Operativa Aplicadas y Calidad, y director del grupo de Ingeniería Estadística Multivariante de la Universitat Politècnica de València. Especializado en el desarrollo y aplicación de técnicas estadísticas multivariantes y su integración con herramientas de *machine learning*, mantiene una intensa actividad docente, investigadora y de transferencia tecnológica en técnicas de análisis de datos y mejora de procesos en diversos ámbitos como la industria, la tecnología y el sector de la salud.

Resumen

Esta monografía proporciona un resumen de las herramientas básicas de estadística descriptiva e inferencia estadística, con un apartado dedicado al análisis de supervivencia. Se presentan 3 ejemplos de aplicación de las técnicas con datos reales del ámbito clínico usando el programa estadístico SPSS. Está pensada para usuarios que quieran recordar de forma sencilla y rápida los conceptos básicos de estadística descriptiva y contrastes de hipótesis, así como tener una guía de qué técnica aplicar en función del problema a resolver. El capítulo dedicado al análisis de supervivencia está especialmente destinado a las personas interesadas en disponer de una guía práctica de cómo realizar este análisis con SPSS. Los ejemplos ilustrados son del ámbito de las Ciencias de la Salud, lo que puede ser especialmente útil, aunque no exclusivo, para todos los interesados de este ámbito.

Preámbulo

Esta monografía proporciona un resumen de las herramientas básicas de estadística descriptiva e inferencia estadística (contrastes de hipótesis, ANOVA y modelos de regresión lineal y logística), con un apartado dedicado al análisis de supervivencia (método de Kaplan-Meier y modelo de regresión de Cox). Así mismo se presentan 3 ejemplos de aplicación de las técnicas con datos reales del ámbito clínico usando el programa estadístico SPSS. Está pensada para usuarios que, tras haber cursado alguna formación básica en herramientas estadísticas de análisis de datos, quieran recordar de forma sencilla y rápida los conceptos básicos de estadística descriptiva y contrastes de hipótesis, así como tener una guía de qué técnica aplicar en función del problema a resolver. El capítulo dedicado al análisis de supervivencia está especialmente destinado a las personas interesadas en disponer de una guía práctica (con su fundamentación conceptual) de cómo realizar este análisis con SPSS. Aunque todas estas técnicas se aplican en muchos campos de la ciencia y la tecnología, los ejemplos ilustrados son del ámbito de las Ciencias de la Salud, lo que puede ser especialmente útil para todos los alumnos y profesionales de este ámbito.

Índice

Capítulo 1
Conceptos básicos de estadística descriptiva

1.1. Variables estadísticas: definición y tipos

En el ámbito de la estadística, el término *variable* hace referencia a cualquier característica o valor numérico que se asocia, se mide o se cuenta para los individuos estudiados. La edad, el sexo, el peso, el número de hijos, el tipo de tratamiento recibido o el número de ciclos son ejemplos de variables.

Existen distintos tipos de variables que, en función de su naturaleza y propiedades, se tienen que tratar y analizar de formas diferentes. Las variables **numéricas** toman valores que describen una cantidad medible con un número y, por lo tanto, son variables **cuantitativas**. Las variables numéricas pueden clasificarse además como continuas o discretas: una variable **continua** es una variable numérica que puede tomar cualquier valor en el conjunto de los números reales, como el peso de una persona (65,1 kg; 70,234 kg; …) o el tiempo de disolución de un fármaco (25,3 s; 0,4217 min; …). Una variable **discreta** es una variable numérica que puede solamente tomar un valor numérico entero, por ejemplo, el número de ganglios linfáticos afectados en un proceso tumoral o el número de comorbilidades (0, 1, 2, …).

Por otro lado, las variables **categóricas** describen una cualidad o característica y suelen representarse por valores no numéricos (aunque también pueden representarse por valores numéricos que han de tratarse como simples códigos). Las variables categóricas pueden clasificarse como ordinales o nominales. Una variable **ordinal** es una variable categórica que toma valores ordenados o clasificados lógicamente, es

decir, estos valores son representativos de diferencias de rango entre los distintos individuos analizados. Algunos ejemplos de variables categóricas ordinales son el tamaño de un tumor primario (T1, T2, T3 T4) o el nivel de respuesta a la quimioterapia (respuesta completa, respuesta parcial, enfermedad estable, progresión). En cambio, una variable **nominal** es una variable categórica que toma valores que no pueden ordenarse según una secuencia lógica, como puede ser el sexo o el grupo sanguíneo. Los datos contenidos en una variable categórica son datos **cualitativos**. La Figura 1 resume esta clasificación jerarquizada entre tipos de variables.

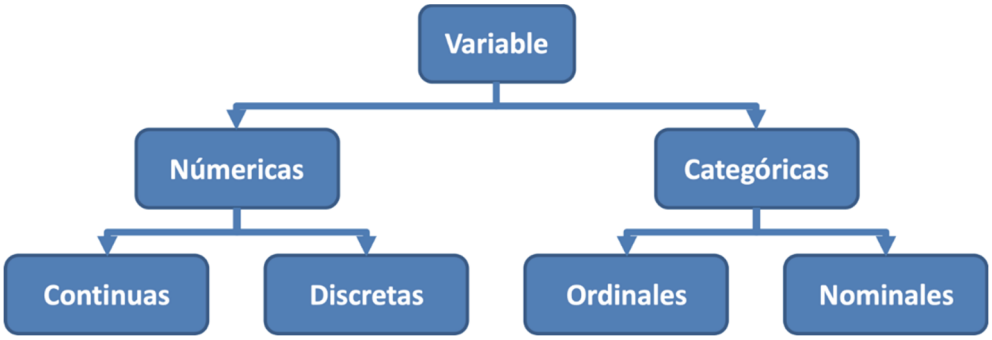


Figura 1 – Clasificación de los distintos tipos de variables

1.2. Descriptores de variables

Para resumir y evaluar la información que contienen las variables recogidas para los diferentes individuos incluidos en un estudio, se puede recurrir a varios tipos de medidas estadísticas en función de la naturaleza de dichas variables.

Para variables numéricas y categóricas ordinales, estas medidas se pueden clasificar en dos grupos en base al tipo de información que proporcionan: **medidas de tendencia central** y **medidas de dispersión**.

Las medidas de tendencia central corresponden a los valores centrales de la variable alrededor de los cuales se distribuyen los valores medidos para el conjunto de individuos. Las medidas de dispersión, por otro lado, cuantifican el grado de variación de los valores de la variable alrededor del valor central a lo largo de la muestra investigada. En base al tipo de variable (numérica o categórica ordinal), las medidas de tendencia central y de dispersión a utilizar varían. Algunas de las más utilizadas son:

a) Medidas de tendencia central (posición)
   1. **Media aritmética** (para variables numéricas). Es el valor obtenido por la suma de todos los valores x_i de la variable dividida entre el número total, n, de valores: $\bar{x} = \sum_{i=1}^{n} x_i/n$. Su calculo es muy sencillo y se interpreta como el "centro de masas" del conjunto de datos. Así, si la variable analizada se distribuye si-

métricamente alrededor de su valor central (es decir, si la distribución de valores a la derecha y a la izquierda de su valor central es similar), la media aritmética constituye una medida adecuada de tendencia central. Sin embargo, está muy afectada por la asimetría de la distribución de la variable y por valores extremos (anómalos) que la variable puede tomar en algunos individuos.

2. **Mediana,** también llamada **segundo cuartil** (para variables numéricas y variables categóricas ordinales). Es el valor que ocupa la posición central de un conjunto de observaciones ordenadas de menor a mayor (el 50 por ciento de las observaciones son mayores que este valor y el otro 50 por ciento son menores). A diferencia de la media aritmética, la mediana es una medida de posición adecuada para distribuciones asimétricas, y robusta (poco afectada) a la presencia de valores anómalos. Media aritmética y mediana coinciden en distribuciones perfectamente simétricas.
3. **Moda** (para variables numéricas y variables categóricas ordinales). Es el valor más frecuente de la variable. A veces, la moda no es única (puede haber más de una moda si múltiples valores se repiten con la misma frecuencia) y puede no situarse en el centro de la distribución de la variable.

De todas estas medidas de posición, las más usadas son la media aritmética para variables con distribución simétrica y la mediana para variables con distribución asimétrica. Es interesante destacar que la moda también podría calcularse para las variables categóricas nominales, pero no tendría ninguna interpretación como medida de tendencia central, sino simplemente como la característica más frecuente.

b) Medidas de dispersión (variabilidad)

1. **Desviación típica** (para variables numéricas). Se calcula 1) sumando los cuadrados de las diferencias de cada valor de la variable en cuestión y su media y 2) calculando la raíz cuadrada del ratio entre esta suma y el número total de valores medidos menos 1: $s = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2/n - 1}$. Constituye una medida del grado de variación de los valores de una variable alrededor de su media aritmética. Al igual que la media aritmética, está muy afectada por la presencia de valores anómalos que la variable puede tomar y por la asimetría de su distribución (por lo que no es una medida de dispersión adecuada cuando la distribución no es simétrica). Una desviación típica baja indica que la mayor parte de los datos de una muestra tienden a estar agrupados cerca de su media, mientras que una desviación típica alta indica que los datos se extienden sobre un rango de valores más amplio. La desviación típica a veces se expresa elevándola al cuadrado, dando lugar a la **varianza** (otra medida de dispersión).
2. **Rango** (para variables numéricas y variables categóricas ordinales). Corresponde al intervalo entre el valor máximo y mínimo de una variable. Cuanto mayor es el rango, más dispersos serán los datos. Proporciona una idea inmediata

del grado de dispersión de una variable, pero solo depende de sus dos valores más "extremos" (su máximo y su mínimo) e ignora sus valores intermedios, por lo que es una medida de dispersión adecuada únicamente para muestras pequeñas.

3. **Primer y tercer cuartiles** (para variables numéricas y variables categóricas ordinales). El primer cuartil es el valor de la variable tal que el 25 por ciento de las observaciones (ordenadas de menor a mayor) son menores que este valor y el otro 75 por ciento son mayores que él. Por su parte, el tercer cuartil es el valor que deja por debajo de sí al 75% de los datos (siendo el 25% mayores que él). Restando el tercer cuartil del primer cuartil se obtiene el **intervalo intercuartílico** que indica el rango que comprende el 50% de los valores centrales de la variable. A diferencia de la desviación típica, el intervalo intercuartílico es una medida de dispersión adecuada para distribuciones asimétricas, y robusta (poco afectada) frente a la presencia de valores anómalos.

De todas estas medidas de dispersión, las más usadas son la desviación típica (y también la varianza) para variables con distribución simétrica, y el intervalo intercuartílico para variables con distribución asimétrica.

Cuando se dispone de variables categóricas nominales (v.g. sexo) no tiene sentido calcular ninguno de los parámetros de posición o dispersión comentados. Para este tipo de variables existen otras medidas que ayudan a caracterizarlas: la **frecuencia** y el **porcentaje**, que también pueden usarse en el caso de variables categóricas ordinales. La frecuencia corresponde al número de veces que se repite un valor de una variable. El porcentaje es una medida de **frecuencia relativa,** y es igual a la frecuencia dividida entre el número total de valores medidos, y multiplicada por 100.

La siguiente tabla contiene un listado con ejemplos de variables medidas en ensayos clínicos junto a los descriptores que pueden ser utilizados para su caracterización estadística:

Tabla 1 – Ejemplos de variables medidas en ensayos clínicos y de los descriptores que pueden ser utilizados para su caracterización estadística.

Variable	Tipo	Medidas de tendencia central	Medidas de dispersión	Otras medidas
Edad	Numérica continua	Media/Mediana/ Moda	Desviación típica/Rango/ Intervalo intercuartílico	-
Sexo	Categórica nominal	-	-	Frecuencia/ Porcentaje/Moda
Consumo de tabaco (Si/No)	Categórica nominal	-	-	Frecuencia/ Porcentaje/Moda
Nº comorbilidades (0, 1, 2, …)	Numérica discreta	Media/Mediana/ Moda	Desviación típica/Rango/ Intervalo intercuartílico	
Histología de un tumor (A/B)	Categórica nominal	-	-	Frecuencia/ Porcentaje/Moda
Extensión de una neoplasia (I a IV)	Categórica ordinal	Mediana/Moda	Rango/Intervalo intercuartílico	Frecuencia/ Porcentaje
Supervivencia libre de progresión (SLP)	Numérica continua	Media/Mediana/ Moda	Desviación típica/Rango/ Intervalo intercuartílico	-
SLP a los 12 meses (Si/No)	Categórica nominal	-	-	Frecuencia/ Porcentaje/Moda
Nivel de respuesta al tratamiento (completa, parcial, estable, progresión)	Categórica ordinal	Mediana/Moda	Rango/Intervalo intercuartílico	Frecuencia/ Porcentaje
Efectos adversos (lista de efectos)	Categórica nominal	-	-	Frecuencia/ Porcentaje/Moda
Efecto adverso específico: neutropenia (Sí/No)	Categórica nominal	-	-	Frecuencia/ Porcentaje/Moda

1.3. Representación gráfica de variables

Junto con el cálculo de los descriptores explicados en la sección anterior, se puede obtener información adicional sobre cómo los valores de las variables medidas se distribuyen a partir de tablas y representaciones gráficas de dichos valores. Una vez más, la selección de estas herramientas depende estrictamente del tipo de variables que se analizan.

Para las variables categóricas se utilizan las siguientes:

1. **Tabla de frecuencias**. Indica el número de veces (frecuencia absoluta) que un valor dado se repite a lo largo de todos los individuos en estudio. Por ejemplo, si se considera el nivel de respuesta a un tratamiento codificado como Bajo, Medio y Alto, la Tabla 2 muestra la tabla de frecuencias con el número de pacientes de la muestra analizada que habrá tenido un nivel de respuesta Bajo, Medio o Alto:

Tabla 2 – Ejemplo de tabla de frecuencias.

Nivel de respuesta	Número de pacientes
Bajo	20
Medio	30
Alto	25

También puede expresarse como frecuencia relativa (porcentaje).

2. **Diagrama de barras**. Es la representación gráfica de la tabla de frecuencias para variables categóricas. Constituye una forma de representar gráficamente un conjunto de datos o valores mediante barras rectangulares de longitud proporcional a la frecuencia de valores medidos de una variable. En el caso del nivel de respuesta al tratamiento de la Tabla 2, un gráfico de este tipo contendrá 3 barras (cada una correspondiente a uno de los tres niveles de respuesta: Bajo, Medio o Alto) con longitud igual al número de pacientes que habrán tenido un nivel de respuesta Bajo, Medio o Alto, respectivamente (Figura 2):

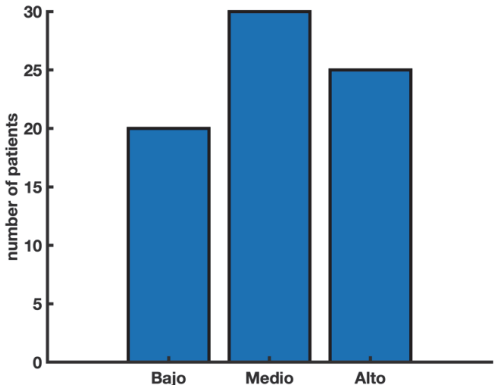


Figura 2 – Diagrama de barras para el ejemplo del nivel de respuesta al tratamiento descrito en la Tabla 2.

También puede expresarse como frecuencia relativa (porcentaje).

3. **Gráfico de tarta**. Permite representar proporciones o porcentajes. En su forma clásica se divide un círculo en segmentos que cubren un área proporcional a la frecuencia relativa de los distintos valores de una variable. En el caso del nivel de respuesta al tratamiento de la Tabla 2, las áreas de los distintos segmentos serán proporcionales a la fracción de pacientes que habrán tenido un nivel de respuesta Bajo, Medio o Alto, respectivamente (Figura 3):

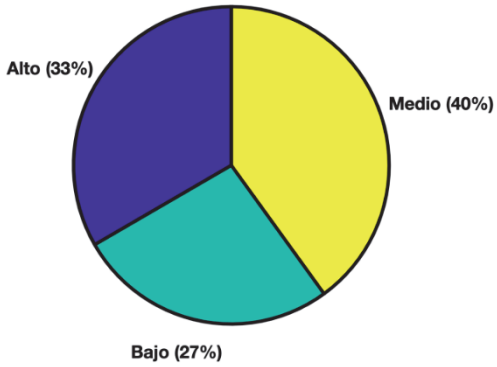


Figura 3 – Gráfico de tarta para el ejemplo del nivel de respuesta al tratamiento descrito en la Tabla 2.

Por otro lado, en el caso del análisis de variables numéricas las herramientas gráficas más usadas son:

1. **Histograma**. Es la representación gráfica de la tabla de frecuencias para variables numéricas. En este caso, como la variable numérica toma muchos valores, hay que definir unos intervalos y calcular la frecuencia (absoluta o relativa) de los valores de dicha variable que caen dentro de cada uno de los intervalos representados en forma de barras. Sirven para obtener una fotografía de la distribución de los valores de la variable a lo largo de su campo

de variabilidad. Para variables discretas, el histograma se sustituye a menudo por un diagrama de barras. El aspecto (perfil) del histograma depende mucho del número de intervalos en los cuales se reparten los datos (que es igual al número de barras representadas). Se recomienda que el número de barras esté dentro del intervalo (5, 15) y sea cercano a la raíz cuadrada del número de datos de la muestra. No se recomienda usar el histograma cuando el tamaño de muestra es pequeño (<50). Un ejemplo de este gráfico se muestra en la Figura 4:

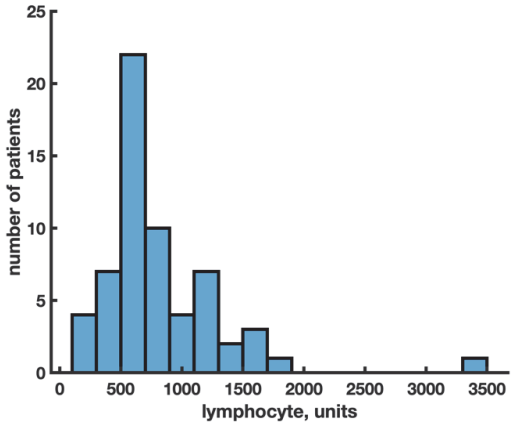


Figura 4 – Histograma del número de linfocitos medidos en una muestra de 61 pacientes de un hospital.

2. ***Box-plot* o diagrama de caja y bigotes**. Representa gráficamente los valores de una variable numérica a través de sus cuartiles. El diagrama se construye mediante: 1) un rectángulo ("caja") delimitado por el primer y el tercer cuartil de la variable en estudio y que contiene una línea que indica la mediana (segundo cuartil) de los datos, y 2) dos líneas ("bigotes") que se extienden desde la caja hasta, como máximo, 1,5 veces el intervalo intercuartílico (es decir la diferencia entre el tercer y primer cuartil). Todos los valores de la variable que se encuentran fuera de los "bigotes" se consideran potencialmente atípicos y se representan como puntos aislados. Los Box-plots pueden usarse con tamaños de muestra pequeños (<50) y proporcionan una visión general de la distribución de los datos (v.g. si la mediana no se sitúa hacia el centro del rectángulo, y los bigotes tienen longitudes muy diferentes la distribución no es simétrica) y son útiles para detectar la presencia de datos anómalos (puntos lejos de los "bigotes"). Facilitan, como se muestra en la Figura 5, la comparación descriptiva de los valores que una variable numérica toma en dos o más subgrupos de observaciones:

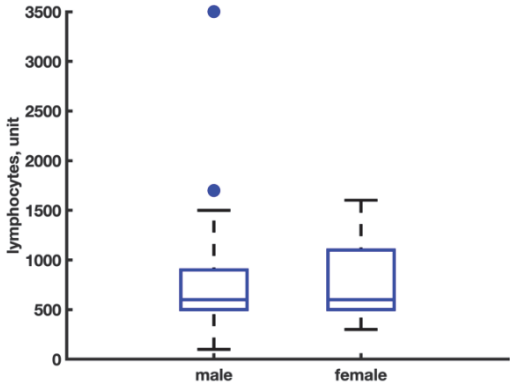


Figura 5 – Diagrama de caja y bigotes del número de linfocitos medidos en una muestra de 61 pacientes (37 hombres y 24 mujeres) de un hospital.

3. **Diagrama de puntos**. Es una de las herramientas gráficas más simples para caracterizar variables numéricas. Es muy útil para detectar la presencia de datos atípicos. Un diagrama de puntos es muy similar a un histograma, pero, en lugar de barras, se representa por cada valor que la variable en estudio asume un número de puntos igual a la frecuencia de dicho valor en los datos medidos (Figura 6). Para variables continuas, generalmente no se consideran valores individuales de la variable investigada sino intervalos de valores (como en un histograma). Su utilización se aconseja cuando se analizan muestras de tamaño pequeño (<50).

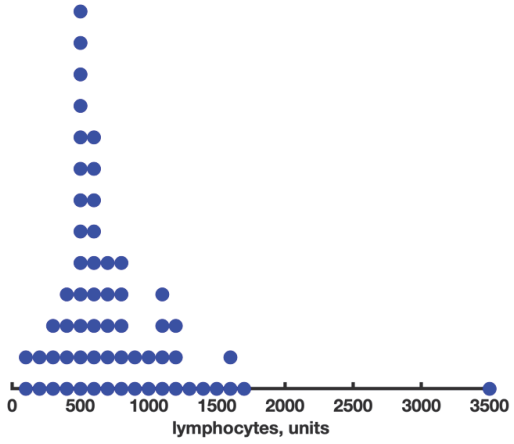


Figura 6 – Diagrama de puntos del número de linfocitos medidos en una muestra de 61 pacientes de un hospital.

A continuación, se muestra un listado de las tabulaciones y herramientas gráficas que se podrían utilizar para un estudio exploratorio de las variables clínicas de la Tabla 1:

Tabla 3 - Ejemplos de variables medidas en ensayos clínicos y de herramientas gráficas que pueden ser utilizadas para su estudio exploratorio.

Variable	Tipo	Representaciones gráficas
Edad (años)	Numérica continua	Histograma/Box-plot/Diagrama de puntos
Sexo	Categórica nominal	Diagrama de barras/Gráfico de tarta
Consumo de tabaco (Si/No)	Categórica nominal	Diagrama de barras/Gráfico de tarta
Nº comorbilidades (0, 1, 2, …)	Numérica discreta	Histograma/Box-plot/Diagrama de puntos
Histología de un tumor (A/B)	Categórica nominal	Diagrama de barras/Gráfico de tarta
Extensión de una neoplasia (I a IV)	Categórica ordinal	Diagrama de barras/Gráfico de tarta
Supervivencia libre de progresión (SLP)	Numérica continua	Histograma/Box-plot/Diagrama de puntos
SLP a los 12 meses (Si/No)	Categórica nominal	Diagrama de barras/Gráfico de tarta
Nivel de respuesta al tratamiento (Bajo, Medio ó Alto)	Categórica ordinal	Diagrama de barras/Gráfico de tarta
Efectos adversos (lista de efectos)	Categórica nominal	Diagrama de barras/Gráfico de tarta
Efecto adverso específico: neutropenia (Sí/No)	Categórica nominal	Diagrama de barras/Gráfico de tarta

Capítulo 2
Inferencia estadística

Todas las herramientas descritas en las secciones 1.2 y 1.3 son útiles para un análisis exploratorio (meramente descriptivo) de las características de la distribución de los valores de las variables en la muestra de individuos objeto de estudio. Sin embargo, la mayoría de las veces el interés de un estudio estadístico va más allá de la simple caracterización descriptiva de cada variable en la muestra analizada, y lo que se busca es conocer características de estas variables en las poblaciones de donde proceden las muestras. Esto es lo que se conoce como **inferencia estadística**. Inferir es extrapolar, generar conclusiones, yendo de lo particular (la muestra observada) a lo general (la población objeto de estudio).

En este contexto existen diversos tipos de estudios inferenciales. En unos se analiza si un factor (que define a dos o más grupos) causa diferencias en los valores poblacionales (v.g. media, varianza o proporción) de una variable (característica de interés). Es el caso, por ejemplo, de los test de comparación de medias y del análisis de la varianza o ANOVA (para la comparación de dos o más medias, respectivamente), de los test de comparación de varianzas, para el caso de variables numéricas, o de los contrastes de homogeneidad (para comparar proporciones en el caso de variables categóricas). Este tipo de estudios permitiría comparar, por ejemplo, si la media o la varianza poblacional del número de linfocitos, o el porcentaje de supervivencia a los 5 años a un cáncer, es diferente entre pacientes que han sido objeto de tratamientos médicos distintos. En otros estudios se explora si existe relación (dependencia) entre dos variables categóricas, como por ejemplo si la aparición de metástasis a los 3 años de tratamiento depende de si el paciente es o no fumador. Por último, existe un tipo de estudios inferenciales que tratan de explicar, mediante un modelo matemático,

cómo varía una cierta característica (variable respuesta, v.g. reducción del tamaño del tumor) en función de los valores de otra(s) característica(s) (variables explicativas o regresores, v.g. dosis de citostático y edad del paciente). Este es el caso de los modelos de regresión. Como se ve, la selección de un tipo de estudio inferencial u otro depende: 1) del objetivo que se quiere alcanzar y 2) del tipo de variables analizadas.

Algunos de estos estudios pueden abordarse utilizando dos tipos de técnicas estadísticas: **pruebas paramétricas** y **pruebas no paramétricas**. Las pruebas o tests paramétricos asumen que las variables en las poblaciones a comparar siguen distribuciones estadísticas específicas (en muchos casos se asume la distribución normal). Por tanto, para que las conclusiones del estudio sean fiables, el usuario debe validar que los datos siguen la distribución supuesta (en muchos casos, la normalidad). Sin embargo, muchas de estas pruebas paramétricas (como los tests de comparación de medias o el ANOVA) pueden usarse con datos que no siguen una distribución normal si se detectan y eliminan los datos anómalos, y se utilizan tamaños de muestra en cada población a comparar parecidos y no pequeños (v.g. >10). Por otra parte, las pruebas no paramétricas no asumen que los datos sigan ninguna distribución estadística específica a priori, pero esto no significa que no estén sujetas a ciertas restricciones (como por ejemplo que las varianzas de los grupos a comparar sean semejantes). Las pruebas paramétricas suelen tener más potencia estadística que las pruebas no paramétricas, por lo tanto, es más probable detectar un efecto significativo cuando realmente existe. Por su parte, las pruebas no paramétricas son generalmente más robustas que las paramétricas a la presencia de datos anómalos.

Hay autores que sugieren usar las pruebas no paramétricas cuando los tamaños de muestra de los grupos a comparar sean pequeños y no pueda validarse la distribución estadística en los grupos a comparar. Sin embargo, la probabilidad de detectar un efecto significativo en caso de que exista puede ser muy pequeña si el tamaño de la muestra es pequeño y se tiene que usar una prueba no paramétrica que es menos eficiente. Por ese motivo, una práctica recomendable para poder usar los tests paramétricos (más eficientes que los no paramétricos) es trabajar con tamaños de muestra de los grupos no pequeños (v.g. >10) y usar métodos de detección de anómalos. También podrían aplicarse transformaciones a la característica a estudiar, como, por ejemplo, la transformación logarítmica en caso de distribuciones con una marcada asimetría (sesgo) positiva.

2.1. Comparación de variables

A continuación, se presentan algunas pruebas paramétricas (y sus alternativas no paramétricas) utilizadas comúnmente en el ámbito clínico para comparar variables.

2.1.1. Comparación de variables categóricas

2.1.1.1. En una misma muestra: ***test Chi cuadrado (χ^2) de independencia****.*

Este test permite determinar si dos **variables categóricas** distintas son **independientes** (hipótesis nula) o están relacionadas (hipótesis alternativa). En el ejemplo de la Tabla 2 considérese que además del nivel de respuesta a la quimioterapia de los pacientes, también se ha observado su sexo. Se pretende estudiar si el nivel de respuesta al tratamiento es independiente del sexo. Los datos se podrían organizar en una tabla de frecuencias como la siguiente:

Nivel de respuesta	Hombres	Mujeres	Total
Bajo	11	9	20
Medio	15	15	30
Alto	12	13	25
Total	38	37	75

El test estima un estadístico de prueba, ${\chi^2}_{\text{calc}}$, como:

$$ {\chi^2}_{\text{calc}} = \sum_i \frac{(O_{i,\text{h}} - E_{i,\text{h}})^2}{E_{i,\text{h}}} + \sum_i \frac{(O_{i,\text{m}} - E_{i,\text{m}})^2}{E_{i,\text{m}}} \quad (1) $$

donde $O_{i,\text{h}}$ es el número de valores **observados** del i-ésimo nivel de respuesta (bajo, medio o alto) entre los hombres, $O_{i,\text{m}}$ es el número de valores observados del i-ésimo nivel de respuesta (bajo, medio o alto) entre las mujeres, $E_{i,\text{h}}$ corresponde al número de valores **esperados** del i-ésimo nivel de respuesta entre los hombres (es decir el producto entre el número de hombres en la muestra ($\sum_i O_{i,\text{h}}$) y el número de personas con el nivel de respuesta *i* en la muestra ($O_{i,\text{h}} + O_{i,\text{m}}$) dividido por el número total de individuos muestreados) y $E_{i,\text{m}}$ corresponde al número de valores esperados del i-ésimo nivel de respuesta entre las mujeres (es decir el producto entre el número de mujeres en la muestra ($\sum_i O_{i,\text{m}}$) y el número de personas con el nivel de respuesta *i* en la muestra ($O_{i,\text{h}} + O_{i,\text{m}}$) dividido por el número total de individuos muestreados). Si las variables son independientes, ${\chi^2}_{\text{calc}}$ se distribuye aproximadamente como una variable χ^2 con $(f-1)(c-1)$ grados de libertad: f y c corresponden, respectivamente, al número de filas y de columnas de la tabla de frecuencias (es decir, en el ejemplo descrito, al número de niveles de respuesta al tratamiento, 3, y al número de sexos, 2). Por eso, si el valor de ${\chi^2}_{\text{calc}}$es mayor que el valor crítico de la distribución χ^2 correspondiente asociado a un nivel de riesgo α dado (normalmente, 0,05), el *p*-valor será menor que el riesgo α impuesto, y se rechaza la hipótesis nula en favor de la hipótesis alternativa (es decir, se concluye que las dos variables en estudio están relacionadas). En caso contrario (*p*-valor mayor que el riesgo α), no existe evidencia estadística suficiente para rechazar la hipótesis nula de independencia entre las variables.

2.1.1.2. En dos o más muestras: ***test Chi cuadrado (χ^2) de homogeneidad***.

Si en el ejemplo anterior se hubiese registrado el nivel de respuesta (Bajo, Medio, Alto) con varios tratamientos distintos, se habría aplicado el test Chi cuadrado de homogeneidad para averiguar **si el nivel de respuesta es igual en los diferentes tratamientos** (hipótesis nula) o distinto (hipótesis alternativa). El cálculo del estadístico de prueba y su comparación con el valor crítico de la distribución Chi cuadrado es semejante al test de independencia.

En resumen, en la prueba de homogeneidad se registra una única variable categórica en dos o más poblaciones y se pretende estudiar si la distribución de la variable categórica es o no la misma en las diferentes poblaciones. Por el contrario, en la prueba de independencia se registran dos variables categóricas en una única población, y se pretende ver si pertenecer a una categoría u otra de una de las variables condiciona de alguna manera el pertenecer a una categoría u otra de la segunda variable.

2.1.2. Comparación de variables numéricas

2.1.2.1. Comprobación de la Normalidad de las variables

Como se ha comentado, algunas pruebas o tests paramétricos asumen que las variables en las poblaciones a comparar siguen una distribución Normal. Por tanto, para que las conclusiones del estudio sean fiables, el usuario debe validar si los datos pueden considerarse normales. Sobre todo, es especialmente crítica la detección de datos anómalos, pues su presencia puede invalidar completamente las conclusiones del estudio. Para realizar esta comprobación, una técnica muy sencilla es representar los valores de la variable a analizar en cada grupo usando un papel probabilístico normal (ppn) y ver si los datos caen aproximadamente en línea recta. Existen pruebas específicas de bondad de ajuste a la distribución normal como el test de Kolmogorov-Smirnov o el de Shaphiro-Wilk que también pueden usarse, y que proporcionan un *p*-valor para decidir si aceptar la hipótesis de normalidad (*p*-valor>riesgo α, normalmente 0.05) o rechazarla (*p*-valor<riesgo α). El ppn y estos dos tests de bondad de ajuste también pueden emplearse con los residuos de técnicas como el ANOVA o los modelos de regresión lineal para comprobar en un solo paso la hipótesis de normalidad de todas las poblaciones a comparar.

2.1.2.2. Análisis de la homogeneidad de la dispersión de las variables: ***F-test (test de Fisher) de comparación de varianzas***

Este test permite determinar si las varianzas de una variable **numérica** en dos poblaciones distintas son iguales (hipótesis nula) o diferentes (hipótesis alternativa). Para el contraste de hipótesis, se calcula el estadístico de prueba F_{calc}:

$$F_{\text{calc}} = \frac{s_{\text{A}}^2}{s_{\text{B}}^2} \tag{2}$$

donde s_{A}^2 y s_{B}^2 son las varianzas de la variable analizada en las dos muestras (en este caso, A y B). Si la hipótesis nula es cierta, F_{calc} se distribuye aproximadamente como una variable F de Fisher con $n_{\text{A}} - 1$ y $n_{\text{B}} - 1$ grados de libertad (n_{A} y n_{B} corresponden, respectivamente, al número de individuos en la muestra A y en la muestra B). Por eso, si el valor de F_{calc} es mayor que el valor crítico de la distribución F de Fisher correspondiente asociado a un nivel de riesgo α dado (normalmente, 0,05), el *p*-valor será menor que el riesgo α, y se rechaza la hipótesis nula en favor de la hipótesis alternativa (es decir, se concluye que la diferencia en las varianzas de la variable en estudio en las dos muestras es estadísticamente significativa). En caso contrario (*p*-valor mayor que el riesgo α), la hipótesis nula no se puede rechazar (en este caso, no se puede afirmar que existe una diferencia estadísticamente significativa entre las varianzas de la variable en las dos muestras).

Cuando se quiere comparar las varianzas en dos o más grupos, usar el **test de Levene**.

2.1.2.3. Comparación de variables numéricas en una misma muestra: ***t-test de dos muestras emparejadas, apareadas o relacionadas***

Este es el caso, por ejemplo, cuando se mide el nivel de un metabolito en un paciente antes y después de recibir un tratamiento. Los individuos de las dos muestras son los mismos, por eso se dice que están emparejados o relacionados. En este caso se calcula una nueva variable (*d*) como diferencia entre los valores de cada individuo en ambas muestras, $d=x_A-x_B$, obteniéndose su media muestral ($\bar{d}$) y su desviación típica muestral (s_d). El *t*-test estima, un estadístico de prueba, t_{calc}, que se calcula como:

$$t_{\text{calc}} = \frac{\bar{d}}{s_d\sqrt{\frac{1}{n}}} \tag{3}$$

siendo *n* el número de individuos de la muestra. El estadístico t_{calc} se compara con el valor crítico de la distribución t de Student con grados de libertad igual al tamaño de muestra menos 1 (*n*-1) para un riesgo de primera especie α (tasa de falsas alarmas) dado (normalmente, 0,05), y se calcula el *p*-valor, siguiéndose el mismo criterio que en el test de muestras independientes para aceptar o rechazar la hipótesis de igualdad de medias en los dos tratamientos.

La alternativa no paramétrica al *t*-test para datos emparejados es la **prueba de Wilcoxon.**

2.1.2.4. Comparación de variables numéricas en dos muestras independientes: ***t-test de dos muestras independientes (no emparejadas)***

Este test permite determinar si las medias de una variable **numérica** son iguales en las poblaciones de donde proceden las dos muestras analizadas (**hipótesis nula**) o diferentes (**hipótesis alternativa**). Asume que 1) los valores de esta variable siguen un modelo estadístico de distribución Normal en las dos poblaciones de donde proceden las muestras, y 2) las varianzas de las variables en estudio en las dos poblaciones de donde proceden dichas muestras son iguales. El *t*-test estima un estadístico de prueba, t_{calc}, que se calcula como:

$$t_{\text{calc}} = \frac{\bar{x}_{\text{A}} - \bar{x}_{\text{B}}}{s_{x_{\text{A}},x_{\text{B}}}\sqrt{\frac{1}{n_{\text{A}}} + \frac{1}{n_{\text{B}}}}} \tag{4}$$

donde $\bar{x}_{\text{A}}$ y $\bar{x}_{\text{B}}$ representan, respectivamente, las medias aritméticas de la variable medida en las dos muestras (en este caso, A y B), n_{A} y n_{B} corresponden al número de individuos en la muestra A y B, y $s_{x_{\text{A}},x_{\text{B}}} = \sqrt{\frac{(n_{\text{A}}-1)s_{x_{\text{A}}}^2 + (n_{\text{B}}-1)s_{x_{\text{B}}}^2}{n_{\text{A}}+n_{\text{B}}-2}}$ (con $s_{x_{\text{A}}}^2$ y $s_{x_{\text{B}}}^2$, las varianzas de la variable medida en las muestras A y B, respectivamente) representa la desviación típica ponderada de la diferencia de medias muestrales. El estadístico t_{calc} se compara con el valor crítico de la distribución t de Student con grados de libertad igual a la suma del número de individuos en las dos muestras menos 2 ($n_{\text{A}} + n_{\text{B}} - 2$) para un riesgo de primera especie α (tasa de falsas alarmas) dado (normalmente, 0,05). A partir de este contraste, es posible encontrar un *p*-valor que, si es menor de dicho riesgo α, permite rechazar la hipótesis nula en favor de la hipótesis alternativa (en este caso, se concluye que la diferencia entre las medias de la variable en las dos muestras investigadas es estadísticamente significativa). Por el contrario, si el *p*-valor es mayor que el riesgo α, la hipótesis nula no se puede rechazar (en este caso, no se puede afirmar que existe una diferencia estadísticamente significativa entre las medias de la variable en las dos muestras).

La alternativa no paramétrica al *t*-test para datos no emparejados es la **prueba U de Mann-Whitney**.

En resumen, se recurre a las pruebas no emparejadas cuando los individuos pertenecientes a los dos grupos en estudio son diferentes (por ejemplo, si dos tratamientos médicos se han aplicados a dos grupos de pacientes distintos). Por otro lado, las pruebas emparejadas se utilizan cuando el mismo grupo de individuos ha sido evaluado en dos escenarios distintos (es decir, cuando, por ejemplo, dos tratamientos médicos han sido aplicados al mismo grupo de pacientes).

2.1.2.5. Comparación de variables numéricas en más de dos muestras independientes: ***análisis de la varianza (ANOVA)***

Esta técnica permite determinar si las medias de una variable **numérica en dos o más poblaciones distintas** son iguales (hipótesis nula) o al menos una es diferente (hipótesis alternativa). Por tanto, se puede considerar como una generalización del *t*-test de dos muestras. Los supuestos sobre los que se basa el ANOVA son: 1) la **independencia** de los individuos en las muestras en estudio, 2) la **homogeneidad** (igualdad) de las varianzas de la variable investigada en las poblaciones y 3) la **normalidad** (distribución Normal) de la variable en las poblaciones estudiadas. De estos tres supuestos lo más crítico es la falta de independencia y la presencia de datos anómalos. Si los tamaños de muestra son parecidos, la falta de igualdad de varianzas y ligeras asimetrías en la distribución de la variable en las poblaciones no comprometen la validez de los resultados. El ANOVA proporciona un *p*-valor que cuantifica la verosimilitud de la hipótesis nula. Si el *p*-valor se encuentra por debajo del nivel de riesgo α impuesto (generalmente, 0,05), la hipótesis nula se considera poco creíble, aceptándose la hipótesis alternativa. Para visualizar entre qué medias muestrales las diferencias son estadísticamente significativas, se puede recurrir a herramientas gráficas como los intervalos LSD (**Least Significant Difference**) de Fisher. Un ejemplo de aplicación del ANOVA sería estudiar si el nivel medio de leucocitos en sangre es el mismo con tres tratamientos distintos. Rechazar la hipótesis nula supondría afirmar que al menos uno de los tratamientos produce un nivel medio de leucocitos en sangre diferente al resto. La probabilidad de equivocarse afirmando que hay diferencias entre esos niveles medios cuando en realidad no las hay es el riesgo de primera especie α (generalmente, 0,05).

Dos alternativas no paramétricas al ANOVA son la **prueba de Kruskal-Wallis** y la **prueba de Mood** (que llevan a cabo **una comparación de medianas** y no de medias).

La siguiente tabla muestra algunos ejemplos de aplicación en el ámbito clínico de las pruebas estadísticas paramétricas y no paramétricas descritas en esta sección:

Tabla 4 – Ejemplos de aplicación de pruebas estadísticas paramétricas y no paramétricas en ámbito clínico.

Estudio	Prueba paramétrica	Prueba no paramétrica
Comparar las medias de una variable numérica (v.g. SLP) en dos grupos de pacientes con tratamientos distintos	*t*-test para muestras independientes	Prueba U de Mann-Whitney
Comparar las medias de una variable numérica (v.g. nivel de linfocitos) en pacientes antes y después del tratamiento	*t*-test para muestras emparejadas	Prueba de Wilcoxon
Comparar las medias de una variable numérica (v.g. reducción del tamaño del tumor) en tres grupos de pacientes con tratamientos distintos	ANOVA	Prueba de Kruskall-Wallis/Prueba de Mood
Comparar las varianzas de una variable numérica (v.g. SLP) en dos grupos de pacientes con tratamientos distintos	*F*-test (test de Fisher) o prueba de Levene	-
Comparar las distribuciones de una variable categórica (v.g. nivel de respuesta al tratamiento) en dos o más grupos de pacientes con tratamientos diferentes	-	Test de homogeneidad χ^2
Determinar la posible relación entre dos variables categóricas (v.g. consumo de tabaco, y presencia de efectos adversos) medidas en la misma muestra de pacientes	–	Test de independencia χ^2

2.2. Estudio de la relación entre una variable respuesta y una o varias variables explicativas

Por otro lado, la herramienta *ad hoc* cuando se trata de identificar y cuantificar mediante un modelo matemático las relaciones existentes entre una o más variables explicativas y una variable respuesta, medidas todas sobre los mismos individuos, es la **regresión lineal**. En general, se distinguen dos tipos de técnicas de regresión que se aplican, respectivamente, cuando la variable respuesta es una variable numérica o categórica. En el primer caso, se habla de **regresión lineal simple** (si sólo hay una variable explicativa) o **regresión lineal múltiple** (si hay más de una variable explicativa). En el segundo caso (variable respuesta categórica) se habla de **regresión logística**.

2.2.1. Variable respuesta numérica: regresión lineal simple o múltiple

Se trata de un modelo estadístico usado para aproximar la relación de dependencia entre una variable respuesta, y, y una o más variables explicativas, x_k. Este modelo se expresa como:

$$y = \beta_0 + \sum_k \beta_k x_k + e \tag{5}$$

donde β_0 es el término constante del modelo y β_k representa el coeficiente por el que hay que multiplicar cada variable explicativa para obtener una estimación o **predicción** de y. Por su parte, e es el término residual que recoge la variación en y que no está explicada por las variables explicativas en estudio y que, por tanto, se asocia con el azar (aquí se incluye todo lo que influye sobre la variable respuesta y no se ha incluido como variable explicativa en el modelo). En el caso de la regresión lineal simple, la ecuación 4 se reduce a:

$$y = \beta_0 + \beta_1 x_1 + e \tag{6}$$

La estimación de los parámetros β_0 y β_k de los modelos de regresión lineal se lleva a cabo mediante el **método de mínimos cuadrados**. Una vez obtenidas, las estimaciones de los coeficientes β_0 y β_k proporcionan información sobre el comportamiento de la variable y frente a las variables explicativas. Más en concreto:

a. β_0 corresponde al valor medio de la variable respuesta y cuando todas las variables explicativas asumen valores de cero;
b. Si $\beta_k = 0$, para cualquier valor de x_k el valor medio de y no varía (es decir, no hay relación entre x_k e y);
c. Si $\beta_k > 0$, al aumentar el valor de x_k, también aumenta el valor medio de y. En este caso, β_k cuantifica el aumento en el valor medio de y al incrementar una unidad de x_k (manteniéndose constante el resto de variables explicativas);
d. Si $\beta_k < 0$, al aumentar el valor de x_k, el valor medio de y disminuye. En este caso, β_k cuantifica la disminución en el valor medio de y al incrementar una unidad de x_k (manteniéndose constante el resto de variables explicativas).

Nota: la interpretación de los coeficientes β_k comentada solo tiene sentido si las variables explicativas son independientes.

Para determinar si la relación lineal entre cada una de las variables explicativas x_k y la variable respuesta y es significativa desde un punto de vista estadístico, se puede recurrir a un contraste de hipótesis específico para cada uno de los coeficientes β_k del modelo de regresión. Dichos contrastes proporcionan un *p*-valor para cada una de las variables explicativas que, si es menor que el nivel de riesgo α impuesto (normalmente, 0,05), permite concluir que la relación lineal es estadísticamente significativa. Por el contrario, si el *p*-valor correspondiente supera el nivel de riesgo α impuesto, se

puede admitir que no hay relación entre la variable explicativa considerada y la variable respuesta y. En este caso, se puede excluir esta variable explicativa y construir un nuevo modelo de regresión sin considerarla en los cálculos.

Una forma muy sencilla de validar un modelo de regresión lineal es comprobar que sus **residuos estandarizados** (es decir las diferencias estandarizadas entre cada valor de y y su predicción $\hat{y} = \beta_0 + \sum_k \beta_k x_k$) se distribuyen normalmente. Para ello, se puede 1) utilizar pruebas estadísticas clásicas como el **test de Kolmogorov-Smirnov**, el **test de Shapiro-Wilk** o el **test de Anderson-Darling**, o 2) representar dichos residuos en **papel probabilístico normal** y ver si aproximadamente caen en una línea recta.

Si las variables explicativas son aleatorias, es decir, sus valores no están prefijados de antemano, otra medida representativa de la relación entre cada x_k y la variable respuesta y es el **coeficiente de correlación lineal**, r_k, que se calcula como:

$$r_k = \beta_k \frac{s_{x_k}}{s_y} \tag{7}$$

donde s_{x_k} y s_y representan, respectivamente, la desviación típica de la variable x_k y de la variable y. Este coeficiente r_k siempre toma valores entre -1 y 1 y resulta que:

a. Si $r_k = 0$, x_k e y no están linealmente relacionadas;
b. Si $r_k = 1$, existe una relación lineal de proporcionalidad **directa exacta** entre x_k e y;
c. Si $r_k = -1$, existe una relación lineal de proporcionalidad **inversa exacta** entre x_k e y;
d. Si $r_k > 0$, existe una relación lineal de proporcionalidad **directa** entre x_k e y (es decir, si x_k aumenta, y tiende a aumentar);
e. Si $r_k < 0$, existe una relación lineal de proporcionalidad **inversa** entre x_k e y (es decir, si x_k aumenta, y tiende a disminuir).

En el modelo de regresión lineal simple, el grado de linealidad de la relación entre x_k e y se mide mediante el **coeficiente de determinación R^2**, que corresponde al cuadrado del coeficiente de correlación r_k y mide el porcentaje de variabilidad de la variable respuesta y explicado por la variable explicativa x_k. En el caso del modelo de regresión múltiple el coeficiente de determinación mide el porcentaje de variabilidad de la variable respuesta y explicado por todas las variables explicativas x_k del modelo.

2.2.2. Variable respuesta categórica: regresión logística

Se trata de un tipo de análisis de regresión utilizado cuando la variable respuesta y es una variable categórica. Es útil, por ejemplo, para modelizar la probabilidad de que un evento ocurra en función de los valores que tomen una o más variables explicativas. La diferencia principal entre la regresión lineal y la regresión logística radica en el

hecho de que la distribución de los valores de la variable respuesta dados unos valores de las variables explicativas se asume que no es Normal (como en la regresión lineal), sino Binomial (si y puede tomar solo 2 valores) o Multinomial (si y puede tomar más de dos valores). Considérese, por ejemplo, el caso de que la variable respuesta pueda tomar solamente dos valores, 1 (el evento de interés ocurre) y 0 (el evento de interés no ocurre), y defínase p como la probabilidad de que esta variable respuesta tome el valor igual a 1, $p = P(y = 1)$, por lo que $P(y = 0) = 1 - p$. La regresión logística ajusta el siguiente modelo:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \sum_k \beta_k x_k \tag{8}$$

Los parámetros β_0 y β_k en la ecuación 7 se pueden estimar mediante diversos métodos (como la estimación por máxima verosimilitud) y su interpretación no es la misma que en los modelos de regresión lineal. Dada la relación no lineal entre las variables explicativas y p, los parámetros β_k no indican directamente el cambio en la probabilidad de que el evento ocurra.

A partir de la ecuación 8 se puede profundizar un poco más en la interpretación de los parámetros:

a. β_0 es el valor de $\ln\left(\frac{p}{1-p}\right)$ cuando todas las variables explicativas asumen valores de cero;
b. En el caso de que la variable explicativa x_k sea cuantitativa (por ejemplo, dosis de una sustancia) β_k cuantifica el cambio aditivo en $\ln\left(\frac{p}{1-p}\right)$ debido al incremento de una unidad de x_k, manteniéndose constantes el resto de variables explicativas. Esto se puede expresar matemáticamente como:

$$\frac{\left(\frac{p}{1-p}\right)_{x_k+1}}{\left(\frac{p}{1-p}\right)_{x_k}} = e^{\beta_k} \tag{9}$$

Esto significa que el término $\frac{p}{1-p}$ (llamado **odds**) y que indica el cociente entre la probabilidad de que ocurra el evento (*p*), frente a que no ocurra (1-*p*) se multiplica por e^{β_k} cuando x_k aumenta en 1 unidad. El término e^{β_k} es el **odds ratio** asociado a la k-ésima variable explicativa. Si β_k>0, el odds ratio>1, lo que indica que la variable x_k es un factor de riesgo para el evento de interés. Por ejemplo, si $\beta_k = 0{,}7$, el odds ratio es $e^{0,7} = 2$, lo que indica que el odds (riesgo asociado al evento de interés) se dobla por cada unidad que aumente la variable explicativa x_k. Por el contrario, si β_k<0, el odds ratio<1, lo que indica que la variable x_k es un factor protector para el evento

de interés. Por ejemplo, si $\beta_k = -0{,}7$, el odds ratio es $e^{-0{,}7} = 0{,}5$, lo que indica que el odds (riesgo asociado al evento de interés) disminuye a la mitad por cada unidad que aumente la variable explicativa x_k.

En el caso de que la variable explicativa x_k sea categórica, por ejemplo, si se comparasen dos tratamientos A ($x_k = 0$) y B ($x_k = 1$):

$$\frac{\left(\frac{p}{1-p}\right)_{Trat\ B}}{\left(\frac{p}{1-p}\right)_{Trat\ A}} = e^{\beta_k} \tag{10}$$

En este caso, por ejemplo, si $\beta_k = 0{,}7$, el odds ratio es $e^{0{,}7} = 2$, lo que indica que el odds (riesgo asociado al evento de interés) del tratamiento B es el doble que el odds del tratamiento A. Por el contrario, si, por ejemplo, $\beta_k = -0{,}7$, el odds ratio es $e^{-0{,}7} = 0{,}5$, lo que indica que el odds (riesgo asociado al evento de interés) del tratamiento B es la mitad que el odds del tratamiento A.

La probabilidad *p* de que ocurra el evento de interés (dados valores específicos de todas las variables explicativas y una vez estimados los parámetros del modelo de regresión logística) se puede calcular según la ecuación:

$$p = \frac{e^{(\beta_0 + \sum_k \beta_k x_k)}}{1 + e^{(\beta_0 + \sum_k \beta_k x_k)}} \tag{11}$$

También en regresión logística, como en el caso de la regresión lineal, para determinar si la relación entre cada una de las variables explicativas y $\ln\left(\frac{p}{1-p}\right)$ es significativa desde un punto de vista estadístico, se puede recurrir a un contraste de hipótesis. Esto puede hacerse calculando un intervalo de confianza (normalmente al 95%) para el odds-ratio, e interpretarlo de la siguiente forma:

- si dicho intervalo contiene al 1, entonces se acepta que $\beta_k = 0$, es decir, la variable explicativa x_k no tiene un efecto estadísticamente significativo sobre la probabilidad de que ocurra el evento;
- si dicho intervalo solo contiene valores mayores que 1, entonces se acepta que $\beta_k > 0$, es decir, la variable explicativa x_k es un factor de riesgo sobre el evento de interés;
- si dicho intervalo solo contiene valores menores que 1, entonces se acepta que $\beta_k < 0$, es decir, la variable explicativa x_k es un factor protector sobre el evento de interés.

Por último, para validar el modelo de regresión logística existen varias opciones: 1) detectar patrones de mal ajuste en representaciones gráficas de formas alternativas de los residuos (como, por ejemplo, los **residuos de Pearson estandarizados** o los **residuos de desviación estandarizados**), 2) calcular formulaciones ajustadas del

coeficiente de determinación (como el **pseudo-** r^2) o 3) utilizar pruebas estadísticas más específicas (como el **test del cociente de verosimilitud**, el **test de desviación** o la **prueba de Hosmer-Lemeshow**).

La siguiente tabla muestra algunos ejemplos de aplicación de la regresión lineal y de la regresión logística en el ámbito clínico:

Tabla 5 – Ejemplos de aplicación de la regresión lineal y de la regresión logística en ámbito clínico.

Objetivo del estudio	Técnica estadística
Cuantificar mediante un modelo matemático la relación entre la edad (variable explicativa) y la capacidad vital forzada (variable respuesta numérica)	Regresión lineal simple
Determinar si la tensión arterial (variable respuesta numérica) se puede predecir a partir del sexo, la edad y el consumo diario de tabaco (variables explicativas)	Regresión lineal múltiple
Cuantificar mediante un modelo matemático la probabilidad de tener un tipo de cáncer (variable respuesta categórica) en función de los hábitos alimentarios (variables explicativas)	Regresión logística
Determinar la influencia de parámetros clínicos como las dosis de fármacos antitumorales suministradas y/o la presencia de comorbilidades (variables explicativas) sobre la probabilidad de muerte en pacientes oncológicos (variable respuesta categórica)	Regresión logística

2.3. Análisis de supervivencia

El análisis de **supervivencia** es una rama de la estadística que permite modelizar el **tiempo hasta que ocurre un evento** como, por ejemplo, la muerte de un paciente o la aparición de metástasis. El análisis de supervivencia intenta responder a preguntas como:

1. ¿Cuántos pacientes sobrevivirán más allá de un determinado tiempo?
2. Los que sobreviven, ¿a qué ritmo morirán?
3. ¿Cómo ciertas variables (edad, sexo, tipo de tratamiento, *etc.*) aumentan o disminuyen la probabilidad de supervivencia?

El **tiempo hasta el evento** viene definido por una fecha de inicio y una fecha de cierre, y se puede medir en las unidades habituales (horas, semanas, años, *etc.*) o bien en otras métricas como el número de ciclos de tratamiento recibidos o de intervenciones quirúrgicas sufridas. Las unidades de medida dependerán en cada caso del objetivo perseguido en el estudio de supervivencia. Es importante destacar que las fechas de inicio pueden ser diferentes para cada individuo, pues los pacientes se

pueden incorporar en el estudio en momentos diferentes. En el ámbito clínico, ejemplos de medidas del tiempo hasta el evento son la Supervivencia Libre de Progresión (SLP) o tiempo desde el inicio de tratamiento hasta que la enfermedad empeora, o el tiempo hasta la aparición de metástasis o hasta la muerte.

En el análisis de supervivencia, el objetivo fundamental es describir y caracterizar la pauta de variabilidad asociada al tiempo hasta el evento medido para un cierto número de individuos. En general dicha caracterización se lleva a cabo mediante la estimación de la **función de supervivencia**, $S(t)$. La función de supervivencia representa la fracción de individuos en la que el evento en estudio (v.g. empeoramiento de la enfermedad, aparición de metástasis o muerte) no se ha producido en el tiempo *t*. $S(t)$ es una función decreciente y a partir de su perfil se puede obtener información sobre la distribución del tiempo hasta el evento. Como muestra la Figura 7, es posible, por ejemplo, determinar el porcentaje de individuos en los que el evento no se ha producido en una cierta fecha o deducir en qué momento el evento se ha producido para un cierto porcentaje de los individuos (por ejemplo, para el 50% de los individuos – **supervivencia mediana**).

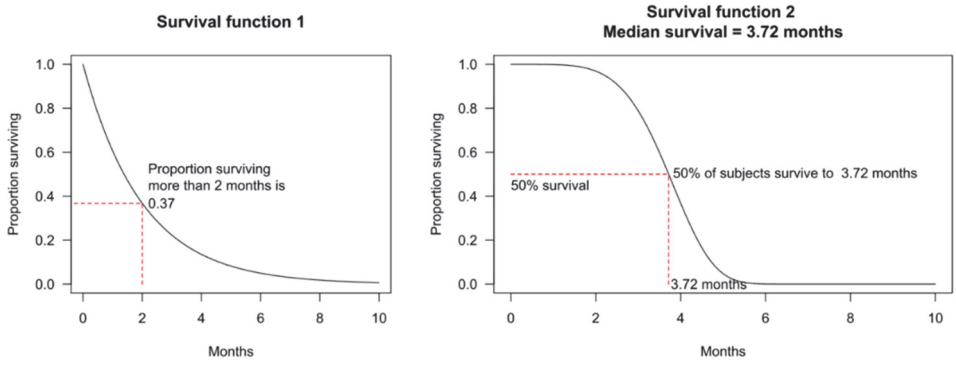


Figura 7 – Ejemplos de funciones de supervivencia.

Esta información es fundamental para contestar a muchas de las cuestiones anteriores y para alcanzar muchos de los objetivos y de los propósitos del análisis de supervivencia. Desafortunadamente, en muchos casos la función de supervivencia no se puede estimar mediante técnicas estadísticas convencionales debido a un problema que suele afectar a los datos recogidos en escenarios de este tipo: la presencia de censura.

2.3.1. Datos censurados

En el contexto del análisis de supervivencia, es muy frecuente la presencia de **censura** en los datos (por ejemplo, no se conoce el valor del tiempo hasta la muerte para todos los pacientes del estudio, pues cuando el estudio termina todavía hay pacientes vivos). Esto ocurre principalmente cuando se quiere evitar una excesiva prolongación

de los ensayos hasta que el evento en estudio se produzca para todos los individuos. Existen 4 distintos tipos de datos censurados:

1. **Censurados por la derecha**, cuando no se conoce el valor exacto del tiempo hasta el evento, sino sólo que es superior a la duración del ensayo. Los datos censurados por la derecha pueden proceder de:
   a. Censurado de tipo I, si el ensayo se detiene al cabo de un tiempo o de un número de ciclos predeterminado;
   b. Censurado de tipo II, si el ensayo se detiene cuando el evento ha ocurrido para un número predeterminado de individuos.
2. **Censurados aleatorios**, cuando lo único que se sabe de los individuos censurados es que su evento no ha ocurrido en una determinada fecha. El censurado aleatorio se produce generalmente:
   a. Cuando para un individuo el evento se produce por una causa diferente de la que se está investigando;
   b. Cuando, por ejemplo, un paciente deja de realizarse controles o deja de participar al estudio.
3. **Censurados por la izquierda**, cuando no se conoce el valor exacto del tiempo hasta el evento, sino solo que el evento ha ocurrido antes de un cierto tiempo. Este censurado se produce cuando no se lleva un seguimiento continuo durante el ensayo, sino solo se realizan controles en ciertos momentos, y se registran, por ejemplo, los pacientes que en el primer control de seguimiento ya habían fallecido.
4. **Censurado en un intervalo**, cuando lo único que se sabe de los individuos censurados es que su tiempo hasta el evento se encuentra en un cierto intervalo ($t_{i\text{-}1}$, t_i). Este censurado se produce cuando no se lleva un seguimiento continuo durante el ensayo, sino solo se realizan controles en ciertos momentos, por ejemplo, cuando durante el ensayo se examina un paciente en un tiempo t_{i-1} y se registra que está vivo, pero al volver a examinarlo en un tiempo t_i ya ha fallecido.

Como ya se ha mencionado, el censurado imposibilita la utilización de técnicas estadísticas convencionales para la estimación de la función de supervivencia. Además, la eliminación de todos los datos censurados supone a menudo un grave error en cuanto puede provocar una subestimación (si se eliminan los datos censurados por la derecha) o una sobreestimación (si se eliminan los datos censurados por la izquierda) de dicha función. Por esta razón, se recurre a métodos alternativos para el análisis de supervivencia con datos parcialmente censurados. Uno de los más utilizados es el método no paramétrico de Kaplan-Meier.

2.3.2. Estimación de la curva de supervivencia: método de Kaplan-Meier

El **método de Kaplan-Meier**, también conocido como **método del producto límite**, es un método no paramétrico que permite estimar la función de supervivencia teniendo en cuenta la presencia de datos censurados. El gráfico de la estimación de Kaplan-Meier de la función de supervivencia, $S(t)$, tiene forma de escalera decreciente con los peldaños delimitados por los tiempos en los que se produce el evento para algunos de los individuos del estudio. La Figura 8 muestra un ejemplo de curva de Kaplan-Meier.

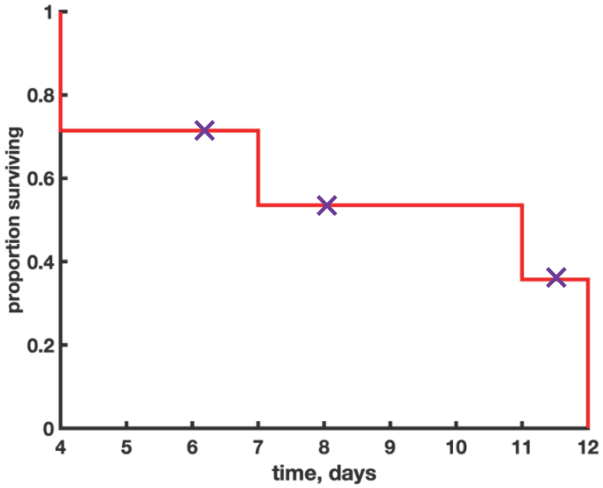


Figura 8 – Ejemplo de curva de Kaplan-Meier. Los símbolos azules identifican el momento en que se produce un censurado aleatorio.

¿Cómo se construye una curva de Kaplan-Meier? Supóngase que se han medido datos de supervivencia para 10 pacientes distintos y que a) el sujeto 1 muere el día 3 (t_1), b) los sujetos 2 y 3 mueren el día 5 (t_2), c) el sujeto 4 muere el día 7 (t_3), d) el sujeto 5 muere el día 8 (t_4), e) los sujetos 6 y 7 mueren el día 9 (t_5), f) el sujeto 8 deja de participar en el estudio a partir del día 10 (censurado aleatorio), g) el sujeto 9 muere al día 11 (t_6) y h) el sujeto 10 se encuentra vivo el día 11 (censurado por la derecha). Los datos pueden expresarse como: 3, 5, 5, 7, 8, 9, 9, 10*, 11, 11* (donde el asterisco indica que el dato está censurado) y resumirse en la siguiente tabla:

i	**1**	**2**	**3**	**4**	**5**	**6**
t_i (días)	3	5	7	8	9	11
d_i	1	2	1	1	2	1
n_i	10	9	7	6	5	2

donde n_i representa el número de pacientes en riesgo (que siguen vivos) justo antes de t_i, y d_i es el número de fallecidos en el intervalo (t_{i-1}, t_i). Está claro que cuando no hay censura, n_i corresponde al número de supervivientes justo antes de t_i, mientras

que cuando hay censura, n_i es el número de supervivientes menos el número de casos censurados justo antes de t_i.

El estimador de Kaplan-Meier de la función de supervivencia en el tiempo t se expresa como:

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i} \tag{12}$$

y se define, por cada tiempo t_i, como el ratio entre el número de pacientes que sobreviven en t_i y el número total de pacientes en riesgo justo antes de t_i. $\hat{S}(t)$ proporciona una medida de la **probabilidad de supervivencia a lo largo del tiempo**. Esta probabilidad en un tiempo t_i se calcula como el producto de las probabilidades estimadas para todos los tiempos anteriores a t_i. Entonces, volviendo al ejemplo:

$$\hat{S}(3) = \frac{10-1}{10} = \frac{9}{10} = 0{,}90 = 90\% \tag{13}$$

$$\hat{S}(5) = \left(\frac{10-1}{10}\right)\left(\frac{9-2}{9}\right) = \left(\frac{9}{10}\right)\left(\frac{7}{9}\right) = \frac{7}{10} = 0{,}70 = 70\% \tag{14}$$

$$\ldots$$

$$\hat{S}(11) = \left(\frac{9}{10}\right)\left(\frac{7}{9}\right)\left(\frac{6}{7}\right)\left(\frac{5}{6}\right)\left(\frac{3}{5}\right)\left(\frac{1}{2}\right) = \frac{3}{20} = 0{,}15 = 15\%. \tag{15}$$

La curva de Kaplan-Meier resultante sería la siguiente:

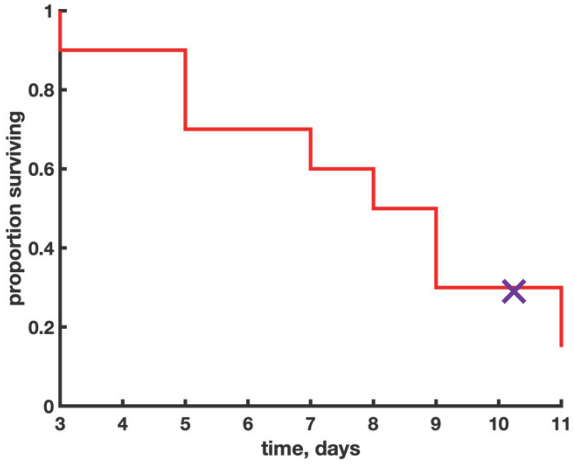


Figura 9 – Curva de Kaplan-Meier para el ejemplo resumido en la sección 2.3.2.

La función de supervivencia se centra en la "no ocurrencia" de un evento (por ejemplo, el paciente no falleció). Sin embargo, si se quisiera obtener información sobre la probabilidad de que un evento ocurra, se podría calcular la **función de riesgo** asociada a la función de supervivencia en estudio. Una función de riesgo representa la probabilidad de que a un individuo que está siendo observado en el tiempo t le suceda el evento de interés en este preciso momento, y se puede interpretar como la tasa de fallo en un determinado momento. Puede ayudar a contestar directamente preguntas como "¿cuál es la probabilidad de que fallezca un paciente operado de cáncer pulmonar a los 16 meses de postoperatorio?" o "¿en qué momento es más probable observar el pico de recidivas?". La función de riesgo en el instante t_i puede estimarse como $\hat{\lambda}(t_i) = d_i/n_i$. También es interesante representar esta tasa de fallo acumulada en el tiempo, dando lugar a lo que se conoce como la **función de riesgo acumulado**, $H(t)$. El estimador de la función de riesgo acumulado en el tiempo t se expresa como:

$$\widehat{H}(t) = \sum_{t_i \leq t} \frac{d_i}{n_i} \tag{16}$$

y corresponde a la suma de los valores de la función de riesgo calculados para todos los tiempos anteriores a t_i. Volviendo al ejemplo anterior:

$$\widehat{H}(3) = \frac{1}{10} = 0{,}10 = 10\% \tag{17}$$

$$\widehat{H}(5) = \left(\frac{1}{10}\right) + \left(\frac{2}{9}\right) = 0{,}10 + 0{,}22 = 0{,}32 = 32\% \tag{18}$$

…

$$\widehat{H}(11) = \left(\tfrac{1}{10}\right) + \left(\tfrac{2}{9}\right) + \left(\tfrac{1}{7}\right) + \left(\tfrac{1}{6}\right) + \left(\tfrac{2}{5}\right) + \left(\tfrac{1}{2}\right) = 0{,}10 + 0{,}22 + 0{,}14 + 0{,}17 + 0{,}4 + 0{,}5 = 1{,}53 = 153\% \tag{19}$$

Entonces, la curva de riesgo acumulado relativa al ejemplo descrito en este apartado sería la siguiente:

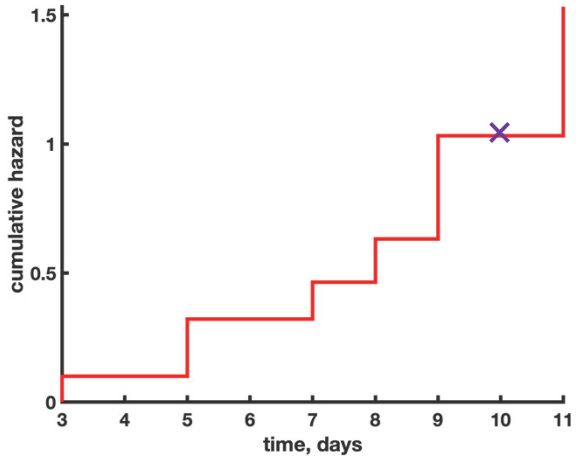


Figura 10 – Curva de riesgo acumulado para el ejemplo resumido en la sección 2.3.2.

2.3.3. Comparación de funciones de supervivencia: log-rank test (o prueba de Mantel-Cox)

En algunas ocasiones resulta de interés **comparar la supervivencia de subgrupos de individuos diferentes** y determinar, por ejemplo, si un grupo de pacientes que sigue un tratamiento A sobrevive más o menos que otro grupo de pacientes que sigue un tratamiento B. Este es el objetivo de uno de los test estadísticos más utilizados en el análisis de supervivencia en ámbitos clínicos: **el log-rank test o prueba de Mantel-Cox**.

El log-rank test compara de forma no paramétrica las estimaciones de la función de supervivencia de dos o más subgrupos de individuos en cada intervalo de tiempo. Considérese, por ejemplo, la comparación de dos grupos: pacientes que siguen el tratamiento A (grupo 1) y pacientes que siguen el tratamiento B (grupo 2), donde:

- $n_{1,i}$, el número de sujetos del grupo 1 en riesgo (es decir, aquellos que siguen vivos porque no han fallecido y no se han censurado) justo antes del tiempo t_i;
- $n_{2,i}$, el número de sujetos del grupo 2 en riesgo (es decir, aquellos que siguen vivos porque no han fallecido y no se han censurado) justo antes del tiempo t_i;
- $d_{1,i}$, el número de sujetos del grupo 1 fallecidos en el intervalo (t_{i-1}, t_i);
- $d_{2,i}$, el número de sujetos del grupo 2 fallecidos en el intervalo (t_{i-1}, t_i).

La hipótesis nula del log-rank test es que las funciones de supervivencia relativas al grupo 1 y al grupo 2 son iguales, $S_1(t) = S_2(t)$. Para el contraste, se calcula el estadístico de prueba Z:

$$Z_{calc} = \sum_j \frac{(\sum_i d_{j,i} - \sum_i E_{j,i})^2}{\sum_i E_{j,i}} \quad (20)$$

Donde *j*=1,2 son los dos tratamientos a comparar, $E_{j,i} = n_{j,i}\frac{d_i}{n_i}$, siendo $n_i = \sum_j n_{j,i}$ y $d_i = \sum_j d_{j,i}$. Bajo la hipótesis nula, Z_{calc} se distribuye como una variable χ^2 con grados de libertad igual al número de subgrupos de individuos analizados menos 1. Por eso, si el valor de Z_{calc} es mayor que el valor crítico de la distribución χ^2 correspondiente asociado a un riesgo de primera especie α dado (normalmente, 0,05), lo que es equivalente a que el *p*-valor de la prueba sea menor que el riesgo α, se rechaza la hipótesis nula, concluyendo que las dos funciones $S_1(t)$ y $S_2(t)$ estimadas resultan ser estadísticamente diferentes. Si la hipótesis nula no se puede rechazar, se concluye que las dos funciones de supervivencia estimadas no se diferencian de forma estadísticamente significativa.

2.3.4. Análisis de curvas de supervivencia: modelo de regresión de Cox

Las técnicas expuestas hasta el momento permiten inferir conclusiones a partir del análisis de una muestra de una población homogénea o comparar la supervivencia de un conjunto de muestras de distintas poblaciones. Frecuentemente, sin embargo, el interés de estudios de este tipo radica en **cuantificar el posible efecto que sobre la supervivencia tienen un conjunto de variables** explicativas sobre las cuales es posible actuar, como, por ejemplo, la dosis de un cierto fármaco. El análisis de este tipo de problemas se puede llevar a cabo mediante modelos que constituyen una **generalización de los modelos de regresión tradicionales**. Uno de estos (y quizás el más utilizado en ámbitos clínicos) es el **modelo de regresión de riesgo proporcional de Cox**. El modelo de Cox es un modelo no paramétrico que expresa la función de riesgo de que el evento en estudio ocurra, λ (tasa de fallo), en un tiempo *t* en función de K variables explicativas (x_k) como:

$$\lambda(t, x_1, \dots, x_K) = \lambda_0(t)e^{\sum_k \beta_k x_k} \quad (21)$$

donde β_k es el coeficiente asociado a la k-ésima variable explicativa, y el término $\lambda_0(t)$ define el **riesgo base**, que cuantifica el riesgo de que el evento ocurra cuando todas las variables asumen el valor 0. Generalmente, se suele asumir que $x_k = 0$ corresponde a una **situación estándar o de referencia**, y que las variables x_k miden desviaciones respecto a esta situación (v.g. respecto a la media de cada variable en la muestra). $\lambda_0(t)$ es la única parte de la ecuación 21 que depende del tiempo; la otra, $e^{\sum_k \beta_k x_k}$, sólo depende de las variables explicativas.

El modelo de Cox no busca estimar $\lambda_0(t)$, que puede ser cualquier función (de hecho no se especifica en el modelo) y se considera idéntica para todos los sujetos, sino modelar la relación entre los riesgos de que el evento ocurra entre dos individuos expuestos a niveles distintos de las variables explicativas. Para ello, el modelo parte de una hipótesis fundamental: que estos riesgos son proporcionales. Para comprender este concepto se puede estudiar el caso en el que haya sólo una variable explicativa categórica nominal asociada al tratamiento recibido por los pacientes, que vale x=0 para el tratamiento A y x=1 para el B. Entonces, la función de riesgo para ambos tratamientos se expresa como:

$$\text{Tratamiento A: } \lambda(t,0) = \lambda_0(t)e^{(\beta\times 0)} = \lambda_0(t) \quad (22)$$

$$\text{Tratamiento B: } \lambda(t,1) = \lambda_0(t)e^{(\beta\times 1)} = \lambda_0(t)\gamma \quad (23)$$

donde el factor $\gamma = e^{\beta} = \lambda(t,1)/\lambda(t,0)$ (llamado ***hazard ratio*** o ratio de riesgos) indica cómo cambia el riesgo de que el evento ocurra cuando el valor de la variable pasa de 0 (tratamiento A) a 1 (tratamiento B). Más en concreto:

1. Si $\gamma > 1$ (o sea $\beta > 0$), el cambio en la variable determina un aumento del riesgo de que el evento ocurra, por lo que se empeora la supervivencia. Por ejemplo, si $\gamma = 1{,}3$ (o sea $\beta = 0{,}26$) el riesgo de que el evento ocurra con el tratamiento B es 1,3 veces el riesgo con el tratamiento A, es decir, un 30% (1,3-1=0,3) superior al del tratamiento A. En este caso el tratamiento B tiene mayor tasa de riesgo y, por tanto, empeora la supervivencia respecto del tratamiento A;
2. Si $\gamma < 1$ (o sea $\beta < 0$), el cambio en la variable determina una reducción del riesgo de que el evento ocurra, por lo que se mejora la supervivencia. Por ejemplo, si $\gamma = 0{,}3$ (o sea $\beta = -1{,}2$) el riesgo de que el evento ocurra con el tratamiento B es 0,3 veces el del tratamiento A, es decir, es un 70% (1-0,3=0,7) inferior al del tratamiento A. En este caso el tratamiento B tiene menor tasa de riesgo y, por tanto, mejora la supervivencia respecto del tratamiento A;
3. Si $\gamma = 1$ (o sea $\beta = 0$), el cambio en la variable no determina una variación del riesgo de que el evento ocurra, por lo que no afecta a la supervivencia. En este caso el tratamiento B tiene igual tasa de riesgo y, por tanto, igual supervivencia que el tratamiento A.

Si se quisiera calcular a partir de las ecuaciones 22 y 23 los valores de supervivencia correspondientes a los dos tratamientos (variable explicativa x=0 y variable explicativa x=1) en el instante de tiempo t_i, dada la relación genérica existente entre la función de riesgo y la función de supervivencia ($S(t) = e^{-\int_0^t \lambda(t)\,dt}$), se obtendría que:

$$S(t_i,0) = S_0(t_i) = e^{-\int_0^{t_i} \lambda_0(t)e^{(\beta\times 0)}dt} = e^{-\int_0^{t_i} \lambda_0(t)dt} \quad (24)$$

$$S(t_i, 1) = S_1(t_i) = e^{-\int_0^{t_i} \lambda_0(t) e^{(\beta \times 1)} dt} = e^{-\int_0^{t_i} \lambda_0(t) \gamma dt} = e^{-\gamma \int_0^{t_i} \lambda_0(t) dt} = \\ = \left[e^{-\int_0^{t_i} \lambda_0(t) dt} \right]^{\gamma} = [S_0(t_i)]^{\gamma} \quad (25)$$

donde $S_0(t_i)$ es el valor de supervivencia estimado en el tiempo t_i cuando la variable explicativa es igual a 0. Por lo tanto, si por ejemplo $S_0(t_i) = 0{,}5$ y $\gamma = 1{,}5$ (o sea $\beta = 0{,}4$), lo que indica que el riesgo de que el evento ocurra en el tratamiento B es 1,5 veces (un 50% mayor) el riesgo con el tratamiento A, resulta que:

$$S_1(t_i) = [S_0(t_i)]^{\gamma} = 0{,}5^{1{,}5} = 0{,}35 \quad (26)$$

es decir, la supervivencia disminuye con el tratamiento B. Si, por el contrario, $\gamma = 0{,}5$ (o sea $\beta = -0{,}69$), lo que indica que el riesgo de que el evento ocurra en el tratamiento B es la mitad (un 50% menor) del riesgo con el tratamiento A, resulta que:

$$S_1(t_i) = [S_0(t_i)]^{\gamma} = 0{,}5^{0{,}5} = 0{,}71 \quad (27)$$

es decir, la supervivencia aumenta con el tratamiento B.

Por otro lado, si la variable explicativa es cuantitativa (v.g. dosis de fármaco) y asume, por ejemplo, valores de 20 y 30 en dos individuos, respectivamente, entonces:

$$\lambda(t, 20) = \lambda_0(t) e^{(\beta \times 20)} = \lambda_0(t) \gamma_{20} \quad (28)$$

$$\lambda(t, 30) = \lambda_0(t) e^{(\beta \times 30)} = \lambda_0(t) \gamma_{30} \quad (29)$$

Si, por ejemplo, se considera que el coeficiente de la variable explicativa $\beta = -0.015$, γ_{20} y γ_{30} serán:

$$\gamma_{20} = e^{-0.015 \times 20} = \lambda(t, 20)/\lambda(t, 0) = 0{,}74 \quad (30)$$

$$\gamma_{30} = e^{-0.015 \times 30} = \lambda(t, 30)/\lambda(t, 0) = 0{,}64 \quad (31)$$

lo que supone una disminución del riesgo de que el evento ocurra conforme aumenta el valor de la variable explicativa. En concreto si $x = 20$, el riesgo es el 74% del riesgo cuando $x = 0$ (disminución del 26%); y si $x = 30$, el riesgo es el 64% del riesgo cuando $x = 0$ (disminución del 36%).

De forma análoga, si $S_0(t_i) = 0{,}5$, en un dado tiempo t_i:

$$S(t_i, 20) = [S_0(t_i)]^{\gamma_{20}} = 0{,}5^{0{,}74} = 0{,}60 \quad (32)$$

$$S(t_i, 30) = [S_0(t_i)]^{\gamma_{30}} = 0{,}5^{0{,}64} = 0{,}64 \quad (33)$$

indicando que la supervivencia aumenta con el valor de la variable explicativa.

Capítulo 3
Ejemplos prácticos con Software SPSS

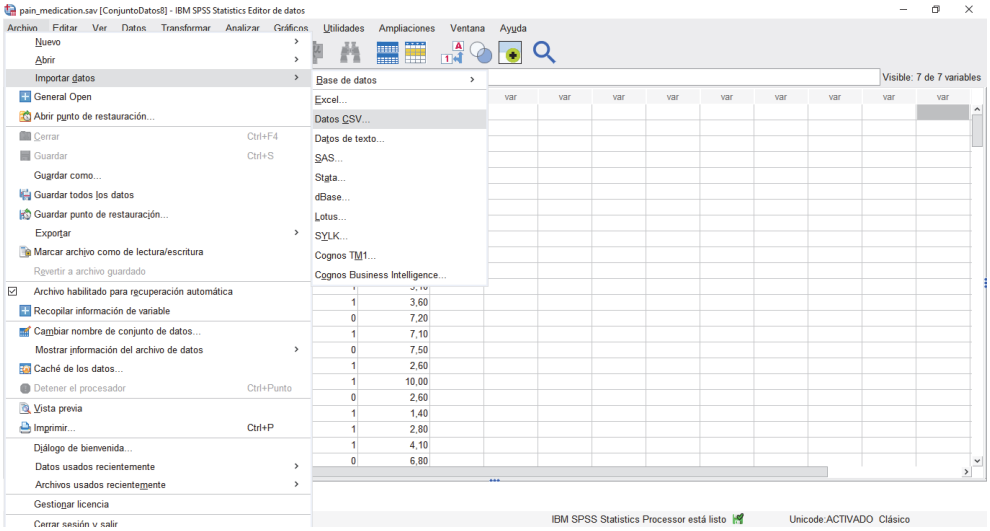
En esta sección, se detallan los pasos fundamentales para ejecutar, a través del *software* IBM SPSS Statistics (IBM, Armonk, Estados Unidos), las técnicas descritas en este documento ilustradas sobre ejemplos con datos reales.

Ejemplo #1: el caso de los antiinflamatorios

En este ejemplo se ilustra cómo calcular estadísticos descriptivos y realizar la representación gráfica de variables, así como la comparación de variables y el análisis de la relación entre variables respuesta y explicativas. Para ello, se analizará una base de datos de 200 pacientes que participaron en un ensayo clínico que tenía como objetivo la determinación de la eficacia de un nuevo medicamento antiinflamatorio para tratar el dolor artrítico crónico. Los datos están organizados en una tabla de 200 filas (pacientes) y 4 columnas (variables). La primera columna (*edad*, variable numérica continua) contiene la edad de los individuos en estudio. La segunda columna (*sexo*, variable categórica nominal) codifica el sexo de los pacientes: 0 (hombre) o 1 (mujer). La tercera (*tratamiento*, variable categórica nominal) proporciona información sobre el tratamiento aplicado: 0 (medicamento estándar) o 1 (nuevo medicamento). La cuarta columna (*tiempo*, variable numérica continua) contiene los valores del tiempo que tarda el fármaco en hacer efecto.

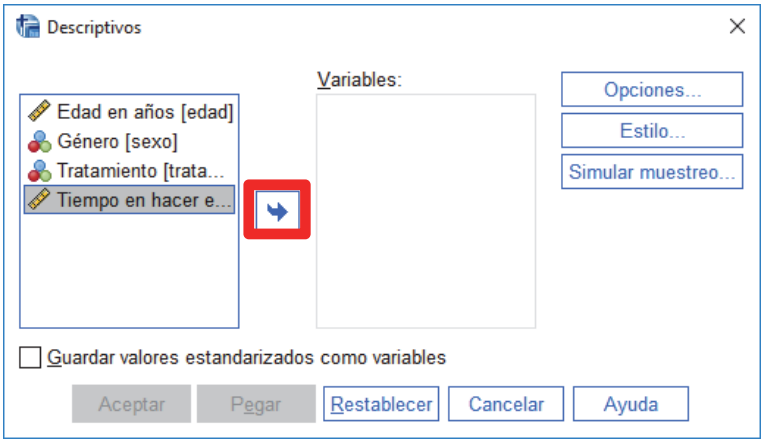
Estadísticos descriptivos y representación gráfica de variables

Para cargar los datos después de haber ejecutado el *software*, hacer click en **Archivo > Importar datos** y seleccionar el tipo específico de fichero que se quiere importar (por ejemplo, .csv):



En el cuadro de dialogo siguiente, buscar el fichero que contiene los datos y pulsar en **Abrir**. Una vez cargados los datos, será posible manipularlos, representarlos y llevar a cabo ciertas pruebas estadísticas para alcanzar los objetivos del análisis.

Por ejemplo, para obtener algunos estadísticos descriptivos de la variable *tiempo*, hacer click en **Analizar > Estadísticos descriptivos > Descriptivos**. Utilizando la flecha azul, seleccionar la variable *tiempo* en el cuadro **Variables** (para analizar al mismo tiempo más de una variable, seleccionar todas aquellas que se quieren considerar):

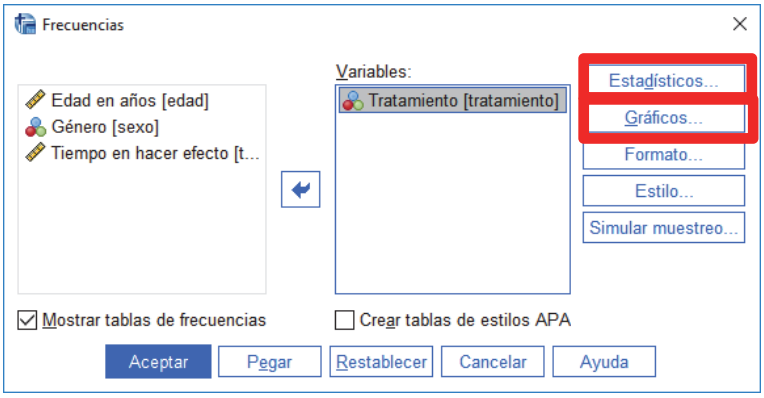


Pulsar en **Opciones** y seleccionar los estadísticos descriptivos de interés (por ejemplo, media, desviación típica, valor mínimo y valor máximo), hacer click en **Continuar** y luego en **Aceptar**. En la ventana principal del *software*, aparecerá una tabla conteniendo el número total de individuos y todos los valores de los estadísticos descriptivos seleccionados en la ventana anterior:

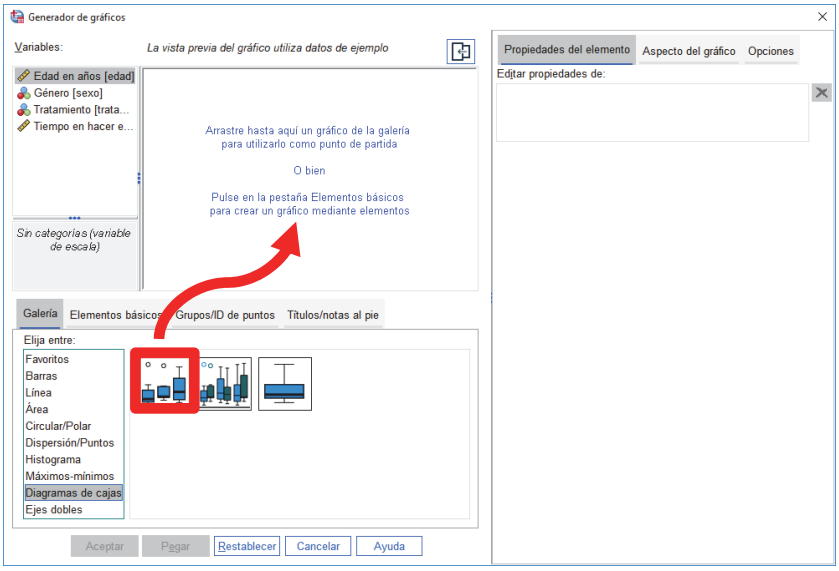
Estadísticos descriptivos

	N	Mínimo	Máximo	Media	Desviación estándar
Tiempo en hacer efecto	200	,60	11,60	4,3660	2,68660
N válido (por lista)	200				

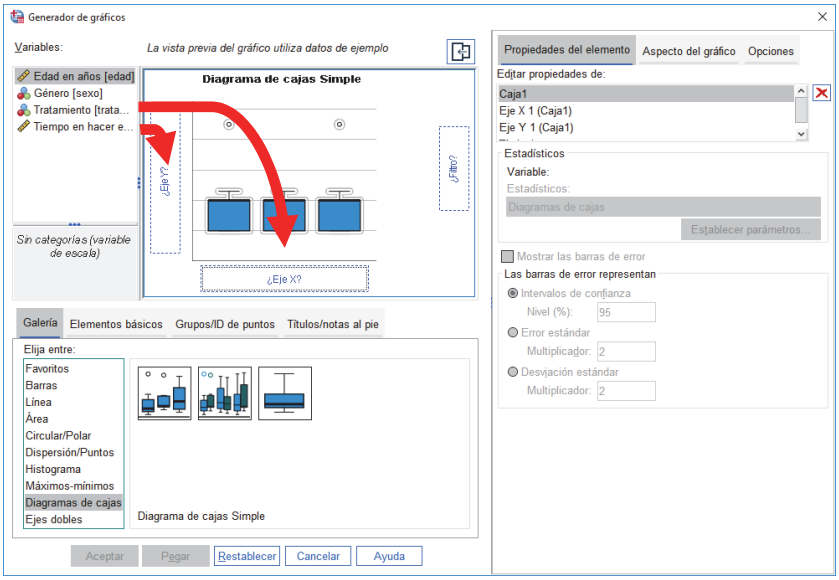
Si se quisiera caracterizar una variable categórica, se puede aplicar el mismo procedimiento pulsando en **Analizar > Estadísticos descriptivos > Frecuencias**. En la ventana que aparecerá el usuario podrá directamente seleccionar en los cuadros **Estadísticos** y **Gráficos** los estadísticos descriptivos (mediana, moda, *etc.*) y las representaciones gráficas (diagramas de barras, diagramas de tarta, tablas de frecuencias, *etc.*) que se deseen:



Para representar gráficamente los valores de la variable *tiempo* en los dos subgrupos de pacientes tratados con fármacos distintos, hacer click en **Gráficos > Generador de gráficos**. Arrastrar el icono asociado al diagrama de cajas y bigotes múltiple desde la galería de gráficos hasta el cuadro de dialogo vacío en la parte alta de la ventana que aparecerá:



Desde el cuadro **Variables**, arrastrar ahora hasta el cuadro **¿Eje X?** la variable *tratamiento* y hasta el cuadro **¿Eje Y?** la variable *tiempo*.



Pulsar en **Aceptar**. En la ventana principal del *software*, se generará un gráfico de caja y bigotes como el representado en la Figura 11, que permite comparar la distribución de los valores del tiempo que tardan los dos fármacos investigados en hacer efecto:

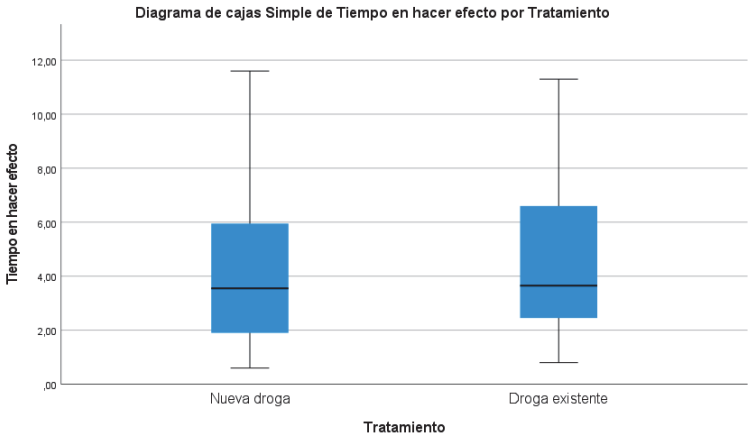
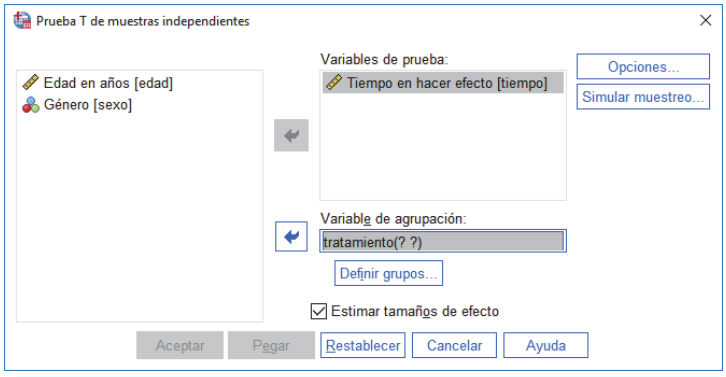


Figura 11 – Diagrama caja-bigotes múltiple del tiempo en hacer efecto de cada uno de los fármacos.

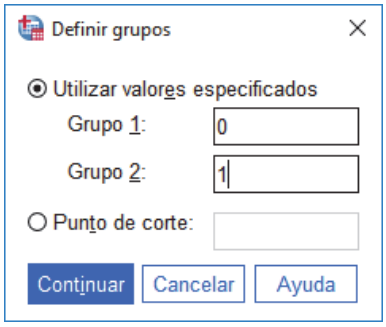
En la Figura 11 se observa que ambas distribuciones muestran asimetría positiva (lo que cuestiona la suposición de normalidad). A pesar de ello, dado que no aparecen datos anómalos y que los tamaños de muestra son grandes, como se ha comentado en el capítulo 2, es recomendable usar pruebas paramétricas, dada su mayor potencia estadística respecto de las alternativas no paramétricas.

Comparación de variables

Para determinar si hay diferencias estadísticamente significativas entre los tiempos medios en las dos muestras de pacientes, se puede recurrir a un *t*-test. Dado que estamos comparando dos grupos de pacientes (cada uno con un tipo de fármaco) se trata de muestras independientes, por lo que usaremos el *t*-test para muestras independientes. Para ello, hacer click en **Analizar > Comparar medias > Prueba T de muestras independientes**. Mediante las flechas azules, seleccionar la variable *tiempo* en el cuadro **Variables de prueba** y la variable *tratamiento* en el cuadro **Variable de agrupación**, como se muestra a continuación:



Pulsar ahora en **Definir grupos** e indicar que los individuos pertenecientes al primer subgrupo de pacientes toman valores de la variable tratamiento iguales a cero, mientras que los individuos pertenecientes al segundo subgrupo de pacientes toman valores de la variable tratamiento iguales a uno:



Pulsar ahora en **Continuar** y luego en **Aceptar**. En la ventana principal del *software*, aparecerá una primera tabla con algunos estadísticos descriptivos de las dos muestras de valores y otra que resume los resultados de la prueba. En este caso, SPSS ejecuta también una prueba de igualdad de varianzas de los tiempos con los dos medicamentos. Los *p*-valores (columna "Significación") asociados a las dos pruebas estadísticas (comparación de medias y de varianzas) para la variable *tiempo* agrupada según el tratamiento recibido por los pacientes son ambos mayores que 0,05. Por esta razón, no se puede rechazar ninguna de las hipótesis nulas, por lo que se concluye que no existen diferencias estadísticamente significativas entre las medias y varianzas de los tiempos que tardan en hacer efecto los dos fármacos:

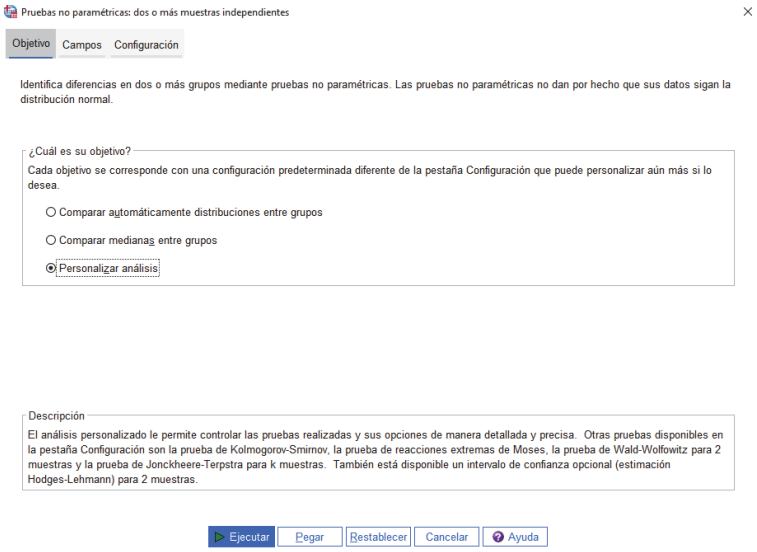
Estadísticas de grupo

	Tratamiento	N	Media	Desviación estándar	Media de error estándar
Tiempo en hacer efecto	Nueva droga	104	4,1490	2,62135	,25704
	Droga existente	96	4,6010	2,74991	,28066

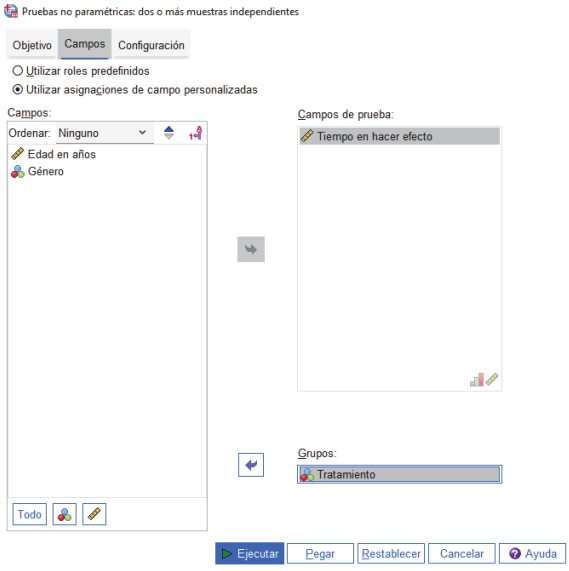
Prueba de muestras independientes

		Prueba de Levene de igualdad de varianzas		prueba t para la igualdad de medias						
									95% de intervalo de confianza de la diferencia	
		F	Sig.	t	gl	Significación	Diferencia de medias	Diferencia de error estándar	Inferior	Superior
Tiempo en hacer efecto	Se asumen varianzas iguales	1,330	,250	-1,190	198	,235	-,45200	,37985	-1,20108	,29707
	No se asumen varianzas iguales			-1,188	194,799	,236	-,45200	,38058	-1,20259	,29859

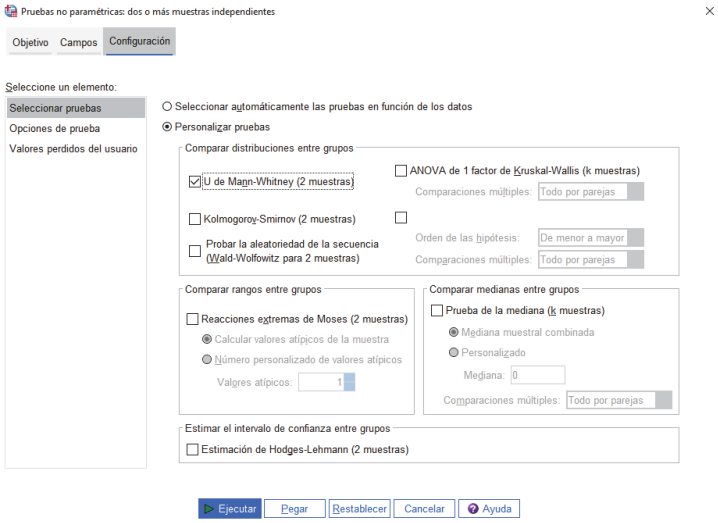
Si en lugar de un *t*-test se quisiera ejecutar una prueba no paramétrica (por ejemplo, la prueba U de Mann-Whitney), hacer click en **Analizar > Pruebas no paramétricas > Muestras independientes**. En la ventana que aparecerá, seleccionar **Personalizar análisis**.



En el cuadro **Campos**, seleccionar la opción **Utilizar asignaciones de campo personalizadas**. Mediante las flechas azules, añadir la variable *tiempo* al listado **Campos de prueba** y la variable *tratamiento* al listado **Grupos**.



En el cuadro **Configuración**, seleccionar las opciones **Personalizar pruebas** y **U de Mann-Whitney** (en este cuadro es posible seleccionar todas las pruebas estadísticas no paramétricas mencionadas en el capítulo 2). Pulsar **Ejecutar**.



Como muestra la siguiente figura, el *p*-valor que resulta de la prueba U de Mann-Whitney (0,295) es ligeramente más alto que el *p*-valor estimado a partir del *t*-test anterior. Los resultados de las dos pruebas son coherentes. Sin embargo, como ya se ha comentado antes, dado que no hay datos anómalos y los tamaños de muestra son grandes, en este caso es mejor usar la prueba paramétrica dada su mayor potencia estadística para detectar diferencias reales.

Resumen de contrastes de hipótesis

	Hipótesis nula	Prueba	Sig.[a,b]	Decisión
1	La distribución de Tiempo en hacer efecto es la misma entre categorías de Tratamiento.	Prueba U de Mann-Whitney para muestras independientes	,295	Conserve la hipótesis nula.

a. El nivel de significación es de ,050.

b. Se muestra la significancia asintótica.

Prueba U de Mann-Whitney para muestras independientes

Tiempo en hacer efecto entre Tratamiento

Resumen de prueba U de Mann-Whitney de muestras independientes

N total	200
U de Mann-Whitney	5420,500
W de Wilcoxon	10076,500
Estadístico de prueba	5420,500
Error estándar	408,878
Estadístico de prueba estandarizado	1,048
Sig. asintótica (prueba bilateral)	,295

Otra forma equivalente de comparar las medias de ambos grupos es mediante el ANOVA. Para ello, hacer click en **Analizar > Comparar medias > ANOVA de un factor**. Mediante las flechas azules, seleccionar la variable *tiempo* en el cuadro **Lista de dependientes** y la variable *tratamiento* en el cuadro **Factor.** Pulsar **Aceptar**. Como muestra la siguiente figura, el *p*-valor que resulta del ANOVA (0,235) corresponde en este caso al *p*-valor estimado a partir del *t*-test anterior (asumiendo varianzas iguales):

ANOVA

Tiempo en hacer efecto

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Entre grupos	10,199	1	10,199	1,416	,235
Dentro de grupos	1426,150	198	7,203		
Total	1436,349	199			

Análisis de la relación entre variables respuesta y explicativas

Supóngase ahora que se quiere analizar la posible relación entre dos variables numéricas como, por ejemplo, la variable *tiempo* que tarda el fármaco en hacer efecto y la variable *edad* del paciente. Esto se podría hacer de forma descriptiva mediante un diagrama de dispersión. Para ello, hacer click en **Gráficos > Generador de gráficos**. Como en el caso anterior, arrastrar el icono asociado al diagrama de **Dispersión/Puntos** desde la galería de gráficos hasta el cuadro de dialogo vacío en la parte alta de la ventana que aparecerá. Desde el cuadro **Variables**, arrastrar ahora hasta el cuadro **¿Eje X?** la variable *edad* y hasta el cuadro **¿Eje Y?** la variable *tiempo*. Después, pulsar en **Aceptar**. En la ventana principal del *software* se generará un gráfico de dispersión a través del cual será posible evaluar si ambas variables están relacionadas.

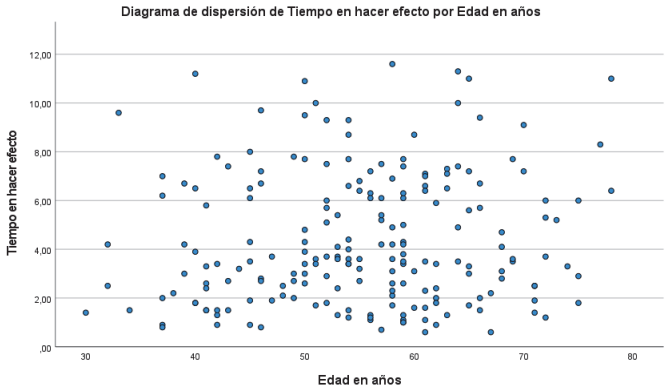
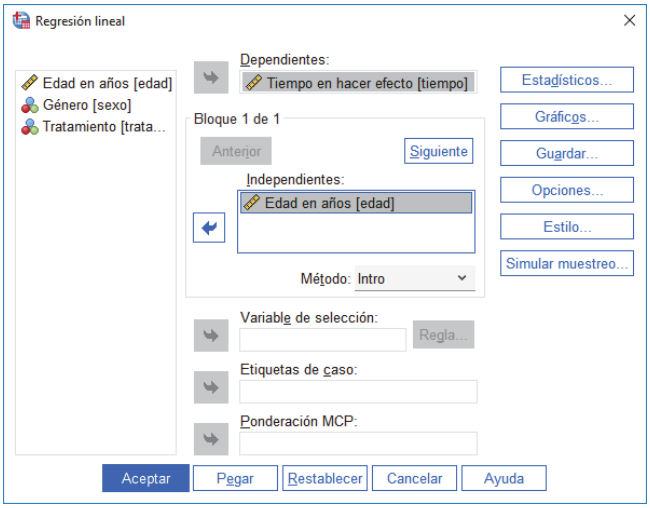


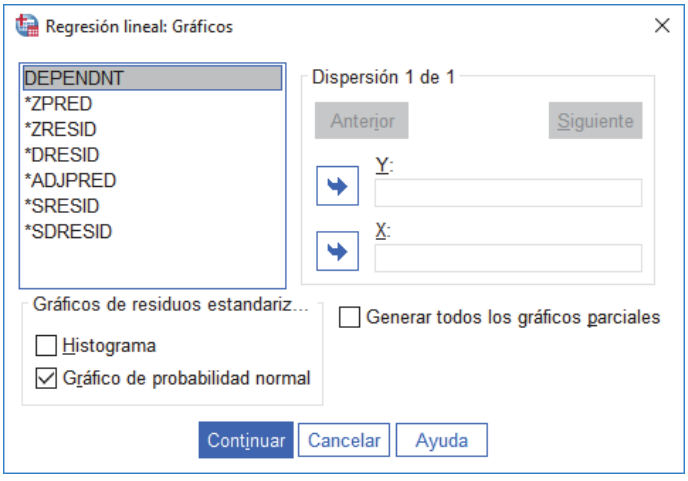
Figura 12 – Diagrama de dispersión entre la variable tiempo en hacer efecto y la variable edad

La Figura 12 muestra que, independientemente de la edad, el tiempo que tardan en hacer efecto los fármacos es similar, lo que sugiere que ambas variables no están relacionadas.

Si se quisiera modelizar matemáticamente la relación entre ambas variables, se podría construir un modelo de regresión entre ambas variables: *edad* y *tiempo*. Para ello, hacer click en **Analizar > Regresión > Lineales**. Mediante las flechas azules, seleccionar la variable *tiempo* en el cuadro **Dependientes** y la variable *edad* en el cuadro **Independientes**:



Pulsar en **Gráficos** y seleccionar la opción **Gráfico de probabilidad normal** que permitirá comprobar si los residuos del modelo de regresión construido se distribuyen normalmente:



Hacer click en **Continuar** y luego en **Aceptar**. En la ventana principal del *software* aparecerán:

1. Una tabla conteniendo, entre otros estadísticos, los valores del coeficiente de correlación, r, y del coeficiente de determinación, R^2:

Resumen del modelo[b]

Modelo	R	R cuadrado	R cuadrado ajustado	Error estándar de la estimación
1	,105[a]	,011	,006	2,67863

a. Predictores: (Constante), Edad en años

b. Variable dependiente: Tiempo en hacer efecto

El R^2 indica que el 1.1% de la variabilidad del tiempo está explicado por la edad. Sin embargo, como se discute a continuación, dado que el efecto de la edad no es estadísticamente significativo, este porcentaje tampoco lo es.

2. Una tabla conteniendo las estimaciones de los coeficientes del modelo de regresión y sus *p*-valores asociados. Como se muestra a continuación, dado que el *p*-valor asociado al parámetro $\beta_1$es 0,141 (>0,05), se concluye con un riesgo α=0,05 que no existe una relación lineal estadísticamente significativa entre la edad de los pacientes y el tiempo que tarda el antiinflamatorio en hacer efecto:

Coeficientes[a]

Modelo		Coeficientes no estandarizados		Coeficientes estandarizados		
		B	Desv. Error	Beta	t	Sig.
1	(Constante)	2,875	1,026		2,802	,006
	Edad en años	,027	,018	,105	1,479	,141

a. Variable dependiente: Tiempo en hacer efecto

Este resultado confirma la sospecha de ausencia de relación del diagrama de dispersión de la Figura 12.

3. La representación gráfica de los residuos estandarizados (diferencia entre el valor observado y el predicho) del modelo de regresión en papel probabilístico normal:

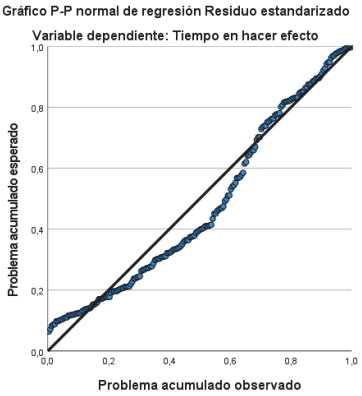


Figura 13 – Papel probabilístico normal de los residuos del modelo de regresión lineal

Su interpretación es muy sencilla: si los puntos representados se alinean aproximadamente sobre una recta (línea negra), se puede asumir que estos residuos se distribuyen normalmente. En este caso se observa una desviación importante respecto al comportamiento ideal, por lo que habría que investigar el comportamiento de los residuos.

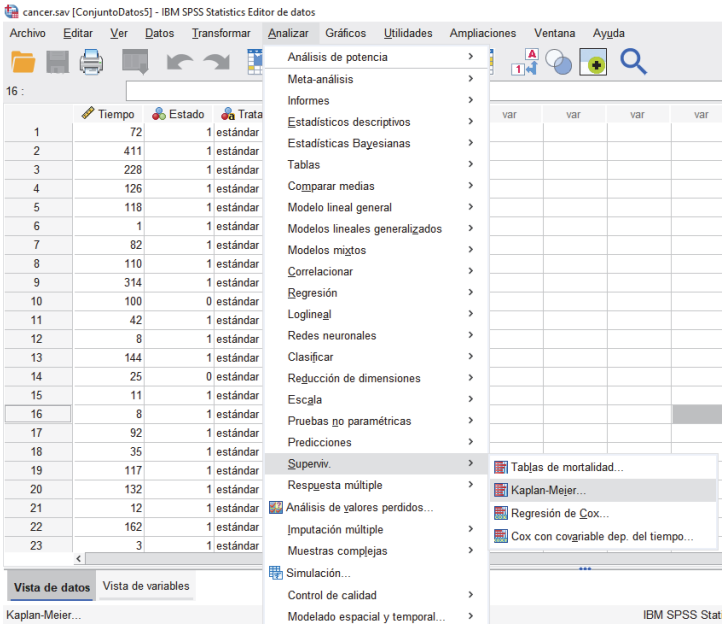
Nota: esta herramienta gráfica se puede utilizar también para averiguar si una variable numérica se distribuye normalmente. Para ello, hacer click en **Analizar > Estadísticos descriptivos > Gráficos P-P** y seleccionar la variable o las variables de interés (mediante la flecha azul) en el cuadro **Variables**. Pulsar luego en **Aceptar**. El gráfico o los gráficos P-P correspondientes aparecerán en la ventana principal del *software*.

En el caso en que la variable respuesta a analizar fuera categórica, siguiendo un procedimiento análogo al de la regresión lineal, se puede construir un modelo de regresión logística a través de los menús **Analizar > Regresión > Logística binaria** (si la variable respuesta sólo toma dos valores posibles) y **Analizar > Regresión > Logística multinomial** (si la variable respuesta puede tomar más de dos valores). Sin embargo, en este caso, como forma de validación, no se recurre al gráfico P-P de los residuos estandarizados, sino a las herramientas alternativas descritas en el capítulo 2. Se accede a algunas de estas herramientas mediante el cuadro **Opciones** de los menús mencionados antes. Los resultados finales contendrán también las estimaciones de los odds-ratio (y sus intervalos de confianza) asociados a cada una de las variables explicativas analizadas

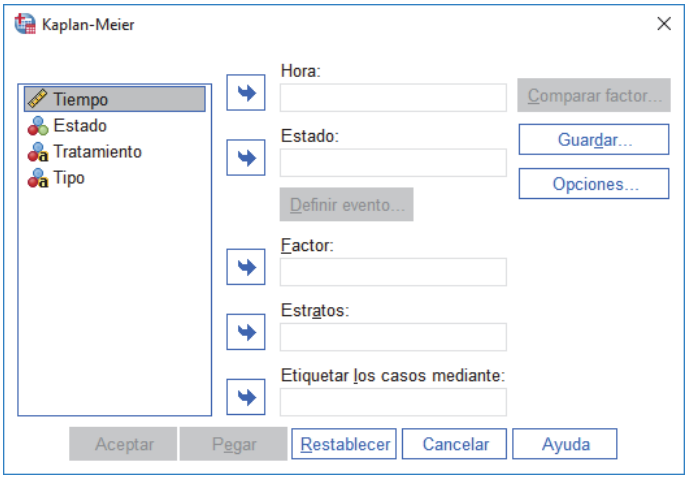
Ejemplo #2: comparación de tiempos de supervivencia de pacientes oncológicos

En este ejemplo se detallan todos los pasos necesarios para construir y representar curvas de supervivencia mediante el método de Kaplan-Meier y comparar curvas de supervivencia mediante la prueba de Mantel-Cox (log-rank test). Para ello, se analizarán los tiempos de supervivencia de 62 pacientes afectados de cáncer de pulmón. Los pacientes se pueden agrupar de dos maneras distintas: 1) los que presentan un carcinoma escamoso frente a los que presentan un carcinoma adenoescamoso, y 2) los que han recibido un tratamiento estándar frente a los que han recibido un nuevo tratamiento. Los datos están organizados en una tabla de 62 filas (pacientes) y 4 columnas (variables). La primera columna (*tiempo*, variable numérica continua) contiene los tiempos de supervivencia (en días) de los pacientes en estudio. La segunda columna (*estado*, variable categórica nominal) toma el valor cero si el dato es censurado y el valor uno si ha ocurrido el evento (muerte). La tercera (*tratamiento*, variable categórica nominal) y la cuarta columna (*tipo*, variable categórica nominal) proporcionan información sobre el tratamiento y el tipo de cáncer, respectivamente. En concreto, en este ejemplo, se estudiará si existen diferencias en la supervivencia entre los dos tipos de carcinoma.

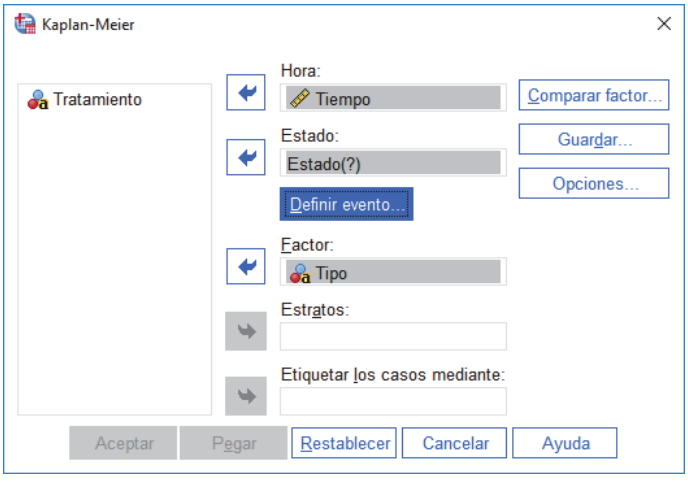
Una vez importado el conjunto de datos como se indica en la sección anterior, se puede acceder a la opción para construir curvas de supervivencia de Kaplan-Meier desde el menú **Analizar > Superviv. > Kaplan-Meier**:



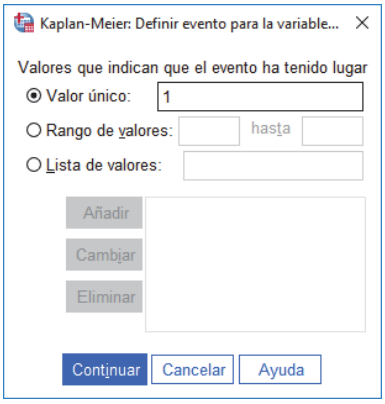
Aparecerá un cuadro de diálogo de este tipo:



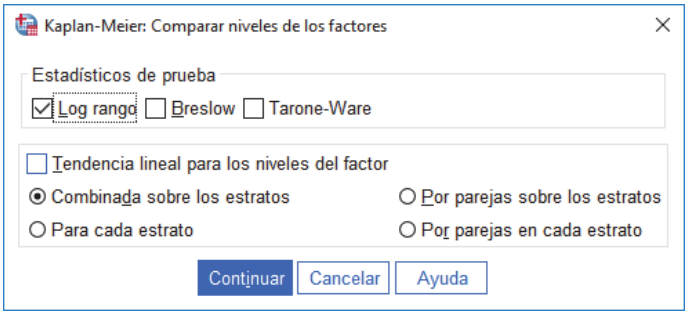
Utilizando las flechas azules, seleccionar la variable *tiempo* en el cuadro **Hora**, la variable *estado* en el cuadro **Estado** y la variable *tipo* en el cuadro **Factor**, como se muestra a continuación:



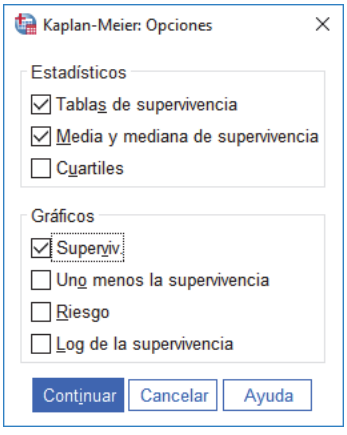
Hacer click ahora en **Definir evento** y especificar que el valor utilizado para indicar si el evento (muerte) ha tenido lugar para los diferentes pacientes analizados es 1:



Pulsar en **Continuar**, luego en **Comparar factor** y seleccionar la opción **Log rango** para que el *software* ejecute la prueba del log-rank test entre las curvas de supervivencia obtenidas para los dos tipos de carcinoma (escamoso y adenoescamoso). Dejar el resto de las opciones por defecto y hacer click en **Continuar**:



Pulsar en **Opciones** y seleccionar las opciones **Tablas de supervivencia**, **Media y mediana de supervivencia** y **Superviv.**:



Hacer click en **Continuar** y luego en **Aceptar**. En la ventana principal del *software* se generarán todos los resultados del análisis. Más en concreto:

1. Una tabla que resume el número total de pacientes en estudio, el número de eventos ocurridos durante el ensayo, y el número y el porcentaje de observaciones censuradas para los dos subgrupos de individuos y para el conjunto de datos global:

Resumen de procesamiento de casos

Tipo	N total	N de eventos	Censurado	
			N	Porcentaje
adeno	27	26	1	3,7%
esca	35	31	4	11,4%
Global	62	57	5	8,1%

2. Las tablas de supervivencia para los grupos de pacientes que contienen 1) el tiempo de supervivencia de cada uno de los individuos analizados, 2) su estado (censurado o no), 3) una estimación de su respectivo valor de supervivencia y del error asociado, y 4) el número de eventos acumulados y el número de casos en riesgo a lo largo del tiempo:

Tabla de supervivencia

Tipo		Hora	Estado	Proporción acumulada que sobrevive en el tiempo		N de eventos acumulados	N de casos restantes
				Estimación	Desv. Error		
adeno	1	3,000	1	,963	,036	1	26
	2	7,000	1	,926	,050	2	25
	3	8,000	1	.	.	3	24
	4	8,000	1	,852	,068	4	23
	5	12,000	1	,815	,075	5	22
	6	18,000	1	,778	,080	6	21
	7	19,000	1	,741	,084	7	20
	8	24,000	1	,704	,088	8	19
	9	31,000	1	,667	,091	9	18
	10	35,000	1	,630	,093	10	17
	11	36,000	1	,593	,095	11	16
	12	45,000	1	,556	,096	12	15
	13	48,000	1	,519	,096	13	14
	14	51,000	1	,481	,096	14	13
	15	52,000	1	,444	,096	15	12
	16	73,000	1	,407	,095	16	11
	17	80,000	1	,370	,093	17	10
	18	83,000	0	.	.	17	9
	19	84,000	1	,329	,091	18	8
	20	90,000	1	,288	,089	19	7
	21	92,000	1	,247	,085	20	6
	22	95,000	1	,206	,080	21	5
	23	117,000	1	,165	,074	22	4

3. Una estimación de las medias y medianas del tiempo de supervivencia, de sus errores y de sus intervalos de confianza asociados para los dos subgrupos de pacientes y para el conjunto de datos global:

Medias y medianas para el tiempo de supervivencia

	Media[a]				Mediana			
			Intervalo de confianza de 95 %				Intervalo de confianza de 95 %	
Tipo	Estimación	Desv. Error	Límite inferior	Límite superior	Estimación	Desv. Error	Límite inferior	Límite superior
adeno	65,556	10,127	45,707	85,404	51,000	6,058	39,126	62,874
esca	229,968	48,510	134,889	325,047	118,000	22,054	74,774	161,226
Global	157,861	29,479	100,081	215,641	84,000	12,338	59,817	108,183

a. La estimación está limitada al tiempo de supervivencia más largo, si está censurado.

4. Los resultados de la prueba del log-rank test, que en este caso indican una diferencia estadísticamente significativa entre las curvas de supervivencia de los dos tipos de carcinoma (*p*-valor < 0,05):

Comparaciones globales

	Chi-cuadrado	gl	Sig.
Log Rank (Mantel-Cox)	11,934	1	<,001

Prueba de igualdad de distribuciones de supervivencia para los distintos niveles de Tipo.

5. Una representación de las curvas de supervivencia estimadas mediante el método de Kaplan-Meier para los dos subgrupos de pacientes. A partir de esta representación se puede observar claramente que los pacientes que sufren un carcinoma adenoescamoso sobreviven menos que los pacientes que sufren un carcinoma escamoso:

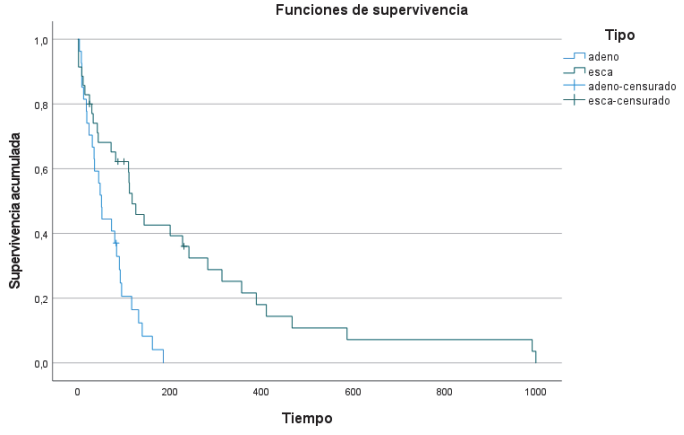
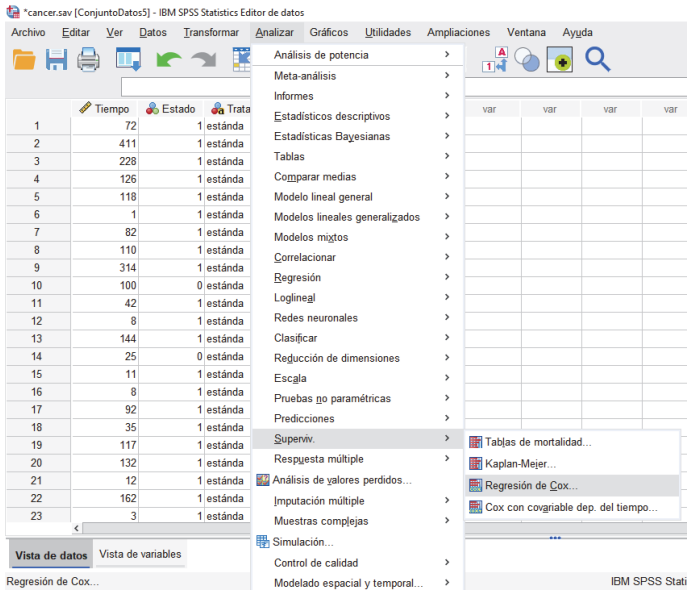


Figura 14 – Comparación de curvas de supervivencia en función del tipo de cáncer estimadas por Kaplan-Meier.

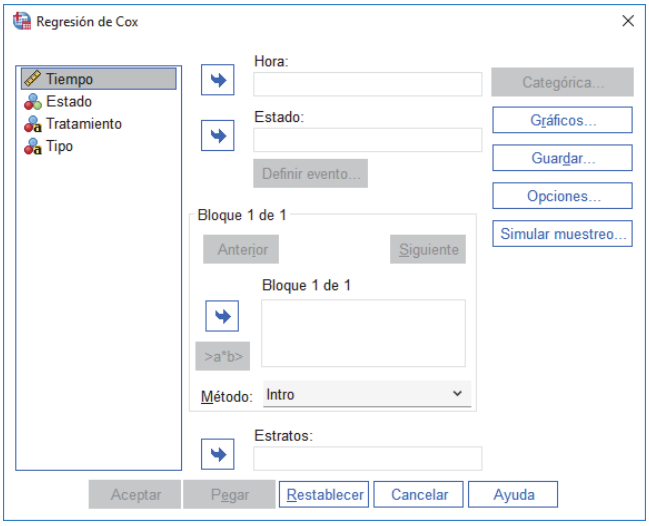
De hecho, la supervivencia mediana (que corresponde al momento en el cual fallece el 50% de los individuos y, por lo tanto, la supervivencia es igual a 0,5) de los pacientes con carcinoma adenoescamoso es alrededor de 100 días, mientras que la de los pacientes con carcinoma escamoso es alrededor de 150 días. Además, se puede notar cómo ninguno de los pacientes afectados por un carcinoma adenoescamoso sobrevive más de 200 días mientras que a los 400 días aproximadamente el 20% de los individuos que sufren un carcinoma escamoso sigue vivo (en riesgo).

Ejemplo #3: relación entre el tipo de cáncer y el tiempo de supervivencia de pacientes oncológicos

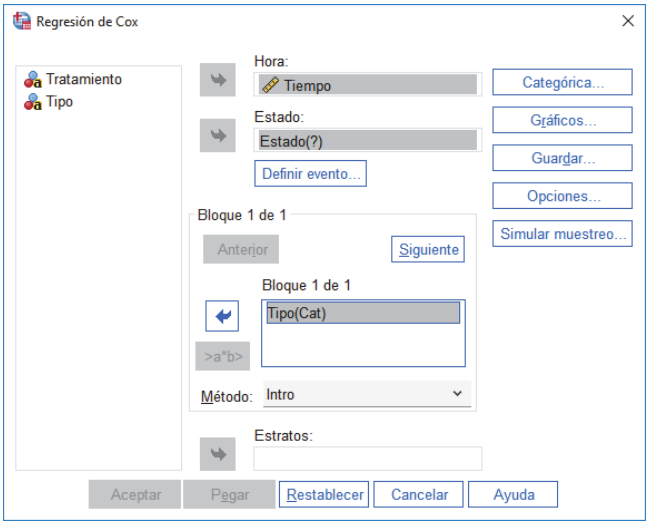
En este ejemplo se ilustra cómo construir un **modelo de regresión de Cox** para modelar el posible efecto del tipo de cáncer sobre el tiempo de supervivencia medido para los pacientes del caso de estudio descrito en el ejemplo #2. Una vez importado el conjunto de datos como se indica en la sección del ejemplo #1, se puede acceder a la opción para construir este modelo desde el menú **Analizar > Superviv. > Regresión de Cox**:



Aparecerá un cuadro de diálogo de este tipo:

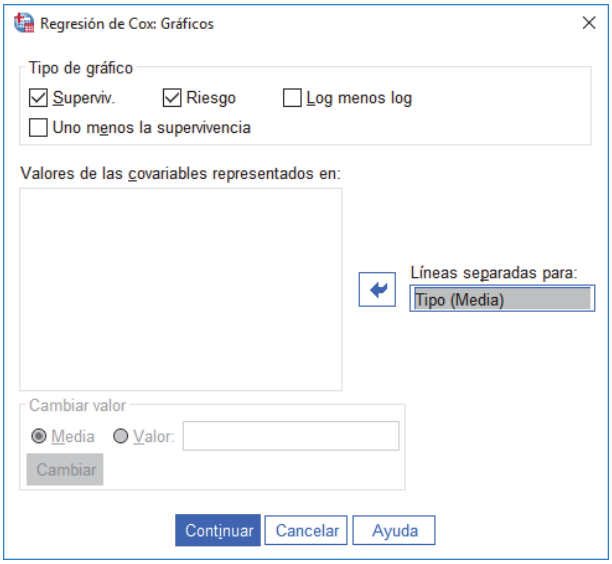


Utilizando las flechas azules, seleccionar la variable *tiempo* en el cuadro **Hora**, la variable *estado* en el cuadro **Estado** y la variable *tipo* en el cuadro **Bloque 1 de 1**, como se muestra a continuación:

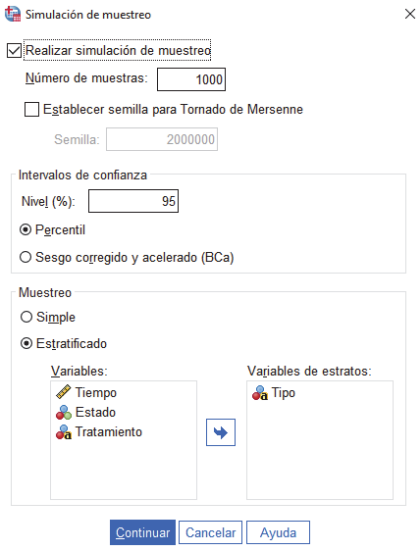


Como ya se hizo en el caso de las curvas de Kaplan-Meier, hacer click ahora en **Definir evento** y especificar que el valor utilizado para indicar si el evento (muerte) ha tenido lugar para los diferentes pacientes analizados es 1. Pulsar en **Continuar**,

luego en **Gráficos** y seleccionar las opciones **Superviv.** y **Riesgo** para representar las funciones de supervivencia y de riesgo estimadas con los datos analizados. Mediante la flecha azul, transferir el valor **Tipo (Media)** desde el cuadro **Valores de las covariables representados en** al cuadro **Líneas separadas para**. Dejar el resto de las opciones por defecto y hacer click en **Continuar**:



Pulsar en **Simular muestreo**. Seleccionar las opciones **Realizar simulación de muestreo** y **Estratificado**. Mediante la flecha azul, añadir la variable *tipo* al listado **Variables de estratos**.



Hacer click en **Continuar** (esto permitirá obtener una estimación de los intervalos de confianza para el coeficiente de regresión de Cox) y luego en **Aceptar**. En la ventana principal del *software* se generarán todos los resultados del análisis. Más en concreto:

1. Una tabla que resume el número total de pacientes en estudio, el número y porcentaje de eventos ocurridos durante el ensayo, y el número y porcentaje de observaciones censuradas para el conjunto de datos global:

Resumen de procesamiento de casos

		N	Porcentaje
Casos disponibles en el análisis	Evento[a]	57	91,9%
	Censurado	5	8,1%
	Total	62	100,0%
Casos eliminados	Casos con valores perdidos	0	0,0%
	Casos con tiempo negativo	0	0,0%
	Casos censurados antes del evento más cercano en un estrato	0	0,0%
	Total	0	0,0%
Total		62	100,0%

a. Variable dependiente: Tiempo

2. Una tabla que contiene, entre otros descriptores, el valor del coeficiente de regresión de Cox, **B**, su *p*-valor, **Sig.**, y el *hazard ratio*, **Exp(B)**, asociados a la variable explicativa analizada:

Variables en la ecuación

	B	SE	Wald	df	Sig.	Exp(B)	95,0% CI para Exp(B)	
							Inferior	Superior
Tipo	1,037	,313	11,003	1	<,001	2,821	1,529	5,207

En este caso, está claro que el efecto del tipo de cáncer sobre el riesgo de muerte de un paciente es estadísticamente significativo (*p*-valor < 0,05). Además, dado el valor del *hazard ratio*, se puede afirmar que el riesgo de muerte de un paciente con carcinoma adenoescamoso (valor de la variable explicativa x igual a 1) es 2,821 veces más alto que el riesgo de muerte de un paciente con carcinoma escamoso (valor de la variable explicativa x igual a 0). El intervalo de confianza para el *hazard ratio* indica que con una confianza del 95% se puede afirmar que el verdadero ratio de riesgos estará en el intervalo (1,529 y 5,207).

3. Una representación de las funciones de supervivencia y de las funciones de riesgo acumulado estimadas mediante el modelo de Cox para los dos subgrupos de pacientes:

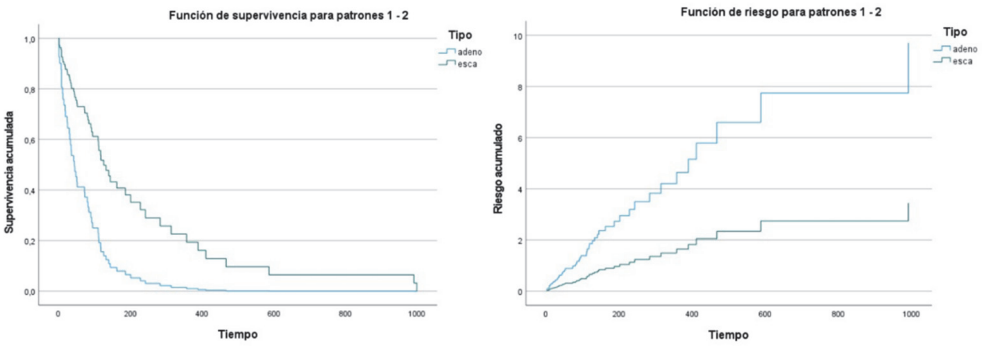


Figura 15 – Curvas de supervivencia y de riesgo acumulado para los dos tipos de cáncer estimadas por el modelo de Cox.

Como se ve en la Figura 15, la supervivencia de los pacientes con carcinoma adenoescamoso es menor (y el riesgo acumulado es mayor) que la de los pacientes con carcinoma escamoso. Estos resultados son semejantes a los ya obtenidos con la prueba de Mantel-Cox (log-rank test) de comparación de curvas de supervivencia descritos en el ejemplo #2. El modelo de regresión de riesgo proporcional de Cox es, sin embargo, un modelo mucho más flexible y general que la prueba de Mantel Cox. Por eso es ampliamente utilizado en los estudios de supervivencia.

Bibliografía

Bowers, D. (2014). Medical Statistics from Scratch. An Introduction for Health Professional. John Wiley & Sons Ltd. 3ª Ed.

Samuels, M. L., Witmer, J. A. & Schaffner, A. (2012). Fundamentos de Estadística para las Ciencias de la Vida. Pearson Educación. 4ª Ed.