Estadística

Introducción a las estadísticas

La estadística nos da métodos para obtener conocimiento a partir de los datos.

¿Para qué se utiliza la estadística?

La estadística se utiliza en todo tipo de **aplicaciones científicas y comerciales**.

Las estadísticas nos brindan un conocimiento más preciso que nos ayuda a **tomar mejores decisiones.**

Las estadísticas pueden enfocarse en hacer **predicciones** sobre lo que sucederá en el futuro. También puede enfocarse en **explicar** cómo se conectan diferentes cosas.

Nota: Las buenas explicaciones estadísticas también son útiles para las predicciones.

Pasos típicos de los métodos estadísticos

Los pasos típicos son:

- **Reuniendo datos**
- **Describir y visualizar datos.**
- **sacar conclusiones**

Es importante tener en cuenta los tres pasos para cualquier pregunta sobre la que deseemos obtener más información.

Saber qué tipos de datos están disponibles puede decirle qué tipo de preguntas puede responder con métodos estadísticos.

Saber qué preguntas desea responder puede ayudar a orientar qué tipo de datos necesita. Es posible que haya muchos datos disponibles, y es importante saber en qué enfocarse.

Estadística

¿Cómo se usa la estadística?

Las estadísticas se pueden utilizar para explicar las cosas de una manera precisa. Puede usarlo para comprender y sacar conclusiones sobre el grupo del que desea saber más. Este grupo se llama **población** .

Una población puede ser muchos tipos diferentes de grupos. Podría ser:

- Todas las personas de un país
- Todos los negocios en una industria
- Todos los clientes de una empresa.
- Todas las personas mayores de 45 años que jueguen al fútbol

y así sucesivamente, solo depende de lo que quieras saber.

La recopilación de datos sobre la población le dará una **muestra** . **Esta es una parte de toda la población**. Luego se utilizan métodos estadísticos en esa muestra.

Los resultados de los métodos estadísticos de la muestra se utilizan para sacar **conclusiones** sobre la población.

Nota: La palabra 'estadística' también puede referirse a fragmentos específicos de conocimiento; como el valor medio de algo.

Conceptos importantes en estadística

- Predicciones y explicaciones
- Poblaciones y Muestras
- Parámetros y estadísticas de muestra
- Métodos de muestreo
- Tipos de datos
- Nivel de medición
- Estadísticas descriptivas
- Variables aleatorias
- Estadísticas univariadas y multivariadas
- Cálculo de probabilidad
- Distribuciones de probabilidad
- Inferencia estadística
- Estimación de parámetros
- Prueba de hipótesis
- Correlación
- Análisis de regresión
- Inferencia causal

Estadística

Estadística y Programación

El análisis estadístico generalmente se realiza con computadoras. Pequeñas cantidades de datos pueden analizarse razonablemente bien sin computadoras.

Históricamente, todos los análisis de datos se realizaban manualmente. Llevaba mucho tiempo y era propenso a errores.

Hoy en día, la programación y el software se utilizan típicamente para el análisis de datos.

En este curso mostraremos ejemplos de código para hacer estadísticas con los lenguajes de programación **Python y R**

Estadística

Estadísticas - Recopilación de datos

1. Reuniendo datos

La recopilación de datos es el **primer paso** en el análisis estadístico.

Digamos, por ejemplo, que desea saber algo sobre **todas las personas en Francia** .

La **población** es entonces toda la población de Francia.

Es demasiado esfuerzo recopilar información sobre todos los miembros de una población (por ejemplo, los más de 67 millones de personas que viven en Francia). A menudo es mucho más fácil reunir un grupo más pequeño de esa población y analizarlo. Esto se llama una **muestra** .

Una muestra representativa

La muestra debe ser **similar** a toda la población de Francia. Debe tener las mismas características que la población. Si solo incluye personas llamadas Jacques que viven en París y tienen 48 años, la muestra no será similar a toda la población.

Entonces, para una buena muestra, necesitará personas de toda Francia, con diferentes edades, profesiones, etc.

Si los miembros de la muestra tienen características similares (como edad, profesión, etc.) a toda la población de Francia, decimos que la muestra es **representativa** de la población.

Una buena **muestra representativa** es crucial para los métodos estadísticos.

Nota: Los datos de una muestra adecuada suelen ser tan buenos datos de toda la población, ¡siempre y cuando sean **representativos!**

Una buena muestra le permite sacar conclusiones precisas sobre toda la población.

Estadística

2. Estadísticas - Descripción de datos

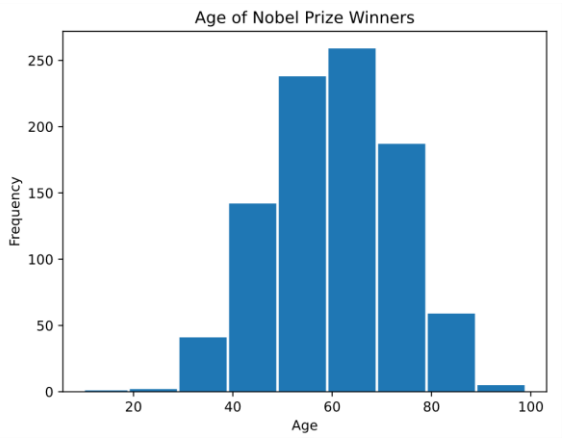
La descripción de los datos suele ser el **segundo paso** del análisis estadístico después de la recopilación de datos.

La información (datos) de su muestra o población se puede visualizar con gráficos o **resumir** con números. Esto mostrará información clave de una manera más simple que solo mirar datos sin procesar. Puede ayudarnos a comprender cómo se **distribuyen** los datos .

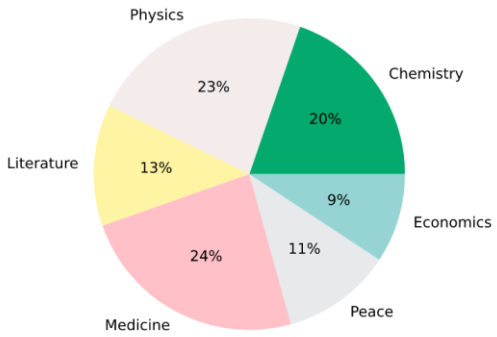
Los gráficos pueden mostrar visualmente la distribución de datos.

Los ejemplos de gráficos incluyen:

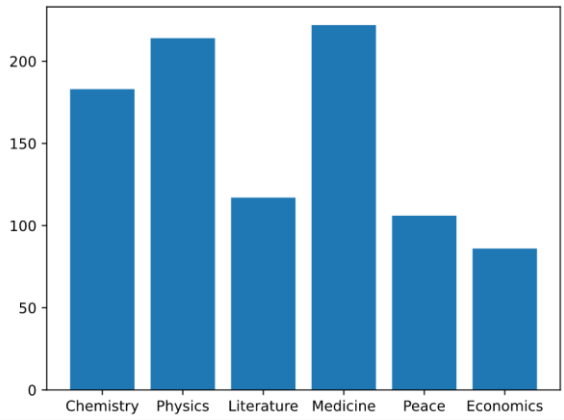
- Histogramas



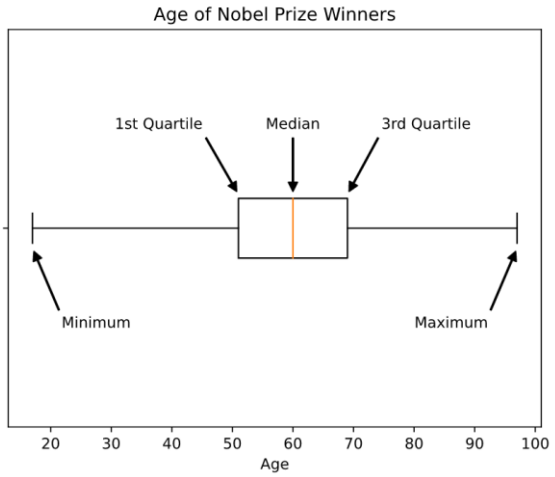
- Gráficos circulares



- Gráficos de barras



- diagramas de caja



Algunos gráficos tienen una estrecha conexión con las estadísticas de resumen numérico. Calcularlos nos da la base de estos gráficos.

Por ejemplo, un **diagrama de caja** muestra visualmente los **cuartiles** de una distribución de datos.

Estadística

Los cuartiles son los datos divididos en cuatro partes de igual tamaño, o cuartos. Un cuartil es un tipo de resumen estadístico.

Resumen estadístico

Las estadísticas de resumen toman una gran cantidad de información y la resumen en unos pocos valores clave.

Los números se calculan a partir de los datos que también describen la forma de las distribuciones. Estas son 'estadísticas' individuales.

Algunos ejemplos importantes son:

- Media, mediana y moda
- Rango y rango intercuartílico
- Cuartiles y percentiles
- Desviación estándar y varianza

Nota: Las estadísticas descriptivas a menudo se presentan como parte del análisis estadístico.

Las estadísticas descriptivas también son útiles para guiar un análisis posterior, dar una idea de los datos y encontrar lo que vale la pena investigar más de cerca.

Estadística

3. Estadísticas - Sacar conclusiones

El uso de estadísticas para sacar conclusiones sobre una población se denomina inferencia estadística.

Inferencia estadística

Las estadísticas de los datos de la **muestra** se utilizan para sacar conclusiones sobre toda la **población** . Este es un tipo de **inferencia estadística** .

La teoría de la probabilidad se utiliza para calcular la certeza de que esas estadísticas también se aplican a la población.

Cuando se usa una muestra, **siempre** habrá cierta incertidumbre sobre cómo se ven los datos para la población.

La incertidumbre se expresa a menudo como **intervalos de confianza** .

Los intervalos de confianza son formas numéricas de mostrar la probabilidad de que el **valor real** de esta estadística se encuentre dentro de un cierto rango para la población.

La prueba de hipótesis es otra forma de verificar si una declaración sobre una población es verdadera. Más precisamente, verifica qué tan probable es que una hipótesis sea cierta según los datos de la muestra.

Algunos ejemplos de declaraciones o preguntas que se pueden verificar con pruebas de hipótesis:

- Los holandeses son más altos que los daneses
- ¿La gente prefiere Pepsi o Coca-Cola?
- ¿Un nuevo medicamento cura una enfermedad?

Nota: Los intervalos de confianza y las pruebas de hipótesis están estrechamente relacionados y describen las mismas cosas de diferentes maneras. Ambos son ampliamente utilizados en la ciencia.

Estadística

Inferencia causal

La inferencia causal se utiliza para investigar si algo causa otra cosa.

Por ejemplo: ¿La lluvia hace crecer las plantas?

Si pensamos que dos cosas están relacionadas, podemos investigar para ver si se **correlacionan** . Las estadísticas se pueden utilizar para averiguar qué tan fuerte es esta relación.

Incluso si las cosas están correlacionadas, descubrir que algo es causado por otras cosas puede ser difícil. Puede hacerse con un buen **diseño experimental** u otras técnicas estadísticas especiales.

Nota: A menudo es difícil lograr un buen diseño experimental debido a preocupaciones éticas u otras razones prácticas.

Estadística

Estadísticas - Predicción y Explicación

Algunos tipos de métodos estadísticos se centran en predecir lo que sucederá.

Otros tipos de métodos estadísticos se centran en explicar cómo se conectan las cosas.

Predicción

Algunos métodos estadísticos no se centran en explicar cómo se conectan las cosas. Sólo la precisión de la predicción es importante.

Muchos métodos estadísticos tienen éxito en la predicción sin dar una idea de cómo están conectadas las cosas.

Algunos tipos de aprendizaje automático permiten que las computadoras hagan el trabajo duro, pero la forma en que predicen es difícil de entender. Estos enfoques también pueden ser vulnerables a errores si las circunstancias cambian, ya que la forma en que funcionan es menos clara.

Nota: Las predicciones sobre eventos futuros se denominan **pronósticos** . No todas las predicciones son sobre el futuro.

Algunas predicciones pueden ser sobre algo más que se desconoce, incluso si no es en el futuro.

Explicación

A menudo se utilizan diferentes métodos estadísticos para explicar cómo se conectan las cosas. Estos métodos estadísticos pueden no hacer buenas predicciones.

Estos métodos estadísticos a menudo explican solo pequeñas partes de la situación total. Pero, si solo quiere saber cómo están conectadas algunas cosas, el resto puede no importar.

Si estos métodos explican con precisión cómo se conectan todas las cosas relevantes, también serán buenos para la predicción. Pero lograr explicar cada detalle a menudo es un desafío.

Algunas veces estamos específicamente interesados en averiguar si una cosa causa otra. Esto se llama **inferencia causal** .

Si estamos ante situaciones complicadas, muchas cosas están conectadas. Para descubrir qué causa qué, necesitamos desenredar todas las formas en que estas cosas están conectadas.

Nota: Las conclusiones sobre la causalidad deben hacerse con cuidado.

Estadística

Estadísticas - Poblaciones y Muestras

Los términos 'población' y 'muestra' son importantes en estadística y se refieren a conceptos clave que están estrechamente relacionados.

Población y Muestras

Población : Todo en el grupo sobre el que queremos aprender.

Muestra : Una parte de la población.

Ejemplos de poblaciones y una muestra de esas poblaciones:

Population	Sample
All of the people in Germany	500 Germans
All of the customers of Netflix	300 Netflix customers
Every car manufacturer	Tesla, Toyota, BMW, Ford

Para un buen análisis estadístico, la muestra debe ser lo más "similar" posible a la población. Si son lo suficientemente similares, decimos que la muestra es **representativa** de la población.

La muestra se utiliza para sacar conclusiones sobre toda la población. Si la muestra no es lo suficientemente similar a toda la población, las conclusiones pueden ser inútiles.

Nota: Muchas palabras tienen significados específicos en estadística.

La palabra 'población' normalmente se refiere a un grupo de personas. En estadística, es cualquier grupo específico sobre el que estamos interesados en aprender.

Estadística

Estadísticas - Parámetros y Estadísticas

Los términos 'parámetro' y (muestra) 'estadística' se refieren a conceptos clave que están estrechamente relacionados en las estadísticas.

También están directamente conectados con los conceptos de poblaciones y muestras.

Parámetros y Estadísticas

Parámetro : Un número que describe algo acerca de toda la **población** .

Estadística de la muestra : un número que describe algo sobre la **muestra** .

Los parámetros son las cosas clave sobre las que queremos aprender. Los parámetros son generalmente desconocidos.

Las estadísticas de muestra nos dan **estimaciones** para los parámetros.

Siempre habrá cierta **incertidumbre** sobre cuán precisas son las estimaciones. Más certeza nos da un conocimiento más útil.

Para cada parámetro sobre el que queremos aprender, podemos obtener una muestra y calcular una estadística de muestra, lo que nos da una estimación del parámetro.

La media, la mediana y la moda son diferentes tipos de promedios (valores típicos en una población).

Por ejemplo:

- La edad típica de las personas en un país.
- Los beneficios típicos de una empresa.
- La autonomía típica de un coche eléctrico

La varianza y **la desviación estándar** son dos tipos de valores que describen cuán dispersos están los valores.

Una sola clase de estudiantes en una escuela generalmente tendría aproximadamente la misma edad. La edad de los estudiantes tendrá **baja** varianza y desviación estándar.

Todo un país tendrá gente de todo tipo y de diferentes edades. La varianza y desviación estándar de la edad en todo el país sería entonces **mayor** que en un solo grado escolar.

Estadística

Estadísticas - Tipos de estudio

Un estudio estadístico puede ser parte del proceso de recopilación de datos.

Hay diferentes tipos de estudios. Algunos son mejores que otros, pero pueden ser más difíciles de hacer.

Principales tipos de estudios estadísticos

Los principales tipos de estudios estadísticos son los estudios **observacionales** y **experimentales** .

A menudo nos interesa saber si algo es la **causa** de otra cosa.

Los estudios experimentales son generalmente mejores que los estudios de observación para investigar esto, pero generalmente requieren más esfuerzo.

Un estudio observacional es cuando se observa y recopila datos sin cambiar nada.

Estudios experimentales

En un estudio experimental, se cambian las **circunstancias** alrededor de la muestra. Por lo general, comparamos dos grupos de una población y estos dos grupos reciben un trato **diferente** .

Un ejemplo puede ser un estudio médico para ver si un nuevo medicamento es efectivo.

Un grupo recibe el medicamento y el otro no. Estas son las diferentes circunstancias alrededor de esas muestras.

Podemos comparar la salud de ambos grupos después y ver si los resultados son diferentes.

Los estudios experimentales pueden permitirnos investigar las relaciones causales. Un estudio experimental bien diseñado puede ser útil ya que puede **aislar** la relación que nos interesa de **otros efectos** . Entonces podemos estar más seguros de que estamos midiendo el verdadero efecto.

Estadística

Estadísticas - Tipos de muestra

Un estudio necesita participantes y hay diferentes formas de reunirlos.

Algunos métodos son mejores que otros, pero pueden ser más difíciles.

Diferentes tipos de métodos de muestreo

Muestreo aleatorio

Una muestra aleatoria es aquella en la que todos los miembros de la población tienen las **mismas posibilidades** de ser elegidos.

El muestreo aleatorio es el mejor. Pero puede ser difícil, o imposible, asegurarse de que sea completamente aleatorio.

Nota: Todos los demás métodos de muestreo se comparan con lo cerca que están de una muestra aleatoria: cuanto más cerca, mejor.

Muestreo de conveniencia

Una muestra de conveniencia es donde se eligen los participantes que son más fáciles de alcanzar.

Nota: El muestreo de conveniencia es el más fácil de hacer.

En muchos casos, esta muestra no será lo suficientemente **similar** a la población, y las conclusiones pueden ser potencialmente inútiles.

Muestreo Sistemático

Una muestra sistemática es aquella en la que los participantes son elegidos por algún sistema regular.

Por ejemplo:

- Las primeras 30 personas en una cola
- Cada tercio en una lista
- Los 10 primeros y los 10 últimos

Estadística

Muestreo estratificado

Una muestra estratificada es donde la población se divide en grupos más pequeños llamados "estratos".

Los 'estratos' pueden, por ejemplo, basarse en datos demográficos, como:

- Diferentes grupos de edad
- Profesiones

La estratificación de una muestra es el primer paso. Otro método de muestreo (como el muestreo aleatorio) se usa para el segundo paso de elegir participantes de todos los grupos más pequeños (estratos).

Muestreo agrupado

Una muestra agrupada es donde la población se divide en grupos más pequeños llamados 'conglomerados'.

Los clusters suelen ser naturales, como diferentes ciudades de un país.

Los conglomerados se eligen aleatoriamente para la muestra.

Todos los miembros de los conglomerados pueden participar en la muestra, o los miembros pueden ser elegidos al azar de los conglomerados en un tercer paso.

Estadística

Estadísticas - Tipos de datos

Los datos pueden ser de diferentes tipos y requieren diferentes tipos de métodos estadísticos para analizarlos.

Diferentes tipos de datos

Hay dos tipos principales de datos: cualitativos (o 'categóricos') y cuantitativos (o 'numéricos'). Estos tipos principales también tienen diferentes subtipos según su **nivel de medición** .

Datos cualitativos

Información sobre algo que se puede clasificar en diferentes categorías que no se puede describir directamente con números.

Ejemplos:

- Marcas
- Nacionalidad
- Profesiones

Con datos categóricos podemos calcular estadísticas como **proporciones** . Por ejemplo, la proporción de indios en el mundo, o el porcentaje de personas que prefieren una marca a otra.

Datos cuantitativos

Información sobre algo que se describe mediante números.

Ejemplos:

- Ingreso
- Años
- Altura

Con datos numéricos podemos calcular estadísticas como la renta **media de un país, o el rango** de estatura de los jugadores de un equipo de fútbol.

Estadística

Estadísticas - Niveles de medición

Los diferentes tipos de datos tienen diferentes niveles de medición.

Los niveles de medición son importantes para saber qué tipos de estadísticas se pueden calcular y cómo presentar mejor los datos.

Niveles de medición

Los principales tipos de datos son Cualitativos (categorías) y Cuantitativos (numéricos). Estos se dividen a su vez en los siguientes niveles de medición.

Estos niveles de medición también se denominan "escalas" de medición.

Nivel Nominal

Categorías (datos cualitativos) sin ningún orden.

Ejemplos:

- Nombres de marca
- Países
- Colores

nivel ordinal

Categorías que se pueden ordenar (de menor a mayor), pero la "distancia" precisa entre cada una no es significativa.

Ejemplos:

- Escalas de calificación de letras de F a A
- rangos militares
- Nivel de satisfacción con un producto.

Considere las calificaciones con letras de la F a la A: ¿La calificación A es exactamente el doble de buena que la B? Y, ¿la calificación B también es el doble de buena que la C?

La distancia exacta entre grados no es clara ni precisa. Si las calificaciones se basan en la cantidad de puntos de una prueba, puede decir que hay una "distancia" precisa en la escala de puntos, pero no las calificaciones en sí.

Estadística

Nivel de intervalo

Datos que se pueden ordenar y la distancia entre ellos es objetivamente significativo. Pero no hay un valor 0 natural donde se origina la escala.

Ejemplos:

- años en un calendario
- Temperatura medida en Fahrenheit

Nota: las personas suelen inventar las escalas de intervalos, como los grados de temperatura.

0 grados Celsius son 32 grados Fahrenheit. Hay distancias constantes entre cada grado (por cada 1 grado Celsius extra, hay 1,8 Fahrenheit extra), pero no están de acuerdo en dónde están los 0 grados.

Nivel de relación

Datos que se pueden ordenar y existe una distancia consistente y significativa entre ellos. Y también tiene un valor 0 natural.

Ejemplos:

- Dinero
- Años
- Tiempo

Los datos que están en el nivel de razón (o "escala de razón") nos brindan la información más detallada. Fundamentalmente, podemos comparar con precisión qué tan grande es un valor en comparación con otro. Esta sería la relación entre estos valores, como el doble o diez veces más pequeño.

Estadística

Tipos de variables

Una variable es algo que puede tener diferentes valores, como el peso, la altura o el color de los ojos, a diferencia de una constante que solo tiene un valor, como la velocidad de la luz.

Los tipos de variables vienen determinados por el dato que representa. Por ejemplo, el peso es una variable cuantitativa cuando se expresa en números como gramos o kilogramos de un objeto. Mientras que si se presenta en términos de "pesado" o "liviano", sería una variable cualitativa, porque presenta una cualidad.

Usamos diferentes tipos de variables en matemáticas, estadística y en la investigación científica. Veamos.

1. Variable cuantitativa continua

Una variable cuantitativa continua es toda variable representada por números que pueden ser expresados por fracciones o decimales como la temperatura, donde encontramos valores como 37 ºC, 37.5ºC o 38.5ºC.

Otros ejemplos de variables son:

- Los niveles de un compuesto en la sangre: los niveles de azúcar en la sangre de una persona con diabetes durante un día puede ser 7.5 mM, 8.3 mM o 5.0 mM.

- La medida de la presión atmosférica: a 0 metros sobre el nivel del mar la presión atmosférica es igual a 1 atm y a 1000 metros sobre el nivel del mar es igual a 0.887 atm.

- La masa de un objeto: los aguacates de un árbol pueden medir 200.5 gramos, 201 gramos o 205.2 gramos.

- La longitud de un objeto: la altura de los árboles en un parque.

Estadística

2. Variable cuantitativa discreta

Una variable cuantitativa discreta solamente puede tomar valores integrales, es decir 1, 2 o 555, pero no 1.5 o 2.25. Ejemplos de este tipo de variables son:

- El número de veces que algo sucede: las veces que llovió cada mes en el año 2020 en Bogotá.

- El número de veces que alguien asume un determinado comportamiento: las veces que personas mayores de 50 años participan en un maratón.

- La cantidad de personas o seres en un grupo: el número de estudiantes en un salón de clases solo pueden ser un valor integral, no puede haber una fracción de un estudiante.

- La cantidad de objetos en un lugar: el número de sillas o de libros en cada salón de clases de una escuela.

Estadística

3. Variable cualitativa dicotómica

La variable cualitativa dicotómica es un dato no numérico que presenta una cualidad, propiedad o condición observable, que nada más presenta dos valores. Por ejemplo:

- El veredicto de un jurado: "culpable" o "no culpable".
- El sexo: "masculino" o "femenino".
- El resultado de un examen de antígeno: "positivo" o "negativo".
- Presencia de una condición: "presente" o "ausente".
- El tipo de hospital: "público" o "privado".

Estadística

4. Variable cualitativa categórica o nominal

Es la variable no numérica que presenta tres o más categorías. Por ejemplo:

- La afinidad por un equipo: en el futbol mexicano puedes ser fanático de "el Atlas Fútbol Club", "el Club América" o "el club León".
- Los deportes olímpicos: "natación", "voleibol", "atletismo", "esgrima" o "gimnasia".
- Los estados de la materia: "sólido", "líquido" o "gaseoso".
- Carreras universitarias: "biología", "derecho", "medicina", "enfermería" o "economía".

Estadística

5. Variables ordinales

Los valores pueden ordenarse, de menor a mayor, de más importante a menos importante, de primero a último, etc. Este tipo de variable la observamos en:

- Clase social: "clase baja", "clase media" o "clase alta".

- Nivel socioeconómico: A/B (clase rica), C+ (clase media alta), C (clase media), D+ (clase media baja), D (clase pobre), E (pobreza extrema).

- Competencia en un idioma: "básico", "intermedio" o "avanzado".

- Grados de un colegio: primer grado, segundo grado, tercer grado, etc.

Estadística

6. Variable independiente

La variable independiente es una variable que se presenta sin necesidad de otra. En ciencia, es la variable manipulada o controlada por el investigador. Es decir, se le puede atribuir valores a voluntad dentro de ciertos límites. Por ejemplo, en el estudio de los efectos de una droga, la variable independiente puede ser cualitativa si hay un grupo control sin droga y un grupo con tratamiento.

Por lo general, los estudios científicos se enfocan en examinar los efectos de una variable independiente. En un estudio se analizó el impacto de cinco intensidades de un campo magnético sobre plantas de cebada. En este caso, la variable independiente fue la intensidad del campo magnético.

Estadística

7. Variable dependiente

La variable dependiente es una variable que es consecuencia de otra. Por ejemplo, la altura de los niños es una variable dependiente de la edad. Un niño de 10 años es más alto que un niño de 5 años.

La variable dependiente es la medida del efecto de la variable independiente. En un estudio se midió la circunferencia de la cintura en dos grupos de mujeres, un grupo control y otro grupo que practicó danza terapia por ocho semanas. La variable dependiente es la circunferencia de la cintura, mientras el régimen de baile es la variable independiente.

La forma más fácil de identificar una variable dependiente es detectando el efecto o la consecuencia de algo, es decir, la variable independiente que es la causa. Como en el caso anterior, la práctica de baile es la causa o variable independiente y la consecuencia es sobre la medida de la cintura o variable dependiente.

Mientras la variable independiente se manipula o fija, la variable dependiente se mide o registra.

8. Variable independiente extraña

Son aquellas variables independientes que no están relacionadas con el propósito del estudio, pero que pueden afectar las variables dependientes.

Un ejemplo de una variable extraña es la inteligencia en un estudio de los efectos de tomar desayuno y los resultados de un examen de matemática en un grupo de estudiantes. Se supone que la variable independiente es tomar o no tomar desayuno antes de realizar un examen de matemáticas. Los resultados del examen serian la variable dependiente, sin embargo, en este caso, la inteligencia de cada niño podría influenciar en el resultado.

Estadística

Estadística descriptiva e inferencial

La estadística descriptiva es el conjunto de métodos estadísticos que describen y/o caracterizan un grupo de datos. La estadística inferencial busca deducir y sacar conclusiones acerca de situaciones generales más allá del conjunto de datos obtenidos.

La estadística es una disciplina que se encarga de procesar y organizar los datos, siendo los datos cualquier medida o valor que se puede obtener a través de experimentos, encuestas, censos u otros medios. El análisis de los datos por lo general se inician con la aplicación de métodos de estadística descriptiva, para luego seguir con métodos de estadística inferencial.

	Estadística descriptiva	Estadística inferencial
Definición	Métodos empleados para resumir las características clave de datos conocidos.	Métodos que implican el uso de datos muestrales para hacer generalizaciones o inferencias acerca de una población.
Objetivos	Caracterizar un grupo de datos Examinar tendencias o distribuciones	Examinar diferencias entre grupos. Examinar si las variables están asociadas. Comparar promedios entre grupos. Predecir una variable a partir de otra.
Métodos de análisis	Medidas de tendencia central: Media Mediana Moda Medidas de variabilidad: -Varianza -Desviación estándar -Rango -Frecuencia	t-test Análisis de varianza Correlación Regresión
Áreas de aplicación	Ciencias naturales y sociales	Ciencias sociales y naturales
Ejemplos	-Características de pacientes que son atendidos en un hospital. -Media y distribución de la edad, peso y altura de los estudiantes de un colegio.	-Predecir la aparición de demencia en personas según su estado cardiovascular. -Probar que un medicamento sirve para el tratamiento de una enfermedad.

Estadística

¿Qué es estadística descriptiva?

La estadística descriptiva es la parte de la estadística que arregla los datos de forma que puedan ser analizados e interpretados. Los métodos de estadística descriptiva nos permiten:

- Determinar la tendencia central de una variable: promedio o media aritmética, mediana o moda.
- Determinar la variabilidad de una variable: desviación estándar, varianza, rangos.
- Determinar cómo es la distribución de una variable: histograma de frecuencias, distribución normal.

Ejemplos de estadística descriptiva

Cuando se quiere caracterizar un grupo de individuos, se usa la estadística descriptiva. Por ejemplo, tenemos los siguientes datos de temperatura corporal en un grupo de hombres y mujeres:

Temperatura corporal en hombres (ºC)	Temperatura corporal en mujeres (ºC)
36,1	36,2
35,9	37,2
36,0	37,3
36,4	37,1
36,3	37,0
36,7	37,2
36,9	36,9
36,8	36,8
37,2	36,4
37,0	37,0

Tal cual como están presentados no podemos sacar ninguna conclusión, pero al aplicar las técnicas de estadística descriptiva, podemos decir entonces que:

- los hombres en este grupo tienen una temperatura promedio de 36,53 ºC con una desviación estándar de 0,45;
- las mujeres en este grupo tienen una temperatura promedio de 36,91 ºC, con una desviación estándar de 0,36.

Estadística

¿Qué es estadística inferencial?

La estadística inferencial o inferencia estadística es la parte de la estadística que busca predecir o deducir características o resultados esperados de una población, basados en los datos obtenidos de una muestra de esa población. Dentro de las técnicas aplicadas en la estadística inferencial existen:

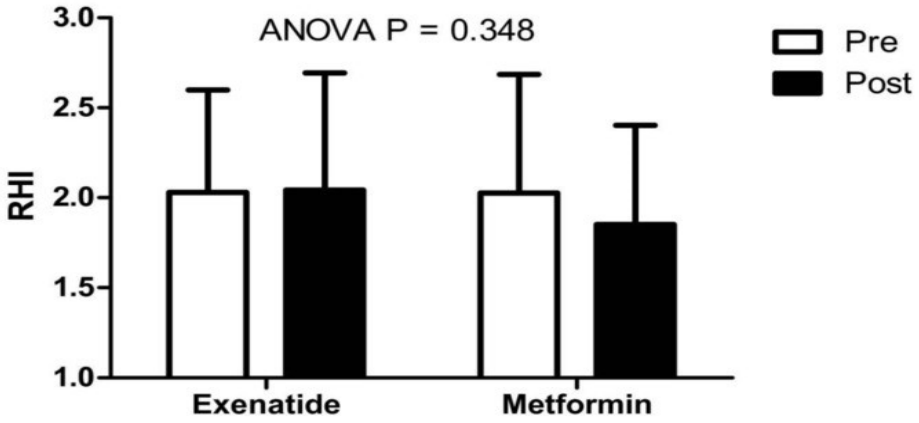
- El t test: se usa para comparar la media aritmética de dos grupos determinando si las diferencias entre los grupos ocurren al azar o de forma sistemática indicando una diferencia real.
- El análisis de varianza o ANOVA: se aplica para comparar a dos o más grupos de variables.
- El análisis de correlación: revela si los valores entre dos variables tienden a cambiar sistemáticamente. Para hacer esas determinaciones se usa el coeficiente de correlación *r* y el valor de *p* o de intervalo de confianza IC.
- El análisis de regresión: permite predecir un valor a partir de otro.

Ejemplos de estadística inferencial

Si queremos determinar si algún comportamiento o estado biológico está asociado a una enfermedad, utilizamos métodos de estadística inferencial. Por ejemplo, en un estudio realizado en Alemania, se evaluó diferentes parámetros de salud a 3109 personas durante casi siete años. Los resultados arrojaron que niveles altos de glucosa en sangre (mayor de 126 mg/dl en ayunas), el fumar y la inactividad física estaban asociados con el desarrollo de demencia.

Cuando se descubre un nuevo medicamento y se quiere demostrar su efectividad en determinada enfermedad, se utiliza estadística inferencial. En este caso, se comparan los efectos de un grupo tratado con el medicamento y otro grupo tratado con un placebo o un medicamento de control.

Estadística



El grupo de Kelly y colaboradores investigaron la función del endotelio en 50 individuos obesos antes (Pre) y después (Post) de tres meses de tratamiento con dos drogas: exenatide y metformin (control). Al analizar los resultados con la técnica de ANOVA (P=0,348; estadística inferencial) descubrieron que el exenatide no tenía efecto en la función endotelial, en comparación con la metformin.

Estadística

Estadísticas - Estadísticas descriptivas

Las estadísticas descriptivas nos dan una idea de los datos sin tener que mirarlos todos en detalle.

Características clave para describir acerca de los datos

Obtener una visión general rápida de cómo se distribuyen los datos es un paso importante en los métodos estadísticos.

Calculamos valores numéricos clave sobre los datos que nos informan sobre la distribución de los datos. También dibujamos gráficos que muestran visualmente cómo se distribuyen los datos.

Características clave de los datos:

- ¿Dónde está el centro de los datos? (ubicación)
- ¿Cuánto varían los datos? (escala)
- ¿Cuál es la forma de los datos? (forma)

Estos pueden describirse mediante estadísticas de resumen (valores numéricos).

El centro de los datos

El centro de los datos es donde se concentran la mayoría de los valores.

Los diferentes tipos de promedios, como la media, la mediana y la moda, son medidas del centro.

Nota: Las medidas del centro también se denominan parámetros de ubicación porque nos dicen algo sobre dónde se 'ubican' los datos en una recta numérica.

La variación de los datos

La variación de los datos es qué tan dispersos están los datos alrededor del centro.

Las estadísticas como la desviación estándar, el rango y los cuartiles son medidas de variación.

Nota: Las medidas de variación también se denominan parámetros de escala .

Estadística

La forma de los datos

La forma de los datos puede referirse a cómo se agrupan los datos a ambos lados del centro.

Estadísticas como sesgo describen si el lado derecho o izquierdo del centro es más grande. El sesgo es un tipo de parámetros de forma .

Tablas de frecuencia

Una forma típica de presentar datos es con tablas de frecuencia .

Una tabla de frecuencia cuenta y ordena los datos en una tabla. Por lo general, los datos deberán clasificarse en intervalos.

Las tablas de frecuencia son a menudo la base para hacer gráficos para presentar visualmente los datos.

Estadística

Visualización de datos

Se utilizan diferentes tipos de gráficos para diferentes tipos de datos. Por ejemplo:

- Gráficos circulares para datos cualitativos
- Histogramas para datos cuantitativos
- Gráficos de dispersión para datos bivariados

Los gráficos a menudo tienen una estrecha conexión con las estadísticas de resumen numérico.

Por ejemplo, los diagramas de caja muestran dónde están los cuartiles .

Los cuartiles también nos dicen dónde están los valores mínimo y máximo, el rango, el rango intercuartílico y la mediana.

Estadística

Estadísticas - Tablas de Frecuencia

Una tabla de frecuencia es una forma de presentar datos. Los datos se cuentan y ordenan para resumir conjuntos más grandes de datos.

Con una tabla de frecuencias puedes analizar la forma en que los datos se distribuyen en diferentes valores.

Tablas de frecuencia

Frecuencia significa el número de veces que aparece un valor en los datos. Una tabla puede mostrarnos rápidamente cuántas veces aparece cada valor.

Si los datos tienen muchos valores diferentes, es más fácil usar intervalos de valores para presentarlos en una tabla.

Aquí está la edad de los 934 ganadores del Premio Nobel hasta el año 2020. En la tabla, cada fila es un intervalo de edad de 10 años.

Age Interval	**Frequency**
10-19	1
20-29	2
30-39	48
40-49	158
50-59	236
60-69	262
70-79	174
80-89	50
90-99	3

Podemos ver que solo hay un ganador entre los 10 y los 19 años. Y que el mayor número de ganadores está en los 60 años.

Nota: Los intervalos para los valores también se denominan "contenedores".

Estadística

Tablas de frecuencias relativas

La frecuencia relativa significa el número de veces que aparece un valor en los datos en comparación con la cantidad total. Un porcentaje es una frecuencia relativa.

Aquí están las frecuencias relativas de las edades de los ganadores del Premio Nobel. Ahora, todas las frecuencias se dividen por el total (934) para dar porcentajes.

Age Interval	Relative Frequency
10-19	0.11%
20-29	0.21%
30-39	5.14%
40-49	16.92%
50-59	25.27%
60-69	28.05%
70-79	18.63%
80-89	5.35%
90-99	0.32%

Estadística

Tablas de frecuencias acumulativas

La frecuencia acumulada cuenta hasta un valor particular.

Aquí están las frecuencias acumuladas de las edades de los ganadores del Premio Nobel. Ahora, podemos ver cuántos ganadores han sido menores de cierta edad.

Age	Cumulative Frequency
Younger than 20	1
Younger than 30	3
Younger than 40	51
Younger than 50	209
Younger than 60	445
Younger than 70	707
Younger than 80	881
Younger than 90	931
Younger than 100	934

También se pueden hacer tablas de frecuencias acumuladas con frecuencias relativas (porcentajes).

Estadística

Estadísticas - Histogramas

Un histograma presenta visualmente datos cuantitativos.

Histogramas

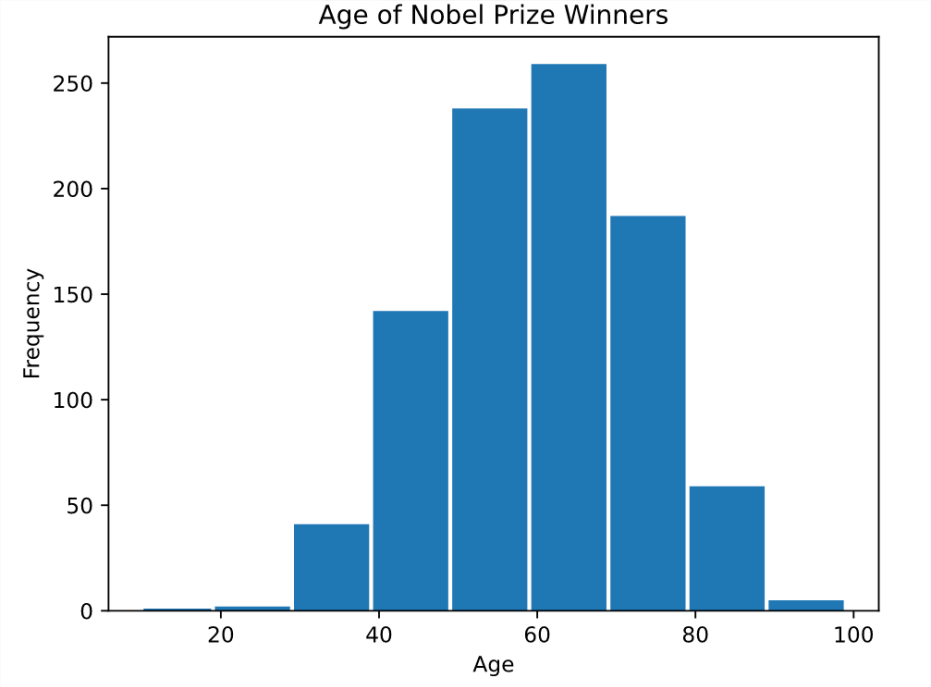
Un histograma es un gráfico muy utilizado para mostrar la distribución de datos cuantitativos (numéricos).

Muestra la frecuencia de valores en los datos, generalmente en intervalos de valores. La frecuencia es la cantidad de veces que ese valor apareció en los datos.

Cada intervalo se representa con una barra, colocada junto a los otros intervalos en una recta numérica.

La altura de la barra representa la frecuencia de valores en ese intervalo.

Aquí hay un histograma de la edad de los 934 ganadores del Premio Nobel hasta el año 2020:



Este histograma utiliza intervalos de edad de 10 a 19, de 20 a 29, etc.

Nota: Los histogramas son similares a los gráficos de barras, que se utilizan para datos cualitativos.

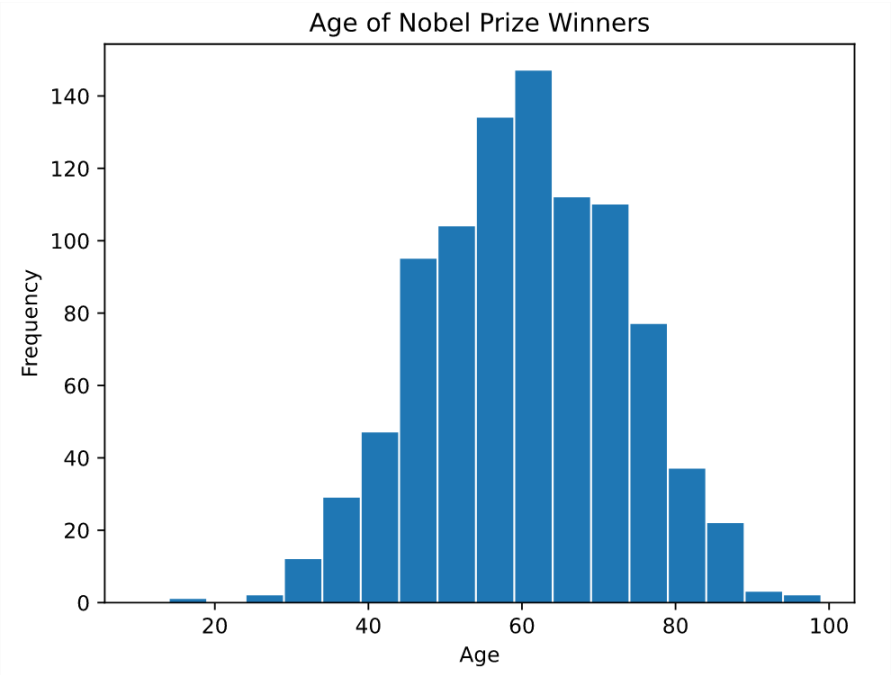
Estadística

Ancho del contenedor

Los intervalos de valores a menudo se denominan "contenedores". Y la longitud de un intervalo se llama 'ancho de contenedor'.

Podemos elegir cualquier ancho. Es mejor con un ancho de contenedor que muestre suficientes detalles sin ser confuso.

Aquí hay un histograma de los mismos datos del ganador del Premio Nobel, pero con anchos de intervalo de 5 en lugar de 10:



Este histograma utiliza intervalos de edad de 15 a 19, de 20 a 24, de 25 a 29, etc.

Los intervalos más pequeños brindan una visión más detallada de la distribución de los valores de edad en los datos.

Estadística

Estadísticas - Gráficos de barras

Un gráfico de barras presenta visualmente datos cualitativos.

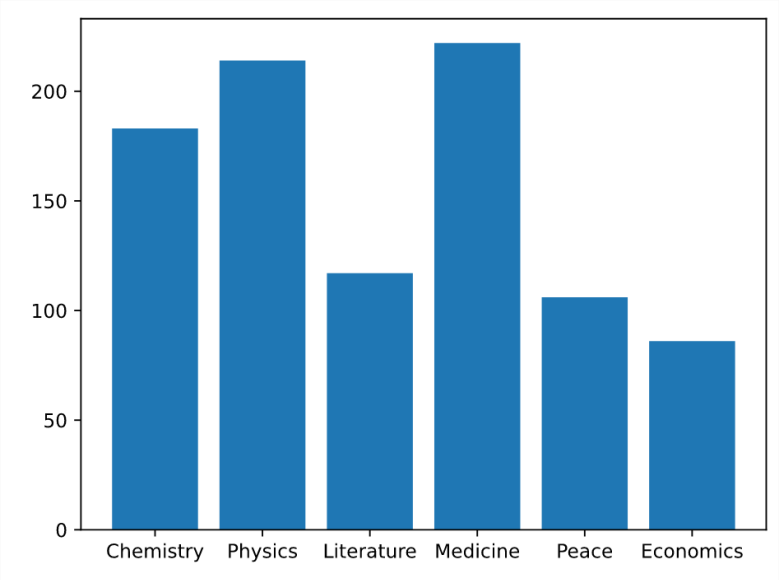
Gráficos de barras

Los gráficos de barras se utilizan para mostrar la distribución de datos cualitativos (categóricos).

Muestra la frecuencia de los valores en los datos. La frecuencia es la cantidad de veces que ese valor apareció en los datos.

Cada categoría se representa con una barra. La altura de la barra representa la frecuencia de los valores de esa categoría en los datos.

Aquí hay un gráfico de barras del número de personas que han ganado un Premio Nobel en cada categoría hasta el año 2020:



Algunas de las categorías han existido por más tiempo que otras. Los ganadores múltiples también son más comunes en algunas categorías. Así que hay un número diferente de ganadores en cada categoría.

Nota: Los gráficos de barras son similares a los histogramas, que se utilizan para datos cuantitativos.

Estadística

Estadísticas - Gráficos circulares

Un gráfico circular presenta visualmente datos cualitativos.

Gráficos circulares

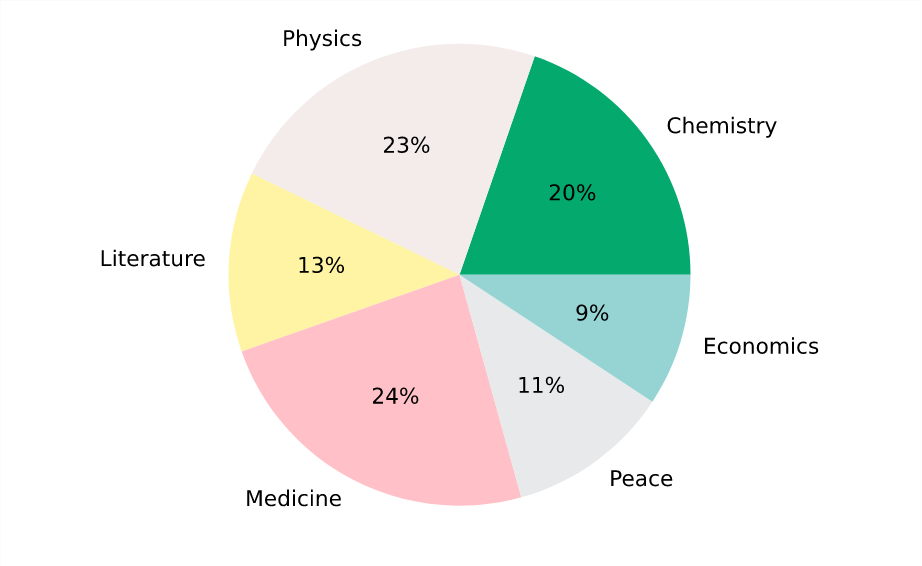
Los gráficos circulares se utilizan para mostrar la distribución de datos cualitativos (categóricos).

Muestra la frecuencia o frecuencia relativa de valores en los datos.

La frecuencia es la cantidad de veces que ese valor apareció en los datos. La frecuencia relativa es el porcentaje del total.

Cada categoría se representa con una porción en el 'pastel' (círculo). El tamaño de cada segmento representa la frecuencia de los valores de esa categoría en los datos.

Aquí hay un gráfico circular de la cantidad de personas que han ganado un Premio Nobel en cada categoría hasta el año 2020:



Este gráfico circular muestra la frecuencia relativa. Por lo tanto, cada segmento tiene el tamaño del porcentaje de cada categoría.

Algunas de las categorías han existido por más tiempo que otras. Los ganadores múltiples también son más comunes en algunas categorías. Así que hay un número diferente de ganadores en cada categoría.

Estadística

Estadísticas - Diagramas de caja

Un diagrama de caja es un gráfico que se utiliza para mostrar las características clave de los datos cuantitativos.

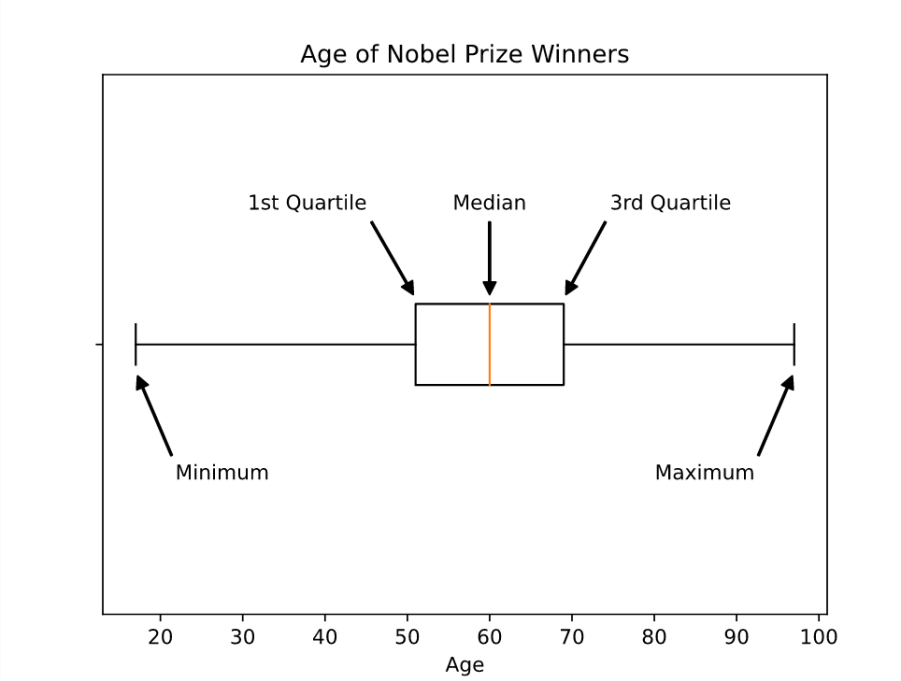
Diagramas de caja

Un diagrama de caja es una buena manera de mostrar muchas características importantes de los datos cuantitativos (numéricos).

Muestra la mediana de los datos. Este es el valor medio de los datos y un tipo de valor promedio.

También muestra el rango y los cuartiles de los datos. Esto nos dice algo sobre cuán dispersos están los datos.

Aquí hay un diagrama de caja de la edad de todos los ganadores del Premio Nobel hasta el año 2020:



La mediana es la línea roja que pasa por el medio de la 'caja'. Podemos ver que esto está justo encima del número 60 en la recta numérica de abajo. Entonces el valor medio de la edad es 60 años.

El lado izquierdo de la caja es el 1er cuartil . Este es el valor que separa el primer trimestre , o el 25% de los datos, del resto. Aquí, esto es 51 años.

El lado derecho de la caja es el tercer cuartil . Este es el valor que separa los tres primeros trimestres , o el 75% de los datos, del resto. Aquí, esto es 69 años.

Estadística

La distancia entre los lados de la caja se denomina rango intercuartílico (RIC) . Esto nos dice dónde está la 'mitad media' de los valores. Aquí, la mitad de los ganadores tenían entre 51 y 69 años.

Los extremos de las líneas del cuadro de la izquierda y la derecha son los valores mínimo y máximo de los datos. La distancia entre estos se llama rango .

El ganador más joven tenía 17 años y el mayor 97 años. Así que el rango de edad de los ganadores fue de 80 años.

Nota: Los diagramas de caja también se denominan "diagramas de caja y bigotes".

Estadística

Estadísticas - Promedio

Un promedio es una medida de dónde se encuentran la mayoría de los valores en los datos.

El centro de los datos

El centro de los datos es donde se encuentran la mayoría de los valores de los datos. Los promedios son medidas de la ubicación del centro.

Hay diferentes tipos de promedios. Los más utilizados son:

- Media
- Mediana
- Moda

Nota: En estadística, los promedios se denominan a menudo "medidas de tendencia central ".

Por ejemplo, usando los valores:

```
40, 21, 55, 21, 48, 13, 72
```

Media

La media generalmente se conoce como 'el promedio'.

La media es la suma de todos los valores de los datos dividida por el número total de valores de los datos:

```
(40 + 21 + 55 + 31 + 48 + 13 + 72)/7 = 38.57
```

Nota: Hay varios tipos de valores medios. El tipo más común de media es la media aritmética .

En este tutorial, 'media' se refiere a la media aritmética.

Estadística

Mediana

La mediana es el 'valor medio' de los datos.

La mediana se encuentra ordenando todos los valores en los datos y eligiendo el valor medio:

```
13, 21, 21, 40, 48, 55, 72
```

La mediana está menos influenciada por los valores extremos en los datos que la media.

Cambiar el último valor a 356 no cambia la mediana:

```
13, 21, 21, 40, 48, 55, 356
```

La mediana sigue siendo 40.

Cambiar el último valor a 356 cambia mucho la media :

```
(13 + 21 + 21 + 40 + 48 + 55 + 72)/7 = 38.57

(13 + 21 + 21 + 40 + 48 + 55 + 356)/7 = 79.14
```

Nota: Los valores extremos son valores en los datos que son mucho más pequeños o más grandes que los valores promedio en los datos.

Moda

La moda son los valores que aparecen con más frecuencia en los datos:

```
40, 21, 55, 21, 48, 13, 72
```

Aquí, 21 aparece dos veces y los otros valores solo una vez. La moda de estos datos es 21.

La moda también se usa para datos categóricos , a diferencia de la mediana y la media. Los datos categóricos no se pueden describir directamente con números, como los nombres:

```
Alice, John, Bob, Maria, John, Julia, Carol
```

Aquí, John aparece dos veces y los otros valores solo una vez. La moda de estos datos es John.

Nota: Puede haber más de una moda si varios valores aparecen la misma cantidad de veces en los datos.

Estadística

Estadísticas - Media

La media es un tipo de valor promedio, que describe dónde se encuentra el centro de los datos.

Media

La media generalmente se conoce como 'el promedio'.

La media es la suma de todos los valores de los datos dividida por el número total de valores de los datos.

La media se calcula para variables numéricas. Una variable es algo en los datos que puede variar, como:

- Años
- Altura
- Ingreso

Nota: Hay varios tipos de valores medios. El tipo más común de media es la media aritmética .

En este tutorial, 'media' se refiere a la media aritmética.

Cálculo de la media

Puede calcular la media tanto para la población como para la muestra .

Las fórmulas son las mismas y usan diferentes símbolos para referirse a la media de la población (μ) y media muestral (μ).

Cálculo de la media poblacional se hace con esta fórmula:

$$\mu = \frac{\sum x_i}{n}$$

Cálculo de la media muestral se hace con esta fórmula:

$$\bar{x} = \frac{\sum x_i}{n}$$

La parte inferior de la fracción (n) es el número total de observaciones.

∑ es el símbolo para sumar una lista de números.

X_i es la lista de valores en los datos:X1,X2,X3,…

La parte superior de la fracción (∑Xi) es la suma de X1,X2,X3,…agregado junto.

Estadística

Entonces, si una muestra tiene 4 observaciones con valores: 4, 11, 7, 14 el cálculo es:

$$\bar{x} = \frac{4+11+7+14}{4} = \frac{36}{4} = \underline{9}$$

Cálculo con Programación

La media se puede calcular fácilmente con muchos lenguajes de programación.

El uso de software y programación para calcular estadísticas es más común para conjuntos de datos más grandes, ya que calcular a mano se vuelve difícil.

Ejemplo

Con Python, use el **mean()** método de la biblioteca NumPy para encontrar la media de los valores 4,11,7,14:

```
import numpy
values = [4,11,7,14]
x = numpy.mean(values)
print(x)
```

Referencia de símbolos de estadísticas

Symbol	Description
μ	The population mean. Pronounced 'mu'.
$\bar{x}$	The sample mean. Pronounced 'x-bar'.
$\sum$	The summation operator, 'capital sigma'.
x	The variable 'x' we are calculating the average for.
i	The index 'i' of the variable 'x'. This identifies each observation for a variable.
n	The number of observations.

Estadística

Estadísticas - Mediana

La mediana es un tipo de valor promedio, que describe dónde se encuentra el centro de los datos.

Mediana

La mediana es el valor medio en un conjunto de datos ordenado de menor a mayor.

Encontrar la mediana

La mediana solo se puede calcular para variables numéricas.

La fórmula para encontrar el valor medio es:

$$\frac{n+1}{2}$$

Dónde (n) es el número total de observaciones.

Si el número total de observaciones es un número impar , la fórmula da un número entero y el valor de esta observación es la mediana.

```
13, 21, 21, 40, 48, 55, 72
```

Aquí, hay 7 observaciones en total, por lo que la mediana es el cuarto valor:

$$\frac{7+1}{2} = \frac{8}{2} = 4$$

El cuarto valor en la lista ordenada es 40 , por lo que es la mediana.

Si el número total de observaciones es un número par , la fórmula da un número decimal entre dos observaciones.

```
13, 21, 21, 40, 42, 48, 55, 72
```

Aquí, hay 8 observaciones en total, por lo que la mediana está entre los valores 4 y 5:

$$\frac{8+1}{2} = \frac{9}{2} = 4.5$$

Los valores cuarto y quinto en la lista ordenada son 40 y 42 , por lo que la mediana es la media de estos dos valores. Es decir, la suma de esos dos valores dividida por 2:

$$\frac{40+42}{2} = \frac{82}{2} = \underline{41}$$

Nota: Es importante que los números estén ordenados antes de poder encontrar la mediana.

Encontrar la mediana con programación

La mediana se puede encontrar fácilmente con muchos lenguajes de programación.

El uso de software y programación para calcular estadísticas es más común para conjuntos de datos más grandes, ya que encontrarlo manualmente se vuelve difícil.

Ejemplo

Con Python, use el método de la biblioteca NumPy median()para encontrar la mediana de los

valores 13, 21, 21, 40, 42, 48, 55, 72:

```
import numpy
values = [13,21,21,40,42,48,55,72]
x = numpy.median(values)
print(x)
```

Estadística

Estadísticas - Moda

La moda es un tipo de valor promedio, que describe dónde se encuentra la mayoría de los datos.

Moda

La moda son los valores que son los más comunes en los datos.

Un conjunto de datos puede tener múltiples valores que son modas.

Una distribución de valores con una sola moda se llama unimodal .

Una distribución de valores con dos modas se llama bimodal . En general, una distribución con más de una moda se denomina multimodal .

La moda se puede encontrar tanto para datos categóricos como numéricos.

Encontrar el modo

Aquí hay un ejemplo numérico :

```
4, 7, 3, 8, 11, 7, 10, 19, 6, 9, 12, 12
```

Tanto el 7 como el 12 aparecen dos veces cada uno, y los otros valores solo una vez. Las modas de estos datos son 7 y 12.

Aquí hay un ejemplo categórico con nombres:

```
Alice, John, Bob, Maria, John, Julia, Carol
```

John aparece dos veces, y los otros valores solo una vez. La moda de estos datos es John.

Encontrar el modo con programación

El modo se puede encontrar fácilmente con muchos lenguajes de programación.

El uso de software y programación para calcular estadísticas es más común para conjuntos de datos más grandes, ya que el cálculo manual se vuelve difícil.

Ejemplo

Con Python, use el multimode()método de la biblioteca de estadísticas para encontrar las modas de los valores 4,7,3,8,11,7,10,19,6,9,12,12:

```
from statistics import multimode
values = [4,7,3,8,11,7,10,19,6,9,12,12]
x = multimode(values)
print(x)
```

Estadística

Estadísticas - Variación

La variación es una medida de cuán dispersos están los datos alrededor del centro de los datos.

La variación de los datos

Las medidas de variación son estadísticas de qué tan lejos están los valores en las observaciones (puntos de datos) entre sí.

Hay diferentes medidas de variación. Los más utilizados son:

- Rango
- Cuartiles y percentiles
- Rango intercuartil
- Desviación Estándar

Las medidas de variación combinadas con un promedio (medida del centro) dan una buena imagen de la distribución de los datos.

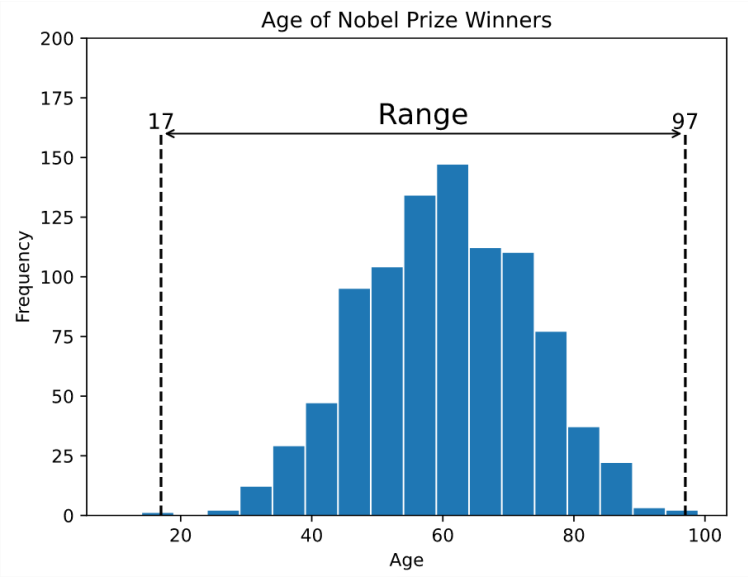
Nota: estas medidas de variación solo se pueden calcular para datos numéricos.

Rango

El rango es la diferencia entre el valor más pequeño y el más grande de los datos.

El rango es la medida más simple de variación.

Aquí hay un histograma de la edad de los 934 ganadores del Premio Nobel hasta el año 2020, que muestra el rango :



El ganador más joven tenía 17 años y el mayor 97 años. El rango de edades para los ganadores del Premio Nobel es entonces de 80 años.

Estadística

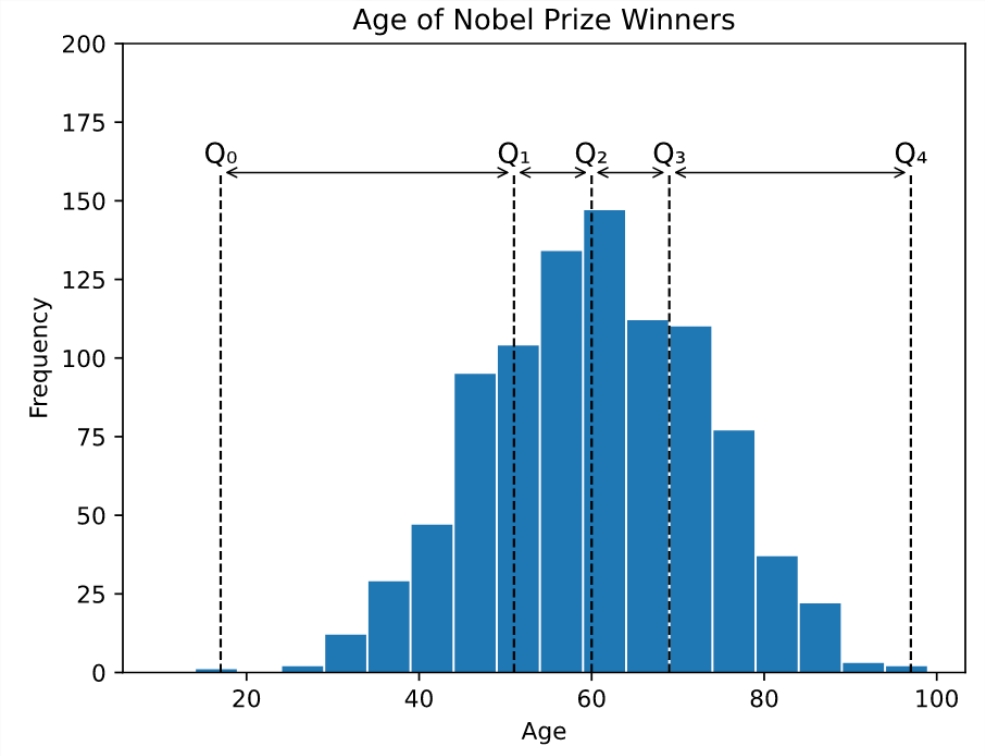
Cuartiles y percentiles

Los cuartiles y percentiles son formas de separar números iguales de valores en los datos en partes.

Los cuartiles son valores que separan los datos en cuatro partes iguales.

Los percentiles son valores que separan los datos en 100 partes iguales.

Aquí hay un histograma de la edad de los 934 ganadores del Premio Nobel hasta el año 2020, que muestra los cuartiles :



Los cuartiles (Q_0 ,Q_1 ,Q_2 ,Q_3 ,Q_4) son los valores que separan cada trimestre.

Entre Q_0 y Q_1 están los valores 25% más bajos de los datos. Entre Q_1 y Q_2 están el siguiente 25%. Y así.

- Q_0 es el valor más pequeño de los datos.
- Q_2 es el valor medio (mediana).
- Q_4 es el valor más grande en los datos.

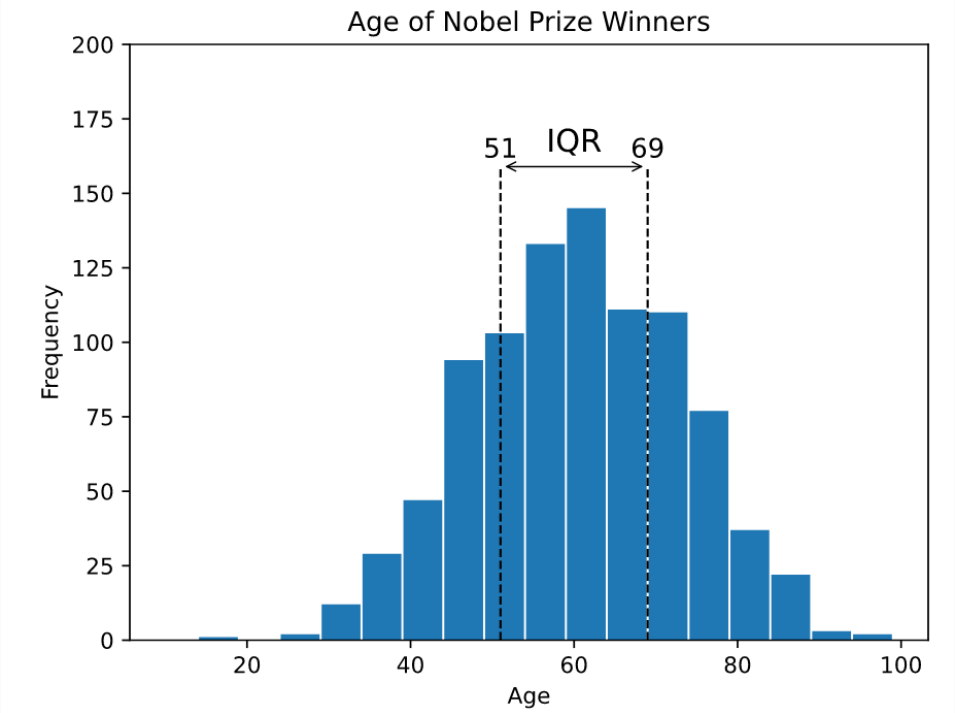
Estadística

Rango intercuartil

El rango intercuartílico es la diferencia entre el primer y el tercer cuartil (Q_1 y Q_3).

La 'mitad media' de los datos está entre el primer y el tercer cuartil.

Aquí hay un histograma de la edad de los 934 ganadores del Premio Nobel hasta el año 2020, que muestra el rango intercuartílico (RIC) :



Aquí, la mitad media tiene entre 51 y 69 años. El rango intercuartílico para los ganadores del Premio Nobel es entonces de 18 años.

Estadística

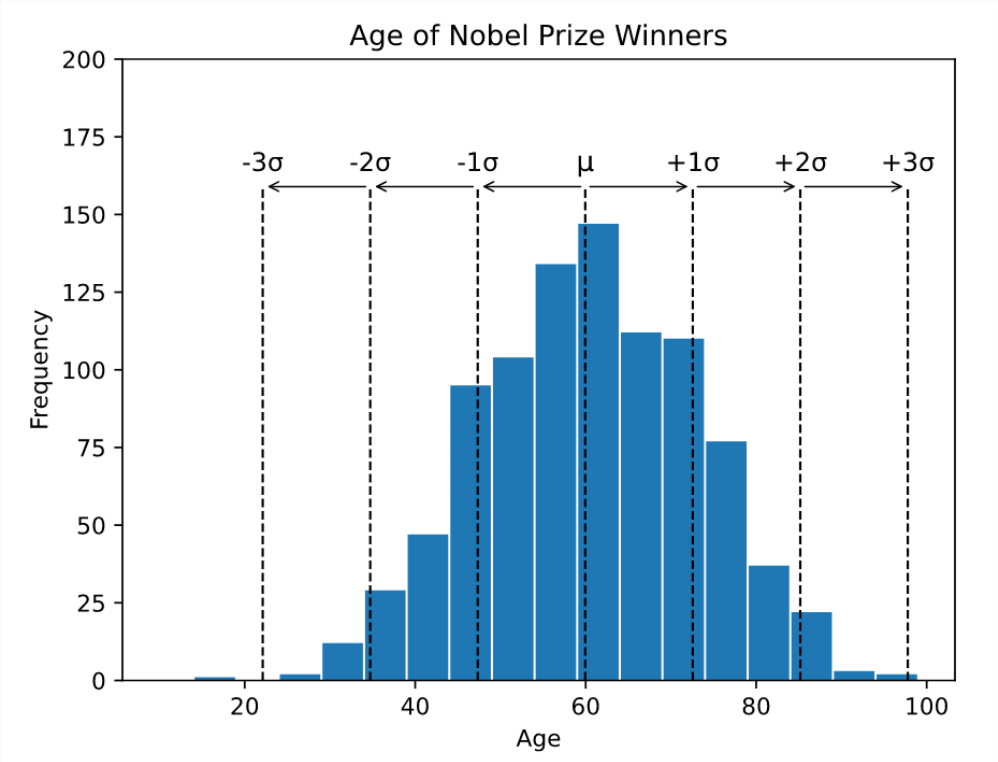
Desviación Estándar

La desviación estándar es la medida de variación más utilizada.

La desviación estándar (σ) mide qué tan lejos está una observación 'típica' del promedio de los datos (μ).

La desviación estándar es importante para muchos métodos estadísticos.

Aquí hay un histograma de la edad de los 934 ganadores del Premio Nobel hasta el año 2020, que muestra las desviaciones estándar :



Nota: Los valores dentro de una desviación estándar (σ) se consideran típicos.

Los valores fuera de tres desviaciones estándar se consideran valores atípicos