

1 Tratamiento de los datos

La estadística descriptiva tendría por objeto la ordenación, clasificación y resumen de dichos datos, con objeto de tener una visión más precisa y conjunta de las observaciones, intentando descubrir de esta manera posibles relaciones entre aquellos, viendo cuáles toman valores parecidos y cuáles difieren en gran medida del resto.

Cuando los datos los tomamos para toda la población, sea o no de personas, nos estamos refiriendo a un censo. Si no referimos a una parte de la población y los datos que se toman del estudio se refieren a un número más o menos reducido de individuos de esa población, hablamos de muestra.

La frecuencia absoluta es precisamente el número de veces que se repite un determinado dato o suceso.

La frecuencia relativa es el cociente entre la frecuencia de cada uno de los sucesos o datos y el número total de elementos observados.

Además de estos tipos de frecuencias existen otros que recogen la suma de frecuencias, frecuencia relativa acumulada y frecuencia absoluta acumulada.

A) Tipos de datos

Dentro de este apartado encontramos los diferentes tipos de preguntas como por ejemplo respuestas de sí o no donde una excluye a la otra, o preguntas con respuestas múltiples. Preguntas abiertas que resultan difíciles de clasificar entre otras.

B) Variables estadísticas

Dentro de ellas hay que distinguir las variables discretas, que son aquellas que asignan un valor entero que no se puede fraccionar, y por ello, en una escala correspondería a un número determinado.

Las variables continuas son aquellas medibles, pero que sí pueden ser fraccionadas en partes, dependiendo del patrón de medida que utilizemos. Es por ello que utilizamos distintas escalas de medida:

- Escala nominal : es aquella en la que los números sirven sólo como etiquetas para identificar los elementos que quieren medir.
- Escala ordinal : lo único que registra es si un elemento tiene más o menos cantidad de aquella característica que mide.
- Escala de intervalo : aquí existe un patrón de medida y una unidad de medida.
- Escala de razón : es similar a la anterior y lo único que la diferencia es que en este caso el origen es natural y no arbitrario como en el anterior.

C) Representación gráfica

Si hablamos de datos cuantitativos utilizamos:

- Histograma : sirve para representar las frecuencias agrupadas en intervalos
- Curva o diagrama de barras acumulativo : Se utiliza para representar frecuencias acumuladas
- Diagrama de barras : sirve para representar frecuencias de forma aislada y sin intervalos
- Diagrama de escalera : sirve para representar frecuencias aisladas pero en forma acumulada.
- Gráfico de dispersión : se limita a representar una nube de puntos de todos los datos obtenidos y en ocasiones sin agrupar
- Diagrama de líneas : es similar al anterior y sirve para representar tendencias a lo largo del tiempo.
- Polígono de frecuencias : se utiliza para variables discretas
- Polígono de frecuencias acumuladas : se utiliza normalmente en variables continuas.
- Áreas de 2 o 3 dimensiones : se utilizan para comparar series temporales diversas
- Diagrama de sectores o circular : sirve para representar en forma de quesitos el

porcentaje o proporción de datos que cumplen una determinada condición.

- Diagrama de barras :ya analizado anteriormente, en este caso sirve para representar variables nominales y ordinales.
- Diagrama de anillos : es similar al de sectores pero con forma de anillo.
- Pictograma :es muy similar al diagrama de barras,pero en lugar de utilizar las barras utiliza dibujos
- Mapa o cartograma :sirve para representar en mapas los datos que requerimos.

2 Análisis e interpretación de los datos

A)Medidas de posición

1. Media aritmética: es el sumatorio de todos los datos observados,dividido entre el número total de observaciones realizadas

2. Mediana: es el valor que queda en el centro cuando utilizamos un conjunto de datos tomados en forma ordenada.

3. Moda: es el valor que más se repite dentro de una distribución de datos.

4. Centiles: son las últimas medidas de posición que vamos a ver,es aquel valor de la variable que divide en partes la distribución de frecuencias , previamente ordenada de forma creciente

5. Cuartiles: la mediana, como hemos visto, divide el conjunto ordenado de observaciones en dos mitades iguales, si dividimos cada mitad de la mediana en otras dos partes iguales, a cada una de esas partes se le conoce como cuartil.

B)Medidas de dispersión

1. Rango o recorrido :es la diferencia entre el valor máximo y mínimo dentro de una dispersión.

2. Varianza : mide la dispersión de los datos,pero a partir de la media y no de la mediana,como en el caso del recorrido intercuadrático.

3. Desviación típica : la varianza tiene el problema de estar expresada en unidades al cuadrado, lo que puede producir una falsa imagen de la dispersión,por ellos se suele utilizar la raíz cuadrada positiva de la misma,es decir, la desviación típica.

3 Probabilidad

La probabilidad estudia este tipo de fenómenos aleatorios.En este sentido, la probabilidad sirve para estudiar ciertos comportamientos que se estructuran constituyendo modelos de comportamiento.

A) Sucesos y experimentos aleatorios

Un suceso es el resultado de realizar un experimento,experiencia o prueba.

Un suceso es aleatorio si los resultados del mismo son conocidos con anterioridad, no es predecible el resultado del experimento o se puede repetir tantas veces se quiera.Otro término que aparece es el espacio muestral, que es el conjunto de todos los posibles sucesos o resultados que se pueden dar a la hora de realizar un experimento.

De esta manera,podemos decir que suceso es cada una de las partes del espacio muestral.Llamamos suceso seguro al que coincide con el espacio muestral y que , por tanto, aparece siempre que se realice la experiencia. Por lo contrario , el suceso imposible es aquel que jamás se realiza y que se representa como \emptyset .

1. Unión de sucesos: Es el conjunto de dos sucesos

2. Intersección de sucesos: Es aquel que se cumple en ambos a la vez.

3. Sucesos incompatibles: Son incompatibles si su intersección es el conjunto imposible.

4. Sucesos contrarios: Son sucesos contrarios si son incompatibles y además la unión de ambos ocupa todo el espacio muestral.

B) Frecuencia y probabilidad

Si repetimos un experimento aleatorio en las mismas condiciones y anotamos sus frecuencias relativas, se observa que dichas frecuencias tienden a estabilizarse en un número que se encuentra entre cero y uno.

Regla de Laplace: Dado un conjunto de sucesos equiprobables entre ellos ,es decir, que

tienen todos ellos la misma probabilidad de salir.

C) Probabilidad de sucesos condicionados y dependencia de sucesos

La probabilidad condicionada es cuando tenemos varios factores, una vez tenemos determinado un factor además tendría que estar determinado por otro.

D) Regla de Bayes

4 Inferencia estadística

A) Introducción

Cuando pretendemos sacar conclusiones sobre el comportamiento de una población bien sea de personas, animales u objetos, lo hacemos usando la estadística descriptiva, a través de una serie de procedimientos y técnicas, que permiten un conocimiento descriptivo de las características básicas.

B) Concepto de inferencia

La inferencia estadística es un proceso a través del cual podemos extraer una serie de conclusiones sobre la población a partir de una muestra dada.

C) Muestra representativa

Un problema que se plantea a la hora de extrapolar los datos de la muestra al conjunto de la población es el de saber si realmente la muestra representa o no a la población objeto de estudio.

D) El tamaño de la muestra

Lo siguiente que tenemos que hacer es elegir un conjunto de unidades muestrales para poder realizar el estudio. Por este motivo a la hora de elegir el tamaño de muestra adecuado hay que tener en cuenta, entre todos los factores:

- El coste económico : si el coste económico de realizar un encuesta no es muy elevado entonces la podremos hacer
- La dispersión de la población : cuanto mayor sea la dispersión de la población, mayor deberá ser el tamaño de la muestra.
- El número de grupos y subgrupos poblacionales : a mayor número de grupos, mayor deberá ser el tamaño de la muestra.
- El margen de error permitido en el estudio : a menor margen de error permitido, mayor deberá ser el tamaño de la muestra.

1. La distribución normal: nos estamos refiriendo al conjunto de datos posibles entre los que se mueve una población determinada.

2. Estimadores: cuando trabajamos con medidas de posición o dispersión como la media, si estos describen las características o atributos de una población lo llamamos parámetros, pero por el contrario nos referimos a características de muestra los llamamos estadísticos.

3. Error estándar o típico de la estimación: si tomamos varias muestras del mismo tamaño al azar y pretendemos calcular con ellas la media, lo que podemos observar es que la media en cada una de ellas es diferente, y ello, sobre todo, cuanto más dispersión haya dentro de esa población, es decir, cuanto mayor sea la varianza y la desviación típica.

4. Intervalo de confianza: representa un espacio estadístico de valores dentro del cual se espera que se encuentre el verdadero valor del parámetro de la población.

5. Poblaciones finitas: antes de determinar el tamaño de la muestra, diremos que hasta ahora el tratamiento de datos realizado sirve para poblaciones grandes, que llamamos, por este motivo poblaciones infinitas.

6. Determinación del tamaño de la muestra: Para determinar el tamaño de la muestra, como hemos visto, se necesita conocer el error de la muestra que estamos dispuestos a soportar y el nivel de confianza que pretendemos conseguir.

7. Afijación: sería el procedimiento a través del cual una vez determinado el tamaño de la población, éste se distribuye entre los diferentes estratos

- Afijación simple : consiste en dividir el tamaño de la muestra entre el número de estratos y obtener de este modo un tamaño igual para cada uno de ellos.

- Afijación proporcional: aquí si se tiene en cuenta la proporción de elementos que hay en cada estrato.
- Afijación óptima :este método afina aún más que los anteriores y tiene en cuenta que los distintos estratos deben tener la misma representatividad y variación