

1. CONTRASTE DE HIPÓTESIS

El uso de muestras tiene el problema de saber si realmente se pueden generalizar los resultados a toda la población simplemente con el uso de las mismas o si, por el contrario, la generalización implica un problema añadido.

En el campo de la investigación comercial tenemos la necesidad de contrastar que la hipótesis de partida con la que contamos se cumple y por tanto puede ser generalizada al resto del universo objeto de estudio.

El contraste de hipótesis tiene por objeto analizar si los resultados obtenidos en un muestreo pueden ser generalizados o no a la población, o si por el contrario son fruto del azar.

Las hipótesis serían en este caso, el punto de partida de cualquier contraste. Hay dos grupos de hipótesis estadísticas:

- Aquellas referidas a un valor concreto o a un intervalo en distribuciones de una variable.
- Aquellas que establecen la igualdad de algún parámetro pero en distribuciones de variables de dos o más poblaciones.

A. Tipos de hipótesis .

Para realizar un contraste de hipótesis lo primero que tenemos que hacer es formular la llamada hipótesis nula, que es la hipótesis que deseamos contrastar. Al trabajar con muestras y no con toda la población, dicha hipótesis nunca podrá ser probada, pero sí que podrá ser rechazada.

Para poder formular una hipótesis nula es necesario tomar un punto de referencia y a partir de él, formular dichas hipótesis.

También debemos señalar que la hipótesis nula hace referencia a que las observaciones realizadas son fruto de la pura casualidad o azar.

Lo siguiente que se necesita para realizar un contraste de hipótesis es formular las hipótesis alternativas. Esta sería la negación de la nula en dos sentidos:

- Por la diferencia con ella cuando desconocemos en qué dirección puede ser falsa, la hipótesis nula. Aquí hablaríamos de un contraste bilateral.
- Por contraste unilateral: cuando sabemos en qué dirección la hipótesis nula es falsa y así podemos afirmar que un producto se vende en mayor proporción a los hombres que a las mujeres, y la hipótesis nula que se vende en la misma proporción.

La hipótesis alternativa hace referencia a la existencia de un efecto real y a que las observaciones estadísticas son precisamente el resultado de este efecto real, y no meramente efecto del azar o la casualidad, como en el caso de la hipótesis nula.

Actualmente el contraste de hipótesis se realiza a través de programas informáticos como pueden ser el SPSS o el SAS, que permiten desarrollar contrastes tanto para pequeñas como para grandes empresas en cualquier tipo de investigación. Rechazar la hipótesis nula supone aceptar la hipótesis alternativa que representa la condición deseada.

B. Obtención de la distribución muestral.

La distribución muestral va a ser la base sobre la que se asienta el verdadero valor de la distribución de la población.

En este capítulo tomaremos la distribución normal y estudiaremos, a partir de ella, el valor por debajo del cual rechazaríamos la hipótesis nula y por tanto aceptaríamos la alternativa.

C. Nivel de significación .

Cuando utilizamos una muestra en lugar de una población, el valor del parámetro, que se corresponde con la población, discrepa con respecto al valor del estadístico de la muestra. Así a la hora de realizar cualquier contraste de hipótesis se produce una diferencia entre el valor del parámetro dado por la hipótesis nula y el valor del estimador calculado en la muestra. Esto es lo que llamamos error relativo de una estimación, que se

produce para cualquier estimador con el que trabajemos, bien sea la media o la proporción. Hay dos tipos de errores:

- Error de tipo I: se produce cuando se rechaza la hipótesis nula siendo cierta.
- Error de tipo II: se produce cuando se acepta la hipótesis nula siendo falsa.

El nivel de significación es el porcentaje de error que estamos dispuestos a soportar al realizar un contraste de hipótesis.

Habitualmente los dos niveles de significación más utilizados en la estadística clásica son 0,05 y 0,01, es decir un 5% y 1%.

D. Región crítica

La región crítica es la zona de la curva de distribución en la que se rechaza la hipótesis nula y en la que está contenido el error. También se llama región de rechazo.

Para poder establecer la región crítica utilizamos cuadros de distribuciones como el de la distribución normal y el de la t de Student. En dichos cuadros observamos el valor crítico, que es el valor de la distribución, que también podemos observar en las tablas de distribución, y que se corresponde con el nivel de significación.

E. Elección de un estadístico para realizar el contraste y toma de decisiones en una población.

- Contraste para una proporción:
 - A la hora de realizar cualquier contraste lo primero que tenemos que hacer es definir la hipótesis nula. Consistirá en ver si la discrepancia que se produce entre la proporción observada y la estimada teórica, está o no por debajo del valor crítico de la distribución para aceptar o no dicha hipótesis nula.
- El contraste para una media:
 - El planteamiento es el mismo que el de la proporción pero aplicado al caso de la media.
 - La toma de decisiones consiste en calcular el valor de la discrepancia y compararlo con el valor crítico de las tablas. Si el valor de la discrepancia es mayor al de las tablas caería dentro de la región crítica y por tanto se tomaría la decisión de rechazar la hipótesis nula, si es inferior caería fuera de la región crítica y por ello no podríamos rechazar las hipótesis nula, aunque tampoco la podríamos aceptar.

F. Elección de un estadístico para realizar el contraste y toma de decisiones en dos poblaciones.

El contraste serviría para comprar una población en dos momentos distintos del tiempo o para comparar dos poblaciones diferentes.

Aceptar la hipótesis nula supondría, en este caso, que los cambios son debidos al azar y rechaza que realmente se producen dichos cambios de forma significativa como consecuencia del cambio de coyuntura.

- Contraste para comparar proporciones.
 - Lo que medimos con este tipo de contraste es la diferencia en términos relativos que se produce entre las dos muestras para de este modo, poder realizar comparaciones entre ellas.
- Contraste para comparar medias.
 - Lo realizaríamos cuando pretendemos ver si las medias obtenidas en dos muestras distintas hacen referencia a cambios significativos en la población.

2. Estudio conjunto de dos variables.

Ahora vamos a analizar si hay relaciones entre dos variables como pueden ser la edad y el nivel de consumo, el nivel de renta y el gasto en un determinado tipo de productos, la zona de residencia y los puntos de venta donde se compran los productos, el sexo y el tipo de productos consumidos, la religión y cierto tipo de consumos. A simple vista, nosotros, por observación, podemos suponer que una familia va a consumir en función de su renta e ingresos, del número de miembros.

Luego, podemos imaginar que existe cierta relación entre el tipo de consumidor que analizamos y lo que consume, aunque esto no sea necesariamente igual en todos ellos.

A. Tablas de contingencia .

La tabla de contingencia es una tabla donde los individuos de una muestra o población se clasifican en función de diversas variables.

Podemos así clasificar a los individuos en función de dos variables.

B. Contraste de independencia entre dos variables .

Se trata de ver si existe o no dependencia entre las variables observadas. Es decir, si consideramos los resultados de la tabla de estado civil y preferencias por el turismo interior, de playa o exterior, lo que intentaríamos es observar si realmente se produce una relación de dependencia o no entre dichas variables. Para ello determinaríamos los grados de libertad.

Un problema que presenta dicho contraste de independencia es que no permite calcular qué ocurriría si aumentase o disminuyese el tamaño de las muestras.

Para ello lo que se utiliza es el llamado coeficiente de asociación que consiste en dividir la distribución Chi cuadrado de Pearson entre la frecuencia total de población, es decir, el tamaño de la muestra local.

C. Regresión .

La regresión es un método que intenta estudiar la relación entre dos ó más variables, una dependiente y otra u otras no dependientes. También se le conoce como regresión lineal.

Hay dos tipos de regresión lineal:

- Regresión lineal simple: método de regresión en el que se tiene en cuenta solamente una variable dependiente y una sola variable independiente.
- Regresión lineal múltiple: método que contempla la relación entre una variable dependiente de varias independientes.

A la variable dependiente se le llama regresora o predictora y a la dependiente, respuesta. La recta de regresión se puede definir como aquella recta que describe como varía la media de la variable dependiente en función de la independiente.

- La covarianza
 - Es una medida de la asociación lineal que se produce entre dos variables y que sirve para sintetizar la información que se produce en un gráfico de dispersión. Sólo sirve para estudios de regresión lineal y con asociación, por lo que no es un método válido en el caso de regresión lineal o sin asociación.
- El coeficiente de correlación
 - Dicho coeficiente tiene una serie de propiedades que vamos a analizar:
 - Cuando el coeficiente de correlación es 1 se produce un ajuste perfecto.
 - El coeficiente no depende del orden en que se tomen las variables.
 - Si el coeficiente de correlación es cero, las variables no están relacionadas entre sí y la covarianza es nula.