... Estadística de las ciencias humanas, guión número 5, autor Alonso Lozano, fecha de emisión 8 de abril de 1980. Regresión.

más fundada y cercana a la realidad que nos ofrece el mundo de cada día en las ciencias humanas. Se cree que las primeras aplicaciones del método proceden del matemático francés Lagendre, que efectuó sus predicciones por los años de la Revolución Francesa. Lagendre, ya en esta época, hacía un cálculo sobre las órbitas de los cometas y ya para su estudio utilizaba, como muchos de nosotros hoy, el método mínimos cuadrados. En el presente siglo ha sido aplicado el método a las ciencias sociales. Por él trataba de determinar la órbita de los cometas a partir de un conjunto de datos observados. Aquí la posición de la órbita aparece asociada a una función en que entran a jugar ciertas variables medibles y parámetros desconocidos de difícil obtención, cuya solución fue aportada por este método. A partir de la época de la Revolución Francesa, la teoría de la órbita de los cometas se ha convertido en una teoría de la vida. En 1830 se inicia la aplicación en gran escala del método.

fundamentalmente en el campo de la genética y por obra de Galton. ¿Qué podemos hacer con este nuevo método en el campo de las ciencias humanas? Los fines que podemos trazarnos al usar el método de la regresión en cualquiera de sus formas son, primero, lograr una explicación fundada y causal de una variable o alcanzar una representación lo mejor posible en función de otra u otras, lo que es tanto como analizar la influencia que sobre una variable ejercen otra o varias, elaborando teorías sobre esas relaciones que intentaremos probar precisamente utilizando la regresión. Segundo, estimar o predecir una variable a partir de otra relacionada con ella. En el caso de la regresión lineal, relacionada linealmente. Claro que en ambos casos todo depende de los supuestos de que para el caso de la regresión lineal,

repartimos o de los datos que tenemos. Pero, ¿qué quiere decir el término regresión? Desde que Galton lo utilizó por vez primera, investigando las características que relacionaban a padres altos y bajos con sus hijos y observó en cada generación su tendencia a la media propia, mayor en los hijos, hemos venido utilizando el término como sinónimo de predicción, pronóstico o estimación. Cuando hagamos una predicción, la haremos en función de una fórmula. Vamos a ver ahora la regresión simple. Esta supone una relación unidimensional, pero puede adoptar diversas formas según una línea recta y entonces usaríamos la fórmula. Y igual a A más BX. Dentro de esto debemos aclarar que la regresión se nos puede predicar por una relación de una forma. La regresión simple es la que se puede presentar bajo dos aspectos. Uno, como regresión teórica, cuya misión es...

formular hipótesis. En las fórmulas y igual a a más bx o x igual a a más bi, partimos de la hipótesis de relación lineal entre las variables x e y. Dos, en la regresión empírica o muestral, trataremos de verificar la hipótesis que queremos estudiar con la expresión que hemos propuesto, y la tarea consiste en predecir los parámetros desconocidos a y b que en ella figuran. Claro que los parámetros se refieren a la población total con la que no es posible trabajar por dificultades obvias. Entonces, ¿qué garantía tienen nuestras predicciones? Bien, nunca estaremos completamente seguros. Para subsanar esto en lo posible, es para lo que aparecerán más tarde, conceptos tales como nivel de confianza, intervalo confidencial y otros.

en la estimación de parámetros, y el coeficiente k de alienación o inseguridad para la correlación. Por dificultades de cantidad, procuramos escoger un grupo que represente a la población, y con él trabajamos sacando conclusiones para el total. Este grupo es la muestra que usamos en inferencia. ¿Podría concretarnos mejor la regresión como realidad predictiva? Claro. Todas las ciencias fijan como parte fundamental la predicción. Para nosotros, en las ciencias humanas, es hasta un arte. Un arte que nos ofrece la oportunidad de, analizando la regresión, alcanzar el objetivo de descubrir la relación y dependencia causal entre variables humanas o de conducta, a la vez que puede también hacernos predecir o estimar una variable dependiente en función de otra independiente. Llámese L, X, Y, S, V. En donde, dado cualquier valor en una variable, se puede ver que la relación entre variables humanas y dependencia causal es de una manera muy diferente. En donde, dado cualquier valor en una variable, se puede ver que la relación entre variables humanas y dependencia causal es de una manera muy diferente. En donde, dado cualquier valor en una variable, se puede ver que la relación entre variables humanas y dependencia causal es de una manera muy diferente.

puede obtener el correspondiente en otra. A cualquier predicción en X o Y, por ejemplo, se le suele poner para distinguirlas como exponente una comita o asterisco, que es precisamente el signo de la predicción que identifica a esa variable. Creo haber observado que a veces los valores pronosticados no coinciden exactamente con los empíricos. ¿A qué es debido? Varia son las causas, sobre todo errores de estimación, pues hay siempre imponderables. Siempre se cumple que Y es igual a Y pronosticada más un error. Si antes hubiéramos partido de obtener la varianza de Y, que vamos a llamar total, ahora veríamos que es igual a la varianza de las puntuaciones pronosticadas, en este caso en Y, dividida por la total de Y. Y siempre es igual al coeficiente de determinación.

Y este no es más que el cuadrado de Pearson. La predicción será más fiable cuanto menor sea el valor de la varianza del error o varianza residual, que hará que el coeficiente de determinación se aproxime a la unidad. Claro, si la correlación es perfecta, o sea, vale 1, da lo mismo pronosticar en X que hacerlo en Y, porque los valores serán coincidentes. Las dos rectas se cortan perpendicularmente cuando R igual a cero. De ahí que el ángulo de incidencia entre ambas rectas, de X o de Y, sea otra forma de indicarnos la intensidad de la relación. Veamos algo sobre el método de mínimos cuadrados. El fundamento lógico, quizá podríamos hallarlo en Wolf, quien en su obra Análisis de la Recta,

demanda que publica el Consejo de Investigaciones Científicas en 1956, página 36, nos dice El principio matemático de los mínimos cuadrados trata como objetivo lograr una representación gráfica aproximada de una cierta función algebraica. Se basa tal principio en algo convencional consistente en admitir que la precisión lograda con esta aproximación será tanto mayor cuanto menor sea la suma de los cuadrados de las desviaciones entre los valores de la función matemática aproximar y un determinado conjunto de observaciones o puntos muestrales del plano X y Y. ¿Los resultados de este método cree usted que son tan interesantes como para aconsejarlo? Bueno, en el campo de las ciencias experimentales los resultados han sido de gran éxito, pero una vez más recomendamos la prudencia al que se desempeñe. ¿Qué se envuelve en el campo de las ciencias humanas? Pues en las ciencias de la naturaleza.

eran experimentos controlados. Aquí trabajamos con datos no experimentales y los resultados no tienen que ser necesariamente iguales. ¿Por qué nos referimos a la recta? Hay formas, como parábolas y otras, pero la aplicación que se impone es la de relación lineal. La ecuación de la recta hace mínima la suma de las diferencias cuadráticas entre la ordenada de cada puntuación y la línea de regresión. Hay una distribución bivariada muy importante en estadística, la normal. En ella se cumple que las medias de las distribuciones de y para todos los valores posibles de x están sobre una línea recta, la recta de regresión. A la vez, las distribuciones de y sobre x son normales, con media sobre la recta de regresión y varianza constante. O sea, la desviación típica de todas las distribuciones de y es la misma. En esto consiste la homosedasticidad, en que todas las distribuciones de y

tienen la misma desviación típica. Por favor, ¿podría aclararnos un poco más el método de mínimos cuadrados? Pues el ajuste por mínimos cuadrados tiene como objetivo proporcionarnos los valores numéricos correspondientes a los parámetros a y b, procurando que el sumatorio de errores al cuadrado sea mínimo. Conocemos que e sub 1, e sub 2 son errores aleatorios. La condición exigida es que esos errores al cuadrado sean mínimos. Aplicando el método a la recta, que es la más usual en nuestras ciencias, tenemos sumatorio de y sub i igual n a, donde n es el número de observaciones o sujetos, más b por sumatorio de x sub i. De aquí deducimos un sistema de dos ecuaciones con dos incógnitas, cuya solución nos dará los valores de los parámetros buscados. ¡Gracias por ver el video!

¿Hasta qué punto podemos decir que presentan analogías y diferencias, correlación y regresión? Tienen aspectos diferentes. 1. Ambas se apoyan en distintos supuestos acerca de los valores en X e Y. Con la regresión elaboramos la ley que explica la variación de una variable dependiente en función de una o varias independientes, mientras la correlación nos da la medida de la asociación entre dos variables aleatorias sin más. 2. El coeficiente de correlación no está definido por variables matemáticas, mientras la regresión lo está por variables matemáticas y por variables aleatorias. 3. Mientras la dependencia o asociación entre dos variables aleatorias nos lo da la correlación, la regresión es una función que selecciona dos o más variables. 4. En las características de la relación entre dos variables aleatorias, las características comunes podemos señalar. 1. Para variables lineales y funciones correspondientes,

el coeficiente de regresión tiene el mismo signo que el de correlación. 2. Ambas analizan conjuntamente dos variables. La correlación viendo el grado de dependencia. La regresión viendo si hay relación entre las mismas. 3. A partir de los parámetros de la ecuación de regresión, que no son sino los coeficientes de regresión, puede ser deducido el coeficiente de correlación. ¿Para qué utilizamos la regresión en las ciencias humanas? En primer lugar, como en la vida moderna el investigador se ve un poco agobiado por la falta de tiempo, la utilización, por ejemplo, de las rectas de regresión, permite poder aplicarlas a otras personas que sean semejantes a aquellas con las que las hemos construido, si son ambos grupos pertenecientes a una misma población. A veces, en investigación, hay confusión en la elección entre correlación y regresión, lo que se evita cuando,

han sido planificados y señalados de modo preciso y claro los objetivos que pretendemos. Desde luego, si el problema que se nos plantea es conocer el grado de asociación entre dos o más variables, lo más apropiado y oportuno es el uso de la correlación, sea cual sea la covariación. Si el objetivo es predecir una variable, ahorrándonos, por ejemplo, una

segunda aplicación de un test, o tenemos que justificar y explicar relaciones de causalidad, lo correcto sería el uso de la regresión, si hay dependencia causal unilateral. En esta emisión hemos pretendido orientar un poco. Si desean completar los contenidos tratados, les aconsejaría la quinta unidad didáctica de estadística Ciencias Humanas, los apuntes de Mariano Yela, de la Universidad Complutense, estadística para psicólogos de J. Amón,

...y de la Universidad Complutense.