This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.
BY: Grupo CDPYE-UGR
SOME RIGHTS RESERVED

Tema 8

Esperanza condicionada. Regresión y correlación

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.

BY: **Grupo CDPYE-UGR**

SOME RIGHTS RESERVED

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.
BY: Grupo CDPYE-UGR
SOME RIGHTS RESERVED

1.- Introducción

Cuando se considera un conjunto de variables aleatorias definidas en relación a un determinado experimento, es usual que existan relaciones entre ellas; en tal caso, el conocimiento del valor tomado por cualquiera de ellas puede proporcionar información sobre el resto, lo que conduce a plantear el problema de regresión. En líneas generales, el problema de regresión consiste en encontrar una función que permita aproximar el valor de determinadas variables, conocidos los valores del resto. Aquí nos limitamos a analizar el problema de regresión bidimensional, en el que se trata de encontrar una función que, aplicada a los valores de una variable aleatoria, proporcione aproximaciones óptimas de los correspondientes valores de otra. Trataremos el denominado problema de regresión mínimo cuadrática que usa como criterio de optimalidad el principio de mínimos cuadrados. La solución a este problema está dada en términos de esperanzas condicionadas y, por este motivo, antes de abordar el problema de regresión, debe estudiarse este concepto.

2.- Esperanza condicionada a una variable aleatoria

2.1.- Definición

Dadas dos variables aleatorias definidas sobre un mismo espacio de probabilidad, la esperanza de una de ellas condicionada a la otra es una variable aleatoria, función de

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.
BY: Grupo CDPYE-UGR

la variable a la que se condiciona, cuyos valores son las esperanzas de las correspondientes distribuciones condicionadas. La existencia de la esperanza condicionada de una variable a otra está asegurada siempre que exista la esperanza marginal. Además, la definición implica directamente que el cálculo de esperanzas condicionadas es similar al de esperanzas, considerando distribuciones condicionadas en lugar de marginales, y, por tanto, las propiedades son similares a las de la esperanza. En particular, la esperanza de una función de una variable aleatoria condicionada a otra se obtiene en términos de la distribución condicionada de la primera variable.

Esperanza condicionada de una variable aleatoria a otra

Esperanza condicionada de una función de una variable aleatoria

2.2.- Propiedades de la esperanza condicionada

Además de las propiedades elementales que se derivan de las correspondientes de la esperanza, destacamos una propiedad importante, de gran utilidad en el estudio de la regresión: la esperanza de la esperanza condicionada de una variable a otra, cuya existencia está asegurada si existe la esperanza marginal de la primera, coincide con ésta.

Propiedades

2.3.- Extensión a vectores aleatorios

La definición de esperanza condicionada se extiende de manera obvia al caso en que se

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.
BY: Grupo CDPYE-UGR

condiciona a un vector aleatorio y, también, en el mismo sentido que la extensión de la esperanza de variables aleatorias a vectores, la extensión es obvia cuando se define la esperanza de un vector condicionado a otro.

Esperanza condicionada de un vector aleatorio a otro

3.- Momentos condicionados

3.1.- Momentos centrados y no centrados

Los momentos condicionados se definen de manera similar a los momentos sin condicionar, pero ahora en términos de las distribuciones condicionadas y con la particularidad de que los momentos condicionados centrados lo son respecto de la esperanza condicionada.

Momentos condicionados

3.2.- Varianza condicionada. Propiedades

Es el momento condicionado centrado de orden dos, destacable por su importancia en el estudio del problema de regresión mínimo cuadrática.

Definición y propiedades

4.- Regresión y correlación bidimensional

Como ya se ha indicado, el problema de regresión consiste en determinar una función,

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.
BY: Grupo CDPYE-UGR

denominada función de regresión, que permita aproximar de forma óptima (en algún sentido) el valor de una variable aleatoria a partir del valor que tome otra variable. Íntimamente ligado al problema de regresión aparece el de correlación, que consiste en obtener indicadores del grado de relación entre las variables; más específicamente, se trata de cuantificar el grado de dependencia de una variable respecto de otra en función de lo bien que se ajusta la distribución conjunta de las dos variables consideradas a la función de regresión óptima.

4.1.- Problema de regresión mínimo cuadrática

El criterio de optimalidad que se usa con mayor frecuencia en el problema de regresión es el de mínimos cuadrados; según este criterio, la función de regresión óptima es aquella que minimiza la media de las desviaciones cuadráticas entre los valores reales de la variable considerada y los aproximados por la función. Las propiedades de la esperanza condicionada permiten concluir que ésta es la función de regresión óptima en el sentido de mínimos cuadrados.

Planteamiento y resolución del problema

4.2.- Curvas de regresión mínimo cuadráticas

La curva definida por la función de regresión óptima se denomina curva de regresión mínimo cuadrática y es, entre todas las curvas del plano, la que mejor se ajusta a los

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.
BY: Grupo CDPYE-UGR
SOME RIGHTS RESERVED

puntos de la distribución conjunta de las variables consideradas, en el sentido de minimizar la media de las desviaciones cuadráticas entre los valores reales y los aproximados por la función de regresión.

Definición y ejemplos

4.3.- Razones de correlación

Las razones de correlación proporcionan una medida del grado de concentración de la distribución teórica en torno a cada una de las curvas de regresión. Tanto su definición, coherente con el problema que se plantea, como sus propiedades, justifican el uso de estos coeficientes para el fin propuesto.

Definición y propiedades

4.4.- Regresión lineal. Rectas de regresión mínimo cuadráticas

A veces, si el cálculo de la esperanza condicionada es complicado y se tiene relativa seguridad de que los puntos de la distribución teórica se ajustan a una determinada forma funcional (exponencial, parabólica, etc.), puede ser útil restringir la búsqueda de la función de regresión óptima a la clase de funciones de dicha forma. Un caso particularmente importante es el de la regresión lineal. Las rectas que mejor se ajustan a los puntos de la distribución en el sentido de mínimos cuadrados (para aproximar cada una de las variables en términos de la otra) se denominan rectas de regresión mí-

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License.
BY: Grupo CDPYE-UGR
SOME RIGHTS RESERVED

mínimo cuadráticas.

Problema de regresión lineal. Rectas de regresión

4.5.- Coeficientes de determinación y correlación lineal

El coeficiente de determinación lineal, que mide el grado de concentración de la distribución en torno a las rectas de regresión, es el análogo a las razones de correlación en la regresión lineal. Sin embargo, en el caso lineal, la medida por excelencia es el coeficiente de correlación lineal que proporciona, además del grado de dependencia lineal entre las variables, el sentido de dicha dependencia.

Coeficiente de determinación lineal

Coeficiente de correlacion lineal