|  |  |  |
| --- | --- | --- |
| ESCUNA~1 | **UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO****FACULTAD DE INGENIERÍA****ESTADÍSTICA APLICADA****MI OCTAVIO ESTRADA CASTILLO****octavioe@unam.mx** |  |
| **REGRESIÓN Y CORRELACIÓN LINEAL** |
| **DEFINICIONES** | Una regresión es un ajuste de un conjunto de puntos dados a un modelo matemático en particular. Cuando el modelo matemático corresponde con una recta, se denomina regresión lineal. Existen muchos tipos de regresión: lineal, exponencial, polinomial, trigonométrica, etcétera. La regresión polinomial se lleva a efecto a través de la serie de Taylor, sin considerar estimadores de incertidumbre en el ajuste. La regresión trigonométrica se lleva a cabo a través de las series de Fourier, también sin considerar estimadores de incertidumbre. Una correlación corresponde con la definición de indicadores que nos permitan determinar qué tan bueno es el ajuste entre el conjunto de puntos dados y el modelo matemático usado. |
| **METODOLOGÍA** | 1. Determinar por muestreo un conjunto n de parejas ordenadas (x, y), de las cuales se sospecha que existe algún tipo de relación funcional. 2. Graficar en un sistema coordenado xy, denominado Diagrama de Dispersión, al conjunto de puntos (x, y) y observar a “ojo de pájaro” si existe algún tipo de relación funcional entre ambas variables. 3. Establecer a que modelo se ajustará el conjunto de puntos: lineal, exponencial, logarítmico, polinomial, trigonométrico, etcétera.4. Por el método de los mínimos cuadrados, esto es, minimizando el Error Cuadrático Medio, determinar los coeficientes o parámetros del modelo matemático de tal manera que se minimice dicho error.5. Se define el modelo matemático que mejor se ajusta, especificando su expresión matemática. 6. Se obtienen estimadores puntuales y por intervalos de los coeficientes o parámetros del modelo, así como de posibles valores de la variable dependiente.7. Se calculan los coeficientes de determinación y de correlación que permitan estimar el grado de ajuste entre el modelo matemático y los datos dados. |



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DESARROLLO MATEMÁTICO** | Del conjunto de puntos observados:

|  |  |
| --- | --- |
| X | Y |
| x1 | y1 |
| x2 | y2 |
| x3 | y3 |
| • | • |
| • | • |
| • | • |
| xn | yn |

 | Se supondrá que el modelo matemático que mejor se ajusta es la ecuación de una recta: Se define el Error Cuadrático Medio (ECM) como: El método que se aplica para obtener los valores de 1 y 0, es minimizar el ECM, a través del criterio de la primera derivada. Al derivar e igualar a cero, se obtiene un sistema de dos ecuaciones lineales, con dos incógnitas, que al ser resuelto, nos proporciona los resultados que se muestran en la siguiente sección. |
| **ESTIMADORES PUNTUALES** |  |  y Un estimador insesgado de la varianza poblacional sería:EL coeficiente de determinación estaría dado por: y el coeficiente de correlación sería: |
| **ESTIMADORES POR INTERVALOS** | El estimador por intervalos para el coeficiente de correlación sería: El estimador por intervalos para un valor de y0 dado un valor de xo de la tabla sería:El estimador por intervalos para un valor y0 de un valor x0 no presente en la tabla sería: Los estimadores por intervalos para la pendiente y para la ordenada al origen serían: |
| **EJEMPLO** | La resistencia a la tensión de un producto de papel, se relaciona con el porcentaje de madera dura en la pulpa. Se tomaron diez muestras en la fábrica de papel de Tuxtepec Oaxaca, y los datos obtenidos se presentan en la siguiente tabla:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| % Madera Dura | 10 | 15 | 15 | 20 | 20 | 20 | 25 | 25 | 28 | 30 |
| Resistencia del Papel | 160 | 171 | 175 | 182 | 184 | 181 | 188 | 193 | 195 | 200 |

1. Grafique un Diagrama de Dispersión del % de madera dura contra la resistencia del papel y a partir de la relación que observe, determine un modelo de regresión a aplicar.
2. Obtenga el modelo de regresión más conveniente y estime los intervalos de confianza más adecuados para el modelo, explicando que tan exacto es el ajuste.
 |
| **DIAGRAMA DE DISPERSIÓN** |  |
| **RESULTADOS** | Estimadores PuntualesSxx= 357.6Syy= 1,300.9Sxy= 671.8R2= 0.9701= 0.985Para x0=20 | Estimadores por Intervalos al 95% de nivel de confianza, en todos los casos:Para x0= 20: |
| **REFRENCIAS Y CRÉDITOS** | 1. Foto de la parte superior derecha de la primer página tomada estrictamente para motivos académicos:

<http://www.fotosimagenes.org/estadistica>1. Ejemplo resuelto tomado del Problema propuesto 10-1, página 422 del libro de Diseño y Análisis de Experimentos, de Douglas C. Montgomery, Editorial LIMUSA SA de CV, 2005.
2. Notas publicadas en la página Web:

<http://www.ingenieria.unam.mx/calyesd/htm/esd_notas.html>Octavio Estrada Castillo, 2009 |