

# INTRODUCCION AL ANALISIS DE REGRESION LINEAL

por MONTGOMERY

Isbn 9702403278

## Indice del Contenido

Prefacio

1. Introducción

1.1 Regresión y formación de modelos

1.2 Recolección de datos

1.3 Usos de la regresión

1.4 El papel de la computadora

2. Regresión lineal simple

2.1 Modelo de regresión lineal simple

2.2 Estimación de los parámetros por mínimos cuadrados

2.2.1 Estimación de  $\beta_0$  y  $\beta_1$

2.2.2 Propiedades de los estimadores por mínimos cuadrados y el modelo ajustado de regresión

2.2.3 Estimación de

2.2.4 Una forma alterna del modelo

2.3 Prueba de hipótesis de la pendiente y de la ordenada al origen

2.3.1 Uso de pruebas t

2.3.2 Prueba del significado de la regresión

2.3.3 Análisis de varianza

2.4 Estimación de intervalo en la regresión lineal simple

2.4.1 Intervalos de confianza de  $\beta_0$ ,  $\beta_1$  y  $\sigma^2$

2.4.2 Estimación de intervalos de la respuesta media

2.5 Predicción de nuevas observaciones

2.6 Coeficiente de determinación

2.7 Algunas consideraciones sobre el uso de la regresión

2.8 Regresión por el origen

2.9 Estimación por máxima verosimilitud

2.10 Caso en el que el regresor x es aleatorio

2.10.1 Distribución conjunta de x y y

2.10.2 Distribución normal conjunta de x y y: el modelo de correlación

Problemas

3. Regresión lineal múltiple

3.1 Modelo de regresión múltiple

3.2 Estimación de los modelos del parámetro

3.2.1 Estimación de los coeficientes de regresión por mínimos cuadrados

3.2.2 Interpretación geométrica de mínimos cuadrados

3.2.3 Propiedades de los estimadores de mínimos cuadrados

3.2.4 Estimación de  $\sigma^2$

3.2.5 Inadecuación de los diagramas de dispersión en regresión múltiple

3.2.6 Estimación de máxima posibilidad

3.3 Prueba de hipótesis en regresión lineal múltiple

3.3.1 Pruebas de la significancia de la regresión

3.3.2 Pruebas sobre coeficientes individuales de regresión

3.3.3 Caso especial de columnas ortogonales en X

3.3.4 Prueba de la hipótesis general lineal

3.4 Intervalos de confianza en regresión múltiple

3.4.1 Intervalos simultáneos de confianza para coeficientes de regresión

3.4.2 Estimación del intervalo de confianza de la respuesta media

- 3.4.3 Intervalos de confianza simultáneos de coeficientes de regresión
  - 3.5 Predicción de nuevas observaciones
  - 3.6 Extrapolación oculta en la regresión múltiple
  - 3.7 Coeficientes normalizados de regresión
  - 3.8 Multicolinealidad
  - 3.9 ¿Por qué los coeficientes de regresión tienen signo incorrecto?
- Problemas

- 4. Comprobación de la adecuación del modelo
    - 4.1 Introducción
    - 4.2 Análisis de residuales
      - 4.2.1 Definición de residuales
      - 4.2.2 Métodos para escalar residuales
      - 4.2.3 Gráficas de residuales
      - 4.2.4 Gráficas de regresión parcial y de residuales parciales
      - 4.2.5 Otras gráficas de residuales y métodos de análisis
    - 4.3 La estadística PRESS
    - 4.4 Detección y tratamiento de puntos atípicos
    - 4.5 Falta de ajuste del modelo de regresión
      - 4.5.1 Una prueba formal de falta de ajuste
      - 4.5.2 Estimación del error puro a partir de vecinos próximos
- Problemas

- 5. Transformaciones y ponderación para corregir inadecuaciones del modelo
    - 5.1 Introducción
    - 5.2 Transformaciones estabilizadoras de la varianza
    - 5.3 Transformaciones para linealizar el modelo
    - 5.4 Métodos analíticos para seleccionar una transformación
      - 5.4.1 Transformaciones de  $y$ : el método de Box-Cox
      - 5.4.2 Transformaciones de las variables regresoras
    - 5.5 Mínimos cuadrados generalizados y ponderados
      - 5.5.1 Mínimos cuadrados generalizados
      - 5.5.2 Mínimos cuadrados ponderados
      - 5.5.3 Algunos asuntos prácticos
- Problemas

- 6. Diagnóstico para balanceo e influencia
    - 6.1 Importancia de detectar observaciones influyentes
    - 6.2 Balanceo
    - 6.3 Medidas de influencia: La  $D$  de Cook
    - 6.4 Medidas de influencia: DFFITS y DFBETAS
    - 6.5 Una medida de eficiencia del modelo
    - 6.6 Detección de grupos de observaciones influyentes
    - 6.7 Tratamiento de observaciones influyentes
- Problemas

- 7. Modelos polinomiales de regresión
  - 7.1 Introducción
  - 7.2 Modelos polinomiales en una variable
    - 7.2.1 Principios básicos
    - 7.2.2 Ajuste polinomial por segmentos (SPLINE)
    - 7.2.3 Términos polinomiales y trigonométricos
  - 7.3 Regresión no paramétrica
    - 7.3.1 Regresión Kernel
    - 7.3.2 Regresión ponderada localmente (Loess)
    - 7.3.3 Precauciones finales
  - 7.4 Modelos polinomiales con dos o más variables

## 7.5 Polinomios ortogonales

### Problemas

## 8. Variables indicadoras

### 8.1 El concepto general de las variables indicadoras

### 8.2 Comentarios sobre el uso de variables indicadoras

#### 8.2.1 Variables indicadoras en función de la regresión con códigos asignados

#### 8.2.2 Variables indicadoras como sustitutos de un regresor cuantitativo

### 8.3 Método de regresión para análisis de varianza

### Problemas

## 9. Selección de variable y formación el modelo

### 9.1 Introducción

#### 9.1.1 El problema de la construcción del modelo

#### 9.1.2 Consecuencias de la mala especificación del modelo

#### 9.1.3 Criterios para evaluar modelos de regresión con subconjuntos de variables

### 9.2 Técnicas computacionales para seleccionar variables

#### 9.2.1 Todas las regresiones posibles

#### 9.2.2 Métodos de regresión por segmentos

### 9.3 Algunas recomendaciones finales para la práctica

### Problemas

## 10. Multicolinealidad

### 10.1 Introducción

### 10.2 Fuentes de multicolinealidad

### 10.3 Efectos de la multicolinealidad

### 10.4 Diagnóstico de multicolinealidad

#### 10.4.1 Examen de la matriz de correlación

#### 10.4.2 Factores de inflación de varianza

#### 10.4.3 Análisis del eigensistema de $X'X$

#### 10.4.4 Otros diagnósticos

### 10.5 Métodos de manejar la multicolinealidad

#### 10.5.1 Recolección de datos adicionales

#### 10.5.2 Reespecificación del modelo

#### 10.5.3 Regresión Ridge

#### 10.5.4 Otros métodos

#### 10.5.5 Comparación y evaluación de estimadores sesgados

### Problemas

## 11. Regresión robusta

### 11.1 La necesidad de una regresión robusta

### 11.2 Estimadores M

### 11.3 Propiedades de los estimadores robustos

#### 11.3.1 Punto de quiebre

#### 11.3.2 Eficiencia

### 11.4 Reseña de otros estimadores robustos de regresión

#### 11.4.1 Estimadores de punto de quiebre alto

#### 11.4.2 Estimadores de influencia limitada

#### 11.4.3 Otros procedimientos

#### 11.4.4 Cálculo de estimadores de regresión robusta

### Problemas

## 12. Introducción a la regresión no lineal

### 12.1 Modelos de regresión lineales y no lineales

#### 12.1.1 Modelos de regresión lineales

#### 12.1.2 Modelos de regresión no lineales

- 12.2 Mínimos cuadrados no lineales
  - 12.3 Transformación a un modelo lineal
  - 12.4 Estimación de parámetros en un sistema no lineal
    - 12.4.1 Linealización
    - 12.4.2 Otros métodos de estimación de parámetros
    - 12.4.3 Valores iniciales
    - 12.4.4 Programas de cómputo
  - 12.5 Inferencia estadística en regresión no lineal
  - 12.6 Ejemplos de modelos de regresión no lineales
- Problemas

- 13. Modelos lineales generalizados
    - 13.1 Introducción
    - 13.2 Modelos de regresión logística
      - 13.2.1 Modelos con una variable de respuesta binaria
      - 13.2.2 Estimación de parámetros en un modelo de regresión logística
      - 13.2.3 Interpretación de parámetros en un modelo de regresión logística
      - 13.2.4 Pruebas de hipótesis sobre los parámetros del modelo
    - 13.3 Regresión de Poisson
    - 13.4 El modelo lineal generalizado
      - 13.4.1 Funciones de cadena y predictores lineales
      - 13.4.2 Estimación de parámetros e inferencia en el modelo lineal general
      - 13.4.3 Predicción y estimación con el modelo lineal general
      - 13.4.4 Análisis de residuales en el modelo lineal generalizado
      - 13.4.5 Sobredispersión
- Problemas

- 14. Otros temas acerca del uso del análisis de regresión
    - 14.1 Modelos de regresión con errores autocorrelacionados
      - 14.1.1 Causas y efectos de la autocorrelación
      - 14.1.2 Detección de la presencia de autocorrelación
      - 14.1.3 Métodos de estimación de parámetros
    - 14.2 Efecto de los errores de medición sobre los regresores
      - 14.2.1 Regresión lineal simple
      - 14.2.2 El modelo de Berkson
    - 14.3 Estimación inversa - El problema de la calibración
    - 14.4 Bootstrap en regresión
      - 14.4.1 Muestreo bootstrap en regresión
      - 14.4.2 Intervalos de confianza de bootstrap
    - 14.5 Árboles de clasificación y regresión (CART)
    - 14.6 Redes neurales
    - 14.7 Experimentos diseñados para regresión
- Problemas

- 15. Validación de los modelos de regresión
    - 15.1 Introducción
    - 15.2 Técnicas de validación
      - 15.2.1 Análisis de los coeficientes y valores predichos del modelo
      - 15.2.2 Recolección de nuevos datos - Corridas de confirmación
      - 15.2.3 División de los datos
    - 15.3 Datos obtenidos con experimentos planeados
- Problemas

APÉNDICE A. Tablas estadísticas

APÉNDICE B. Conjuntos de datos para los ejercicios

APÉNDICE C. Material suplementario

- C.1 Resumen de pruebas básicas en estadística prueba
  - C.2 Resumen de la teoría de modelos lineales
  - C.3 Resultados importantes acerca de SSR y SSRes
  - C.4 El teorema de Gauss-Markov,  $\text{Var}(e) = \sigma^2 I$
  - C.5 Aspectos computacionales de la regresión múltiple
  - C.6 Un resultado acerca de la inversa de una matriz
  - C.7 Desarrollo del estadístico
  - C.8 Desarrollo de  $S^2$  (cuadrado)
  - C.9 Una prueba para valores atípicos basada en R de Student
  - C.10 El teorema de Gauss-Markov,  $\text{Var}(E) = V$
  - C.11 El sesgo en SSRes cuando el modelo está subespecificado
  - C.12 Cálculo de diagnóstico de influencia
  - C.13 Modelos lineales generalizados
- Referencias

Índice