

**Práctica N°4**

***Ivan Fernando Mujica Mamani***

*Análisis Estadístico I*

*Maestría en Ciencia de Datos v.2*

**1. Ejercicio 1**

*Una industria algodonera, interesada en maximizar el rendimiento de la semilla de algodón, quiere comprobar si dicho rendimiento depende del tipo de fertilizante utilizado para tratar la planta. A su disposición tiene 5 tipos de fertilizantes. Como puede haber diferencia entre las parcelas, el experimentador decide efectuar un diseño en bloques aleatorizados. Para ello, divide el terreno en 4 bloques y cada bloque en 5 parcelas, fumigando dentro de cada bloque cada una de las parcelas con un fertilizante. Al recoger la cosecha se mide el rendimiento de la semilla, obteniéndose las siguientes observaciones:*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Bloques | | | |
| Fertilizante | A | B | C | D |
| 1 | 87 | 86 | 88 | 83 |
| 2 | 85 | 87 | 95 | 85 |
| 3 | 90 | 92 | 95 | 90 |
| 4 | 89 | 97 | 98 | 88 |
| 5 | 99 | 96 | 91 | 90 |

Se pide probar si el rendimiento de la semilla de algodón difiere significativamente dependiendo del tipo de fertilizante utilizado. Y si los bloques de terreno son significativamente distintos.

**Respuesta**

Calculamos la siguiente tabla.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | 1 | 2 | 3 | 4 | 5 | Total | Promedio |
| Bloques | A | 87 | 85 | 90 | 89 | 99 | 450 | 90 |
|  | B | 86 | 87 | 92 | 97 | 96 | 458 | 91.6 |
|  | C | 88 | 95 | 95 | 98 | 91 | 467 | 93.4 |
|  | D | 83 | 85 | 90 | 88 | 90 | 436 | 87.2 |
| total |  | 344 | 352 | 367 | 372 | 376 | 1811 | 362.2 |
| promedio |  | 86 | 88 | 91.75 | 93 | 94 | media global | 90.55 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  | **Total** |
| 12.6025 | 30.8025 | 0.3025 | 2.4025 | 71.4025 | 117.5125 |
| 20.7025 | 12.6025 | 2.1025 | 41.6025 | 29.7025 | 106.7125 |
| 6.5025 | 19.8025 | 19.8025 | 55.5025 | 0.2025 | 101.8125 |
| 57.0025 | 30.8025 | 0.3025 | 6.5025 | 0.3025 | 94.9125 |

La suma total de cuadrados es:

Calculamos la suma de los cuadrados debidos a los tratamientos, utilizando el promedio de los tratamientos menos la media global al cuadrado.

Calculamos la suma de los cuadrados debidos a los bloques, utilizando el promedio de los bloques menos la media global al cuadrado

Calculamos la suma de cuadrados debidos al error

Entonces tenemos que :

|  |  |  |
| --- | --- | --- |
| **Suma de cuadrados** | **Grado de Libertad** | **Cuadrado Medio** |
| 186.2 | 4 | 46.6 |
| 103.75 | 3 | 34.6 |
| 131 | 12 | 10.9 |
| 420.95 | 19 | 22.2 |

*Para*

*Para*

Concluimos que son diferentes entre tratamientos, pero no entre bloques.

**2. Ejercicio 2**

Un fabricante de calzado desea mejorar la calidad de las suelas, las cuales se pueden hacer con uno de los cuatro tipos de cuero A, B, C y D disponibles en el mercado. Para ello prueba los cueros con una máquina que hace pasar los zapatos por una superficie abrasiva; la suela de estos se desgasta al pasarla por dicha superficie.

Como criterio de desgaste se usa la pérdida de peso después de un número fijo de ciclos. Se prueban en orden aleatorio 24 zapatos, seis de cada tipo de cuero. Los datos (en miligramos) sobre el desgaste de cada tipo de cuero se muestra en la siguiente tabla

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Tipo de Cuero** | **desgaste** | | | | | | **Promedio** |
| **A** | **264** | **260** | **258** | **241** | **262** | **255** | **256.7** |
| **B** | **208** | **220** | **216** | **200** | **213** | **206** | **210.5** |
| **C** | **220** | **219** | **263** | **225** | **230** | **228** | **230.8** |
| **D** | **217** | **226** | **215** | **227** | **220** | **222** | **220.7** |

Se pide

1. Probar si los tratamientos son diferentes.
2. Probar las comparaciones con Comparaciones de rangos múltiples.
3. Aplicar la prueba LSD (Diferencia mínima significativa).
4. Aplicar el método de Tukey (HSD).

Realizamos los cálculos iniciales

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tipo de cuero** | **Desgaste** | | | | | | **Total** | **promedio** |
| **A** | 264 | 260 | 258 | 241 | 262 | 255 | **1540** | **256.667** |
| **B** | 208 | 220 | 216 | 200 | 213 | 206 | **1263** | **210.500** |
| **C** | 220 | 263 | 219 | 225 | 230 | 228 | **1385** | **230.833** |
| **D** | 517 | 226 | 215 | 224 | 220 | 222 | **1624** | **270.667** |
|  | **Promedio total** | | | |  |  | **5812** | **242.167** |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Varianza Muestral** | | | | | | **Suma** | **Varianza Muestral** |
| 53.78 | 11.11 | 1.78 | 245.44 | 28.44 | 2.78 | 343.33 | 68.67 |
| 6.25 | 90.25 | 30.25 | 110.25 | 6.25 | 20.25 | 263.50 | 52.70 |
| 117.36 | 1034.69 | 140.03 | 34.03 | 0.69 | 8.03 | 1334.83 | 266.97 |
| 60680.11 | 1995.11 | 3098.78 | 2177.78 | 2567.11 | 2368.44 | 72887.33 | 14577.47 |

Calculamos la suma de cuadrados de tratamientos:

SCT= 1261.5+6016.6667+770.6666.7+4873.5=12922.333

Calculamos la suma de cuadrados de los errores:

SCE=343.333+263.5+1334.83+72887.3=74829

|  |  |  |
| --- | --- | --- |
| **SUMA DE CUADRADOS** | **GRADOS DE LIBERTAD** | **CUADRADOS MEDIOS** |
| 12922.333 | 3 | 4307.44 |
| 74829 | 20 | 3741.45 |
| 87751.333 | 23 |  |

Entonces decimos que

f=4307.44/3741.45=0.3528

Para la suma de cuadrados totales

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| (1i-y)2 | (2i-y)2 | (3i-y)2 | (4i-y)2 | (5i-y)2 | (6i-y)2 | **Totales** |
| 476.69444 | 318.027778 | 250.6944444 | 1.3611111 | 393.36111 | 164.6944444 | 1604.8333 |
| 1167.3611 | 491.361111 | 684.6944444 | 1778.0278 | 850.69444 | 1308.027778 | 6280.1667 |
| 491.36111 | 434.027778 | 536.6944444 | 294.69444 | 148.02778 | 200.6944444 | 2105.5 |
| 75533.361 | 261.361111 | 738.0277778 | 330.02778 | 491.36111 | 406.6944444 | 77760.833 |
|  |  |  |  |  | Total | 87751.333 |

Entonces declaramos que los tratamientos son idénticos

Pruebas Multiples, prueba de Fisher

A vs B

Estadístico de Prueba

t=1.3072

p = 0.205948

t con n -k grados de libertad con alpha/2 de probabilidad: 2.085963447

t0=-2.086 , t1 = 2.08596

No se rechaza Ho de que A y B sean en media iguales

A vs B

Diferencia absoluta = 46.16667

LSD = 73.66582

Rechazamos si la diferencia absoluta es mayor o igual a LSD, entonces no se rechaza

A vs B

Diferencia absoluta = 46.167

HSD = 57.93384

Rechazamos si la diferencia absoluta es mayor o igual a HSD, entonces no se rechaza

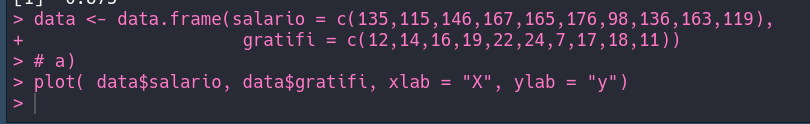
**3. Ejercicio 6, página 554**

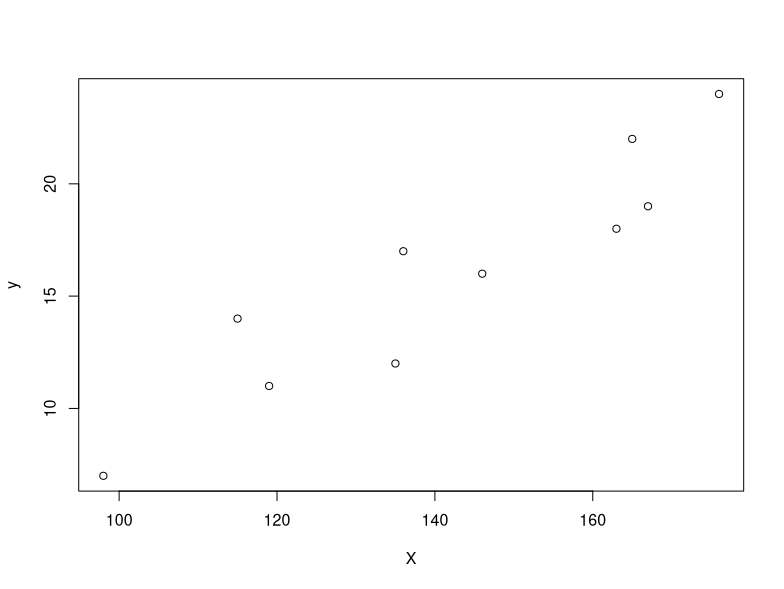
Wageweb realiza estudios sobre datos salariales y presenta resúmenes de éstos en su sitio de la Red. Basándose en datos salariales desde el 1 de octubre de 2002 Wageweb publicó que el salario anual promedio de los vicepresidentes de ventas era $142 111 con una gratificación anual promedio de $15 432 (Wageweb.com, 13 de marzo de 2003). Suponga que los datos siguientes sean una muestra de salarios y bonos anuales de 10 vicepresidentes de ventas. Los datos se dan en miles de dólares.

|  |  |  |
| --- | --- | --- |
| **Vicepresidente** | **Salario** | **Gratificación** |
| 1 | 135 | 12 |
| 2 | 115 | 14 |
| 3 | 146 | 16 |
| 4 | 167 | 19 |
| 5 | 165 | 22 |
| 6 | 176 | 24 |
| 7 | 98 | 7 |
| 8 | 136 | 17 |
| 9 | 163 | 18 |
| 10 | 119 | 11 |

**A. Trace un diagrama de dispersión con estos datos tomando como variable independiente los salarios.**

Aplicacion en R:

****



**B. ¿Qué indica el diagrama de dispersión del inciso a) acerca de la relación entre salario y gratificación?**

Gratificacion o bonos anuales contra salario de vicepresidentes, se observa una aparente relación directa entre salarios y los bonos anuales.

**C. Use el método de mínimos cuadrados para obtener la ecuación de regresión estimada.**

**1)** Encontramos X\*Y y , utilizando la siguiente tabla:

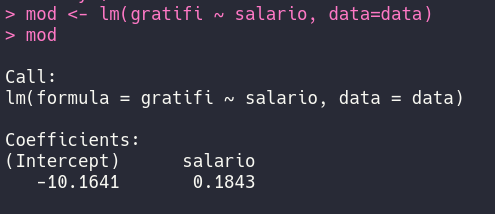
|  |  |  |  |
| --- | --- | --- | --- |
| **X** | **Y** | **X-Y** | **X\*X** |
| 135 | 12 | 1620 | 18225 |
| 115 | 14 | 1610 | 13225 |
| 146 | 16 | 2336 | 21316 |
| 167 | 19 | 3173 | 27889 |
| 165 | 22 | 3630 | 27225 |
| 176 | 24 | 4224 | 30976 |
| 98 | 7 | 686 | 9604 |
| 136 | 17 | 2312 | 18496 |
| 163 | 18 | 2934 | 26569 |
| 119 | 11 | 1309 | 14161 |

**2)** Encontramos la suma de cada columna

|  |  |  |  |
| --- | --- | --- | --- |
| **X** | **Y** | **X-Y** | **X\*X** |
| 135 | 12 | 1620 | 18225 |
| 115 | 14 | 1610 | 13225 |
| 146 | 16 | 2336 | 21316 |
| 167 | 19 | 3173 | 27889 |
| 165 | 22 | 3630 | 27225 |
| 176 | 24 | 4224 | 30976 |
| 98 | 7 | 686 | 9604 |
| 136 | 17 | 2312 | 18496 |
| 163 | 18 | 2934 | 26569 |
| 119 | 11 | 1309 | 14161 |
| **1420** | **160** | **23834** | **207686** |

**3)** Usamos la siguiente ecuación para encontrar a y b:

**4)** Sustituimos a y b en la ecuación de la regresión:

**D. Dé u****na interpretación de la ecuación de regresión estimada.**

La relación entre el salario recibido por los vicepresidentes y los bonos anuales es directa, a medida que aumenta el salario aumenta los bonos anuales.

Se osbserva que por cada unidad del salario correspondido al vicepresident los bonos anuales se incrementan en 0.18 $us.

**E. ¿Cuál será la gratificación de un vicepresidente que tenga un salario anual de $120.000?**

*= 120\*0.1843-10.1641*

*= 11.9 Mil $.*

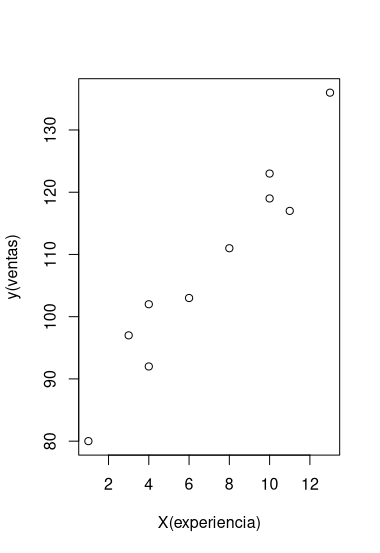
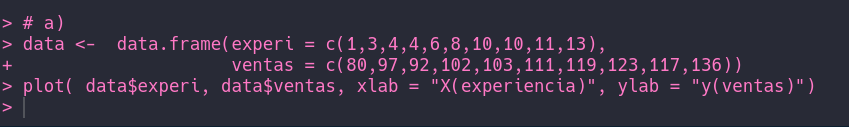
**4 . Ejercicio 9, página 556**

Un gerente de ventas recolectó los datos siguientes sobre ventas anuales y años de experiencia

|  |  |  |
| --- | --- | --- |
| **vendedor** | **Años de experiencia** | **Ventas anuales** |
| 1 | 1 | 80 |
| 2 | 3 | 97 |
| 3 | 4 | 92 |
| 4 | 4 | 102 |
| 5 | 6 | 103 |
| 6 | 8 | 111 |
| 7 | 10 | 119 |
| 8 | 10 | 123 |
| 9 | 11 | 117 |
| 10 | 13 | 136 |

**A. Elabore un diagrama de dispersión con estos datos, en el que la variable independiente sean los años de experiencia.**

Aplicación en R:



**B. Dé la ecuación de regresión estimada que puede emplearse para predecir las ventas anuales cuando se conocen los años de experiencia.**

**1)**encontramos para cada valor en la siguiente tabla.

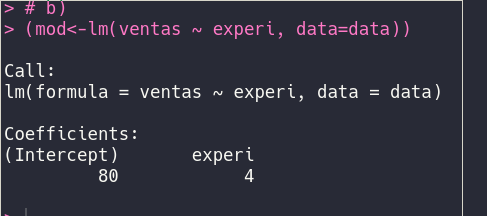
|  |  |  |  |
| --- | --- | --- | --- |
| **X** | **Y** | **X\*Y** | **X\*X** |
| 1 | 80 | 80 | 1 |
| 3 | 97 | 291 | 9 |
| 4 | 92 | 368 | 16 |
| 4 | 102 | 408 | 16 |
| 6 | 103 | 618 | 36 |
| 8 | 111 | 888 | 64 |
| 10 | 119 | 1190 | 100 |
| 10 | 123 | 1230 | 100 |
| 11 | 117 | 1287 | 121 |
| 13 | 136 | 1768 | 169 |

**2)** Calculamos la sumatoria de cada columna.

|  |  |  |  |
| --- | --- | --- | --- |
| **X** | **Y** | **X\*Y** | **X\*X** |
| 1 | 80 | 80 | 1 |
| 3 | 97 | 291 | 9 |
| 4 | 92 | 368 | 16 |
| 4 | 102 | 408 | 16 |
| 6 | 103 | 618 | 36 |
| 8 | 111 | 888 | 64 |
| 10 | 119 | 1190 | 100 |
| 10 | 123 | 1230 | 100 |
| 11 | 117 | 1287 | 121 |
| 13 | 136 | 1768 | 169 |
| **70** | **1080** | **8128** | **632** |

**3)** Utilizamos la siguiente ecuación para encontrar los valores de a y b.

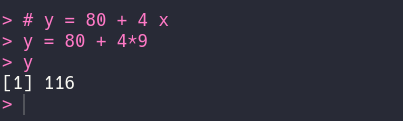
Aplicacion en R:



**4)** Sustituimos los valores de a y b en la ecuación de regresión.

**C. Use la ecuación de regresión estimada para pronosticar las ventas anuales de un vendedor de 9 años de experiencia.**

Aplicacion en R:



Las ventas anuales previstas son 116 mil $ aproximados.

**5. Ejercicio 20.**

*Consumer Reports publica pruebas y evaluaciones sobre televisores de alta definición. Para cada modelo se elaboró una evaluación general basada principalmente en la calidad de la imagen. Una evaluación más alta indica un mejor funcionamiento. En los datos siguientes se dan evaluación general y precio de televisores de plasma de 45 pulgadas (Consumer Reports, marzo 2006).*

|  |  |  |
| --- | --- | --- |
| **Marca** | **Precio** | **Puntuación en la valuación** |
| dell | 2800 | 62 |
| Hisense | 2800 | 53 |
| Hitachi | 2700 | 44 |
| JVC | 3500 | 50 |
| LG | 3300 | 54 |
| Maxent | 2000 | 39 |
| Panasonic | 4000 | 66 |
| Phillips | 3000 | 55 |
| Proview | 2500 | 34 |
| Samsung | 3000 | 39 |

**A. Use estos datos para obtener una ecuación de regresión estimada que pueda emplearse para estimar la puntuación en la evaluación general de una televisión de 42 pulgadas dado el precio.**

**1)** Encontramos X\*Y, utilizando la siguiente tabla

|  |  |  |  |
| --- | --- | --- | --- |
| **X** | **Y** | **X\*Y** | **X\*X** |
| 2800 | 62 | 173600 | 7840000 |
| 2800 | 53 | 148400 | 7840000 |
| 2700 | 44 | 118800 | 7290000 |
| 3500 | 50 | 175000 | 12250000 |
| 3300 | 54 | 178200 | 10890000 |
| 2000 | 39 | 78000 | 4000000 |
| 4000 | 66 | 264000 | 16000000 |
| 3000 | 55 | 165000 | 9000000 |
| 2500 | 34 | 85000 | 6250000 |
| 3000 | 39 | 117000 | 9000000 |

**2)** Encontramos las sumatorias de cada columna.

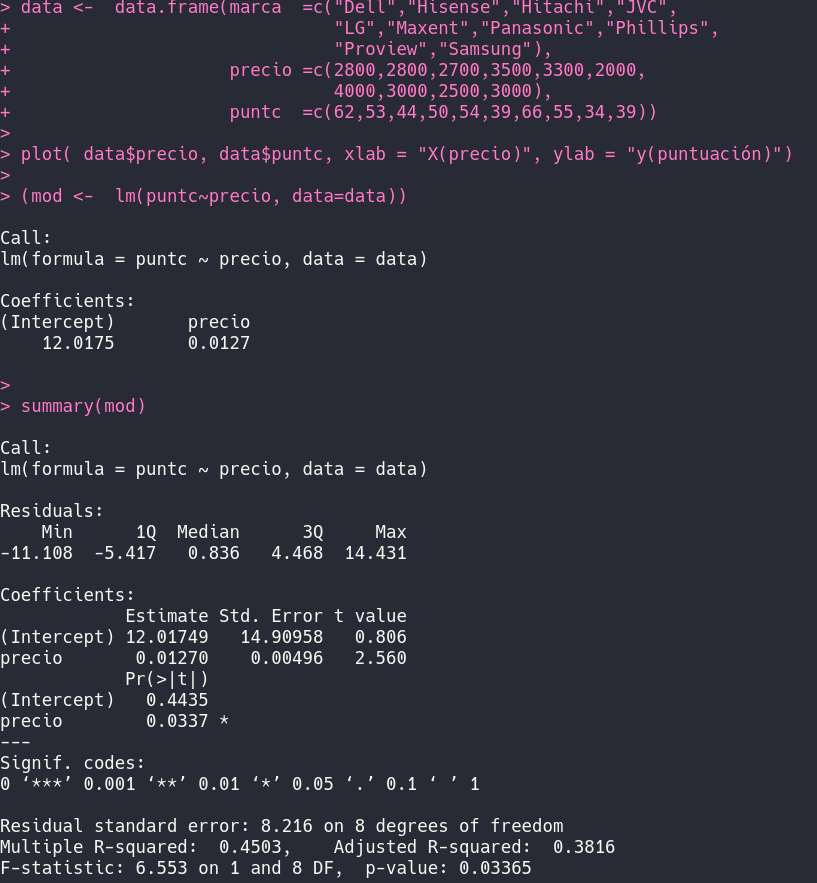
|  |  |  |  |
| --- | --- | --- | --- |
| **X** | **Y** | **X\*Y** | **X\*X** |
| 2800 | 62 | 173600 | 7840000 |
| 2800 | 53 | 148400 | 7840000 |
| 2700 | 44 | 118800 | 7290000 |
| 3500 | 50 | 175000 | 12250000 |
| 3300 | 54 | 178200 | 10890000 |
| 2000 | 39 | 78000 | 4000000 |
| 4000 | 66 | 264000 | 16000000 |
| 3000 | 55 | 165000 | 9000000 |
| 2500 | 34 | 85000 | 6250000 |
| 3000 | 39 | 117000 | 9000000 |
| **29600** | **496** | **1503000** | **90360000** |

**3)** Utilizamos la siguiente formula para hallar los valores de a y b.

**4)**

Sustituimos a y b en la ecuación de regresión:

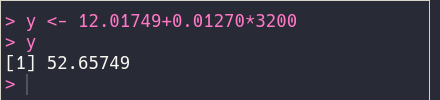
*Aplicacion en R:*



**B. Calcule . ¿Proporcionó un buen ajuste la ecuación de regresión estimada?**

El modelo puede explica solo el 45% de la variación de la variable precio.

**C. Estime la puntuación en la evaluación general de un televisor cuyo precio es $3200.**



La puntuación estimada es de 52.7