# ESTADÍSTICA
## GUIA No 6

# REGRESIÓN Y CORRELACIÓN

## 1. Regresión lineal y no lineal

### 1.1 Modelos de Regresión.

Los modelos de regresión son modelos matemáticos utilizados para realizar predicciones o pronósticos cuando se tienen distribuciones bidimensionales, es decir, distribuciones donde cada dato es un par (x,y) conformado por valores de dos variables X y Y.

Si se representan las parejas (x,y) en un sistema coordenado (X,Y) se genera una nube de puntos llamada diagrama de dispersión o dispersograma; sobre la nube de puntos puede trazarse una curva que se ajuste lo mejor posible a ella, llamada curva de ajuste o línea de regresión.

De acuerdo con la disposición que tengan los datos en el plano cartesiano se busca establecer un modelo adecuado. Existe una gran variedad de modelos de regresión, entre ellos se encuentran el modelo lineal y los modelos no lineales (cuadrático, polinómico, logístico, exponencial, geométrico o potencial, logarítmico, Gompertz, etc.).

### 1.2 Ajuste de curvas por el método de mínimos cuadrados.

La **regresión**, por el método de mínimos cuadrados, es una técnica que tiene como objetivo obtener una curva que minimice la discrepancia entre los puntos dados y la curva de ajuste.

Supóngase que se tiene una muestra de n puntos $(x_1,y_1)$, $(x_2,y_2)$, … , $(x_n,y_n)$, donde $x_k$ son los valores que toma una variable X, $y_k$ son los valores que toma una variable Y, y $(x_k,y_k)$ son las dos mediciones para el objeto observado $O_k$. Al graficar este conjunto de puntos en un sistema (X,Y) se obtendrá un diagrama de dispersión o dispersograma. La tendencia de la nube de puntos motivará la elección de una función $y=f\left(x|a_1,a_2,\ldots,a_m\right)$, aquella cuya gráfica se ajusta mejor al dispersograma. Las letras $a_1$, $a_2$, … , $a_m$ son los parámetros de la función. Para obtener estos parámetros se procede de la siguiente manera:

Cada valor $x_k$ tendrá dos imágenes o valores: un valor es el $y_k$ dado en la muestra y se llamará **valor observado**, el otro valor, llamado **valor teórico**, es $y_k^*$ y se obtiene al sustituir $x_k$ en la ecuación $y=f\left(x|a_1,a_2,\ldots,a_m\right)$, es decir, $y_k^*=f\left(x_k|a_1,a_2,\ldots,a_m\right)$.

Para cada valor $x_k$ se calcula la diferencia $d_k=y_k^*-y_k$, la cual se llamará desviación o residuo. Con estos residuos se construye la expresión $\varphi(a_1,a_2,\ldots,a_m)=\sum_{k=1}^{k=n}(d_k)^2$ que se llamará función de error total.

Ahora, se deben hallar los valores de los parámetros $a_k$ para los cuales la expresión $\varphi(a_1,a_2,\ldots,a_m)=\sum_{k=1}^{k=n}(d_k)^2=\sum_{k=1}^{k=n}\left(y_k^*-y_k\right)^2=\sum_{k=1}^{k=n}\left(f\left(x_k|a_1,a_2,\ldots,a_m\right)-y_k\right)^2$ se minimiza. Como $\varphi=\varphi(a_1,a_2,\ldots,a_m)$

es una función dependiente de los m parámetros $a_1$, $a_2$,… y $a_m$, sus valores se pueden encontrar resolviendo el sistema $\frac{\partial\varphi}{\partial a_1}=0,\ \frac{\partial\varphi}{\partial a_2}=0,\ \ldots,\ \frac{\partial\varphi}{\partial a_m}=0$, el cual es llamado sistema de ecuaciones normales de regresión de Gauss.

**a) Modelo de regresión lineal simple.**

El modelo de regresión lineal es una ecuación de la forma Y=mX+b, cuya gráfica es una recta en el plano $R^2$. A continuación se obtendrá este modelo por el método de los mínimos cuadrados.

Supóngase que se tiene un conjunto de n parejas $(x_1,y_1)$, $(x_2,y_2)$, .., $(x_n,y_n)$. Se debe obtener una recta que se ajuste al conjunto de puntos, de tal manera que se obtenga un gráfico como el que muestra en la figura 4:

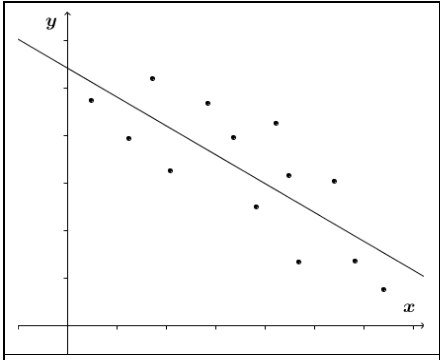


Fig. 4

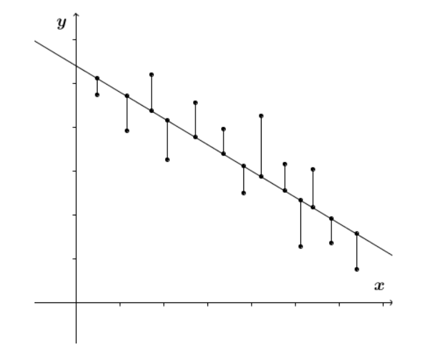


Fig 5. Distancias verticales.

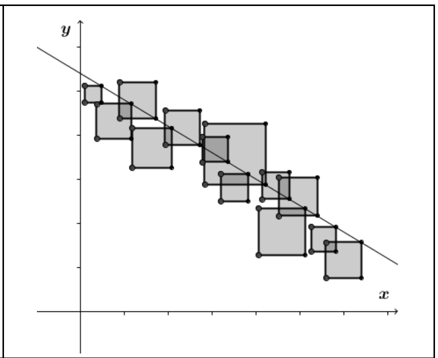


Fig 6. Residuos al cuadrado.

Para cada valor $x_k$ se tiene un valor observado $y_k$ y un valor teórico $y_k^* = mx_k + b$, así los residuos $d_k$ (diferencias o distancias verticales) serán los números $d_k = y_k^* - y_k = mx_k + b - y_k$. Estos residuos se elevan al cuadrado y se suman para formar la función de error total:

$$\varphi(m,b) = \sum_{k=1}^{k=n} (d_k)^2 = \sum_{k=1}^{k=n}\left(y_k^* - y_k\right)^2 = \sum_{k=1}^{k=n}\left(mx_k + b - y_k\right)^2$$

Para hallar los valores de los parámetros m y b que minimizan el error, se resuelve el sistema de ecuaciones $\begin{cases} \dfrac{\partial\varphi}{\partial m} = 2\displaystyle\sum_{k=1}^{k=n} x_k\left(mx_k + b - y_k\right) = 0 \\ \dfrac{\partial\varphi}{\partial b} = 2\displaystyle\sum_{k=1}^{k=n}\left(mx_k + b - y_k\right) = 0 \end{cases}$, el cual es equivalente al sistema $\begin{cases} m\displaystyle\sum_{k=1}^{k=n} x_k^2 + b\sum_{k=1}^{k=n} x_k = \sum_{k=1}^{k=n} x_k y_k \\ m\displaystyle\sum_{k=1}^{k=n} x_k + nb = \sum_{k=1}^{k=n} y_k \end{cases}$, denominado sistema de ecuaciones normales de regresión.

De la solución del sistema resultan las siguientes expresiones para los parámetros m y b:

$$m = \frac{n\displaystyle\sum_{k=1}^{k=n} x_k y_k - \sum_{k=1}^{k=n} y_k \sum_{k=1}^{k=n} x_k}{n\displaystyle\sum_{k=1}^{k=n} x_k^2 - \left(\sum_{k=1}^{k=n} x_k\right)^2} \quad \text{y} \quad b = \frac{\displaystyle\sum_{k=1}^{k=n} y_k - m\sum_{k=1}^{k=n} x_k}{n}.$$

Si los promedios de los $x_k$, $y_k$, $x_ky_k$ y $x_k^2$ se denotan $\overline{X}$, $\overline{Y}$, $\overline{YX}$ y $\overline{X^2}$ respectivamente, entonces $m=\dfrac{\overline{X\cdot Y}-\overline{X}\cdot\overline{Y}}{\overline{X^2}-\left(\overline{X}\right)^2}$ y $b=\dfrac{\overline{X^2}\cdot\overline{Y}-\overline{X\cdot Y}\cdot\overline{X}}{\overline{X^2}-\left(\overline{X}\right)^2}$.

**b) Modelo de regresión cuadrática.**

El modelo de regresión cuadrática es una ecuación de la forma $Y=aX^2+bX+c$, cuya gráfica es una parábola en el plano $R^2$. A continuación se obtendrá este modelo por el método de los mínimos cuadrados.

Supóngase que se tiene un conjunto de n parejas $(x_1,y_1)$, $(x_2,y_2)$, .., $(x_n,y_n)$. Para cada valor $x_k$ se tiene un valor observado $y_k$ y un valor teórico $y_k^* = ax_k^2+bx_k+c$, así los residuos $d_k$ (diferencias o distancias verticales) serán los números $d_k = y_k^* - y_k = ax_k^2+bx_k+c-y_k$. Estos residuos se elevan al cuadrado y se suman para formar la función de error total:

$$\varphi(a,b,c)=\sum_{k=1}^{k=n}\left(d_k\right)^2=\sum_{k=1}^{k=n}\left(y_k^*-y_k\right)^2=\sum_{k=1}^{k=n}\left(ax_k^2+bx_k+c-y_k\right)^2$$

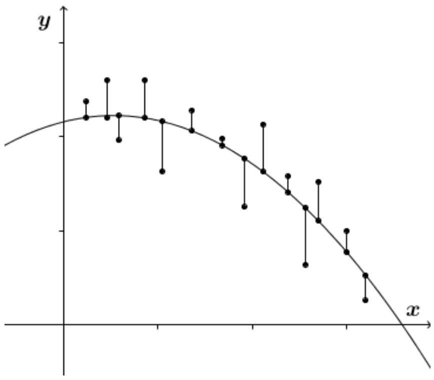

Figura 7. Distancias verticales.

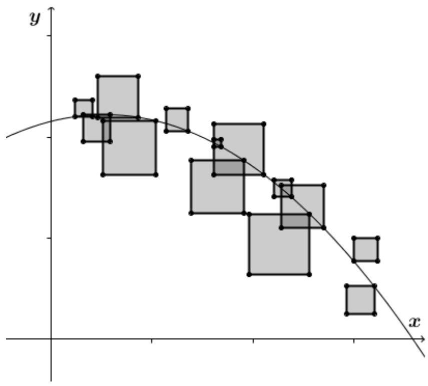

Figura 8. Residuos al cuadrado.

Para hallar los valores de los parámetros a, b, y c, se calculan las derivadas de la función $\varphi(a,b,c)=\sum_{k=1}^{k=n}\left(ax_k^2+bx_k+c-y_k\right)^2$ con respecto a cada parámetro y se igualan a cero, obteniéndose:

$$\begin{cases}\dfrac{\partial\varphi}{\partial a}=-2\sum_{k=1}^{k=n}\left(y_k-ax_k^2-bx_k-c\right)x_k^2=0\\ \dfrac{\partial\varphi}{\partial b}=-2\sum_{k=1}^{k=n}\left(y_k-ax_k^2-bx_k-c\right)x_k=0\\ \dfrac{\partial\varphi}{\partial c}=-2\sum_{k=1}^{k=n}\left(y_k-ax_k^2-bx_k-c\right)=0\end{cases}\Rightarrow\begin{cases}a\sum_{k=1}^{k=n}x_k^4+b\sum_{k=1}^{k=n}x_k^3+c\sum_{k=1}^{k=n}x_k^2=\sum_{k=1}^{k=n}y_kx_k^2\\ a\sum_{k=1}^{k=n}x_k^3+b\sum_{k=1}^{k=n}x_k^2+c\sum_{k=1}^{k=n}x_k=\sum_{k=1}^{k=n}y_kx_k\\ a\sum_{k=1}^{k=n}x_k^2+b\sum_{k=1}^{k=n}x_k+cn=\sum_{k=1}^{k=n}y_k\end{cases}$$

Al dividir por n cada ecuación del sistema de ecuaciones normales, resultan expresiones más simples, y por tanto expresiones más simples para los parámetros a, b y c.

El sistema simplificado es: $\begin{cases}a\cdot\overline{X^2}+b\cdot\overline{X}+c=\overline{Y}\\ a\cdot\overline{X^3}+b\cdot\overline{X^2}+c\cdot\overline{X}=\overline{X\cdot Y}\\ a\cdot\overline{X^4}+b\cdot\overline{X^3}+c\cdot\overline{X^2}=\overline{X^2\cdot Y}\end{cases}$

**c) Modelos no lineales y Método de linealización de datos.**

Antes de presentar el método de linealización, veamos algunos modelos no lineales y sus correspondientes funciones de error:

| Nombre: | Modelo: | Función de error: |
|---|---|---|
| Hiperbólico | $Y=\frac{1}{aX+b}$ | $\varphi(a,b)=\sum_{k=1}^{k=n}\left(\frac{1}{ax_k+b}-y_k\right)^2$ |
| Geométrico | $Y=aX^b$ | $\varphi(a,b)=\sum_{k=1}^{k=n}\left(ax_k^b-y_k\right)^2$ |
| Exponencial | $Y=ce^{aX}$ | $\varphi(a,c)=\sum_{k=1}^{k=n}\left(ce^{ax_k}-y_k\right)^2$ |
| Logarítmico | Y=aln(X)+b | $\varphi(a,b)=\sum_{k=1}^{k=n}\left(a\ln(x_k)+b-y_k\right)^2$ |
| Logístico | $Y=\frac{b}{1+ce^{aX}}$ | $\varphi(a,b,c)=\sum_{k=1}^{k=n}\left(\frac{b}{1+ce^{ax_k}}-y_k\right)^2$ |

Como se puede ver, las funciones de error tienen formas muy complejas que con seguridad, conducirán a sistemas de ecuaciones normales muy difíciles de resolver. Por esta razón, se requiere realizar cambios de variables en el modelo a fin de que el sistema de ecuaciones normales resultante sea lineal. A continuación se muestran algunos ejemplos de cambios de variables convenientes de acuerdo al modelo, en algunos modelos se requiere aplicar logaritmos y por tanto se requiere que los datos sean positivos.

**Modelo hiperbólico:**

Para el modelo $Y=\frac{1}{aX+b}$, es conveniente escribirlo como $aX+b=\frac{1}{Y}$, luego hacer los cambios $u=\frac{1}{Y}$ y v=X, resultando el modelo lineal **u=av + b**.

**Modelo Potencial:**

Para el modelo $Y=aX^b$, se aplican logaritmos y resulta la ecuación logarítmica ln(Y)=ln(a)+bln(X). Después de hacer los cambios u=ln(Y), v=ln(X) y c=ln(a) resulta la ecuación lineal **u = bv + c**.

**Modelo Exponencial:**

Para el modelo $Y=ce^{aX}$, se aplican logaritmos y resulta la ecuación logarítmica ln(Y)=ln(c)+aX. Después de hacer los cambios u=ln(Y), v=X y b=ln(c) resulta la ecuación lineal **u = b + av**.

Para el modelo Y=aln(X)+b, se aplican los cambios u=Y, v=ln(X) y resulta la ecuación lineal **u = av + b**.

**Modelo Logístico:**

Para el modelo $Y=\dfrac{b}{1+ce^{aX}}$, primero se escribe como $1+ce^{aX}=\dfrac{b}{Y}$, luego se escribe como $ce^{aX}=\dfrac{b}{Y}-1$, se aplican logaritmos y resulta la ecuación logarítmica $\ln\left(\dfrac{b}{Y}-1\right)=aX+\ln(c)$. Después de hacer los cambios $u=\ln\left(\dfrac{b}{Y}-1\right)$, v=X y d=ln(c) resulta la ecuación lineal **u = av + d**.

### 1.3. Coeficiente de correlación generalizado.

La correlación es un mecanismo que permite determinar si un par de variables cuantitativas aleatorias guardan o no algún tipo de relación o dependencia.

Para determinar qué tan relacionadas linealmente se encuentran un par de variables X e Y, debemos calcular las desviaciones típicas de las variables X e Y de manera independiente.

Si $\mu_X$ y $\mu_Y$ son los promedios de los valores de las variables X e Y respectivamente, entonces:

$$\delta_X=\sqrt{\frac{1}{n}\sum_{k=1}^{k=n}(x_k-\mu_X)^2}=\sqrt{\frac{1}{n}\sum_{k=1}^{k=n}(x_k)^2-\mu_X{}^2} \qquad \delta_Y=\sqrt{\frac{1}{n}\sum_{k=1}^{k=n}(y_k-\mu_Y)^2}=\sqrt{\frac{1}{n}\sum_{k=1}^{k=n}(y_k)^2-\mu_Y{}^2}$$

Calculamos además la covarianza de las variables X e Y de la siguiente manera:

$$\delta_{XY}=\frac{1}{n}\sum_{k=1}^{k=n}(x_k-\mu_X)\cdot(y_k-\mu_Y)=\frac{1}{n}\sum_{k=1}^{k=n}(x_k\cdot y_k)-(\mu_X\cdot\mu_Y)$$

Ahora, determinamos el coeficiente de correlación con la siguiente expresión: $r=\dfrac{\delta_{XY}}{\delta_X\cdot\delta_Y}$

Los resultados de r se encuentran en el intervalo [-1,1].

El valor r=1 significa que existe una correlación perfecta, es decir que una de las variables depende de la otra de manera directa y lineal. Si 0<r<1 significa que existe correlación positiva es decir que se ajusta a un modelo de la forma Y=mX+b con m>0 como se muestra en la figura1.

Si r=0 no existe correlación lineal, aunque las variables pueden estar relacionadas de manera no lineal. En la figura 2 se ilustra una nube de puntos en la cual no hay correlación.

Si r=-1 entonces existe una correlación negativa perfecta, es decir que los datos reales corresponden a los datos que se obtienen del modelo. Si -1<r<0 existe una correlación negativa, es decir que se ajusta a un modelo de la forma Y=mX+b, con m<0, como se muestra en la figura 3.

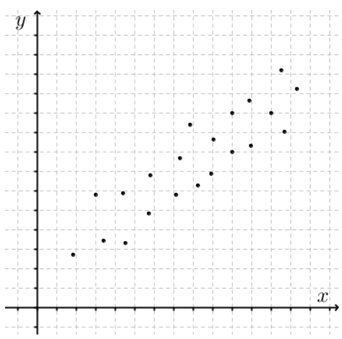

Figura 1

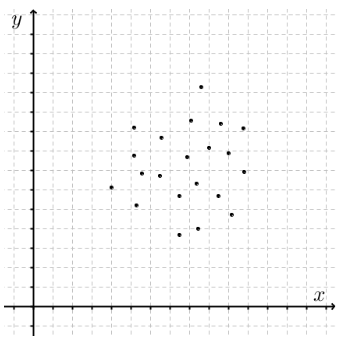

Figura 2

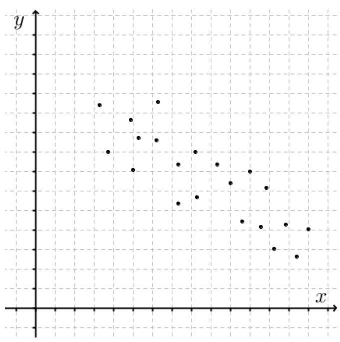

Figura 3

Dado que el coeficiente de correlación lineal se aplica únicamente cuando se presume que los datos se encuentran relacionados de manera lineal, un coeficiente de correlación general para cualquier tipo de regresión puede definirse como el cociente entre la variación total de Y con la variación esperada para esa variable, de la siguiente manera:

$$r = \frac{\sum_{k=1}^{k=n}(y_{est} - \mu_Y)^2}{\sum_{k=1}^{k=n}(y_k - \mu_Y)^2}$$

Donde $y_{est}$ corresponde al valor estimado de Y para un valor dado de X, obtenido de la curva de regresión de Y sobre X.

## Ejercicios

**Ejercicio 1. Un problema de medicina legal.**

El Instituto Nacional de Medicina Legal y Ciencias Forenses, recoge y analiza la información sobre lesiones fatales y no fatales de causa externa para darlas a conocer y coadyuvar con la formulación y seguimiento a las políticas públicas diseñadas con el fin de reducir el fenómeno de la violencia en Colombia.

Es así como a través de los puntos de atención en donde el INMLCF tiene presencia (información directa) y por medio de los médicos rurales en los lugares del país en donde no se cuenta con unidad básica (información indirecta), se ingresa la información que permite caracterizar las lesiones de causa externa mediante variables sociodemográficas, de caracterización del hecho y espacio-temporales. La siguiente tabla contiene datos de los homicidios durante los años 2004 a 2009 en Bogotá:

| **Departamento del hecho** | **2004** | **2005** | **2006** | **2007** | **2008** | **2009** | **Total** |
|---|---|---|---|---|---|---|---|
| Bogotá, D.C. | 1600 | 1689 | 1336 | 1401 | 1465 | 1649 | 9140 |

Teniendo en cuenta el número de homicidios en Bogotá entre los años 2004 y 2009, determine el coeficiente de correlación. En el año 2012, ¿cuántos homicidios podemos suponer que se presentarán en la ciudad de Bogotá?

**Ejercicio 2: Un problema de frecuencia cardiaca**.

En la siguiente tabla se muestra el número de latidos que registran 10 personas cada 6 segundos en relación con el tiempo que gastan en recorrer un kilómetro.

| **Latidos por minuto** | **Tiempo (en minutos)** |
|---|---|
| 52 | 4,7 |
| 55 | 5,0 |
| 58 | 5,8 |
| 60 | 5,7 |
| 65 | 6,1 |
| 70 | 6,4 |
| 72 | 6,8 |
| 75 | 7,4 |
| 80 | 7,2 |
| 82 | 7,8 |

a) Encuentre la recta de regresión lineal.
b) ¿Cuántos latidos por minuto, se espera que registre una persona que gasta 8,5 minutos para recorrer un kilómetro?

**Ejercicio 3: Un problema de población de roedores**.
La tabla muestra la población de cierto tipo de roedores en una zona selvática de Suramérica, registrada al finalizar cada semestre durante seis años.

| Semestre | Población |
|---|---|
| Semestre 1 de 2005 | 12 |
| Semestre 2 de 2005 | 34 |
| Semestre 1 de 2006 | 62 |
| Semestre 2 de 2006 | 96 |
| Semestre 1 de 2007 | 134 |
| Semestre 2 de 2007 | 176 |
| Semestre 1 de 2008 | 222 |
| Semestre 2 de 2008 | 271 |
| Semestre 1 de 2009 | 324 |
| Semestre 2 de 2009 | 379 |
| Semestre 1 de 2010 | 437 |
| Semestre 2 de 2010 | 498 |

a) Encuentre el modelo exponencial que mejor se ajusta a los datos.
b) Pronostique la población para el segundo semestre del 2014.

**Ejercicio 4: Ingresos y Publicidad.**
La tabla muestra la inversión anual en publicidad para promover cierta marca de jabones y los ingresos brutos recibidos por ventas anuales, durante 12 años.

| Año | Inversión (en millones) | Ingresos (en millones) |
|---|---|---|
| 2000 | 10 | 200 |
| 2001 | 20 | 230 |
| 2002 | 30 | 260 |
| 2003 | 40 | 290 |
| 2004 | 50 | 240 |
| 2005 | 60 | 280 |
| 2006 | 70 | 320 |
| 2007 | 80 | 380 |
| 2008 | 90 | 320 |
| 2009 | 100 | 380 |
| 2010 | 110 | 440 |
| 2011 | 120 | 500 |

a) Encuentre el modelo Geométrico que mejor se ajusta a los datos.
b) Pronostique los ingresos para cuando se inviertan 150 millones de pesos en publicidad.

**Ejercicio 5: ¿Cómo elegir el modelo apropiado?**
Retomemos el ejemplo anterior: La tabla muestra la inversión anual en publicidad para promover cierta marca de jabones y los ingresos brutos recibidos por ventas anuales, durante 12 años.

| Inversión (en millones) | Ingresos (en millones) |
|---|---|
| 10 | 200 |
| 20 | 230 |
| 30 | 260 |
| 40 | 290 |
| 50 | 240 |
| 60 | 280 |
| 70 | 320 |
| 80 | 380 |
| 90 | 320 |
| 100 | 380 |
| 110 | 440 |
| 120 | 500 |

Comparar los modelos lineal, cuadrático, geométrico, exponencial y logarítmico, y determinar cuál es el más apropiado para realizar pronósticos de los ingresos.

**Ejercicio 6: Censo electoral**

Según datos de la Registraduría Nacional del Estado Civil, la abstención ha tenido altibajos del año 1958 al año 2006, como se muestra en la tabla. Determine entre los modelos lineal, cuadrático e hiperbólico, cuál es el más apropiado para pronosticar la abstención en futuras elecciones.

| Elecciones presidenciales 1958 - 2006 | | | |
|---|---|---|---|
| **Año** | **Censo** | **Votación total** | **Abstención** |
| 1958 | 5.365.191 | 3.108.567 | 42.29% |
| 1962 | 5.404.765 | 2.634.840 | 51.25% |
| 1966 | 6.611.352 | 2.649.258 | 59.93% |
| 1970 | 7.683.785 | 4.036.458 | 47.47% |
| 1974 | 8.925.330 | 5.218.855 | 41.53% |
| 1978 | 12.580.851 | 5.075.719 | 59.66% |
| 1982 | 13.734.093 | 6.834.250 | 50.24% |
| 1986 | 15.611.274 | 7.228.676 | 53.70% |
| 1990 | 14.237.110 | 6.047.576 | 57.52% |
| 1994* | 17.146.597 | 7.427.742 | 56.68% |
| 1998* | 20.857.801 | 12.310.107 | 59.01% |
| 2002 | 24.208.311 | 11.249.734 | 53.53% |
| 2006 | 26.731.700 | 12.041.737 | 54.95% |

* Segunda vuelta.

**Ejercicio 7: Paracaidismo**

Para conocer la relación entre la velocidad de caída de un paracaidista y la fuerza de fricción hacia arriba, se han realizado las siguientes mediciones:

| | | | | | | |
|---|---|---|---|---|---|---|
| **Velocidad** | 1.5 | 2.5 | 3.5 | 4.5 | 5.5 | 6.5 |
| **Fricción** | 6.0 | 15.5 | 30.0 | 46.5 | 66.5 | 90.5 |

La velocidad se mide en centímetros por segundo y la fricción en dinas. Elabore el diagrama de dispersión y obtenga la curva parabólica que mejor se ajusta a los datos.

**Ejercicio 8: Población de roedores**

Cuando una población P(t) tiende a cierto valor L a medida que transcurre el tiempo t, la gráfica de P es una curva conocida como curva logística y definida como $P(t) = \frac{L}{1 + ce^{at}}$. Encuentre la curva logística que se ajusta a la siguiente tabla, considerando que la población tiende a L=1200.

| **t** | **0** | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|---|
| P(t) | 180 | 350 | 600 | 800 | 950 | 1100 |

**Ejercicio 9: Créditos hipotecarios**

Las tasas de interés aplicadas a los créditos hipotecarios en los últimos años han influido en el número de construcciones iniciadas en el país. Los datos que se muestran a continuación muestran las tasas de interés por trimestre y la cantidad de construcciones iniciadas en este semestre en una ciudad dada:

| **Mes** | **X = Tasa de interés (%)** | **Y = Cantidad de Construcciones iniciadas** |
|---|---|---|
| Enero | 11,5 | 260 |
| Febrero | 11,4 | 250 |
| Marzo | 11,6 | 241 |
| Abril | 12,4 | 256 |
| Mayo | 12,8 | 270 |
| Junio | 13,2 | 220 |
| Julio | 13,5 | 190 |
| Agosto | 13,0 | 195 |
| Septiembre | 12,7 | 200 |
| Octubre | 12,9 | 210 |
| Noviembre | 12,5 | 230 |
| Diciembre | 12,0 | 245 |

a) Encuentre la recta de regresión lineal para estos datos.

b) Pronostique el número de construcciones iniciadas que corresponden a una tasa de interés de 11,0%.

**Ejercicio 10: Nutrición**

Un nutricionista planteó una dieta de adelgazamiento para un grupo de pacientes. En la tabla aparecen los datos sobre el promedio de kilos perdidos desde el inicio de la dieta (variable Y) y la cantidad de semanas que llevan siguiendo la dieta (Variable X).

| **X** | 3 | 5 | 6 | 8 | 11 | 13 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|
| **Y** | 2.5 | 5.5 | 5.5 | 8.5 | 10.5 | 13.5 | 15.0 | 15.0 |

a) Determine el modelo de regresión que explica el peso perdido en función del tiempo que se lleva siguiendo la dieta.

b) Según el modelo considerado, ¿qué peso esperaría perder una persona que siga la dieta durante 8 semanas?

**Ejercicio 11: El mejor modelo**

Las funciones $Y = 4.5 - 0.121X$, $Y = 0.0054X^2 - 0.3065X + 5.3333$, $Y = 8.4242X^{-0.5644}$, $Y = 4.9408e^{-0.0516X}$ y $Y = 6 - 1.4427\ln(X)$, son modelos que se ajustan a los datos que se muestran en la tabla. Determine cuál de las cinco curvas es la que mejor se ajusta.

| **X** | 2 | 4 | 8 | 16 | 32 |
|---|---|---|---|---|---|
| **Y** | 5 | 4 | 3 | 2 | 1 |

**Ejercicio 12: Población**

A continuación se presenta una tabla incompleta de la población en las ultimas 11 décadas de una ciudad cosmopolita de acuerdo a censos realizados. Se debe llenar la tabla con ayuda de un modelo de regresión apropiado.

| X: Año | Y: Población |
|---|---|
| 1900 | 1650 |
| 1910 | 2400 |
| 1920 | 4250 |
| 1930 | 8250 |
| 1940 | 17.750 |
| 1950 | ? |

| X: Año | Y: Población |
|---|---|
| 1960 | 60.550 |
| 1970 | 125.500 |
| 1980 | 270.550 |
| 1990 | 520.400 |
| 2000 | 1.150.000 |
| 2010 | ? |

**Ejercicio 13: Oferta y demanda**

En mercadeo, la ecuación de demanda relaciona el precio de un artículo con la cantidad de unidades que los consumidores están dispuestos a comprar a ese precio. Generalmente a mayor precio la demanda baja y a menor precio la demanda sube. En la siguiente tabla se muestra el precio p de un artículo (en miles de pesos) y la demanda q (en miles de unidades) en diferentes periodos. Encuentre la curva de regresión que mejor se ajusta a los datos, como una función hiperbólica de p en términos de q.

| **q** | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
|---|---|---|---|---|---|---|---|---|---|
| **p** | 100 | 70 | 50 | 55 | 50 | 45 | 40 | 35 | 35 |

**Ejercicio 14: Problema de Población**

La siguiente tabla muestra la población de Estados Unidos cada 20 años, desde el año 1820 hasta el año 2000. Encuentre las curvas de regresión exponencial y geométrica asociadas a la tabla de datos, y determine de dos maneras diferentes cuál es la más apropiada para pronosticar la población aproximada en el año 2020.

| **Año** | **1820** | **1840** | **1860** | **1880** | **1900** | **1920** | **1940** | **1960** | **1980** | **2000** |
|---|---|---|---|---|---|---|---|---|---|---|
| Pobl | 1,0 | 1,7 | 3,1 | 5,0 | 7,6 | 10,6 | 13,2 | 17,9 | 22,7 | 28,1 |

**Ejercicio 15: Evolución de una Virosis**

Se desea estudiar la evolución de una virosis en un instituto que tiene 520 alumnos, por tanto el contagio sólo puede extenderse, como máximo a 520 alumnos. Dos años antes una virosis similar evolucionó de la manera como se muestra en la tabla, donde S representa el número de semanas transcurridas desde el momento que apareció la virosis y A representa la cantidad de alumnos contagiados. Aprovechando los datos históricos, encuentre la curva logística que mejor se ajusta a la propagación de la virosis y determine en qué semana se habrá enfermado la mitad de los alumnos.

| **S** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **A** | 14 | 28 | 34 | 56 | 68 | 114 | 158 | 226 | 326 | 400 | 438 | 474 |