

## **Laboratorio N.4**

Introduccion a Los Metodos Estadisticos  
Prueba de Hipotesis y Regresion

Diana Carolina Arias Sinisterra Cod. 1528008

Kevin Steven Garcia Chica Cod. 1533173

Cesar Andres Saavedra Vanegas Cod. 1628466

**Universidad Del Valle**

Facultad De Ingenieria

Estadistica

Diciembre

2017

# Índice

<b>1. Prueba De Hipotesis</b>	<b>3</b>
1.1. Situación 1 . . . . .	3
1.2. Situación 2 . . . . .	4
1.3. Punto A. . . . .	4
1.4. Punto B. . . . .	4
1.5. Situación 3 . . . . .	5
1.6. Punto A. . . . .	5
1.7. Punto B. . . . .	5
1.8. Situación 4 . . . . .	6
1.9. Situación 5 . . . . .	7
1.10. Punto A. . . . .	7
1.11. Punto B. . . . .	7
1.12. Punto C. . . . .	7
1.13. Situación 6 . . . . .	8
1.14. Punto C. . . . .	8
1.15. Punto D. . . . .	8
1.16. Situación 7 . . . . .	9
1.17. Punto A. . . . .	9
1.18. Punto B. . . . .	9
<b>2. Regresion</b>	<b>10</b>
2.1. Situación 1 . . . . .	10
2.2. Punto A. . . . .	10
2.3. Punto B. . . . .	11
2.4. Punto C. . . . .	12
2.5. Punto D. . . . .	12
2.6. Punto E. . . . .	13
2.7. Punto F. . . . .	13
2.8. Punto G. . . . .	14
2.9. Situación 2 . . . . .	15
2.10. Punto A. . . . .	15
2.11. Punto B. . . . .	15
2.12. Punto C. . . . .	15

## Índice de figuras

1. Gráfica de dispersión con recta de ajuste entre las variables X y Y . . . . .	11
--	----

# **1. Prueba De Hipotesis**

## **1.1. Situación 1**

**1.2. Situación 2**

**1.3. Punto A.**

**1.4. Punto B.**

**1.5. Situación 3**

**1.6. Punto A.**

**1.7. Punto B.**

## **1.8. Situación 4**

**1.9. Situación 5**

**1.10. Punto A.**

**1.11. Punto B.**

**1.12. Punto C.**

**1.13. Situación 6**

**1.14. Punto C.**

**1.15. Punto D.**



**1.16. Situación 7**

**1.17. Punto A.**

**1.18. Punto B.**

## 2. Regresion

### 2.1. Situación 1

### 2.2. Punto A.

- Reduccion porcentual del total de solidos: El total de sólidos disueltos (a menudo abreviado como TDS, del inglés: Total Dissolved Solids) es una medida del contenido combinado de todas las sustancias inorgánicas y orgánicas contenidas en un líquido en forma molecular, ionizada o en forma de suspensión micro-granular (sol coloide). Los TDS (Total dissolved solids) son la suma de los minerales, sales, metales, cationes o aniones disueltos en el agua. Esto incluye cualquier elemento presente en el agua que no sea (H<sub>2</sub>O) molécula de agua pura y sólidos en suspensión. (Sólidos en suspensión son partículas ó sustancias que ni se disuelven ni se asientan en el agua, tales como pulpa de madera.) En general, la concentración de sólidos disueltos totales es la suma de los cationes (carga positiva) y aniones (cargado negativamente) iones en el agua.
- Reduccion porcentual de demanda bioquimica de oxigeno: La demanda bioquímica de oxígeno (DBO) es un parámetro que mide la cantidad de dióxígeno consumido al degradar la materia orgánica de una muestra líquida.

### 2.3. Punto B.

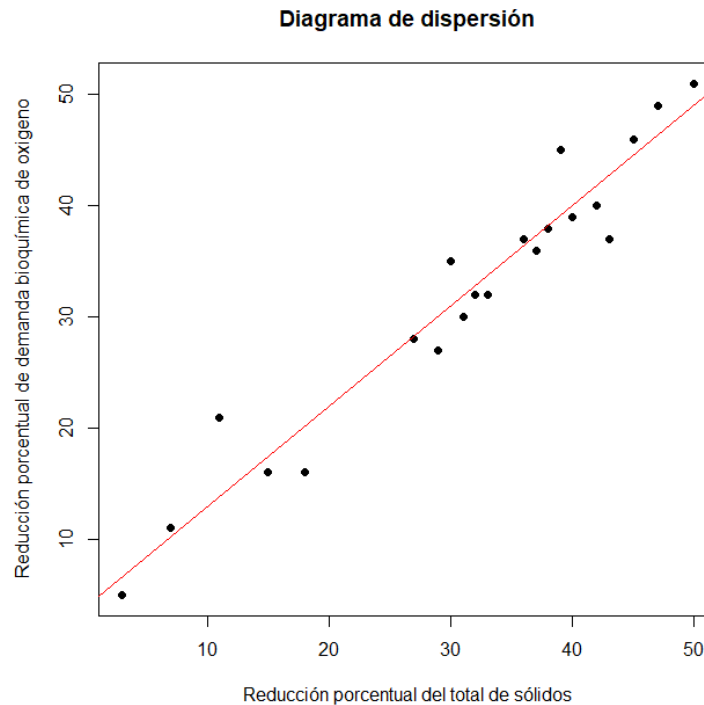


Figura 1: Gráfica de dispersión con recta de ajuste entre las variables X y Y

En esta imagen podemos ver que hay una correlación positiva bastante fuerte, por lo que esperamos que el coeficiente de correlación sea cercano a 1. También podemos ver que las distancias de los datos o los puntos a la recta de ajuste son considerablemente pequeñas, por lo cuál creemos que el  $R^2$  es también cercano a 1, diciéndonos esto, que el modelo ajustado representara en un alto porcentaje la variación total de Y.

Procedemos a calcular los dos valores mencionados:

■ Coeficiente de correlación:  $\hat{\rho} = r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$

$\bar{x} = 31.095238$  y  $\bar{y} = 31.9523809$ , entonces:

$$\hat{\rho} = r = \frac{3202.095238}{\sqrt{3543.809524} \cdot \sqrt{3086.952381}} = \frac{3202.095238}{3307.502267} = 0.9681$$

■ Coeficiente de determinación  $R^2$ :  $R^2 = r^2 = 0.9372$

Tal y como esperabamos, el coeficiente de correlación lineal, nos arrojo un resultado de 0.9681 que es bastante alto. Este resultado nos dice que a medida que X aumenta, Y tambien aumenta en una proporción aproximada de 0.9681. Con respecto al  $R^2$ , también nos arrojó

un valor que esperábamos (0.9372) que es bastante alto; este valor nos dice que el 93.72 % de la variabilidad total de la variable Y, es explicada por la variable X. Entonces, según lo anterior, concluimos que estas dos variables tienen una relación positiva demasiado fuerte, es decir, cuando aumenta X, Y también aumentará, y por el  $R^2$  concluimos que X es una buena variable explicativa para Y.

## 2.4. Punto C.

Para evaluar la asociación lineal entre las variables X y Y, tendremos que plantear las hipótesis sobre  $\rho$  de la siguiente manera:

$H_0: \rho = 0$  (No hay correlación lineal)

$H_1: \rho \neq 0$  (Hay correlación lineal)

$$T_\rho = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

Reemplazando los valores, teniendo en cuenta que por el punto anterior  $r = 0.9681$ , tenemos:

$$T_\rho = \frac{0.9681}{\sqrt{\frac{1 - (0.9681)^2}{21 - 2}}} = 16.84139$$

Tomando  $\alpha = 0.05$ ,  $T_{(0.975; 19)} = 2.093$

Como  $T_\rho = 16.84139 > 2.093$ , rechazamos  $H_0$  y concluimos que con una confianza del 95 %, si existe correlación lineal entre las dos variables.

## 2.5. Punto D.

El modelo ajustado es de la forma:  $Y_i = \beta_0 + \beta_1 X_i + e_i$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{24067 - 21(31.095238)(31.9523809)}{23849 - 21(31.095238)^2} = \frac{3202.095336}{3543.809648} = 0.903574$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 31.9523809 - (0.903574 \cdot 31.095238) = 3.855532$$

El modelo ajustado queda:  $Y_i = 3.855532 + 0.903574 X_i + e_i$

### INTERPRETACIÓN:

$\hat{\beta}_0 = 3.855532$  : Sin tener en cuenta la variabilidad de la reducción porcentual del total de sólidos, se espera una reducción porcentual de demanda bioquímica de oxígeno de 3.855532.

$\hat{\beta}_1 = 0.903574$  : Por cada unidad adicional de la reducción porcentual del total de sólidos, se espera que la reducción porcentual de demanda bioquímica de oxígeno aumente en promedio en 0.903574 unidades.

## 2.6. Punto E.

Los indicadores utilizados para evaluar la bondad de ajuste de un modelo son el  $\rho$  y el  $R^2$ , como ya los obtuvimos en el punto b, vamos a interpretarlos.

$\rho = 0.9681$ : Este valor nos indica que existe una correlación positiva bastante fuerte (casi perfecta), es decir, cuando  $x$  aumenta,  $y$  aumenta casi en la misma proporción. Cuando tenemos una correlación tan alta entre las dos variables, podemos estar seguros de que el modelo ajustado será un buen modelo ( $R^2$  tendiendo a 1) para explicar  $Y$  en términos de  $X$ .

$R^2 = 0.9372$ : Este valor nos dice que el 93.72% de la variabilidad total de variable  $Y$  (reducción porcentual de demanda bioquímica de oxígeno) es explicada por la variable  $X$  (reducción porcentual del total de sólidos). Lo que en el fondo nos dice que el modelo es bastante bueno, ya que las distancias de los datos o los puntos a la recta de regresión ajustada, son muy pequeñas.

## 2.7. Punto F.

**Para  $\beta_0$ :**  $\langle \beta_0 \rangle_{(1-\alpha)\%} = \langle \hat{\beta}_0 \pm t_{(\frac{\alpha}{2}, n-2)} \sqrt{V(\hat{\beta}_0)} \rangle$

$$V(\hat{\beta}_0) = \frac{\hat{\sigma}^2 \sum_{i=1}^n x_i^2}{nS_{xx}}$$

$$\hat{\sigma}^2 = \frac{1}{n-2} [S_{yy} - \hat{\beta}_1 S_{xy}] = \frac{1}{21-2} [3086.952381 - (0.903574 \cdot 3202.095238)] = \frac{1}{19} (193.6223784) = 10.1906515$$

$$\text{Entonces, reemplazando: } V(\hat{\beta}_0) = \frac{10.1906515 \cdot 23849}{21 \cdot 3543.809524} = 3.265746$$

Ahora, con  $\alpha = 0.05$ :

$$\langle \beta_0 \rangle_{0.95\%} = \langle 3.855532 \pm 2.093 \cdot \sqrt{3.265746} \rangle$$

$$(0.073193; 7.6378708)$$

**Para  $\beta_1$ :**  $\langle \beta_1 \rangle_{(1-\alpha)\%} = \langle \hat{\beta}_1 \pm t_{(\frac{\alpha}{2}, n-2)} \sqrt{V(\hat{\beta}_1)} \rangle$

$$V(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{xx}}$$

Ya sabemos que  $\hat{\sigma}^2 = 10.1906515$

$$\text{Entonces: } V(\hat{\beta}_1) = \frac{10.1906515}{3543.809524} = 0.00287562$$

Ahora, con  $\alpha = 0.05$ :

$$\langle \beta_1 \rangle_{0.95\%} = \langle 0.903574 \pm 2.093 \cdot \sqrt{0.00287562} \rangle$$

$$(0.791337; 1.0158107)$$

## 2.8. Punto G.

### ESTIMACIÓN PUNTUAL:

$E[Y|X] = E[\beta_0 + \beta_1 x_1 + e] = E[\beta_0] + E[\beta_1 x_1] + E[e]$ , sabemos que  $E[e] = 0$ , por los supuestos del modelo lineal.

Entonces:  $E[\hat{Y}|X] = \hat{\beta}_0 + \hat{\beta}_1 x_1$

Por consiguiente:  $E[Y|X = 30] = \hat{\beta}_0 + \hat{\beta}_1 \cdot 30 = 3.855532 + 0.903574(30) = 30.962752$

El valor esperado de la reducción de la demanda bioquímica de oxígeno cuando se reduce a 30 % el porcentaje total de sólidos es de 30.962752 %.

### ESTIMACIÓN POR INTERVALOS:

$$\langle E[Y|X = 30] \rangle_{0.95\%} = \langle E[Y|\hat{X} = 30] \rangle \pm t_{(\frac{\alpha}{2}; n-2)}$$

**2.9. Situación 2**

**2.10. Punto A.**

**2.11. Punto B.**

**2.12. Punto C.**