

CAPÍTULO II

2. MARCO TEÓRICO

2.1. Introducción

En este capítulo se presenta una recopilación de toda la información relacionada directamente o indirectamente al tema de investigación.

Esta información se la ha dividido en dos partes, estas son:

Sigatoka negra: Esta primera parte incluye una breve reseña histórica de la Sigatoka negra, una descripción detallada del hongo *Mycosphaerella Fijiensis* causante de esta enfermedad además formas de evaluar la enfermedad.

Análisis estadístico: En esta parte se presenta la teoría relacionada al análisis univariado y de regresión además se define la estructura de las pruebas de hipótesis involucradas en el análisis y se menciona algo sobre el software estadístico utilizado en el estudio.

2.2. Sigatoka negra

2.2.1. Reseña histórica⁽⁴⁾

La Sigatoka negra se detectó por primera vez en el continente americano en el valle de Ulúa-Honduras en el año de 1972, en bananos y plátanos, actualmente se encuentra difundida en las principales áreas productoras de *musáceas* en todo el mundo.

En Cuba se detectó por primera vez en el año 1990.

⁽⁴⁾ Centro de Investigaciones Biotecnológicas del Ecuador

En América Central se detectó en Belice en el año 1976, en Guatemala en el año 1977, en Costa Rica en el año 1979, en El Salvador y Nicaragua en el año 1979, en Panamá y México en el año 1980 y en Sur América en Colombia en el año 1981, en Venezuela en el año 1991, en Perú en el año 1994 y en Brasil en el año 1998.

En el Ecuador se detectó por primera vez en año 1987 en las haciendas Timbre, Flamingo y Victoria, localizadas en la provincia de Esmeraldas y en la actualidad se ha extendido a todas las áreas bananeras del Ecuador y ya se ha reportado su presencia en pequeñas áreas cultivadas en la provincia insular de Galápagos.

2.2.2. Nombres comunes de la enfermedad

Algunos de los nombres comunes con los que se conoce a la enfermedad son Sigatoka negra del plátano, Raya negra de la hoja, Black Sigatoka, Black leaf streak.

2.2.3. *Mycosphaerella Fijiensis*⁽⁵⁾

Organismo causal:

Existen algunas especies del hongo *Mycosphaerella*, todas estas atacan al género musa. El agente causal de la Sigatoka negra es el hongo *Mycosphaerella Fijiensis*, fue en América Latina donde se le dio el nombre de Sigatoka negra.

⁽⁵⁾ Evaluating bananas

Sintomatología:

El único órgano afectado por el hongo *Mycosphaerella Fijiensis* es la hoja, el cual destruye el follaje rápidamente si no se aplican medidas de control.

Asimismo, como resultado del ataque de este patógeno se observan síntomas característicos en los frutos.

Aunque la Sigatoka negra fue descrita por primera vez el año 1964, los síntomas en hojas fueron detallados recién el año 1969 por Meredith y Lawrence.

En 1982, Fouré redefinió los síntomas mostrados durante el desarrollo de la enfermedad en seis estadíos, que son descritos a continuación.

- **Estadio 1.** Es el primer síntoma externo de la enfermedad, las plantas afectadas por Sigatoka negra presentan puntos oscuros, café rojizo menor de 0,25 mm de longitud, visible a simple vista en el envés de las hojas, generalmente más abundante cerca del margen derecho de la hoja (vista de frente) y hacia el ápice. Esto se asemeja a la primera etapa de la Sigatoka amarilla.

- **Estadio 2.** En este síntoma las estrías aparecen entre 10 a 14 días después de la infección.

Dicha infección aparece como una raya, generalmente de color café y visible en el envés de la hoja; más adelante, este síntoma también aparece como una raya en la parte de arriba del limbo, su color amarillo se asemeja al estadio 1 de la Sigatoka amarilla.

Este color cambiará progresivamente a café y más adelante a negro en la parte de arriba de la hoja; sin embargo mantendrá el color café en el envés de la hoja.

- **Estadio 3.** Se diferencia del anterior en sus dimensiones, la raya se hace más grande y bajo ciertas condiciones (poco inóculo y condiciones climáticas desfavorables), puede alcanzar una longitud de 2 cm. a 3 cm de longitud.
- **Estadio 4.** Aparece en el envés de la hoja como una mancha café y en la parte de arriba como una mancha negra.
- **Estadio 5.** Ocurre cuando la mancha elíptica se vuelve totalmente negra y se ha extendido al envés de la hoja. Esta mancha tiene un halo amarillo que la rodea y su centro se empieza a aplanar.
- **Estadio 6.** Ocurre cuando el centro de la mancha se seca, adquiere un color gris claro y lo rodea un anillo bien definido de

color negro, rodeado a su vez por un halo de color amarillo brillante. Estas manchas se podrán observar aún después de que la hoja se ha secado ya que el anillo persiste.

Como resultado se tiene la destrucción del follaje, además destruye su capacidad de respiración y fotosíntesis, la producción de fruta es reducida drásticamente, sin embargo, también ocasiona un llenado deficiente y la maduración prematura de los dedos en campo, durante el transporte o almacenaje, daño que es considerado como el más importante.

Este daño, estaría determinado por la acción de toxinas del patógeno, más que debido a la defoliación.

Si el ataque se produce sobre todo en las hojas jóvenes de la planta, la pérdida de la producción es total.

Condiciones ambientales

La Sigatoka negra es favorecida por humedad y temperatura altas.

Debido a que *Mycosphaerella Fijiensis* es un hongo de reproducción sexual (ascoporas) y asexual (conidias), la infección se produce más rápidamente por medio de las ascoporas debido a propagación a través del viento. Sin embargo, las conidias también juegan un rol muy importante durante periodos secos.

Las conidias se forman bajo condiciones de alta humedad, especialmente si hay una película de agua sobre la superficie de la

hoja. Se forman durante los primeros estadios de desarrollo de la enfermedad, especialmente 2, 3 y 4.

Las ascosporas son producidas en los pseudotecios formados en las lesiones maduras (estadios 5 y 6), que son comunes en las hojas más viejas de las plantas o en hojas que se están secando sobre la superficie del suelo.

La infección exitosa es promovida por periodos extendidos de alta humedad y agua libre sobre las hojas. La máxima germinación ocurre cuando hay agua libre. Las conidias germinan a través de un rango de humedad relativa más amplio que va desde 92 a 100%, que las ascosporas (98-100%). La temperatura óptima para la germinación de las esporas es de alrededor 27 °C.

Dentro de una plantación de plátano, las condiciones culturales que determinen la ocurrencia de ambientes húmedos, entre ellas drenajes deficientes y acumulación de agua en el suelo, densidades de plantas muy altas, presencia de malezas, falta de deshierpes, favorecerán significativamente el desarrollo de la enfermedad al permitir los procesos de germinación y penetración de las esporas.

2.2.4. Formas de evaluar la enfermedad

Es necesario tener un control del estado sanitario de la finca para prevenir daños severos en el cultivo y su producción. Además deben hacerse evaluaciones periódicas (semanales o quincenales) sobre la evolución y severidad de la Sigatoka negra, a continuación se presentan 2 formas que son los más utilizados para este control:

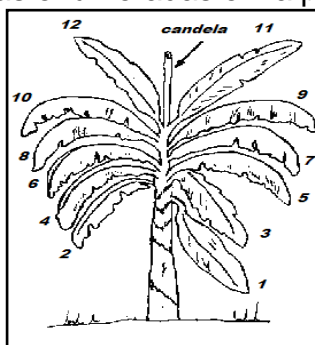
- Preaviso biológico.
- Evaluación de severidad (Stover).

2.2.4.1. Preaviso Biológico

Esta forma de evaluar consiste en la detección temprana de los síntomas de las hojas más jóvenes de la planta, debido a que las manchas de la Sigatoka negra aparecen más temprano atacando las hojas número 2, 3 y 4.

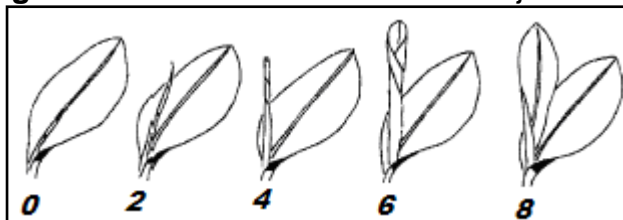
La primera variable que se mide para obtener las variables que evalúan el desarrollo y severidad es la emisión foliar, para ello se marcan de abajo hacia arriba el número de hojas emitidas y se identifica la hoja candela tal como se muestra en la figura 2.1.

Figura 2.1 Hojas enumeradas en la planta de banano



El valor que toma variable emisión foliar está compuesta por una parte entera y una parte decimal. La parte entera está determinada por el número de hojas brotadas que tiene la planta y la parte decimal esta definida por el estado evolutivo de la hoja candela, para esta observación se consideran los estados de desarrollo descritos por Brun (1963) que se presentan en la figura 2.2.

Figura 2.2 Estados evolutivo de la hoja candela



Las variables restantes que permiten evaluar el desarrollo y severidad son las que cualitativamente miden el estado de las tres primeras hojas (HIII, HIII, HIV) contadas a partir de la hoja candela, para esto se toman en cuenta el número de pizcas o estrías presentes en las respectivas hojas y se las califica tomando en cuenta las siguientes consideraciones:

- +1 si existen más de 50 pizcas que indican la presencia de SN
- 1 si existen menos de 50 pizcas que indican la presencia de SN
- +2 si existen más de 50 estrías que indican la presencia de SN
- 2 si existen menos de 50 estrías que indican la presencia de SN

La hoja de cálculo que se utiliza para registrar estas variables tiene el formato mostrado en la Figura 2.3

Al introducir las variables antes mencionadas en la tabla mostrada en la figura siguiente se obtienen las variables que permiten evaluar la evolución y severidad de la Sigatoka negra las cuales serán analizadas en el estudio estas son: SH2, SH3, SH4, EE y SEV.

Figura 2.3 Hoja de cálculos relacionados a la Sigatoka negra

TABULACION DE DATOS DE PREAVISO DE SIGATOKA NEGRA																			
HDA:		Carmita				LOTE:		sector C		Semana:		50							
Prov.		Balao				Fecha:		16-dic-05		Evaluador:									
10/01/2004 Lote 12													ESTADO DE ENFERMEDAD		HOJAS DAÑADAS		POR COMPONENTE		
Planta	E F P	E F A	C A N	R E F	C C	H J E	H J N	HII	HIII	HIV	CHE	Grados sintomas		II	III	IV	II	III	IV
1	17,2	17,8	8	0,6	16	5	8	-2	2	2	-1	0	4	0	0	0	160	0	
2	17,2	18,0	0	0,8	0	6	10	-2	2	2	1	0	5	2	0	300	80		
3	17,0	17,6	6	0,6	12	5	10	-2	2	2	-2	0	5	4	0	400	240		
4	16,8	17,2	2	0,4	4	6	8	-2	2	2	2	0	1	9	0	100	720		
5	15,4	16,2	2	0,8	4	6	9	-1	1	2	-3	0	0	0	0	0	0		
6	16,8	17,2	2	0,4	4	6	8	1	1	2	3	0	0	0	0	0	0		
7	17,2	17,6	6	0,4	12	5	9	1	-2	2	-4	0	0	0	0	0	0		
8	17,8	18,6	6	0,8	12	6	11	1	2	2	4	0	0	0	0	0	0		
9	18,8	19,2	2	0,4	4	5	10	-1	2	2									
10	17,8	18,4	4	0,6	8	5	9	-1	-2	2				SUMA	0	960	1040		
11	15,8	16,2	2	0,4	4	6	8	1	-2	2				SB			2000		
12	17,2	18,0	0	0,8	0	6	9	-1	-2	2				SEV			1792		
13	15,2	15,8	8	0,6	16	5	8	1	2	2				EE			2427		
14	15,8	16,2	2	0,4	4	5	8	2	2	2									
15	17,6	18,2	2	0,6	4	5	8	-2	2	2									
			8,6	104															
REFa=	1,23																		
N=	7																		
REFp=	1,48	REFx	1,354285714																
CE	208																		

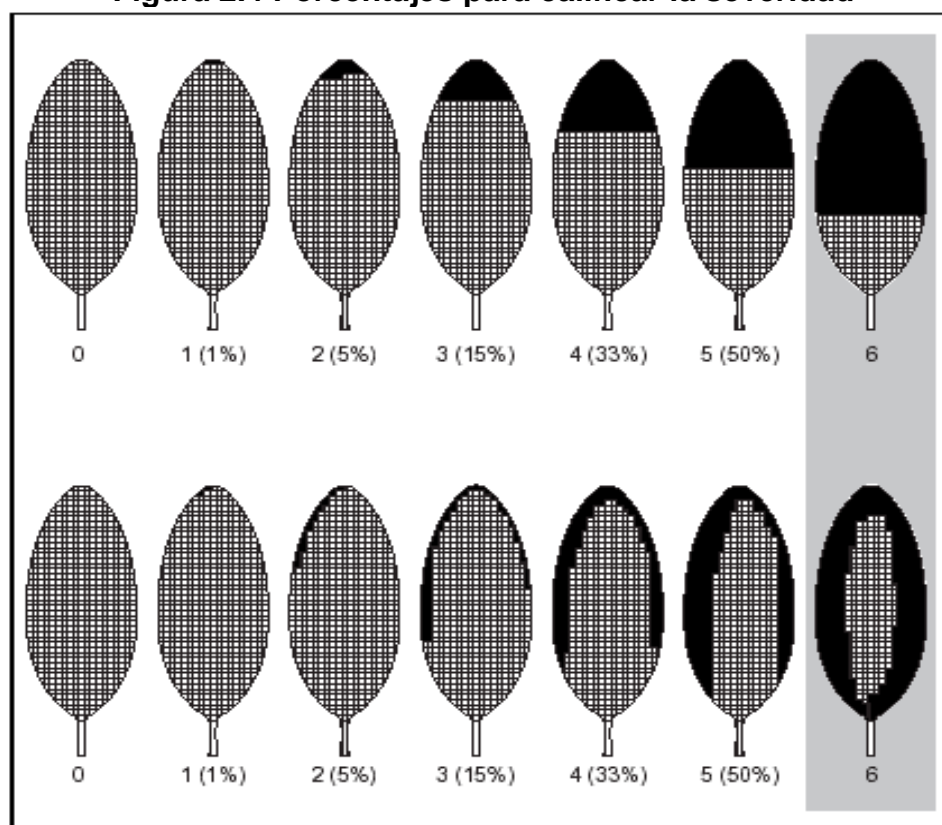
Fuente: CIBE

2.2.4.2. Severidad

Esta forma de evaluar consiste en examinar todas y cada una de las hojas de la planta de banano considerando el porcentaje del área afectada sobre la cada hoja el cual se estima visualmente.

La calificación según los porcentajes toma valores entre 1-6 como se presenta en la Figura 2.4 mostrada a continuación:

Figura 2.4 Porcentajes para calificar la severidad



2.3. Análisis Estadístico

2.3.1. Análisis Univariado⁽⁶⁾

El análisis univariado en un sentido amplio, se refiere a todos los métodos estadísticos que analizan la distribución de una sola variable.

Para el análisis en este estudio se consideraran las siguientes medidas descriptivas:

De posición

- Cuartiles
- Mínimo
- Máximo

Centralización o tendencia central

- Media
- Mediana

Dispersión

- Varianza y desviación estándar

Forma

- Sesgo o asimetría
- Curtosis o puntiagudez

Referente a los gráficos que se consideraran en el estudio se tiene:

- Diagrama de caja

⁽⁶⁾ Sitio Web Docente

- Histograma

2.3.2. Análisis de Regresión

2.3.2.1. Reseña histórica⁽⁷⁾

El término regresión fue introducido por Sir Francis Galton (1822-1911) cuando estudiaba la eugénica, término también introducido por él para definir la mejora de la raza humana en base a los caracteres hereditarios.

Esto consistía en estudiar la altura de los hijos en relación a la altura de los padres, y probó que la altura de los hijos altos “regresaba” hacia la altura media de la población a lo largo de sucesivas generaciones, es decir, hijos de padres extraordinariamente altos tendían a ser en promedios más bajos que sus padres y viceversa.

Actualmente el término regresión se utiliza siempre que se busca predecir una variable en función de otra.

2.3.2.2. Regresión⁽⁸⁾

Si, basados en una relación teórica, se tiene observaciones relacionadas (X e Y), para suponer que las variaciones de Y son causadas por las variaciones X , podemos utilizar el análisis de regresión para probar la relación matemática que existe entre estas variables.

⁽⁷⁾ Sitio Web IES

⁽⁸⁾ Estadística Matemática con Aplicaciones

Al conjunto de observaciones Y se los denotará como variable endógena, explicada, **dependiente** o respuesta.

Al conjunto de observaciones X se los denotará como variable exógena, explicativa, **independiente**, predictora, o regresora.

Teniendo en cuenta que un modelo es un conjunto de restricciones sobre la distribución conjunta de las variables dependientes e independientes, entonces los objetivos de un modelo de regresión puede ser dos:

Predictivo, en el que el interés del investigador es predecir lo mejor posible la variable dependiente, usando un conjunto de variables independientes, es decir, obtener una ecuación que nos permita predecir la variable dependiente.

Estimativo, en el que el interés se centra en estimar la relación de una o más variables independientes con la variable dependiente, es decir cuantificar la relación entre las variables X y la variable Y .

2.3.2.3. Regresión lineal simple

La técnica de regresión lineal simple es utilizada cuando se pretende explicar una variable respuesta cuantitativa en función de una variable explicativa también cuantitativa. Consideremos esta variable aleatoria dependiente Y , que supondremos relacionada con la otra variable (no necesariamente aleatoria) que llamaremos independiente denotada por X .

A partir de una muestra de n observaciones para las que se dispone de los valores de ambas variables, $\{X_i, Y_i; i = 1, 2, \dots, n\}$ el problema que subyace a la técnica de la regresión lineal simple es el de encontrar una recta que se ajuste a la nube de los n puntos $\{X_i, Y_i\}$ una vez dispuestos en un plano x e y , la misma que pueda ser utilizada para predecir los valores de Y a partir de los de X .

El modelo intenta aproximar la variable respuesta mediante una función lineal de la variable explicativa de la siguiente forma:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad ; \quad i = 1, 2, \dots, n$$

Donde el término Y_i es el escalar que se corresponde a la observación i -ésima de la variable dependiente del modelo.

Por su parte, X_i es la observación i -ésima de la variable independiente del modelo.

Los coeficientes β_0 y β_1 son desconocidos y deberán ser estimados.

Siendo β_0 el término independiente (constante o “intercept”) y β_1 el coeficiente de regresión de la variable explicativa (pendiente o “slope”) y el término ε_i es una perturbación estocástica agregada al modelo para recoger todos los posibles errores de medida tanto en las variables X e Y así como los errores en la especificación lineal del modelo, es decir recogerá todos aquellos factores que por error no se han incluido en el modelo y que pueden afectar a la variable dependiente del modelo.

Nótese que, en definitiva, el término de perturbación ε_i indica en que medida las variables X e Y se apartan de la relación lineal.

Donde se supondrá:

$$\varepsilon \rightarrow N(0, \sigma^2)$$

La regresión consiste en estimar los coeficientes (parámetros) de la ecuación tal que la sumatoria de los errores al cuadrado sea mínima.

$$\text{Mín} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2$$

Resolviendo este problema mediante cálculo diferencial, se obtienen los estimadores de mínimos cuadrados de los coeficientes de la recta de regresión:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} ; \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Al igual que ocurre con otros estimadores, existirá cierta incertidumbre en el cálculo de las estimaciones, que se podrá reflejar mediante intervalos de confianza para ambos valores, contruidos bajo la hipótesis de normalidad de los residuos, mediante las expresiones:

$$IC(1-\alpha)\% (\beta_0) = \left(\hat{\beta}_0 \pm t_{\alpha/2}^{n-2} S \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}} \right)$$

$$IC(1-\alpha)\% (\beta_1) = \left(\hat{\beta}_1 \pm t_{\alpha/2}^{n-2} \frac{S}{\sqrt{S_{XX}}} \right)$$

Donde:

$$S_{XX} = \sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i \right)^2}{n} \quad \text{y} \quad S = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2}{n-2}}$$

Las desviaciones de los estimadores de los coeficientes del modelo están dados por:

$$\sigma_{\hat{\beta}_0} = S \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}} \quad \sigma_{\hat{\beta}_1} = \frac{S}{\sqrt{S_{XX}}}$$

Coefficiente de determinación:

$$R^2 = \frac{SCR}{SCT}, \quad 0 \leq R^2 \leq 1$$

Coefficiente de determinación ajustado:

$$R_a^2 = 1 - \left(\frac{n-1}{n-p} * \frac{SCE}{SCT} \right), \quad 0 \leq R^2 \leq 1$$

Donde:

$$SCR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

$$SCE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$$SCT = \sum_{i=1}^n (Y_i - \bar{Y}_i)^2$$

2.3.2.4. Regresión Múltiple

Generalizando la notación usada para el modelo de regresión lineal simple, a partir de una muestra de n observaciones de la cual se dispone los valores de una variable respuesta Y además de los valores de k variables explicativas X_1, X_2, \dots, X_k .

El modelo de regresión lineal múltiple con k variables intenta aproximar la variable respuesta mediante una función lineal de las variables explicativas de la siguiente forma:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad ; \quad i = 1, 2, \dots, n \quad (1)$$

De modo que los coeficientes (parámetros) de la ecuación se estiman siguiendo el criterio de mínimos cuadrados:

$$\text{Mín} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_{i1} - \beta_2 X_{i2} - \dots - \beta_k X_{ik})^2$$

La obtención aquí de las expresiones de los estimadores mínimo cuadráticos de dichos coeficientes exigen reescribir la expresión (1) utilizando notación matricial. Así, (1) quedaría:

$$Y = X\beta + \varepsilon$$

Donde:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Resolviendo este problema mediante el cálculo diferencial, se obtienen los estimadores de mínimos cuadrados a través de la siguiente expresión matricial:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Y mantienen una interpretación análoga al caso de la regresión lineal simple (representa el incremento por término medio en la variable respuesta por cada unidad adicional en la variable).

Estimación de sigma:

$$s^2 = \frac{SCE}{n - p}$$

Donde:

n es el número de observaciones

p el número de parámetros (k+1)

SCE igual a la siguiente expresión:

$$SCE = (Y - X\hat{\beta})^T (Y - X\hat{\beta})$$

Matriz de covarianza:

$$\Sigma = s^2 (X^T X)^{-1}$$

Coefficiente de determinación:

$$R^2 = \frac{SCR}{SCT}, \quad 0 \leq R^2 \leq 1$$

Coeficiente de determinación ajustado:

$$R_a^2 = 1 - \left(\frac{n-1}{n-p} * \frac{SCE}{SCT} \right), \quad 0 \leq R^2 \leq 1$$

Donde:

$$SCR = \beta^T X^T Y - n\bar{Y}^2$$

$$SCE = Y^T Y - \beta^T X^T Y$$

$$SCT = Y^T Y - n\bar{Y}^2$$

2.3.3. Supuestos de mínimos cuadrados

Normalidad en los residuos

La distribución ε debe ser normal de media cero y varianza constante, es decir: $\varepsilon \rightarrow N(0, \sigma^2)$

El modo más sencillo de comprobar esto es obteniendo una impresión visual a partir de un gráfico de los residuos frente a la variable dependiente Y Q-Q plot, además haciendo pruebas de normalidad para los residuos, como el de Kolmogorov-Smirnov o se puede complementar este análisis mediante gráficos de probabilidad normal.

No autocorrelación en los residuos

Los residuos no se correlacionan entre si, es decir:

$$Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \text{para } i \neq j$$

Este supuesto se puede verificar mediante gráficos de autocorrelación y contrastes de independencia como el de Durbin-Watson.

También se puede identificar cuando los signos y valores de los parámetros estimados cambian mucho al añadir o eliminar variables al modelo.

Esto se puede verificar mediante gráficos de autocorrelación y contrastes de independencia como el de Durbin-Watson.

No multicolinealidad

El rango de la matriz $n \times K$ (matriz X), es K con probabilidad 1.

Para entender este supuesto recuerde que el rango de una matriz es el número de columnas linealmente independientes.

El supuesto dice que ninguna de las K columnas de X puede ser expresada como una combinación lineal de otras columnas de X , esto quiere decir que el supuesto asegura que $n > K$.

Se puede identificar cuando los signos y valores de los parámetros estimados cambian mucho al añadir o eliminar variables al modelo.

Además se puede eliminar este problema al excluir variables que estén correlacionadas.

2.3.4. Pruebas de hipótesis

2.3.4.1. Anova

2.3.4.1.1. Anova (comparación de medias)

El análisis de varianza (Anova) es utilizado para comparar más de dos medias de muestras diferentes que se presumen provienen de una misma población.

Supuestos

Observaciones independientes

Homogeneidad de varianzas

Normalidad.

Contraste de Hipótesis

$$H_0 : \mu_0 = \mu_1 = \dots \mu_k = 0$$

$$H_1 : \text{Al menos un } \mu_i \neq 0 \quad ; \quad i = 0,1,2, \dots, k$$

$$EP: F^* = \frac{MCR}{MCE}$$

$$RR: F \geq F_{(p-1),(n-p)}^{1-\alpha}$$

Tabla anova

TABLA 2.1 Tabla Anova				
Fuente	Grados de libertad	Sumas cuadráticas	Medias cuadráticas	F
Tratamiento	p-1	SCR	MCR =SCR/(p-1)	F=MCR/MCE
Error	n-p	SCE	MCE =SCE/(n-p)	
Total	n-1	SCT		

Donde:

$$SCR = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$$

$$SCE = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$$SCT = \sum_{i=1}^n (Y_i - \bar{Y})^2$$

2.3.4.1.2. Anova (regresión)

El análisis de varianza en regresión (múltiple) es utilizado para verificar si todos los coeficientes en el modelo son ceros, es decir, confirmar si alguna de las variables independientes aporta explicación de la variable dependiente.

Contraste de hipótesis

$$H_0 : \beta_0 = \beta_1 = \dots \beta_k = 0$$

$$H_1 : \text{Al menos un } \beta_i \neq 0 \quad ; \quad i = 0, 1, 2, \dots, k$$

$$EP : F^* = \frac{MCR}{MCE}$$

$$RR : F \geq F_{(p-1), (n-p)}^{1-\alpha}$$

Tabla anova

TABLA 2.2 Tabla Anova de regresión				
Fuente	Grados de libertad	Sumas cuadráticas	Medias cuadráticas	F
Regresión	p-1	SCR	MCR=SCR/(p-1)	F=MCR/MCE
Error	n-p	SCE	MCE=SCE/(n-p)	
Total	n-1	SCT		

Donde:

$$SCR = \beta^T X^T Y - n\bar{Y}^2$$

$$SCE = Y^T Y - \beta^T X^T Y$$

$$SCT = Y^T Y - n\bar{Y}^2$$

2.3.4.2. Prueba de Kolmogorov Smirnov

En esta prueba se está interesado en conocer el grado de concordancia entre la distribución de frecuencia muestral ($f_0(x, \theta)$) y la distribución de frecuencia teórica ($f(x, \theta)$), esta prueba trabaja con distribución de frecuencia acumulada y pertenece al campo de la estadística no paramétrica.

Contraste de hipótesis

$$H_0 : f(x, \theta) = f_0(x, \theta)$$

$$H_1 : f(x, \theta) \neq f_0(x, \theta)$$

$$EP : D_{MAX}(x) = \text{Max}|F_X(x) - S_n(x)|$$

$$RR : D_{MAX}(x) \leq D_{MAX}(\alpha, n)$$

Donde:

$F_X(x)$ *distribución de frecuencia acumulada teórica*

$S_n(x)$ *distribución de frecuencia acumulada muestral*

α *nivel de significancia*

n *número de observaciones*

2.3.4.3. Prueba t

En esta prueba se está interesado en conocer la significancia estadística de cada variable dentro del modelo, para esto se establece el siguiente contraste de hipótesis:

$$H_0 : \beta_i = 0 ; \text{ para } i = 1, 2, 3, \dots, p$$

$$H_1 : \beta_i \neq 0$$

$$EP : t^* = \frac{\hat{\beta}_i}{\hat{\sigma}_\beta}$$

$$RR : t \geq_{n-2}^{1-\alpha}$$

2.3.4.4. Prueba de Durbin Watson

En esta prueba se está interesado en conocer si existe autocorrelación positiva o negativa entre las observaciones para esto se plantea el siguiente contraste de hipótesis:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

$$EP : D^* = \frac{\sum_{i=2}^n (\epsilon_i - \epsilon_{i-1})^2}{\sum_{i=1}^n \epsilon_i^2}$$

$$RR : D \leq D_L^{2\alpha} \vee D \geq D_U^{2\alpha}$$

Donde D_U, D_L son leídos de la tabla Durbin Watson

2.3.5. Software utilizado

SPSS 11.0 for Windows es un conjunto de programas orientados a la realización de análisis estadísticos aplicados a las ciencias sociales. Nos permite realizar análisis y gráficos estadísticos sin tener que conocer la mecánica de los cálculos ni la sintaxis de los comandos del sistema. Comparado con otros programas, es más intuitivo y fácil de aprender. Su desventaja es que es menos flexible y con menos procedimientos avanzados que otros programas comerciales.

SPSS es bueno a la hora de organizar y analizar datos. Se puede ordenar datos, calcular nuevos datos y realizar una gran variedad de análisis estadísticos. En teoría el tamaño de los ficheros de datos que SPSS puede manejar no está limitado por lo que puede trabajar con ficheros grandes. Esta versión también permite el manejo cómodo de ficheros, la personalización de los informes, y el cortar y pegar en otros programas.

Para ejecutarlo, se busca el programa en Inicio. pulsa dos veces sobre el icono del programa, en el escritorio de Windows.

Hay dos tipos de archivos asociados a SPSS:

- Archivos de datos: tienen extensión .sav y están en formato SPSS.
- Archivos de texto: tienen extensión .sps (archivos de sintaxis) o .spo (archivos de resultados).

2.3.5.1. SPSS breve descripción

Los pasos básicos en el análisis de datos consisten en:

- Introducir los datos, manualmente o recurriendo a un archivo ya existente.
- Seleccionar un procedimiento estadístico.
- Seleccionar las variables para el análisis, las variables que podemos usar en cada procedimiento se muestran en un cuadro de diálogo del que se seleccionan.
- Ejecutar el procedimiento y ver los resultados, los resultados aparecen en una ventana de resultados y se pueden guardar como archivos con extensión .spo. Los gráficos se pueden modificar en la ventana del editor de gráficos que se presenta cuando se da doble clic sobre dicho gráfico.

2.3.5.2. Vistas en SPSS

La primera presentación es de una tabla de datos, donde se deberán introducir los datos de cada problema o leerlos de un fichero. Corresponde al Editor de datos.

Con el editor de datos podemos crear nuevos archivos o modificar los existentes. No se puede tener más de un archivo de datos abierto al mismo tiempo en la misma sesión de SPSS. Dentro del editor de datos, dos vistas son posibles:

- **Vista de datos**, muestra los valores de datos reales o las etiquetas de valor definidas:

Las filas son casos. Cada fila representa un caso u observación.

Las columnas son variables. Cada columna representa una variable o característica que se mide.

Las casillas contienen valores numéricos o de cadena, siendo éste un valor único de una variable para cada caso.

A diferencia de una hoja de cálculo, las casillas del editor de datos no pueden contener fórmulas.

- **Vista de variables**, contiene descripciones de los atributos de cada variable del archivo de datos. Aquí:

Las filas son variables.

Las columnas son atributos o características de las variables.

Cambiamos de una vista a otra a través de las pestañas en la parte inferior de la ventana.

Otras ventanas irán apareciendo a medida que vayamos realizando nuestro análisis, podrán contener gráficos (Editor de gráficos), informes con los resultados, etc.